Machine Learning Techniques for Biomedical Natural Language Processing: A Comprehensive Review

ESSAM H. HOUSSEIN, **REHAB E. MOHAMED**, **AND ABDELMGEID A. ALI**
Faculty of Computers and Information, Minia University, Minia 61519, Egypt

Corresponding author: Essam H. Houssein (essam.halim@mu.edu.eg)

ABSTRACT The widespread use of electronic health records (EHR) systems in health care provides a large amount of real-world data, leading to new areas for clinical research. Natural language processing (NLP) techniques have been used as an artificial intelligence strategy to extract information from clinical narratives in electronic health records since they include a great amount of valuable clinical information. However, in a free-form text such as electronic health records, many clinical data are still hidden in a clinical narrative format. Therefore, the performance of biomedical NLP techniques is required to unlock the full potential of EHR data to convert a clinical narrative text automatically into structured clinical data. In this way, biomedical NLP applications can be used to direct clinical decisions, identify medical problems, and effectively postpone or avoid the occurrence of a disease. This review discusses the current literature on the secondary use of electronic health record data for clinical research on chronic diseases and addresses the potential, challenges, and applications of biomedical NLP techniques. We review some of the biomedical NLP methods and systems used over EHRs and give an overview of machine learning and deep learning methodologies used to process EHRs and improve the understanding of the patient's clinical records and the prediction of chronic diseases risk, providing a great chance to extract previously unknown clinical information. Moreover, this review summarizes the utilizing of Deep Learning and Machine Learning techniques in biomedical NLP tasks based on chronic diseases related EHR data. Finally, this review presents the future trends and challenges in the biomedical NLP.

INDEX TERMS Artificial intelligence (AI), clinical information, deep learning, electronic health records (EHR), machine learning, natural language processing (NLP).

I. INTRODUCTION

There is a significant impact of Natural Language Processing (NLP) and Machine Learning techniques on processing digital data. The reliance on digital data is increasing, so it is essential to use the value of data in different research fields. Extracting information from the clinical text can be applied to various applications such as automatic terminology management, de-identification of the clinical text, data mining, identification of research subject, prediction of the onset and progress of different chronic diseases, analysis of the disease medication and its side effect, etc. Although NLP-based machine learning techniques have a better performance in the field of biomedicine and healthcare, more experience

The associate editor coordinating the review of this manuscript and approving it for publication was Shen Yin.

is required in the analysis of the narrative clinical text [1]. Therefore, it is necessary to intensively review the problems and challenges of extracting information from the clinical text to develop new opportunities in this field of research [2].

Biomedical NLP is a field of research that includes natural language processing, bioinformatics, medical informatics, and computer linguistics [1]. Extracting valuable information from a free clinical text embedded in unstructured data is a significant task of NLP that can support decision making, reporting on administration, and research. Applying biomedical NLP applications in EHRs has a considerable effect on several domains of healthcare and biomedical research.

Healthcare-related NLP paved the way to medical language processing. Usually, most of the biomedical data exist in an unstructured form, which is the result of dictated transcriptions, direct entry, or using speech recognition

applications. Consequently, data pre-processing is required for information extraction because the summarization and decision-support tasks cannot be performed using the input data in its narrative form. Preprocessing includes document structure analysis, tokenization, part-of-speech tagging, spell checking, sentence splitting, Word Sense Disambiguation (WSD), and some form of parsing. Situation dependent features like event subject identification, temporality, and negation play a crucial role in the inappropriate interpretation of the extracted information [3].

There are various information extraction techniques, such as rule-based techniques, pattern matching techniques, and machine learning and statistical techniques. Then, the extracted information can be used to analyze the clinical text as well as improve the EHR and the decision support systems and to be related to concepts in the standard terminologies. The biomedical natural languages processing involves the methods and studies of how NLP can be applied to the biomedical and electronic medical record texts and literature.

Recently, deep learning techniques have achieved better performance by applying their techniques to different general NLP tasks such as language modeling, (Part of Speech) POS tagging, named entity recognition, paraphrase detection, and sentiment analysis compared to traditional machine learning (ML) techniques. Because of the frequent use of acronyms and non-standard clinical terminology by healthcare professionals, unorganized structure of the document, and the need for complete de-identification and anonymization to protect the privacy of patient data, clinical reports typically face specific challenges compared to general-domain text. Eventually, addressing and solving these challenges could promote further research and improvement for various biomedical applications such as clinical decision support, identification of patient cohorts, patient engagement support, public health management, pharmacovigilance, medications, and summarization of clinical texts.

A. MOTIVATION

Historically, extracting clinical information from narrative clinical texts is done manually by clinical experts, which caused several issues such as lack of scalability and high cost. These issues have particularly affected chronic diseases since clinical notes are more than structured data; for example, the number of clinical notes compared to structured data for chronic diseases such as rheumatoid arthritis, Parkinson's disease, and Alzheimer's disease is graphically quantified by Wei *et al.* [4].

NLP approaches have a significant roles in addressing and solving several challenges of various clinical tasks such as automatic extraction of relevant clinical information that may postpone or avoid the onset of disease for instance. The main objective of this study is to provide a detailed overview of NLP in the clinical domain, including methodologies, system architecture, applications, and the challenges that clinical NLP methodologies meet in understanding clinical narratives, as outlined below:

- We have identified the NLP in general and biomedical NLP with its methods and technologies.
- Then we have presented the application areas of machine learning/deep learning in the biomedical NLP.
- We have provided an overview of the most popular biomedical NLP systems and their general architecture.
- We have identified the usage of NLP applications in clinical notes to identify chronic diseases and understand the challenges currently facing them.
- Next, we have discussed a literature review of the application of various NLP techniques to narrative clinical notes on chronic diseases, including the analysis of difficulties faced by NLP methodologies in clinical narrative comprehension.
- Finally, we conclude this review paper by describing existing challenges currently faced and open issues associated with the processing of the biomedical and clinical text and providing the NLP domain with sufficient resources and opportunities to extract new methodologies.

B. CRITERIA FOR SEARCH AND SELECTION

We searched for previous studies released from 2009 to 2021 using Google Scholar, PubMed, and the Web of Science. All searches used the keywords ''electronic health records'' or ''electronic medical records'' or ''EHR'' or ''EMR,'' in combination with either ''machine learning'' or the name of a particular technique of machine learning in conjunction with 'chronic diseases'. Figure 1a shows the number of publications related to applying machine learning to EHR per year. Figure 1b shows the number of publications related to the use of EHR in chronic diseases per year.

In the rest of this review paper, we propose an overview of the most significant and noticeable articles and researches that focus on EHR using the machine learning and deep learning techniques.

We start with a general review of NLP in general, NLP in biomedicine and healthcare with its methods, technologies and potential tasks/usecases in the biomedical and healthcare domains in Section II and Section III, followed by application areas of machine learning in the biomedical NLP in Section IV. Then we provide an overview of NLP systems and system architecture in Section V. Next, we discuss a literature review of recent related works about applying NLP on chronic diseases in Section VI. Then we look at current open issues and challenges in the domain of the biomedical NLP in Section VII. Finally, Section VIII demonstrates the conclusion of the review paper by identifying current challenges and open issues.
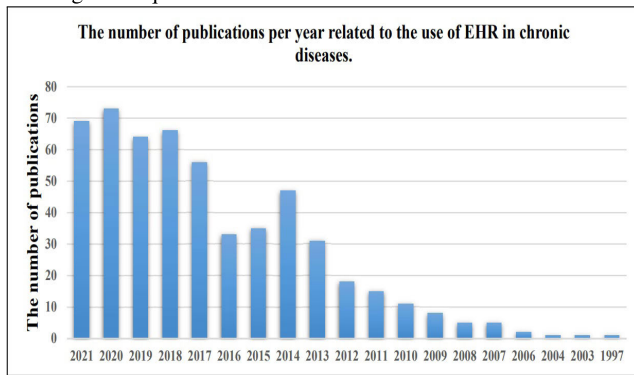
II. BASICS AND BACKGROUND
A. NATURAL LANGUAGE PROCESSING OVERVIEW

NLP is a sub-field that combines computer science, Artificial Intelligence (AI), and linguistics, where the aim is to process and interpret human language to carry out several tasks
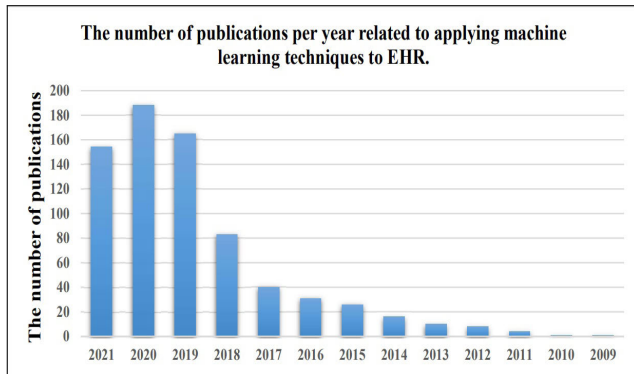
(a) The number of publications per year related to applying machine learning techniques to EHR.

(b) The number of publications per year related to the use of EHR in chronic diseases.

FIGURE 1. The number of publications per year related to EHR.

(e.g., automatic answering questions and, translation of languages). NLP is widely considered an AI-complete problem because of the various complexities involved in describing, understanding, and utilizing linguistic, social, global, or visual information. NLP usually requires processing at different levels of an input text such as tokenization, morphological analysis, syntactic analysis, semantic analysis, and discourse processing [5]. NLP is a specialized branch of AI that focuses primarily on perception and human-generated data-text or speech-based. The technology has several subdisciplines, including Natural Language Query (NLQ), Natural Language Generation (NLG), and Natural Language Understanding (NLU) [5].

B. NATURAL LANGUAGE PROCESSING IN BIOMEDICINE AND HEALTHCARE

Many challenges are facing natural language processing when it is applied to general language, but some critical issues are especially relevant to the biomedical and healthcare domains. There is a wealth of electronic information concerning the healthcare domain, including publications, e-health records, and the Internet. Subsequently, there are many critical aspects relating to biomedical information, most of which are in textual form, in terms of controlling and using such information which is necessary to health research promotion,

quality improvement and cost reduction. NLP is important because it is required to convert narrative clinical texts into structured data that can be used in computer applications [6].

The adoption of electronic health records systems in hospitals has increased significantly in the last ten years, by providing incentives of $30 billion to hospitals and physicians practices for the adoption of EHR systems, partly because of the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act [3]. The basic EHR system is used by 84% of hospitals, which has increased 9-fold since 2008, according to the most recent study from the Office of the National Coordinator for Health Information Technology (ONC) [7]. Furthermore, the use and adoption of basic and certified EHRs by office-based physicians has increased from 42% to 87%. Data of each encountered patient are stored by EHR systems such as demographic information, diagnosis, laboratory examinations, drugs, radiological images, clinical notes, etc.

Generally, the use of Electronic Health Records (EHR) Systems in both the hospital and outpatient care settings has increased significantly [7]. The use of EHR in hospitals and clinics has the potential to enhance patient care by reducing errors, and improving efficiency, the quality of treatment, while providing researchers with a rich data source [8]. The functionality of EHR systems can vary and are usually classified into basic EHRs without clinical notes, basic EHR with clinical notes, and comprehensive systems [7]. Even basic EHR systems can provide a wide range of patient information such as medical history, diseases, and medication use while lacking more advanced features. As the EHR was mainly developed for administrative activities at the hospital, there are many classification schemes and controlled vocabulary for recording patient medical data and events. Table 1 shows some codes from International Statistical Classification of Diseases and Related Health Problems (ICD) containing diagnosis codes, codes from Current Procedural Terminology (CPT) containing procedure codes, codes from Logical Observation Identifiers Names and Codes (LOINC) containing laboratory notes, and codes from RxNorm containing drug codes.

Such codes may differ between organizations, with partial mappings managed by tools such as the United Medical Language System (UMLS) and the Systemized Nomenclature of Medicine - Clinical Terms (SNOMED CT). With the availability of various classification schemas, coordinating and analyzing data through terminologies and across organizations is a field of ongoing research.

Diverse types of patient information are stored in EHR systems, including demographics, diagnostics, physical examinations, sensor measurement, lab results, prescribed or managed medicines, and clinical notes. One difficulty is to deal with the complexity of EHR data with its different types of data, including the following:

(i) **Numerical Quantities** like the index of body mass.
(ii) **Date/Time Items** like the date of birth or admission time.

TABLE 1. Example of classification schema for diagnoses, procedures, laboratory examinations, and drugs.

Classification Schema	Number of Codes	Examples	
ICD-10 (Diagnosis)	68,000	**Code**	**Associated Diagnosis**
		J9600	Acute respiratory failure
		I509	Heart failure
		I5020	Systolic heart failure
CPT (Procedures)	9,641	**Code**	**Associated Procedures**
		72146	MRI Thoracic Spine
		67810	Eyelid skin biopsy
		19301	Partial mastectomy
LOINC (Laboratory)	80,868	**Code**	**Associated Laboratory**
		4024-6	Salicylate, Serum
		56478-1	Ethanol, Blood
		3414-0	Buprenorphine Screen
RxNorm (Drugs)	116,075	**Code**	**Associated Drugs**
		161	Acetaminophen
		7052	Morphine
		1819	Buprenorphine

(iii) **Categorical Values** like ethnicity or controlled vocabularies codes like ICD-10 (formerly ICD-9) diagnoses or CPT procedures.

(iv) **Free-Text Natural Language** like progress reports or discharge summaries. Those types of data can also be ordered chronologically.

(v) **Time-Series Derived** like signals of vital perioperative sign or multimodal patient history.

While other biomedical data such as medical images or genomic information are present and treated in important recent researches [9]–[11], we concentrate in this review paper on the five types of data that exist in most modern EHR systems. In the field of chronic diseases, new methods are needed to support and advance evidentiary medicine, given the increasing incidence of such conditions all over the world. There is a powerful and successful impact of the secondary use of EHRs in processing clinical data for biomedical and translational applications.

Several research studies have discovered a secondary use of EHRs in bioinformatics and healthcare applications [12], [13], although it is designed primarily to enhance operational healthcare performance. In particular, patient-relevant data stored in EHR systems were used for biomedical tasks such as extracting medical concepts [2], [14], modeling patient trajectories [15], diagnosing diseases [16], [17], supporting clinical decisions [18], etc.

Processing EHRs using machine learning and deep learning methods contributes to a better and more deep understanding of clinical patient trajectories which track the patient status from one health state to another being diagnosed with a specific clinical condition and risk prediction of chronic diseases, giving a unique opportunity to get unknown clinical information. However, A wide range of clinical history, remains locked in free-form texts behind clinical narratives. As a result, the unlocking of the full potential of EHR data depends on the development of NLP techniques to automatically convert the clinical text from its narrative nature to a structured form that can direct clinical decisions and potentially postpone or prevent the onset of diseases [19].

EHR processing and modeling are significant challenges due to its high dimensionality, noise, heterogeneity, sparse design, incompleteness, random errors, and systematic biases. In addition, a vast amount of information about patient clinical history is usually stored in free-text clinical narratives [20] since the most widely and descriptive method for recording clinical events remains written text. The development of NLP techniques integrated into machine learning algorithms is essential for the automatic conversion of clinical free-text into a structured data format. NLP has been used for a broader variety of applications in the clinical domain, including the detection of medical concepts from nursing documentation [21], discharge summaries [22] and radiology reports [23] as much potentially useful clinical information for pharmacoepidemiological research is contained in unstructured free-text documents. Routine health data such as Scottish Morbidity Records (SMR01) frequently use generic 'stroke' codes. Free-text Computerised Radiology Information System (CRIS) reports have potential to provide this missing detail. In order to increase the number of stroke-type-specific diagnoses by augmenting SMR01 with data derived from CRIS reports and to assess the accuracy of this methodology. However, applying NLP-based frameworks to a narrative clinical text has not been widely used in clinical activities and tasks to direct decision-support systems or administrative processes.

C. TASKS OF NLP IN THE HEALTHCARE DOMAIN
There are several Tasks of the clinical NLP:

- **Word Sense Disambiguation (WSD):** is the process of automatically assign an accurate meaning (sense) to an ambiguous word in a specific context. The biomedical NLP tasks require the ability to accurately understanding ambiguous words within a specific context which is a critical issue. According to the medical word sense disambiguation, there is a list of all possible meanings (senses) for each ambiguous word. There are many ambiguous terms in clinical notes. There are a variety of interpretations for the abbreviation "PCA," including

principal component analysis, patient-controlled analgesia, and prostate cancer. WSD is a critical issue in the medical domain [24], [25] [26], [27] because it is an essential step for the analysis of clinical notes [28].

- **Name Entity Recognition (NER):** is a subtask of IE (Information Extraction). One of the most important tasks in biomedical NLP is to turn unstructured text into computer-readable structured data [29]. NER is the task of identifying expressions that denote named entities (such as diseases, medications, and lab tests) in clinical notes. Many techniques can be used in NER such as [30] dictionary-based approach, rule-based approach, statistical approach, deep learning approach, hybrid approach [31].

- **Adverse Drug Events (ADEs) Detection:** Both medical research and hospital medical treatment benefit from detecting adverse drug events (ADEs) and medication-related information in clinical notes. ADEs are known as diseases occurring from medical interventions of medicines such as prescription errors, overdoes, adverse drug reactions, and allergic reactions [32]. EHRs have a wealth of information on ADEs which is hidden in unstructured data such as discharge summaries, procedural notes, medical history, laboratory results [33]–[35]. The process of identifying and detecting the information related to ADEs from narrative clinical notes is very difficult and time-consuming. So there is a need for the NLP system for automatically processing narrative EHRs and detecting drugs, ADEs, and their interactions [36].

- **Information Extraction (IE):** is an important biomedical NLP task that facilitates the use of EHRs for clinical decision support, quality improvement, or clinical and translation research by automatically extracting and encoding clinical concepts from narratives notes. In the general domain, IE is commonly recognized as a specialized area in empirical NLP and refers to the automatic extraction of concepts, entities, and events, as well as their relations and associated attributes from free text [34], [37]

- **Relation Extraction (RE):** is an important subtask of information extraction (IE) that focuses on identifying and detecting semantic relationships between clinical concepts in clinical notes [38], [39]. For example, in this clinical note ''an MRI revealed a C5-6 disc herniation with cord compression'', the lab test ''MRI'' indicates two diseases ''a C5-6 disc herniation'' and ''cord compression''. Many types of relations are mentioned by previous researches such as disease-attribute pair extraction [40], [41], temporal relation identification [42], adverse drug event detection [43], [44], etc. clinical NLP domain has recently launched several shared tasks related to relation extraction from clinical notes such as Integrating Biology and the Bedside (i2b2) challenges [45], the Semantic Evaluation (SemEval) challenges [46] and the most recent 2018 National NLP

Clinical Challenge (n2c2) [47]. These open shared tasks and challenges provide many resources and methods for medical RE tasks [40].

III. BIOMEDICAL NLP METHODS

This review paper gives an overview of the most recent articles based on most of the main biomedical NLP methods employing dictionary-based, rule-based, and machine learning techniques. Figure 2 shows the number of publications in the EHR domain applying deep/machine learning methods and rule-based techniques per year. Although the use of machine-learning methods is growing compared to rule-based methods, the performance and efficiency of machine learning algorithms can be highlighted by using rule-based methods as a benchmark because we are still seeing a shift from rule-based methods to machine learning algorithms [48].
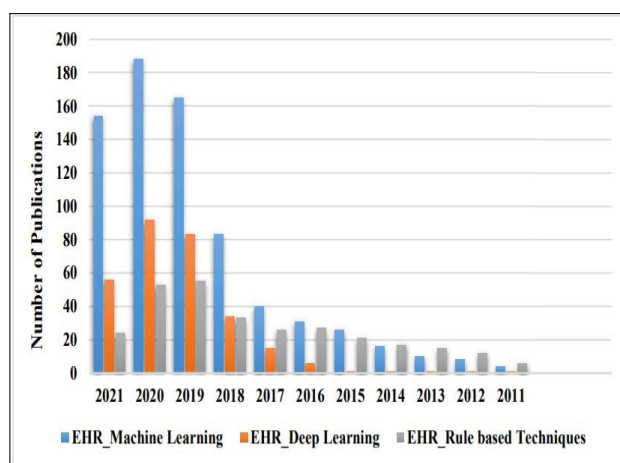

FIGURE 2. The number of publications per year in the EHR domain applying deep/machine learning methods and rule-based techniques.

Recently, biomedical NLP researches have shown the significant performance of the methods based on deep learning. The effective performance of Recurrent Neural Network (RNN) in biomedical texts for the NER (Name Entity Recognition) task was proposed by Sahu and Anand [49]. They developed a model which is a combination of a bidirectional Long Short-Term Memory Network (BiLSTM) and Conditional Random Field (CRF) applying character-level word embedding. Habibi *et al.* [50] combined the BiLSTM-CRF model developed by Lample *et al.* [51] and the word embedding model developed by Pyysalo *et al.* [52]. To generate important features such as orthographic features of biomedical organisms, Habibi *et al.* [50] used character-level word embedding to show that the characteristic word embedding is successful in biomedical NLP tasks.

A. RULE-BASED TECHNIQUES
Rule-based techniques are based on a set of specific textual relationship rules that called patterns that encode similar

structures in the expression of relationships. These rules are represented over words or POS tags as regular expressions. In such systems, the rules extend as patterns by adding more constraints to resolve a few issues, including checking negation of relations and determining the direction of relationships. The rules are generated in two ways: manually and automatically generated from the training dataset. The efficiency of the rule-based system can be enhanced to a certain extent using an extension with additional rules, but it tend to produce much FP information. Therefore, rules-based systems usually provide high precision but low recall because the rules for a particular dataset cannot be created for other data sets. However, the recall of such systems can be improved by relaxing the constraints or by learning rules automatically from training data [53].

Although the architecture of dictionary-based systems is simple, they cannot be applied to manage unknown entities or ambiguous words, resulting in low recall [54], [55]. It also requires a considerable amount of manual labor to develop and maintain a comprehensive and up-to-date dictionary. Although the rule-based method is more flexible, the features are handcrafted to fit a model into a dataset [56], [57]. Both rules and dictionary-based methods can achieve high precision [58], but they can generate wrong predictions when the out-of-vocabulary problem occurs if a new word, which is not found in the training data, appears in a sentence. The issue of out-of-vocabulary arises in the biomedical field in particular because it is common to register a new biomedical term such as the name of a new drug.

There are several methods based on rule-based techniques such as:

- Dictionary Lookup [59]–[61].
- Domain ontology-based terminology recognition [62]–[65].
- Set of manual rules [66], [67].
- Regular expressions patterns [68], [69].

B. MACHINE LEARNING TECHNIQUES

There are two main categories of learning techniques of machine learning algorithms: supervised learning and unsupervised learning. Supervised learning techniques provide a function that maps from inputs x to outputs y:

$$y = f(x) \qquad (1)$$

There are two main types of supervised learning techniques: classification and regression, and the most widely learning algorithms are logistic regression algorithm and support vector machine algorithm. On the other hand, the purpose of unsupervised learning techniques is to learn about the input x distribution features. There are two main methods of unsupervised learning:- cluster analysis and principal component. The input representation is an essential task for all machine learning frameworks. Machine-learning techniques input is a set of attributes known as features that are extracted for each data point. Such features are handcrafted based on domain knowledge in traditional machine learning where automatic

data-oriented feature extraction is an essential aspect of deep learning techniques.

Up to the last few years, machine learning methods such as logistic regressions, support vector machines (SVM), and random forests were employed as key methods for analyzing and processing rich EHR data [70]. Most modern NLP platforms are built on models refined through machine learning techniques [71], [72]. Machine learning techniques are based on four components: a model; data; a loss function, which is a measure of how well the model fits the data; and an algorithm for training (improving) the model [73].

1) DEEP LEARNING TECHNIQUES

Deep learning is a subfield of machine learning methods based on multi-layered neural network architectures with hierarchical data representations learning, as shown in Figure 3. Machine learning techniques require time-consuming and hard work for data representation feature extraction [74], While learning multiple levels of representations can be automatically done by deep learning techniques with increasing order of abstractions [75].

There are several factors contributed in the development of deep learning such as the availability of extensive unlabeled data along with rapid computing resources based on powerful graphics processing units (GPUs), new algorithms and frameworks and adaptations/transformations of learned data features/representations to similar or a new domain of interest.
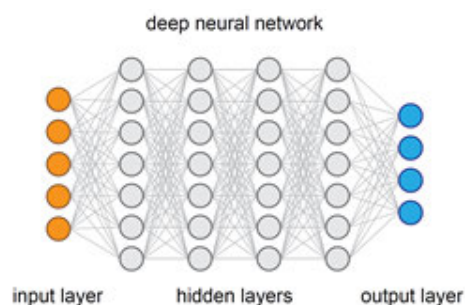

FIGURE 3. A deep neural network architecture.

Several non-linear classification problems with hierarchical inputs that naturally occur, such as language and images, can be solved by deep learning methods. Recently, deep learning techniques can be applied to NLP applications providing better results than techniques based on linear models such as support vector machines (SVMs) or logistic regression [76].

The most popular deep learning architectures are illustrated in this section by highlighting their key equation that demonstrates their operation method. Data representation is the primary task of deep learning. Using a machine learning algorithm, input features must be hand-crafted from the dataset based on the researcher's experience and the domain of knowledge to identify specific patterns of prior interest.

The development method of designing, reviewing, choosing and testing suitable features can be complicated and
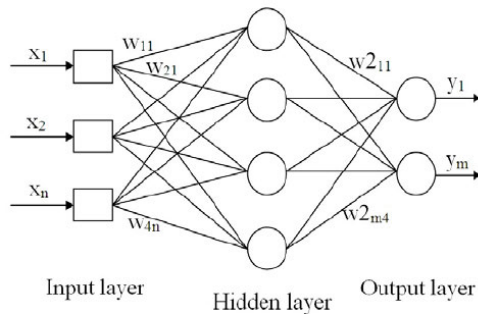
FIGURE 4. Multilayer perceptron architecture.

time-consuming. It can also be regarded as a "black art" [77] requiring creativity, trial-and-error, and sometimes luck. On the other hand, learning the optimal feature directly from the given dataset is performed by deep learning techniques without any handcrafting. Through deep learning, complex data representation is sometimes represented as compositions of other, simpler representations.

Recurrent deep learning architecture is a complex unsupervised hierarchical representation. Many of the major deep learning algorithms and architectures are based on the artificial neural network (ANN) architecture. ANNs consist of multiple interconnected nodes (neurons) organized in layers, as shown in Figure 3. Hidden units are neurons that do not appeared in the input or output layers and store several weights W, which are updated with the training of the model.

The optimization of ANN weights is performed by minimizing loss function as shown in Equation 2, such as a negative log-likelihood.

$$E(\theta, D) = -\sum_{(t=0)}^{D} \left[\log P(Y = y_t | x_t, \theta) \right] + \lambda \parallel \theta \parallel_p \quad (2)$$

The summation of the log loss over the given training dataset D is minimized by the first term. While the minimization of the learned model parameters θ^t p-norm controlled by a tunable parameter λ is performed by the second term, which is known as the regularization technique used to prevent the model overfitting and to improve its ability to be applied to new problems. Usually, the backpropagation technique is used for loss function optimization by minimizing the final layer loss over the network [75].

Many open-source sources are supported by different programming languages such as TensorFlow, Theano, Keras, Torch, PyTorch, Caffe6, CNTK, and Deeplearning4j to deal with deep learning algorithms. In the following subsections, we give an overview of the most common deep learning techniques which can be applied to biomedical NLP applications, such as supervised and unsupervised techniques.

A) **Multilayer perceptron (MLP)**
A multi-layer perceptron is a multiple-hidden layered type of ANN that completely connects each neuron in the layer i to each neuron in the layer i + 1. These networks are usually limited to specific hidden layers,

and unlike recurrent or undirected architectures, the data flows only in one direction. From the definition of the single-layer ANN, as shown in Equation 3, the outputs weighted sum from the previous layer is calculated by each hidden unit, followed by a non-linear activation σ of the calculated sum. Where d is the number of the previous layer units, x_j is the output of jth node of the prior layer, and ω_{ij} and b_{ij} are the weight and bias terms of each x_j. The most common nonlinear activation functions are usually sigmoid or tanh, but recently rectified linear units (ReLU) are used by modern networks [75].

$$h_i = \sigma \left(\sum_{(j=1)}^{d} x_j \omega_{ij} + b_{ij} \right) \quad (3)$$

The network can learn the relationship between the input X and the output Y after optimizing the weights of the hidden layer during training. With the addition of more hidden layers, the input data is supposed to be more abstractly represented due to the non-linear activations of each hidden layer. Although MLP is one of the simplest architecture, other architectures combine fully connected neurons in their final layers.

B) **Convolutional neural networks (CNN)**
Recently, the most popular method is Convolutional Neural Networks (CNN) especially when applied in the image processing domain. CNNs require local raw data connectivity. A one-dimensional time series is also a set of local signal segments. Equation 4 demonstrates one-dimensional convolution where the input is x and the weighting function or convolution filter is w.

$$C_{1d} = \sum_{(a=-\infty)}^{\infty} x(a) \, \omega(t-a) \quad (4)$$

In Equation 5, where X is a 2-D grid (e.g., image) and K is a kernel, is demonstrated in a two-dimensional convolution in which a kernel or filter can pass a matrix of weights through the entire input to the feature maps.

$$C_{2d} = \sum_m \sum_n X(m, n) K(i-m, j-n) \quad (5)$$

The generated number of parameters is small since the filters are usually smaller than the input; therefore, CNNs have limited interactions. CNNs facilitates parameter sharing since all filters are applied across the entire input. The convolution layer in CNN contains several convolutional filters which receive the same input from the previous layer to extract different lower-level features. For features aggregation, these convolution layers are usually pooled or subsampled. Figure 5 provides an example of two convolutionary layers of the CNN architecture, followed by a pooling layer.

C) **Recurrent neural networks (RNN)**
In the case of a simple-spatial structure of input data (like image pixels), CNNs are the appropriate approach, but in the case of sequential organized data (such as
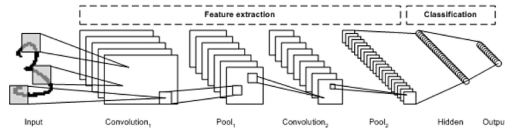
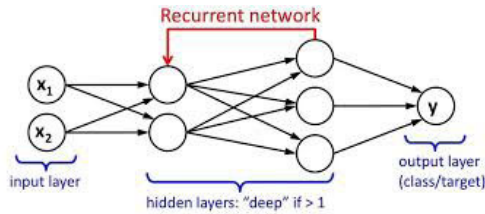**FIGURE 5.** Example of data classification using CNN.

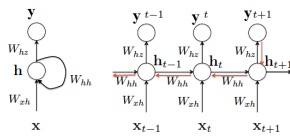**FIGURE 6.** Recurrent neural networks architecture.

FIGURE 7. Recurrent neural networks architecture.

time-series data or the natural language), recurrent neural networks (RNN) is the best method. The generated features are shallow when the CNN is fed with single-dimensional sequences, which means that the feature representations include only very close localized relationships between some neighbors [75]. RNNs are designed to manage this time dependencies for a long time. The RNNs are used to update the hidden ht state sequentially, based not only on the activation of the current input x_t at time t, but also on the previously hidden state h_{t-1}, which in turn was updated from x_{t-1}, h_{t-2}, and so on as shown in Figure 7. Thus, after processing a whole sequence, the final hidden state includes information from all its previous sequences.

Long-term memory (LTM) and Gated Recurrent Unit (GRU) models, both known as Gated RNNs, belong to the popular RNN architectures. While standard RNNs consist of interconnected hidden cells, a particular cell containing an internal recurrence circuit and a gate system that controls the information flow is substituted for each unit in the gated RNN. The gated RNNs have demonstrated better performance when modeling LSTM [75].

D) **Autoencoders (AE)**

The autoencoder (AE) is one of the deep learning models that demonstrate the concept of unsupervised representation learning. First they were considered as a tool for pre-training supervised deep learning models, but they are still useful for completely unsupervised tasks like phenotype discovery. Autoencoders are used to convert input data into a lower-dimensional space called

z. After that, the encoded representation is decoded by reconstructing an estimated representation of the input x, called $\grave{x}$.

The process of encoding and reconstruction for an autoencoder with one hidden layer are illustrated in equations 6 and 7, respectively. The weights of encoding and decoding processes are W and $\grave{W}$ and the encoded representation z is considered more accurate when minimizing the reconstruction error $\| x - \grave{x} \|$.

$$z = \sigma \left(Wx + b \right) \tag{6}$$

$$\hat{x} = \sigma \left(\grave{W}z + \grave{b} \right) \tag{7}$$

A single input is fed through the network as the encoded representation of the input after the AE has been trained with the innermost hidden layer activations. The main task of AEs is to convert and encode the input data to only represent the most significant derived dimensions. Therefore, AEs are similar to traditional dimensionality reduction techniques such as principal component analysis (PCA) and singular value decomposition (SVD), but they provide a major impact for solving complex problems due to nonlinear transformations through the activation functions of each hidden layer.

Many models of AEs have been developed, such as variation autoencoders (VAE), de-noising autoencoders (DAE) [78], and sparse autoencoders (SAE) [75].

E) **Restricted Boltzmann machine (RBM)**

The restricted Boltzmann machine (RBM) is another deep learning model that demonstrates the concept of unsupervised representation learning. RBMs are similar to autoencoders in that they estimate the probability distribution of the input data, but they do so in a stochastic manner. Therefore RBMs are considered to be generative models as they attempt to model the underlying process by which the data was generated. The energy-based model with visible binary units $\vec{v}$ and hidden units $\vec{h}$, with energy functions defined in the Equation 8 is called the canonical RBM [75].

$$E \left(v, h \right) = -b^T - c^T h - Wv^T h \tag{8}$$

However an RBM has no connected visible or hidden units, all the units in a standard Boltzmann machine (BM) are fully connected. It generates the learned representation of the input data in a form of h. RBMs can be stacked hierarchically for the development of a deep belief network (DBN) for supervised learning tasks.

IV. APPLICATION OF MACHINE LEARNING AND DEEP LEARNING TECHNIQUES IN THE BIOMEDICAL NLP DOMAIN

Early EHR analyses were based on simpler and more conventional statistical techniques [79]. Recently, machine learning techniques, such as: Logistic Regression [80], Support Vector Machines (SVM) [81], Cox Proportional Hazard Model [82]

TABLE 2. A recent review of the applications of EHR research using machine learning techniques.

Model	Biomedical NLP application on chronic disease	Reference
SVM	Predicting heart disease. Identification of diabetes. Classification of breast radiology reports.	[88][89][90]
Naïve Bayes	Prediction of heart disease. Classification of smoking status. Identification of multiple sclerosis. Classification of obesity and cancer.	[91] [92] [93] [94] [95] [96] [20] [97]
CRFs	Prediction of heart disease. Identification of diabetes. Identification of multiple sclerosis. Classification of breast radiology reports. Identification tumor characteristics in radiology reports	[98] [88] [99] [90] [100]
Random forests	Prediction of heart disease. Cancer classification. Identifying tumor characteristics Hypertension identification	[101] [102]

and Random Forest [83] have been applied to EHR data for mining reliable predictive patterns.

There are critical issues based on the statistical model when being applied to EHR data analyses [84]–[86]. Such issues can be overcome by applying modeling techniques that can be used to analyze and extract complex nonlinear variables interactions [75], [87] which come from each entire patient's medical history such as mixed and multimodal data obtained in random times [86].

The support vector machine algorithm is the most popular machine learning method that has been applied to medical reports for the prediction of heart disease [88], [89], the identification of diabetes EHR progress notes and the classification of breast radiology reports according to BI-RADS [90].

The second most popular machine learning method is Naïve Bayes which has been applied to medical records for the prediction of heart disease [91], [92], the classification of smoking status [93], for the identification of multiple sclerosis [94], and the EHR records classification for obesity [95] and cancer [20], [96], [97].

Conditional random fields (CRFs) are the third most common machine learning method, which has been applied to medical records for the prediction of heart diseases [88], [98], for the identification of diabetes EHR progress notes [99], for breast radiology reports classification [90], and identifying tumor characteristics in radiology reports [100].

Finally, random forests have been used for heart disease prediction, cancer classification [101], and identification of hypertension [102].

Table 2 outlines the most recent biomedical models using machine learning techniques with their major application.

The drawback of machine learning is in handling high-scale data, their adoption of several statistical and structural assumptions, and their use of hand-crafted features/markers make the use of such statistical models in analyzing the EHR data is impractical, despite its simplicity and interpretability required for biomedical applications [84]–[86]. Recent

breakthroughs in these areas have led to vastly improved NLP models that are powered by deep learning, a subfield of machine learning [103].

Through the deep hierarchical construction of features and the efficient capture of long-range data dependencies, deep learning techniques have recently achieved significant progress in many fields [75]. There is an increased number of researches which apply deep learning techniques to EHR data for biomedical tasks [62], [104], due to the growing development of deep learning methods and the increasing number of patient data providing enhanced results and requiring less time-consumption preprocessing and feature extraction compared with traditional methods.

Modern biomedical NLP systems can identify and model more complex relationships and concepts [105] by using the main deep learning architectures such as feed-forward neural networks (FFNN), convolutional neural networks (CNN), and recurrent neural networks (RNN) that can be applied for the analysis and modeling of HER. Vector-embedding approaches are used for data preprocessing by encoding words before feeding them into a model. These approaches understand that words may have different meanings depending on context (for example, the meanings of "patient," "shot," and "virus" differ depending on context) and treat them as points in a conceptual space rather than separated entities. The emergence of transfer learning has improved the performance of the models, which involves taking a model trained to perform one task and using it as the starting model for training on a similar task [106].

Convolution neural networks (CNNs) have an effective performance in a wide range of NLP biomedical tasks, for example:

1) CNNs have an effective success in the development of a model for classifying biomedical articles to identify cancer hallmarks associated with an abstract article [107].
2) CNNs are used to learn the representation of time expressions for clinical temporal relationship extraction [108].

TABLE 3. A recent review of the applications of EHR research using deep learning techniques.

Application	Task	Input Data	Models	References
Information Extraction	1. Single Concept Extraction	Clinical Notes	LSTM, Bi-LSTM, GRU, CNN	[125] [117] [126]
	2. Temporal Event Extraction		RNN+ Word Embedding	[127]
	3. Relation Extraction		AE	[128]
	4. Abbreviation Extraction		Custom Word Embedding	[129]
Representation Learning	1. Concept Representation	Medical Codes	RBM, Skip-gram, AE, LSTM	[130] [128]
	2. Patient Representation		RBM, Skip-gram, GRU, CNN, AE	[62] [104] [130] [128] [131] [132]
Outcome Prediction	1. Static Prediction	Mixed	AE, LSTM,RBM	[62] [104] [130][133] [134]
	2. Temporal Prediction		LSTM	[84] [135] [131] [136] [137]
Phenotyping	1. New Phenotype Discovery	Mixed	AE, LSTM,RBM	[62] [132] [138][139]
	2. Improving Existing Definitions		LSTM	[140] [141]
De-Identification	Clinical Text de identification	Clinical Notes	Bi-LSTM, RNN+ Word Embedding	[142] [118]

3) CNNs can be applied to model the appropriate article for the biomedical article retrieval task [109].
4) CNNs can be applied to biomedical reports to identify protein-protein interaction relations [110].
5) CNNs can be used with an attention mechanism to extract drug-drug interactions [111].
6) CNNs can be used for classifying free-text radiology reports using the pulmonary embolism results [112].
7) CNNs can effectively support the classification of patient portal messages [113].
8) CNNs can be applied to biomedical text for named entities recognition [114].

In the case of automated coding in radiology reports by using an International Classification of Disease (ICD-10) system, CNN models have contributed to achieving improved efficiency compared with machine learning classifiers [115]. There is also a semi-supervised CNN architecture that can be used in social media to automatically detect adverse drug events (ADE), inspired by the previously mentioned accomplishments of CNNs for different clinical NLP applications, unlike conventional systems [116] that usually employ lexicon-and machine learning-based techniques that depend on expert annotations for ADE detection by producing large quantities of labeled data to train supervised machine learning algorithms.

Many clinical events can be detected from free text EHR notes by applying Recurrent Neural Network (RNN) architectures such as disorders, medications, tests, adverse drug effects [117], and patient data de-identification from EHRs [118]. Bidirectional RNNs/LSTMs have been successfully applied to several biomedical NLP tasks such as building models for the prediction of the missing punctuation in medical reports [119], the identification of biomedical events [120], the modeling of relational and contextual similarities between the named entities in biomedical articles to understand important information to provide appropriate treatment suggestions [121], the extraction of clinical concepts from EHR reports [122], and the recognition of named entities in clinical texts [123]. Many recent researches develop models using the embedded graph information for adverse drug reaction detection in social media data [124] by applying bidirectional LSTM transducer. RNNs are used to develop recognition models for disease name learning

with term- and character-level embedding features [49] when they are used in conjunction with CNNs. We provide, in this section, an overview of the recent state of the art of biomedical applications as a consequence of the rapid and recent development of deep learning techniques being applied to electronic health records (EHR). Table 3 outlines the most recent biomedical models using deep learning techniques with their major application, subtask definitions, and type of input data according to current research's logical classification.
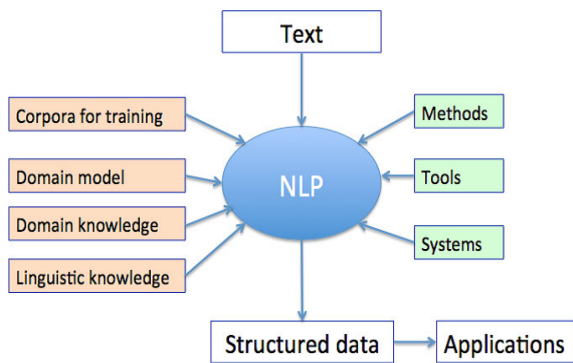
V. BIOMEDICAL NLP SYSTEMS
We give an overview of NLP systems and their architecture in this section.

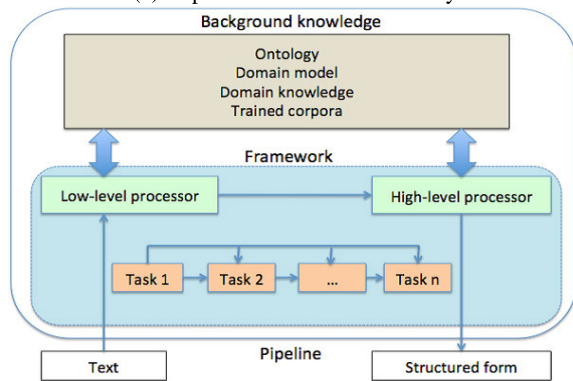General architecture of the Biomedical NLP System

Friedman and Elhadad's discussion [6] illustrates NLP and its different aspects and parts, as shown in Figure 8a.

As shown in Figure 8a, the left part consists of the trained corpora, domain model, domain knowledge, and linguistic knowledge; the right part includes techniques, tools, systems, and applications. Thus aspects of NLP can be divided into two parts.

Figure 8b provides an overview of the general architecture of the NLP system, in which there are two primary components of the NLP system: background knowledge corresponding to the left part of the figure, and a framework that incorporates NLP tools and modules corresponding to the right part of the figure. The two primary components of biomedical NLP systems and their tasks are illustrated below, which are how NLP tools incorporated into a pipeline designed on top of a particular framework. Regarding the framework, which is a software platform for controlling and managing pipeline components like loading, unloading, and handling, the framework's components may be integrated, combined, or used in the system as plug-ins. There are two levels of the framework of biomedical NLP systems:- low-level and high-level processors. Basic NLP tasks are carried out by low-level processors such as part-of-speech tagging, segment tagging, sentence boundary detection, and chunking of noun phrases. Semantic level processing, such as named entities recognition (e.g., disease/disorder, sign/symptoms, medicines), relationship identification, and timeline extraction, is performed by high-level processors.

(a) Aspects of biomedical NLP systems.

(b) A general biomedical NLP system architecture.

FIGURE 8. Biomedical NLP system.

A. BIOMEDICAL NLP BACKGROUND KNOWLEDGE

The Unified Medical Language System (UMLS) Biomedical and linguistic knowledge are essential components in the development of biomedical NLP systems. The **Unified Medical Language System (UMLS)** was developed in 1986 and applied to the biomedical NLP systems. There are three key components of the UMLS: the *Metathesarus, the Semantic Network*, and *the SPECIALIST lexicon*. The UMLS can be known as the ontology of biomedical concepts and their relationships for practical applications.

Furthermore, background knowledge includes domain models and trained corpora that are used to be applied to particular fields like radiology/pathology reports and discharge summaries. Annotated corpora will be labeled by human annotators manually and will be used to train machine/deep linguistic classifiers and to evaluate rule-based systems.

The Metathesarus of UMLS currently comprises over one million biomedical terms and five million concept names have been derived from more than several biomedical controlled vocabularies, such as RxNorm, MeSH, ICD-10, and SNOMED CT.

The UMLS Semantic Network categorizes all UMLS Metathesaurus concepts consistently depending on their semantic types to minimize Metathesaurus complexity. It currently contains 135 main categories and 54 relationships between categories. For example, the "Disease" category

has a relationship "associated with" with the "Finding" category, and the "Hormone" category has a relationship "Affects" with the "Disease" category.

The UMLS SPECIALIST lexicon contains information for biomedical terms on their syntax, morphology, and spelling [143]. Currently, it contains over 200,000 biomedical terms and is used for biomedical NLP tasks by the UMLS lexical tools.

B. TOOLS AND FRAMEWORKS OF THE BIOMEDICAL NLP
1) NLP TOOLS/METHODS
For the construction of NLP tools, there are two main methods/techniques. The first technique is rule-based, mainly focused on rules and dictionary look-up. The second technique is the machine learning method based on annotated corpora to train learning algorithms. The Rule-based approach was often adopted by early systems because their design and implementation was very simple. Currently, many biomedical NLP systems have shifted away from using rule-based methods and depend on machine learning approaches due to their progress and the growing number of annotated corpora, while new annotated training data may have a high cost to generate. Machine learning approaches often deliver better results than rule-based methods, as demonstrated in many challenges of biomedical NLP. At the same time, most recent NLP systems have been designed from integrating rule-based and machine learning methods, which have been called hybrid systems [6].

2) NLP FRAMEWORKS
It is possible to incorporate the framework into the NLP system itself or to use the available common architectures. GATE (General Architecture for Text Engineering) and UIMA (Unstructured Information Management Architecture) are the two most common generalized architectures, which consist of open-source software.

GATE, which was initially developed in 1995 at Sheffield University, is commonly applied in the NLP domain. It contains basic NLP tools for low-level processing (e.g., tokenizers, penetration splitters, and part-speak taggers) packed into a CREOLE wrapper and a high-level processor for named entity recognition packaged into an ANNIE which is an information extraction system. It can incorporate current techniques of NLP and machine learning such as Weka, RASP, SVM Light, and LIBSVM. GATE was used as a basis by many clinical NLP systems, such as HITEx and caTIES, for the extraction of cancer information.

UIMA belongs to the Apache Software Foundation software and it was initially designed since 2006 by IBM. Its objective is to promote the reuse of analytical components and to reduce the duplication of analytical development. The pluggable architecture of UIMA allows you to easily plug-in your analysis components and combines them with others." IBM's 2011 Jeopardy challenge Watson system has developed UIMA's framework, which is recognized as the best-known foundation. The functionality of UIMA is broader

TABLE 4. Major biomedical NLP systems.

Biomedical NLP System	Purpose	Creator	Source of Clinical Dataset	Encoding
LSP-MLP	NLP system for extraction and summarization of signs/symptoms and drug information, and potential drugs and side effects recognition.	New York University	Progress Note. Clinical Note. X-ray Report. Discharge Summary.	SNOMED
MedLEE	A semantically driven system used for: i. Extracting information from clinical narratives reports. ii. Participating in an automated decision-support system. iii. Allowing NLP queries	Columbia University	Radiology. Mammography. Discharge Summary.	UMLS's CUI
MetaMap	A highly configurable program to map biomedical text to UMLS Meta-thesaurus concepts.	National Library of Medicine	Biomedical Text Candidate and Mapping Concepts from UMLS	UMLS's CUI
cTAKES	Mayo clinical Text Analysis and Knowledge Extraction System.	Mayo Clinic and IBM	Discharge Summary. Clinical Note Clinical Named Entities: (diseases/disorders, signs/symptoms, anatomical sites, procedures, medications) Relation, Co-reference Smoking Status Classifier Side Effect Annotator	UMLS's CUI and RxNorm
SPRUS/ SymText/ MPLUS	A semantically driven IE system. NLP system with syntactic and probabilistic semantic analysis driven by Bayesian Networks.	University of Utah	Radiology Concepts from findings in radiology reports.	ICD-9
SPECIALIST	A part of the UMLS project with the SPECIALIST lexicon, semantic network, and UMLS Meta-thesaurus.	National Library of Medicine (NLM)	UMLS	-

than that of GATE, since UIMA can be used to analyze audio/video data in addition to textual data. Several biomedical NLP systems, such as cTAKES, MedKAT/P, and MedEx, use the UIMA framework for cancer-specific characteristics extraction [144], [145] and medication extraction.

This section provides a general overview of the biomedical NLP system architecture by explaining the most significant and influenal NLP systems in the biomedical NLP field. Two of the common systems for extracting UMLS concepts from clinical texts are the Linguistic String Project-Medical Language Processor (LSP-MLP) [146] and the Language Extraction and Encoding System (MedLEE) [147]. The Mayo clinical Analysis and Knowledge Extraction System (cTAKES) [148], Special Purpose Understanding System (SPUS) [149], SymText (Symbolic Text Processor) [150] and SPECIALIST language-processing system [151] are the major systems developed by few dedicated research groups for maintaining the extracted information in the clinical domain. Another important system widely used in the clinical domain is MetaMap [152]. Among all, MetaMap is found to be useful with patients' HER for automatically providing

relevant health information. Table 4 presents the characteristics of the major biomedical NLP systems discussed in this section.

C. THE ENSEMBLE METHODS FOR THE BIOMEDICAL NLP TOOLS

The ensemble approach improves the portability of biomedical NLP systems by combining the strengths of individual tools. An ensemble is a meta-algorithm that incorporates various basic models into a predictive model, and in several machine learning tasks, this combination has demonstrated superior results [153]–[155].

The ensemble approach has been widely applied to various clinical and biomedical issues such as identification of biomarker [156], protein-protein interaction [157], causal molecular networks inference [158] and gene expression based disease diagnosis [159].

Many studies have explored the ensemble of NLP tools for medical concept recognition.

For example, Torii *et al.*, developed BioTagger-GM by combining recognition results from individual systems and

TABLE 5. The classification and the related number of articles reviewed for chronic diseases.

Chronic diseases	Diseases related to chronic diseases	Number of Research Papers
The circulatory system Diseases	Congestive heart disease. Coronary artery disease. Heart disease. Heart failure. Hypertension. Peripheral arterial disease. Pulmonary disease.	38
Neoplasms	Breast cancer. Colorectal cancer. Prostate cancer. Lymphoma.	34

using a voting schema and achieved the best performance in the BioCreAtIvE II challenge to recognize gene/protein names from literature [160], [161].

Doan *et al.* demonstrated that the ensemble classification results which incorporate single classification models into a voting system could perform better than a single classification model in identifying medical information from clinical text using the 2009 i2b2 (Informatics for Integrating Biology and the Bedside) challenge datasets [162].

Kang *et al.*, merged two dictionaries-based systems with five statistical systems into a simple voting scheme and achieved a third-place finish in the 2010 i2b2/VA challenge to extract medical problems, examinations and medications [163].

Kuo *et al.* combined cTAKES and MetaMap to develop an ensemble pipeline that improved the efficiency of NLP tools in extracting clinical data terms, but with high variation depending on the cohort [164].

VI. LITERATURE REVIEW AND RELATED WORKS

Throughout this section, we discuss some articles and surveys that constitute our literature review, including a list of all related works for applying machine learning to biomedical NLP, especially on chronic diseases.

Diseases Classification: About 106 studies have been analyzed and were mainly linked to 43 specific chronic diseases. One objective was to clarify the application of NLP and its related clinical notes for particular types of conditions. Therefore, using the International Classification of Diseases, 10th Revision (ICD-10) [19], the 43 specific chronic diseases were then classified into ten types of diseases, as shown in Table 5. Figure 9 shows the number of EHR-related publications on chronic diseases per year.

A. DISEASES OF THE CIRCULATORY SYSTEM

A) **Cardiovascular Diseases** Heart disease is one of the major death causes, while prediction and prevention have recently developed. The identification of risk factors is a necessary first step in predicting and preventing heart disease. Many studies have been proposed to
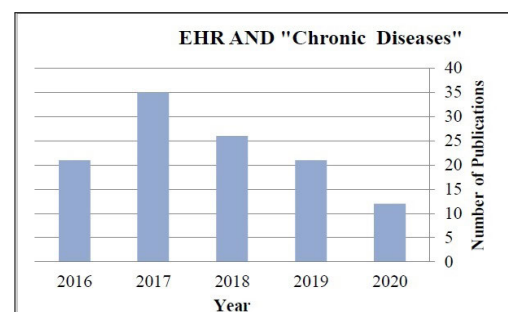
FIGURE 9. The number of EHR-related publications per year on chronic diseases.

determine heart disease-related risk factors, but no one has tried to identify all risk factors. A challenge for biomedical NLP, in 2014, was released by the National Center for Computer Science for Integrating Biology and Beside (i2b2) that involved a track (track 2) for determining risk factors of heart disease in the clinical texts over time. The purpose of this track was to classify information on cardiovascular risks, as well as to monitor the quality of the historical medical records. It was important to classify tags and characteristics associated with the existence and development of the disease, risk factors, and medications in inpatient medical history. Table 6 summarizes the number of papers related to diseases of the circulatory system.

B) **Peripheral and coronary arterial disease** Millions of people worldwide were affected by Peripheral arterial disease (PAD), which is a type of chronic disease. For automated determination of PAD status using predetermined criteria in clinical reports, the NLP algorithm should be used as a determining PAD status from clinical notes, which is labor-intensive and time-consuming by manual chart review. Many researchers have used NLP to identify peripheral arterial disease (PAD) cases and critical limb ischemia in clinical records. It is also used by recent genome-wide

TABLE 6. Summary of the studies that use NLP methods for cardiovascular diseases.

Study	Objectives and Contributions	Methods	Dataset	Results	Weaknesses
Chen et al., [98]	Identification of heart disease risk information.	Hybrid pipeline system based on both: machine learning-based rule-based approaches using: SVM, libshortText, and CRFsuite.	The i2b2/UTHealth Challenge.	F1-score of 92.68%	The proposed system did not perform very well for coronary artery disease (CAD), obesity status, and smoking status.
Torri et al., [91]	Detection of the risk factor for cardiac disease.	A hybrid of several ML and rule-based techniques.	The 2014 i2b2/UTHealth Challenge.	F1 score of 91.85% Recall of 94.09% Precision of 89.72%	The proposed system was not feasible to obtain objective evaluation metrics on the training set, and the current evaluation results were solely based on the one test set.
Karystianis et al., [165]	Extraction of the risk factors for detecting cardiac disease.	Knowledge-driven and the system implement local lexicalized rules.	The 2014 i2b2/UTHealth Challenge.	F1 score of 88% Recall of 90% Precision of 86%	The proposed system achieved lower performance with CAD which proved to be the most challenging class to recognize (F-score of 73.63%).
Yang et al., [88]	Extraction of the identification of heart disease risk factors.	Machine learning, Rule-based methods, Dictionary-based keyword.	The 2014 i2b2/UT Health NLP Challenge.	F-measure of 91.5%	Poorest classification accuracy is obtained for CAD (0.787). The main reason for that is due to the difficulties in the identification of sentence-level CAD clinical facts, event, test, and symptom.
Kogan et al., [166]	Assessment of stroke severity.	Several machine learning models.	The Optum© de-identified Integrated Claims-Clinical dataset.	R2 (coefficient of determination) of 0.57. R (Pearson correlation coefficient) of 0.76. A root-mean-square error of 4.5.	The current EHR database has information which could be critical for model performance including imaging of brain scans was not available. As with all studies based on real-world data, there is the potential for missing records. Healthcare information in the database was not available until January 2007.
Garg et al., [167]	Automatic classification of the Ischemic Stroke subtype.	Machine Learning and NLP algorithms	The Northwestern Enterprise Data Warehouse (EDW).	Kappa of .25	The proposed method relies on the level of documentation and detail in the EHR. The proposed system did not include the entire EHR (e.g., cardiac imaging, laboratory, procedures). The proposed system did not include CT-based radiology reports to reduce variability in the dataset.
Kim et al., [168]	Identifying AIS patients by automatically classifying brain MRI reports.	Supervised ML-based NLP algorithms.	All brain MRI reports from a single academic institution.	F1-measure of 93.2% Accuracy of 98.0%	The proposed system used text corpus which was created at a single institution. The proposed system only included brain MRI reports with conventional stroke MRI sequence.
Grechishcheva et al., [169]	Developed a study of risk markers identification.	NLP algorithms.	Almazov National Medical Research Center.	high accuracy.	One of the weak sides of the algorithm is its speed. For current corpus of data, it took 6,100 seconds to remove marginal parts of speech, short words and form a normal form for left ones.

PAD research to identify medications, diseases, signs/symptoms, anatomical locations, and procedures. Table 7 summarizes the number of papers related to peripheral and coronary arterial disease.

C) **Hypertension** One of the main health problems is hypertension (HTN) and high blood pressure (HBP) diseases. It is estimated that by 2025, adults with hypertension will increase by 60 One of the major risks for cardiovascular and kidney diseases is HTN. Any

HTN-relevant patient knowledge has significant application in cohort discovery and the development of predictive prevention and monitoring models. Most of this important medical knowledge typically takes the form of non-structured clinical records distributed over multiple EHR systems. Extracting patient-relevant information from unstructured clinical notes usually takes a lot of resources and consumes time. In particular, manual extracting of HTN information is a significant

TABLE 7. Summary of the studies that use NLP methods for peripheral and coronary arterial disease.

Study	Objectives and Contributions	Methods	Dataset	Results	Weaknesses
Afzal et al., [63]	Identification of critical limb ischemia using NLP.	NLP algorithm for PAD identification.	The Mayo clinical data warehouse.	F1-score CLI-NLP of 90%	The proposed algorithm used data for a PAD cohort from a single medical center and future studies should apply and validate this algorithm to other institutions to make the findings generalizable.
Afzal et al., [170]	Identifying PAD cases from narrative clinical notes.	NLP Algorithm.	The Mayo clinical data warehouse.	accuracy of 91.8%	A limitation of this study is that data were retrieved from the data warehouse of a single academic medical center.
Leeper et al., [64]	Applying Text-Mining to clinical notes for Profiling the Cilostazol Safety.	NLP Algorithm.	The Stanford Translational Research Integrated Database Environment (STRIDE).	No association between the use of Cilostazol and any serious cardiovascular adverse event including stroke is observed.	A limitation of this study is that it could have missed comorbidities due to false negatives from lower sensitivity (73%). The outcome measures may not have captured events occurring outside of the hospital or that led to hospitalizations in other institutions.
Buchan et al., [89]	Automatic prediction of coronary artery disease from clinical narratives.	Naive Bayes, MaxEnt, and SVM	The 2014 i2b2 Heart Disease Risk Factors Challenge data set.	F1 score of 77.4%	One limitation of this experiment is embedded in the selection of patients who do not develop CAD. Another limitation of proposed study is embedded in the subsample of patients who develop CAD. The relatively small size of the dataset is a limitation of proposed study.

TABLE 8. Summary of the studies that use NLP methods for hypertension.

Study	Objectives and Contributions	Methods	Dataset	Results	Weaknesses
Jonnagaddala et al., [171]	Extraction of hypertension information from unstructured clinical reports.	HTN-System	The 2014 i2b2/UTHealth Shared-Task 2.	Recall of 87. 70%. Precision of 78. 63%. F-measure of 82.92%.	The HTN-System had a good overall performance on the corpus, but the results may vary depending on the corpus. The performance of the HTN-System is very similar to the performance of other systems.
Boytcheva et al., [172]	Analysis of clinical narratives to extract numerical blood pressure values.	A hybrid model for generating rules automatically for IE from clinical data.	Bulgarian NHIF repository of nameless outpatient records in XML format.	Precision of 92% Recall of 98%	some limitations in the access to healthcare documentation, administrative and geographical scalability is not applicable. limitations of using small training excerpt.
Teixeira et al., [102]	Identification of hypertensive individuals.	Natural Language Processing (NLP) Algorithm.	The Vanderbilt University Medical Center EHR.	Random forests using billing codes, medications, vitals, and concepts had the best performance with a median area under the receiver operator characteristic curve (AUC) of 0.976.	The proposed system evaluated the portability at only a single additional site. Other institutions may differ from both Vanderbilt and Marshfield Clinic. Proposed algorithm also did not detect the date of onset of hypertension, which could be clinically interesting in a number of circumstances.

issue, which is time-consuming as HTN information is usually reported in multiple records for one patient. Another important issue besides the manual extraction is coding HTN information to standard ontologies like SNOMED-CT. There are simple mining techniques of clinical texts that can be applied to extracting HTN information from unstructured clinical reports. Table 8 summarizes the number of papers related to hypertension disease.

D) **Heart failure identification** Heart failure is a chronic disease usually caused by some deficiency in structure or function. The quick and accurate prediction of heart failure mortality is important for improving patient health care and preventing death. But, due to the weak feature representation of heart failure data, prediction of death caused by heart failure is a significant challenge using simple models. Table 9 summarizes the number of papers related to heart failure disease.

B. NEOPLASMS

EHR provides important cancer-related knowledge which can be valuable for biomedical research because extracting and structuring this knowledge is provided by NLP methods. This section discusses many studies related to cancer, such as the identification of multiple cancer types, the extraction of tumor characteristics and tumor-related information, cancer patient trajectories, cancer recurrence, and cancer

TABLE 9. Summary of the studies that use NLP methods for heart failure identification.

Study	Objectives and Contributions	Methods	Dataset	Results	Weaknesses
Liu et al., [173]	Heart Failure Readmission prediction from Clinical Notes.	Convolutional Neural Networks (CNN).	The MIMIC III database.	F1 score of 75.6%	This study did not test different word embeddings trained from clinical notes vs PubMed.
Wang et al., [174]	Prediction of Heart Failure Mortality using feature rearrangement based on deep learning technique.	Deep Learning System.	Shanghai Shuguang HF (SSHF) data.	the proposed framework achieves the best performance for Heart Failure mortality prediction. Extensive experimental results compared with other machine learning methods demonstrate that the proposed method has the highest average accuracy.	The performance of the FRDLS (feature rearrangement based deep learning system) for heart failure mortality prediction relies on two hyper-parameters which require some time to adjust. The prediction label in this study is binary to "death" or "alive", making the complex HF problem simplified.
Topaz et al., [175]	Identify patients suffering from heart failure with ineffective self-management status.	NLP Algorithm.	The Partners Healthcare System EHR data.	F-measure of 86.3% Precision of 95% Recall of79.2%	The performance of proposed system was evaluated with only one type of note—discharge notes—in one healthcare system. So the ability to assess the system's performance for different HF self-management domains was limited because of the small number of positive cases in the testing data.
Garvin et al., [176]	Automation of heart failure quality measures using NLP.	NLP System.	The VA's Corporate Data Warehouse (CDW).	Sensitivity (SN) of 98.9% Positive predictive value of 98.7%	First, it is likely that some clinical information was not documented in the patient charts and therefore could not be captured by the NLP system. Second, although the CHIEF (Congestive Heart Failure Information Extraction Framework) system performed well using VA text notes, it might not perform as well in non-VA settings. Third, documents from only eight medical centers were used in this research; therefore, the CHIEF might under-perform initially when used with documents from other VA medical centers.
Vijayakrishnan et al., [65]	Recognition of signs/ symptoms of HF in EHRs.	EMR Data Extraction.	The Geisinger Health System HER data.	A total of 892,805 affirmed criteria were documented over an average observation period of 3.4 years. Among eventual HF cases, 85% had $1 criterion within 1 year before their HF diagnosis, as did 55% of control subjects.	First, unlike the Framingham investigators, the proposed application was unable to accurately account for other, non-cardiac, causes for a patient to have experienced minor signs and symptoms. The variability in the documentation of HF signs and symptoms by various clinicians.

stage identification. Table 10 summarizes the number of papers related to neoplasms disease.

Regarding breast neoplasm, a significant data source for epidemiologic research is the EHR system. In studies related to population, structured EHR data such as diagnosis and procedure codes are usually used. They do not accurately capture some conditions such as breast cancer recurrence that is only recorded in unstructured clinical reports. A typical method for extracting information from EHR data is manual processing, which consumes time, costly and causes inherent privacy risks, restricting the amount of available information for the study. NLP methods can be used to solve this issue by processing unstructured texts and they were used as an alternative or a supplement to manual chart abstraction. NLP has been successfully applied to several biomedical applications such as analyzing results from imaging and pathology reports, recognizing persons based on cancer examinations,

selecting clinical trials, detecting postoperative surgical complications, and performing pharmacogenomics and translational research. NLP has also demonstrated recent progress in identifying breast and prostate malignancies recorded in pathology reports. NLP-based algorithms, in some cases, do as the same as manual processing, or even better. Table 11 summarizes the number of papers related to breast cancer disease.

VII. OPEN ISSUES AND CHALLENGES
One of the primary healthcare issues is broadly acknowledged as the risk of chronic diseases such as cancers, diabetes, and hypertension. Although considerable development has been achieved in discovering new therapies and prevention methods, it remains a challenge, and the magnitude of this challenge is growing, having a significant effect on the quality of life and the cost of healthcare. Therefore, effective

TABLE 10. Summary of the studies that use NLP methods for neoplasms disease.

Study	Objectives and Contributions	Methods	Dataset	Results	Weaknesses
Osborne et al., [177]	Effective identification of reportable cases of cancer nationally authorized.	Natural Language Processing Machine learning.	CRCP-DUAL dataset.	Accuracy of 0.872 Precision of 0.843 Recall of 0.848.	CRCP-NLP (The Cancer Registry Control Panel) was compared only to the original manual review process, due to resource constraints. CRCP-NLP rely on proxy data, such as monthly state case submissions, to provide an estimate. Unfortunately, yearly changes in coding practices as well as multiple changes in registrar personnel and leadership make comparisons difficult. Finally, CRCP uses rules for document segmentation generated solely from UAB training data. Thus, the reliability of the used set of regular expressions is expected to be reduced at other institutions to the extent that provider and sectioning practices differ.
AAlAbdulsalam at el.[178]	TNM stage mentions are extracted and classified from the Utah Cancer Registry records automatically.	A hybrid model integrating pattern-matching for extraction and supervised machine learning for classification.	Utah Cancer Registry Data.	Extraction accuracy: 95.5%–98.4% classification accuracy: 83.5%–87%.	Although regular expressions were more robust for extracting TNM (tumor characteristics (T), lymph node involvement (N), and tumor metastasis (M)) mentions, the used range of features used with the CRF classifier were still limited and potential improvements may be observed if other more sophisticated feature patterns were used such as character N-grams. In addition, other machine learning algorithms could yield better performance than CRF(The Conditional Random Fields) and further investigation is required.
Si Y. et al., [41]	Frame-based NLP system based on a Bi-LSTM-CRF neural network to extract cancer-related information in clinical narratives.	Deep Learning-based Approach.	MIMIC-III.	F1 between 87.18% and 96.33% Element classification F1 up to 93.02%.	A limitation of this study is that its evaluation was limited to the gold standard for each process in the pipeline instead of a multi-step evaluation and optimizing the pipeline.
Datta.[179]	This study presented a literature review of biomedical NLP related to cancer.	NLP Algorithm.	–	The most common scope of the papers (36 of 78) was cancer diagnostics, with recent work on the extraction of information on treatment and the diagnosis of breast cancer.	The first limitation of this review is that, given the rapid pace of NLP development and publication, it is likely that papers meeting inclusion criteria were omitted. Specifically, papers published starting in late 2018 were likely missed. Second, although the proposed review have tried to accurately interpret and represent all the possible information types extracted in the literature as frames, there might be a few which are not captured or mis-represented in proposed final frame list. Finally, there may be inconsistencies in assigning the frames and associated elements across all the papers.

strategies and methodologies are required to supplement and expand beyond existing evidence-based therapies, which can mitigate the severity of chronic conditions. The secondary use of EHRs for processing patient data, promoting medical research, and enhancing the clinical decision making is a promising path. Methods based on EHR processing and modeling contribute to a better understanding of patient clinical trajectories and improving stratification of the patient and risk prediction. Effective extraction of unknown clinical knowledge is provided by using machine learning and especially deep learning for processing EHRs. The longitudinal structure of chronic diseases provides a broad continuous stream of data that can identify useful clinical trends and direct clinical decisions in a way that delays or avoids the onset of the disease.

Because of the various difficulties involved in the production of clinical reports, progress in NLP research in the biomedical domain is sluggish and lagging relative to progress in general NLP. The main reasons for the challenges to the development of biomedical NLP are that the access to shared data is very difficult, the annotated datasets that can be used for training and benchmarking are insufficient, the annotation agreements and standards are inadequate, reproducibility is formidable, partnerships are restricted, and user-centered development and scalability are missing. The i2b2/VA Challenge shared tasks, tackle these obstacles by providing participants with annotated datasets for potential solutions.

The development of biomedical NLP has several issues and challenges that faced the process of clinical notes for chronic disease detection. It is worth noting that these challenges remain until now as presented in [53], [106]:

1) **Domain knowledge:** Adequate knowledge of the domain is the most important requirement for an NLP

TABLE 11. Summary of the studies that use NLP methods for breast cancer disease.

Study	Objectives and Contributions	Methods	Dataset	Results	Weaknesses
Carrell et al., [180]	Identification of breast cancer recurrences using the NLP system form clinical text.	The recurrent detection system for breast cancer based on NLP.	(COMBO) Study Commonly Used Medications and Breast Cancer Recurrence.	Reduce the number of charts that require manual analysis by 90%.	First, the used NLP modules may require adaptation to accommodate language usage and document sectioning in other institutional settings. Second, NLP development costs limit its applicability to large or repeated tasks where it is cost effective relative to 100% manual abstraction. Third, NLP requires access to machine-readable clinical text and does not work with print documents or their scanned copies. Fourth, proposed study cohort was limited to women with early stage (I or II) breast cancers; the algorithm has not been tested for recurrence in women with initial late stage disease or ductal carcinoma in situ (stage 0 Fifth, reference standard corrections were limited to the review of charts where NLP and the reference standard were discordant.
Castro et al., [90]	The Breast Imaging-and Data System (BRI-RADS) Automated Categories of Breast Radiology Extraction is discussed.	Rule-based NLP algorithm. Supervised machine learning approach.	Text Information Extraction System (TIES) of Pittsburgh University.	F-1 measure of 0.95 F-1 measure of 0.91 for BIRADS 0-6 0.93 for BIRADS 3-5.	An important potential limitation of proposed methods was the using of the BROK software to sample by BI-RADS category during the development of the corpus used for development of the used machine-learning base annotator. For the BI-RADS token annotator, the generalizability of proposed findings may also be limited because the detection step of used annotator strongly relies on the accurate performance of the pre-processing steps.
Miller et al., [181]	tested a generalized a co-reference resolution tool in clinical texts tested in both domains of colon and breast cancer.	Mention-pair coreference system that operates over pairs of mentions with best-first conflict resolution	The THYME colon cancer dataset. The DeepPhe breast cancer dataset.	The mention-synchronous system performs similarly on in-domain data but performs much better on new data.	There is still a performance gap when testing in new domains.
Mykowiecka et al., [182]	An IE system based on the rules tested on mammography records.	IE techniques based on natural language.	Warsaw health care institutions database.	Precision and Recall were above 80%	A known limitation of rule-based systems is the necessity of foreseeing all possible ways of expressing the information to be extracted and lower precision.
Bozkurt et al., [183]	Recognize lesions in free-text mammography reports.	NLP information extraction Method.	Set of 500 hospital-report mammography reports.	Precision of 94.9% Recall of 90.9% F measure of 92.8%	First, since the bulk of the text is automatically generated by the reporting system, the variability in narrative reporting style was not represented, which could make the proposed results seem better than when applied to reports not generated using the proposed system. Another limitation is that if the radiologist edits the text report, corresponding changes are not made in the structured database. Another limitation is that although the regular expressions that used in the proposed system seemed to work well for detecting BI-RADS descriptors, this approach is specific to BI-RADS and might not work as well if radiologists use other types of terminologies in reporting mammograms.

researcher involved in the development of systems and methodologies for processing biomedical records. The primary importance of domain knowledge stems from the fact that the output of the system is made available for application in healthcare. Thus, the system is always required to have sufficient recall, accuracy, and F-measurement for the intended biomedical application, with the necessary performance modification. Interestingly, it is possible to apply the NLP techniques capturing the domain knowledge available in the free text. The NLP approach for the automated capture of ontology-related domain knowledge, for example, uses a two-phase methodology to extract terms from the

linguistic representations of concepts in the initial phase followed by the extraction of semantic relations.

2) **Confidentiality of the biomedical text:** A sample of training data is required to develop and evaluate an NLP system. The training dataset is a vast array of electronic patient records in textual formats in a clinical context. The privacy of patient data is protected by The Health Insurance Portability and Accountability Act (HIPAA) in the United States. It is necessary to de-identify personal information to make the records accessible for research purposes. However, automated recognition of details such as names, addresses, telephone numbers, etc., is a highly challenging task, which

often needs manual review. Eighteen personal information identifiers, i.e., the identification of protected health information (PHI) in the clinical report, which should be excluded as required by HIPAA, is very complicated and time-consuming. In 2006, the challenge of i2b2 de-identification took the most significant effort to develop and evaluate automated de-identification tasks. The available approaches to de-identification include (1) rule-based methods that use dictionaries and manually crafted rules to match PHI patterns, (2) machine learning methods that learn to identify PHI patterns based on training datasets automatically, and (3) hybrid methods, which combine both techniques.

3) **Abbreviations:** The clinical text will include several medical abbreviations. The abbreviations are often readily interpreted by healthcare experts due to their domain knowledge. Nevertheless, when a clinical NLP system tries to derive clinical information from the free text, abbreviations are found to be extremely ambiguous. For example, the clinical text abbreviation PT could mean a patient, prothrombin, physical therapy, and so on. The correct interpretation of clinical abbreviations is often challenging and involves two major tasks: detecting abbreviations and choosing the proper expanded forms. Dictionary lookup and morphology-based matching are the most widely used methods for detecting abbreviations in the clinical domain, and machine-learning approaches are used to select the right extended type. Researchers have contributed several methods to identify abbreviations present in clinical texts, construct a knowledge base for clinical abbreviations, and disambiguate ambiguous abbreviations, in addition to developing clinical NLP systems such as MedLEE, MetaMap, etc. to extract medical concepts and associated abbreviations from clinical texts.

4) **Diverse formats:** There is no standardized format for the biomedical text, especially with patients' medical reports: (1) the clinical text often contains the information in a free-text format like a pseudo table, i.e., text intentionally made to appear as a table. Although the contents of the pseudo table are easy to interpret by a human, for a general NLP system, the identification of the formatting features is complicated. (2) Although the importance of report sections and subsections relevant to many applications, the section headers are either ignored or combined with similar headers on many occasions. (3) Another issue often found in the clinical text is the missing or incorrect punctuation, e.g., to indicate the end of a sentence, a new line can be used instead of a point. The Clinical Text Architecture (CTA), which tries to define the criteria for the clinical report structure, effectively addresses the issue of different formats of the clinical text.

5) **Expressiveness:** The biomedical domain language is hugely expressive. There are many ways to describe the same medical concept, e.g., cancer can be expressed as a tumor, lesion, mass, carcinoma, metastasis, neoplasm, etc. Likewise, the modifiers of the concept can also be described with many different terms, e.g., the modifiers for certainty information would match more than 800 MedLEE lexicons, thus making the retrieval process more complicated.

6) **Intra- and interoperability:** A biomedical NLP system is expected to work well in various healthcare and biomedical applications and to be easily integrated into a biomedical information system. In other words, the system needs to handle a biomedical text in different formats. For example, the formats of discharge summaries, diagnostic reports, and radiology reports are different. The output of the NLP system can also be stored in the clinical database. However, it is almost unlikely to map the same to the clinical database scheme because of the complexity and nested relationships of the output. Additionally, the output from the NLP system must be available for comparison for a variety of automated applications through widespread deployment across the institutions. To achieve this, the output must be mapped onto a standardized vocabulary system such as UMLS, ICD-10, and SNOMED-CT, and onto a standard domain representation. Finally, it is considered essential to interpret the biomedical information and the relationships between concepts to construct a representational model. For example, "treats" is one of the relationships between a drug and a disease.

7) **Interpreting information:** Interpretation of clinical information available in a report requires the knowledge of the report structure and additional medical knowledge to associate the findings with possible diagnoses. The complexity of interpreting information depends on the type of report and section, e.g., it is easier to obtain information on the vaccination being administered than to get information from a radiological report containing patterns of lights (patchy opacity). An NLP system that interprets light patterns to specific diseases should contain medical knowledge related to the findings.

Despite the recent advances and developments, these recent limitations have affected the use of NLP technology [106]:

1) **The availability, consistency and characteristics of the training data** The availability, consistency and characteristics of the training data are very essential for building NLP models [184]. For the training and implementation of an effective NLP models, the access and availability of appropriately annotated datasets are very important. For example, the designing of NLP algorithms that can perform a systematic synthesis of published research on a specific topic or an analysis and data extraction from EHR needs unrestricted access to databases of publisher or primary care/hospital. While the number of biomedical datasets and pre-trained models that are publicly available has increased over recent

years, the availability of public health concepts is still restricted [185].

2) **The ability to de-bias data** The ability to de-bias data which means the ability to inspect, explain and ethically modify data is an important issue for training and using NLP models in healthcare domain. If data biases are not taken into consideration in the development (e.g. data annotation), deployment (e.g. use of pre-trained platforms) and evaluation of NLP models, the results of NLP models can be compromised [186]. However, it should be noted that this does not guarantee the same effect across morally appropriate levels, even if datasets and assessments are modified for biases. For example, it must take into account particular age group and socioeconomic groups that use social media sites when using the health data available. A Facebook-trained monitoring system could be biased towards health data and linguistic issues unique to people older than that in Snapchat's data [187]. Recently, several agnostic model tools have been developed to assess and correct injustices in machine learning models in accordance with the efforts of the government and academic communities to identify unacceptable development of AI [188]–[192].

3) **The limited access to dataset** Recently, the limited access to data is a major issue that barriers the progress of NLP system in healthcare domain [6], [19]. Health data are generally regulated regionally in Canada, and there is reluctance to provide access to these systems and incorporation with other datasets without restrictions due to security and confidentiality issues (e.g. data linkage). Public understanding of the privacy and data access has also caused critical issues. A new study of social media users revealed that most people found analyzing their social media data in order to find ''intrusive and exposing'' problems of mental health is not accepted [193]. Before key public health NLP activities such as the real-time analysis of national disease patterns can be carried out, jurisdictions must collectively identify a reasonable scope and access to data sources of public health (e.g. EHR and administrative data). Future NLP applications which analyzing personal EHR rely on their ability to incorporate varying privacy in models, both during and after training to avoid breaches of privacy and data misuse [194]. The current methods for accessing full text publications often restrict access to essential data. Total automation and synthesis of PICO-specific information requires unlimited access to journal databases or new data storage modelling [195]. The available clinical datasets are MIMIC-II, the Informatics for Integrating Biology and the Bedside (i2b2) datasets, PhenoCHF, Temporal Histories of Your Medical Event (THYME), and Cancer Deep Phenotype Extraction (DeepPhe).

4) **The assessment and evaluation of NLP models** Finally, as with any emerging technology, validation and evaluation of NLP models should be taken into account to ensure that they operate as expected and keep up with the changing ethical views of society. These NLP technology must be tested to ensure it performs as intended and to take bias into account [196]. Although many methods today publish equal or better-than-human scores on tasks of textual analysis, it is important not to equate high scores with a real understanding of language. But it is also important not to consider the lack of true understanding of the language as an inefficiency. Models with a ''relatively poor'' depth of understanding can still be highly effective at information extraction, classification and prediction tasks, particularly with the increasing availability of labelled data.

VIII. CONCLUSION AND FUTURE RESEARCH ISSUES

We have discussed in this review paper, an overview of NLP in general and NLP in biomedicine and healthcare with its methods and technologies and its potential tasks and use-cases in the biomedical and healthcare domains. Then we have presented the application areas of machine learning/deep learning in the biomedical NLP. We have provided an overview of the most popular biomedical NLP systems and their general architecture. Next, we have discussed a literature review of the application of various NLP techniques to narrative clinical notes on chronic diseases, including the analysis of difficulties faced by NLP methodologies in clinical narrative comprehension. Finally, we conclude this review paper by describing existing challenges currently faced and open issues associated with the processing of the biomedical and clinical text and providing the NLP domain with sufficient resources and opportunities to extract new methodologies.

In this review paper, we have discussed essential challenges such as domain knowledge, the confidentiality of clinical texts, abbreviations, diverse formats, expressiveness, intra-operability and interoperability, and information interpreting. These discussions provide an opportunity to understand the complexity of the clinical text processing and various approaches available. An important area of research related to the understanding of the challenges involved in processing the clinical text is the development of methodologies for processing the diverse format of clinical texts. Each format, on its own, is a challenge for NLP researchers and can be explored using traditional and hybrid methodologies. Our review has shown that biomedical NLP methods need to be modified and updated beyond the extraction of clinical terms to concentrate more on the interpretation of concepts (i.e., not only understanding of relationships between concepts but also combining the clinical data, domain knowledge, and general knowledge in the reasoning process).

In conclusion, NLP provides a powerful methods for unlocking information about chronic diseases from unstructured clinical narratives. Despite of developing new standards and better encoding EHR with clinical terminology standards, there is still a narrative aspect, which makes the biomedical NLP methods essential for clinical research informatics.

There have also been widespread application of different techniques and models to biomedical literature and all of these NLP techniques are important and can be applied to effectively mining EHRs to support essential clinical research activities. New deep learning techniques have contributed with a significant progress across various tasks and will be increasingly adopted to analysis big data of EHRs effectively and efficiently, further advancing disease management, quality improvement, and all aspects of clinical research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies with human participants or animals performed by any of the authors.

CREDIT AUTHOR STATEMENT

Essam H. Houssein: Conceptualization, Methodology, Formal analysis, Writing - review & editing. Rehab Emad: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing. Abdelmgeid A. Ali: Supervision, Writing - review & editing. All authors read and approved the final article.

REFERENCES

[1] S. A. Hasan and O. Farri, "Clinical natural language processing with deep learning," in *Data Science for Healthcare*. Springer, 2019, pp. 147–171.

[2] G. K. Savova, K. C. Kipper-Schuler, J. F. Hurdle, and S. M. Meystre, "Extracting information from textual documents in the electronic health record: A review of recent research," *Yearbook Med. Informat.*, vol. 17, no. 1, pp. 128–144, 2008.

[3] D. Demner-Fushman, W. W. Chapman, and C. J. McDonald, "What can natural language processing do for clinical decision support?" *J. Biomed. Inform.*, vol. 42, no. 5, pp. 760–772, 2009.

[4] W.-Q. Wei, P. L. Teixeira, H. Mo, R. M. Cronin, J. L. Warner, and J. C. Denny, "Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. e1, pp. e20–e27, Apr. 2016.

[5] W. contributors. (2020). *Natural Language Processing—Wikipedia, the Free Encyclopedia*. Accessed: Oct. 4, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Natural_language_processing

[6] C. Friedman, T. C. Rindflesch, and M. Corn, "Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine," *J. Biomed. Inform.*, vol. 46, no. 5, pp. 765–773, 2013.

[7] J. Henry, Y. Pylypchuk, T. Searcy, and V. Patel, "Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2015," *ONC Data Brief*, vol. 35, pp. 1–9, May 2016.

[8] L. A. Knake, M. Ahuja, E. L. McDonald, K. K. Ryckman, N. Weathers, T. Burstain, J. M. Dagle, J. C. Murray, and P. Nadkarni, "Quality of EHR data extractions for studies of preterm birth in a tertiary care center: Guidelines for obtaining reliable data," *BMC Pediatrics*, vol. 16, no. 1, p. 59, Dec. 2016.

[9] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[10] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[11] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology," *Mol. Syst. Biol.*, vol. 12, no. 7, p. 878, Jul. 2016.

[12] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, "Secondary use of EHR: Data quality issues and informatics opportunities," *Summit Transl. Bioinf.*, vol. 2010, p. 1, Oct. 2010.

[13] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: Towards better research applications and clinical care," *Nature Rev. Genet.*, vol. 13, no. 6, pp. 395–405, 2012.

[14] M. Jiang, Y. Chen, M. Liu, S. T. Rosenbloom, S. Mani, J. C. Denny, and H. Xu, "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 601–606, Apr. 2011.

[15] S. Ebadollahi, J. Sun, D. Gotz, J. Hu, D. Sow, and C. Neti, "Predicting patient's trajectory of physiological data using temporal trends in similar patients: A system for near-term prognostics," in *Proc. AMIA Annu. Symp.*, 2010, p. 192.

[16] D. Zhao and C. Weng, "Combining PubMed knowledge and EHR data to develop a weighted Bayesian network for pancreatic cancer prediction," *J. Biomed. Informat.*, vol. 44, no. 5, pp. 859–868, Oct. 2011.

[17] P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, and D. S. Lee, "Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes," *J. Clin. Epidemiol.*, vol. 66, no. 4, pp. 398–407, 2013.

[18] G. J. Kuperman, A. Bobb, T. H. Payne, A. J. Avery, T. K. Gandhi, G. Burns, D. C. Classen, and D. W. Bates, "Medication-related clinical decision support in computerized provider order entry systems: A review," *J. Amer. Med. Inform. Assoc.*, vol. 14, no. 1, pp. 29–40, 2007.

[19] S. Sheikhalishahi, R. Miotto, J. T. Dudley, A. Lavelli, F. Rinaldi, and V. Osmani, "Natural language processing of clinical notes on chronic diseases: Systematic review," *JMIR Med. Informat.*, vol. 7, no. 2, Apr. 2019, Art. no. e12239.

[20] K. Jensen, C. Soguero-Ruiz, K. Oyvind Mikalsen, R.-O. Lindsetmo, I. Kouskoumvekaki, M. Girolami, S. Olav Skrovseth, and K. M. Augestad, "Analysis of free text in electronic health records for identification of cancer patient trajectories," *Sci. Rep.*, vol. 7, no. 1, p. 46226, May 2017.

[21] L. L. Popejoy, M. A. Khalilia, M. Popescu, C. Galambos, V. Lyons, M. Rantz, L. Hicks, and F. Stetzer, "Quantifying care coordination using natural language processing and domain-specific ontology," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. e1, pp. e93–e103, Apr. 2015.

[22] H. Yang, I. Spasic, J. A. Keane, and G. Nenadic, "A text mining approach to the prediction of disease status from clinical discharge summaries," *J. Amer. Med. Inform. Assoc.*, vol. 16, no. 4, pp. 596–600, Jul. 2009.

[23] R. W. V. Flynn, T. M. Macdonald, N. Schembri, G. D. Murray, and A. S. F. Doney, "Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes," *Pharmacoepidemiol. Drug Saf.*, vol. 19, no. 8, pp. 843–847, Aug. 2010.

[24] Y. Chen, H. Cao, Q. Mei, K. Zheng, and H. Xu, "Applying active learning to supervised word sense disambiguation in MEDLINE," *J. Amer. Med. Inform. Assoc.*, vol. 20, no. 5, pp. 1001–1006, Sep. 2013.

[25] H. Liu, "A multi-aspect comparison study of supervised word sense disambiguation," *J. Amer. Med. Inform. Assoc.*, vol. 11, no. 4, pp. 320–331, Apr. 2004.

[26] M. J. Schuemie, J. A. Kors, and B. Mons, "Word sense disambiguation in the biomedical domain: An overview," *J. Comput. Biol.*, vol. 12, no. 5, pp. 554–565, Jun. 2005.

[27] H. Xu, M. Markatou, R. Dimova, H. Liu, and C. Friedman, "Machine learning and word sense disambiguation in the biomedical domain: Design and evaluation issues," *BMC Bioinf.*, vol. 7, no. 1, pp. 1–16, Dec. 2006.

[28] Q. Dong and Y. Wang, "Enhancing medical word sense inventories using word sense induction: A preliminary study," in *Heterogeneous Data Management, Polystores, and Analytics for Healthcare*. Springer, 2020, pp. 151–167.

[29] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, and G. Wang, "Data processing and text mining technologies on electronic medical records: A review," *J. Healthcare Eng.*, vol. 2018, Apr. 2018, Art. no. 4302425.

[30] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "A brief survey of text mining: Classification, clustering and extraction techniques," 2017, *arXiv:1707.02919*. [Online]. Available: http://arxiv.org/abs/1707.02919

[31] A. Śniegula, A. Poniszewska-Marańda, and L. Chomątek, "Towards the named entity recognition methods in biomedical field," in *Proc. Int. Conf. Current Trends Theory Pract. Inform.* Springer, 2020, pp. 375–387.

[32] L. T. Kohn, J. M. Corrigan, and M. S. Donaldson, "Institute of medicine (US) committee on quality of health care in America," in *To Err Is Human: Building a Safer Health System*. Washington, DC, USA: National Academies, 2000.

[33] J. A. Casey, B. S. Schwartz, W. F. Stewart, and N. E. Adler, "Using electronic health records for population health research: A review of methods and applications," *Annu. Rev. Public Health*, vol. 37, no. 1, pp. 61–81, Mar. 2016.

[34] Y. Wang, "Clinical information extraction applications: A literature review," *J. Biomed. Inform.*, vol. 77, pp. 34–49, Jan. 2018.

[35] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. F. Jones, R. Forshee, M. Walderhaug, and T. Botsis, "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review," *J. Biomed. Informat.*, vol. 73, pp. 14–29, Sep. 2017.

[36] L. Chen, Y. Gu, X. Ji, Z. Sun, H. Li, Y. Gao, and Y. Huang, "Extracting medications and associated adverse drug events using a natural language processing system combining knowledge base and deep learning," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 1, pp. 56–64, Jan. 2020.

[37] S. Fu, D. Chen, H. He, S. Liu, S. Moon, K. J. Peterson, F. Shen, L. Wang, Y. Wang, A. Wen, Y. Zhao, S. Sohn, and H. Liu, "Clinical concept extraction: A methodology review," *J. Biomed. Informat.*, vol. 109, Sep. 2020, Art. no. 103526.

[38] Y. Shinyama and S. Sekine, "Proceedings of the main conference on human language technology conference of the north American chapter of the association of computational linguistics," Assoc. Comput. Linguistics, Stroudsburg, PA, USA, Tech. Rep., 2006.

[39] B. Rink, S. Harabagiu, and K. Roberts, "Automatic extraction of relations between medical concepts in clinical texts," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 594–600, Sep. 2011.

[40] Q. Wei, Z. Ji, Y. Si, J. Du, J. Wang, F. Tiryaki, S. Wu, C. Tao, K. Roberts, and H. Xu, "Relation extraction from clinical narratives using pre-trained language models," in *Proc. AMIA Annu. Symp.*, 2019, p. 1236.

[41] Y. Si and K. Roberts, "A frame-based nlp system for cancer-related information extraction," in *Proc. AMIA Annu. Symp.*, 2018, p. 1524.

[42] W. Sun, A. Rumshisky, and O. Uzuner, "Evaluating temporal relations in clinical text: 2012 i2b2 challenge," *J. Amer. Med. Inform. Assoc.*, vol. 20, no. 5, pp. 806–813, 2013.

[43] J. Xu, H.-J. Lee, Z. Ji, J. Wang, Q. Wei, and H. Xu, "UTH_CCB system for adverse drug reaction extraction from drug labels at TAC-ADR 2017," in *Proc. TAC*, 2017, pp. 1–6.

[44] E. Aramaki, Y. Miura, M. Tonoike, T. Ohkuma, H. Masuichi, K. Waki, and K. Ohe, "Extraction of adverse drug effects from clinical records," in *Proc. MEDINFO*. Amsterdam, The Netherlands: IOS Press, 2010, pp. 739–743.

[45] Ö. Uzuner, B. R. South, S. Shen, and S. L. DuVall, "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text," *J. Amer. Med. Informat. Assoc.*, vol. 18, no. 5, pp. 552–556, Jun. 2011.

[46] S. Bethard, G. Savova, W.-T. Chen, L. Derczynski, J. Pustejovsky, and M. Verhagen, "SemEval-2016 task 12: Clinical TempEval," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 1052–1062.

[47] S. Henry, K. Buchan, M. Filannino, A. Stubbs, and O. Uzuner, "2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 1, pp. 3–12, Jan. 2020.

[48] R. M. Cronin, D. Fabbri, J. C. Denny, S. T. Rosenbloom, and G. P. Jackson, "A comparison of rule-based and machine learning approaches for classifying patient portal messages," *Int. J. Med. Informat.*, vol. 105, pp. 110–120, Sep. 2017.

[49] S. Kumar Sahu and A. Anand, "Recurrent neural network models for disease name recognition using domain invariant features," 2016, *arXiv:1606.09371*. [Online]. Available: http://arxiv.org/abs/1606.09371

[50] M. Habibi, L. Weber, M. Neves, D. L. Wiegandt, and U. Leser, "Deep learning with word embeddings improves biomedical named entity recognition," *Bioinformatics*, vol. 33, no. 14, pp. i37–i48, Jul. 2017.

[51] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," 2016, *arXiv:1603.01360*. [Online]. Available: http://arxiv.org/abs/1603.01360

[52] S. Moen and T. S. S. Ananiadou, "Distributional semantics resources for biomedical text processing," in *Proc. LBM*, 2013, pp. 39–44.

[53] K. Raja and S. Jonnalagadda, "Natural language processing and data mining for clinical text," *Healthcare Data Anal.*, vol. 36, p. 219, Jan. 2015.

[54] K. M. Hettne, R. H. Stierum, M. J. Schuemie, P. J. M. Hendriksen, B. J. A. Schijvenaars, E. M. V. Mulligen, J. Kleinjans, and J. A. Kors, "A dictionary to identify small molecules and drugs in free text," *Bioinformatics*, vol. 25, no. 22, pp. 2983–2991, Nov. 2009.

[55] M. Song, H. Yu, and W.-S. Han, "Developing a hybrid dictionary-based bio-entity recognition technique," *BMC Med. Informat. Decis. Making*, vol. 15, no. S1, p. S9, Dec. 2015.

[56] K.-I. Fukuda, T. Tsunoda, A. Tamura, and T. Takagi, "Toward information extraction: Identifying protein names from biological papers," in *Proc. PAC Symp. Biocomput.*, vol. 707, 1998, pp. 707–718.

[57] D. Proux, F. Rechenmann, L. Julliard, V. Pillet, and B. Jacq, "Detecting gene symbols and names in biological texts a first step toward pertinent information extraction," *Genome Inform.*, vol. 9, pp. 72–80, Sep. 1998.

[58] S. Lee, D. Kim, K. Lee, J. Choi, S. Kim, M. Jeon, S. Lim, D. Choi, S. Kim, A.-C. Tan, and J. Kang, "BEST: Next-generation biomedical entity search tool for knowledge discovery from biomedical literature," *PLoS ONE*, vol. 11, no. 10, Oct. 2016, Art. no. e0164680.

[59] Y. Ni, J. Wright, J. Perentesis, T. Lingren, L. Deleger, M. Kaiser, I. Kohane, and I. Solti, "Increasing the efficiency of trial-patient matching: Automated clinical trial eligibility pre-screening for pediatric oncology patients," *BMC Med. Informat. Decis. Making*, vol. 15, no. 1, p. 28, Dec. 2015.

[60] A. M. Small, D. H. Kiss, Y. Zlatsin, D. L. Birtwell, H. Williams, M. A. Guerraty, Y. Han, S. Anwaruddin, J. H. Holmes, J. A. Chirinos, R. L. Wilensky, J. Giri, and D. J. Rader, "Text mining applied to electronic cardiovascular procedure reports to identify patients with trileaflet aortic stenosis and coronary artery disease," *J. Biomed. Informat.*, vol. 72, pp. 77–84, Aug. 2017.

[61] Y. Lu, C. J. Vitale, P. L. Mar, F. Chang, N. Dhopeshwarkar, R. A. Rocha, and L. Zhou, "Representation of information about family relatives as structured data in electronic health records," *Appl. Clin. Informat.*, vol. 5, no. 2, pp. 349–367, 2014.

[62] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, May 2016.

[63] N. Afzal, V. P. Mallipeddi, S. Sohn, H. Liu, R. Chaudhry, C. G. Scott, I. J. Kullo, and A. M. Arruda-Olson, "Natural language processing of clinical notes for identification of critical limb ischemia," *Int. J. Med. Informat.*, vol. 111, pp. 83–89, Mar. 2018.

[64] N. J. Leeper, A. Bauer-Mehren, S. V. Iyer, P. LePendu, C. Olson, and N. H. Shah, "Practice-based evidence: Profiling the safety of cilostazol by text-mining of clinical notes," *PLoS ONE*, vol. 8, no. 5, May 2013, Art. no. e63499.

[65] R. Vijayakrishnan, S. R. Steinhubl, K. Ng, J. Sun, R. J. Byrd, Z. Daar, B. A. Williams, C. deFilippi, S. Ebadollahi, and W. F. Stewart, "Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record," *J. Cardiac Failure*, vol. 20, no. 7, pp. 459–464, Jul. 2014.

[66] Z. Tian, S. Sun, T. Eguale, and C. M. Rochefort, "Automated extraction of VTE events from narrative radiology reports in electronic health records: A validation study," *Med. Care*, vol. 55, no. 10, p. e73, 2017.

[67] B. E. Chapman, S. Lee, H. P. Kang, and W. W. Chapman, "Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm," *J. Biomed. Informat.*, vol. 44, no. 5, pp. 728–737, Oct. 2011.

[68] F. P.-Y. Lin, A. Pokorny, C. Teng, and R. J. Epstein, "TEPAPA: A novel in silico feature learning pipeline for mining prognostic and associative factors from text-based electronic medical records," *Sci. Rep.*, vol. 7, no. 1, pp. 1–13, Dec. 2017.

[69] C. Nath, M. S. Albaghdadi, and S. R. Jonnalagadda, "A natural language processing tool for large-scale data extraction from echocardiography reports," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0153749.

[70] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[71] O. Baclic, M. Tunis, K. Young, C. Doan, H. Swerdfeger, J. Schonfeld, P. Data, and I. Hub, "Natural language processing (NLP) a subfield of artificial intelligence," *CCDR*, vol. 46, no. 6, pp. 1–10, 2020.

[72] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, 2011.

[73] N. J. Nilsson and N. J. Nilsson, *Artificial Intelligence: A New Synthesis*. San Mateo, CA, USA: Morgan Kaufmann, 1998.

[74] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[75] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[76] Y. Goldberg, "A primer on neural network models for natural language processing," *J. Artif. Intell. Res.*, vol. 57, pp. 345–420, Nov. 2016.

[77] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[78] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, 2008, pp. 1096–1103.

[79] F. E. Harrell, Jr., K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati, "Regression modelling strategies for improved prognostic prediction," *Statist. Med.*, vol. 3, no. 2, pp. 143–152, Apr. 1984.

[80] I. T. M. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 366–374, 2008.

[81] R. J. Carroll, A. E. Eyler, and J. C. Denny, "Naïve electronic health record phenotype identification for rheumatoid arthritis," in *Proc. AMIA Annu. Symp.*, 2011, p. 189.

[82] J. Hippisley-Cox and C. Coupland, "Predicting risk of emergency admission to hospital using primary care data: Derivation and validation of QAdmissions score," *BMJ Open*, vol. 3, no. 8, Aug. 2013, Art. no. e003482.

[83] F. Rahimian, G. Salimi-Khorshidi, A. H. Payberah, J. Tran, R. Ayala Solares, F. Raimondi, M. Nazarzadeh, D. Canoy, and K. Rahimi, "Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records," *PLOS Med.*, vol. 15, no. 11, Nov. 2018, Art. no. e1002695.

[84] N. Wickramasinghe, "Deepr: A convolutional net for medical records," *J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 22–30, 2017.

[85] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," *J. Biomed. Inform.*, vol. 69, pp. 218–229, May 2017.

[86] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, and M. Sun, "Scalable and accurate deep learning with electronic health records," *NPJ Digit. Med.*, vol. 1, no. 1, p. 18, 2018.

[87] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[88] H. Yang and J. M. Garibaldi, "A hybrid model for automatic identification of risk factors for heart disease," *J. Biomed. Informat.*, vol. 58, pp. S171–S182, Dec. 2015.

[89] K. Buchan, M. Filannino, and Ö. Uzuner, "Automatic prediction of coronary artery disease from clinical narratives," *J. Biomed. Inf.*, vol. 72, pp. 23–32, Aug. 2017.

[90] S. M. Castro, E. Tseytlin, O. Medvedeva, K. Mitchell, S. Visweswaran, T. Bekhuis, and R. S. Jacobson, "Automated annotation and classification of BI-RADS assessment from radiology reports," *J. Biomed. Informat.*, vol. 69, pp. 177–187, May 2017.

[91] M. Torii, J.-W. Fan, W.-L. Yang, T. Lee, M. T. Wiley, D. S. Zisook, and Y. Huang, "Risk factor detection for heart disease by applying text analytics in electronic medical records," *J. Biomed. Informat.*, vol. 58, pp. S164–S170, Dec. 2015.

[92] J. C. Denny, N. N. Choma, J. F. Peterson, R. A. Miller, L. Bastarache, M. Li, and N. B. Peterson, "Natural language processing improves identification of colorectal cancer testing in the electronic medical record," *Med. Decis. Making*, vol. 32, no. 1, pp. 188–197, Jan. 2012.

[93] J. Jonnagaddala, S.-T. Liaw, P. Ray, M. Kumar, H.-J. Dai, and C.-Y. Hsu, "Identification and progression of heart disease risk factors in diabetic patients from longitudinal electronic health records," *BioMed Res. Int.*, vol. 2015, pp. 1–10, 2015.

[94] H. S. Chase, L. R. Mitrani, G. G. Lu, and D. J. Fulgieri, "Early recognition of multiple sclerosis using natural language processing of the electronic health record," *BMC Med. Informat. Decis. Making*, vol. 17, no. 1, p. 24, Dec. 2017.

[95] R. L. Figueroa and C. A. Flores, "Extracting information from electronic medical records to identify the obesity status of a patient based on comorbidities and bodyweight measures," *J. Med. Syst.*, vol. 40, no. 8, p. 191, Aug. 2016.

[96] S. N. Kasthurirathne, B. E. Dixon, J. Gichoya, H. Xu, Y. Xia, B. Mamlin, and S. J. Grannis, "Toward better public health reporting using existing off the shelf approaches: A comparison of alternative cancer detection approaches using plaintext medical data and non-dictionary based feature selection," *J. Biomed. Informat.*, vol. 60, pp. 145–152, Apr. 2016.

[97] G. Napolitano, A. Marshall, P. Hamilton, and A. T. Gavin, "Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction," *Artif. Intell. Med.*, vol. 70, pp. 77–83, Jun. 2016.

[98] Q. Chen, H. Li, B. Tang, X. Wang, X. Liu, Z. Liu, S. Liu, W. Wang, Q. Deng, S. Zhu, Y. Chen, and J. Wang, "An automatic system to identify heart disease risk factors in clinical texts over time," *J. Biomed. Informat.*, vol. 58, pp. S158–S163, Dec. 2015.

[99] N.-W. Chang, H.-J. Dai, J. Jonnagaddala, C.-W. Chen, R. T.-H. Tsai, and W.-L. Hsu, "A context-aware approach for progression tracking of medical concepts in electronic medical records," *J. Biomed. Informat.*, vol. 58, pp. S150–S157, Dec. 2015.

[100] W.-W. Yim, S. W. Kwan, and M. Yetisgen, "Classifying tumor event attributes in radiology reports," *J. Assoc. Inf. Sci. Technol.*, vol. 68, no. 11, pp. 2662–2674, Nov. 2017.

[101] S. N. Kasthurirathne, B. E. Dixon, J. Gichoya, H. Xu, Y. Xia, B. Mamlin, and S. J. Grannis, "Toward better public health reporting using existing off the shelf approaches: The value of medical dictionaries in automated cancer detection using plaintext medical data," *J. Biomed. Informat.*, vol. 69, pp. 160–176, May 2017.

[102] P. L. Teixeira, W.-Q. Wei, R. M. Cronin, H. Mo, J. P. VanHouten, R. J. Carroll, E. LaRose, L. A. Bastarache, S. T. Rosenbloom, T. L. Edwards, D. M. Roden, T. A. Lasko, R. A. Dart, A. M. Nikolai, P. L. Peissig, and J. C. Denny, "Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals," *J. Amer. Med. Inform. Assoc.*, vol. 24, no. 1, pp. 162–171, Jan. 2017.

[103] M. Zhou, N. Duan, S. Liu, and H.-Y. Shum, "Progress in neural NLP: Modeling, learning, and reasoning," *Engineering*, vol. 6, no. 3, pp. 275–290, Mar. 2020.

[104] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Medical concept representation learning from electronic health records and its application on heart failure prediction," 2016, *arXiv:1602.03686*. [Online]. Available: https://arxiv.org/abs/1602.03686

[105] B. Tang, Z. Pan, K. Yin, and A. Khateeb, "Recent advances of deep learning in bioinformatics and computational biology," *Frontiers Genet.*, vol. 10, p. 214, Mar. 2019.

[106] O. Baclic, M. Tunis, K. Young, C. Doan, and H. Swerdfeger, "Challenges and opportunities for public health made possible by advances in natural language processing," *Canada Communicable Disease Rep.*, vol. 46, no. 6, pp. 161–168, Jun. 2020.

[107] S. Baker, A.-L. Korhonen, and S. Pyysalo, "Cancer hallmark text classification using convolutional neural networks," in *Proc. 5th Workshop Building Evaluating Resour. Biomed. Text Mining (BioTxtM)*, 2017, pp. 1–9.

[108] C. Lin, T. Miller, D. Dligach, S. Bethard, and G. Savova, "Representations of time expressions for temporal relation extraction with convolutional neural networks," in *Proc. BioNLP*, 2017, pp. 322–327.

[109] S. Mohan, N. Fiorini, S. Kim, and Z. Lu, "Deep learning for biomedical information retrieval: Learning textual relevance from click logs," in *Proc. BioNLP*, 2017, pp. 222–231.

[110] Y. Peng and Z. Lu, "Deep learning for extracting protein-protein interactions from biomedical literature," 2017, *arXiv:1706.01556*. [Online]. Available: http://arxiv.org/abs/1706.01556

[111] M. Asada, M. Miwa, and Y. Sasaki, "Extracting drug-drug interactions with attention CNNs," in *Proc. BioNLP*, 2017, pp. 9–18.

[112] M. C. Chen, R. L. Ball, L. Yang, N. Moradzadeh, B. E. Chapman, D. B. Larson, C. P. Langlotz, T. J. Amrhein, and M. P. Lungren, "Deep learning to classify radiology free-text reports," *Radiology*, vol. 286, no. 3, pp. 845–852, 2017.

[113] L. Sulieman, D. Gilmore, C. French, R. M. Cronin, G. P. Jackson, M. Russell, and D. Fabbri, "Classifying patient portal messages using convolutional neural networks," *J. Biomed. Informat.*, vol. 74, pp. 59–70, Oct. 2017.

[114] G. Crichton, S. Pyysalo, B. Chiu, and A. Korhonen, "A neural network multi-task learning approach to biomedical named entity recognition," *BMC Bioinf.*, vol. 18, no. 1, p. 368, Dec. 2017.

[115] S. Karimi, X. Dai, H. Hassanzadeh, and A. Nguyen, "Automatic diagnosis coding of radiology reports: A comparison of deep learning and conventional classification methods," in *Proc. BioNLP*, 2017, pp. 328–332.

[116] R. Feldman, O. Netzer, A. Peretz, and B. Rosenfeld, "Utilizing text mining on online medical forums to predict label change due to adverse drug reactions," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 1779–1788.

[117] A. N. Jagannatha and H. Yu, "Bidirectional RNN for medical event detection in electronic health records," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, p. 473.

[118] S. Yadav, A. Ekbal, S. Saha, and P. Bhattacharyya, "Deep learning architecture for patient data de-identification in clinical records," in *Proc. Clin. Natural Lang. Process. Workshop (ClinicalNLP)*, 2016, pp. 32–41.

[119] W. Salloum, G. Finley, E. Edwards, M. Miller, and D. Suendermann-Oeft, "Deep learning for punctuation restoration in medical reports," in *Proc. BioNLP*, 2017, pp. 159–164.

[120] P. V S S Rahul, S. Kumar Sahu, and A. Anand, "Biomedical event trigger identification using bidirectional recurrent neural network based models," 2017, *arXiv:1705.09516*. [Online]. Available: http://arxiv.org/abs/1705.09516

[121] H. He, K. Ganjam, N. Jain, J. Lundin, R. White, and J. Lin, "An insight extraction system on biomedical literature with deep neural networks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2691–2701.

[122] R. Chalapathy, E. Zare Borzeshi, and M. Piccardi, "Bidirectional LSTM-CRF for clinical concept extraction," 2016, *arXiv:1611.08373*. [Online]. Available: http://arxiv.org/abs/1611.08373

[123] I. J. Unanue, E. Z. Borzeshi, and M. Piccardi, "Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition," *J. Biomed. Inform.*, vol. 76, pp. 102–109, Dec. 2017.

[124] G. Stanovsky, D. Gruhl, and P. Mendes, "Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, vol. 1, 2017, pp. 142–151.

[125] A. Jagannatha and H. Yu, "Structured prediction models for RNN based sequence labeling in clinical text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, p. 856.

[126] Y. Wu, M. Jiang, J. Lei, and H. Xu, "Named entity recognition in Chinese clinical text using deep neural network," *Stud. Health Technol. Inform.*, vol. 216, p. 624, Jan. 2015.

[127] J. Alan Fries, "Brundlefly at SemEval-2016 task 12: Recurrent neural networks vs. joint inference for clinical temporal information extraction," 2016, *arXiv:1606.01433*. [Online]. Available: http://arxiv.org/abs/1606.01433

[128] X. Lv, Y. Guan, J. Yang, and J. Wu, "Clinical relation extraction with deep learning," *Int. J. Hybrid Inf. Technol.*, vol. 9, no. 7, pp. 237–248, Jul. 2016.

[129] Y. Liu, T. Ge, K. S. Mathews, H. Ji, and D. L. McGuinness, "Exploiting task-oriented resources to learn word embeddings for clinical abbreviation expansion," 2018, *arXiv:1804.04225*. [Online]. Available: http://arxiv.org/abs/1804.04225

[130] T. Tran, T. D. Nguyen, D. Phung, and S. Venkatesh, "Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM)," *J. Biomed. Inform.*, vol. 54, pp. 96–105, Apr. 2015.

[131] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *J. Amer. Med. Informat. Assoc.*, vol. 24, no. 2, pp. 361–370, Mar. 2017.

[132] S. Mehrabi, S. Sohn, D. Li, J. J. Pankratz, T. Therneau, J. L. Sauver, H. Liu, and M. Palakal, "Temporal pattern and association discovery of diagnosis codes using deep learning," in *Proc. Int. Conf. Healthcare Informat.*, Oct. 2015, pp. 408–416.

[133] Z. Liang, G. Zhang, J. X. Huang, and Q. V. Hu, "Deep learning for healthcare decision making with EMRs," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2014, pp. 556–559.

[134] H. Li, X. Li, M. Ramanathan, and A. Zhang, "Identifying informative risk factors and predicting bone disease progression via deep belief networks," *Methods*, vol. 69, no. 3, pp. 257–265, Oct. 2014.

[135] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 301–318.

[136] Y. Cheng, F. Wang, P. Zhang, and J. Hu, "Risk prediction with electronic health records: A deep learning approach," in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, Jun. 2016, pp. 432–440.

[137] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Dec. 2018.

[138] B. K. Beaulieu-Jones and C. S. Greene, "Semi-supervised learning of the electronic health record for phenotype stratification," *J. Biomed. Informat.*, vol. 64, pp. 168–178, Dec. 2016.

[139] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data," *PLoS ONE*, vol. 8, no. 6, Jun. 2013, Art. no. e66341.

[140] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*. [Online]. Available: http://arxiv.org/abs/1301.3781

[141] Z. Che, D. Kale, W. Li, M. T. Bahadori, and Y. Liu, "Deep computational phenotyping," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2015, pp. 507–516.

[142] F. Dernoncourt, J. Y. Lee, O. Uzuner, and P. Szolovits, "De-identification of patient notes with recurrent neural networks," *J. Amer. Med. Inform. Assoc.*, vol. 24, no. 3, pp. 596–606, May 2017.

[143] A. T. McCray, S. Srinivasan, and A. C. Browne, "Lexical methods for managing variation in biomedical terminologies," in *Proc. Annu. Symp. Comput. Appl. Med. Care*, 1994, p. 235.

[144] H. Xu, S. P. Stenner, S. Doan, K. B. Johnson, L. R. Waitman, and J. C. Denny, "MedEx: A medication information extraction system for clinical narratives," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 1, pp. 19–24, 2010.

[145] S. Doan, L. Bastarache, S. Klimkowski, J. C. Denny, and H. Xu, "Integrating existing natural language processing tools for medication extraction from discharge summaries," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 5, pp. 528–531, Sep. 2010.

[146] N. Sager, M. Lyman, C. Bucknall, N. Nhan, and L. J. Tick, "Natural language processing and the representation of clinical data," *J. Amer. Med. Inform. Assoc.*, vol. 1, no. 2, pp. 142–160, 1994.

[147] N. Sager, M. Lyman, N. T. Nhan, and L. J. Tick, "Medical language processing: Applications to patient data representation and automatic encoding," *Methods Inf. Med.*, vol. 34, nos. 1–2, pp. 140–146, 1995.

[148] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 5, pp. 507–513, 2010.

[149] L. Christensen, P. Haug, and M. Fiszman, "MPLUS: A probabilistic medical language understanding system," in *Proc. Workshop Natural Lang. Process. Biomed. Domain (ACL)*, 2002, pp. 29–36.

[150] P. Haug, S. Koehler, L. M. Lau, P. Wang, R. Rocha, and S. Huff, "A natural language understanding system combining syntactic and semantic techniques," in *Proc. Annu. Symp. Comput. Appl. Med. Care*, 1994, p. 247.

[151] A. T. McGray, J. L. Sponsler, B. Brylawski, and A. C. Browne, "The role of lexical knowledge in biomedical text understanding," in *Proc. Annu. Symp. Comput. Appl. Med. Care*, 1987, p. 103.

[152] A. R. Aronson and F.-M. Lang, "An overview of MetaMap: Historical perspective and recent advances," *J. Amer. Med. Inform. Assoc.*, vol. 17, no. 3, pp. 229–236, 2010.

[153] T. G. Dietterich, "Ensemble learning," in *The Handbook of Brain Theory and Neural Networks*, vol. 2, M. A. Arbib, Ed. 2002, pp. 110–125.

[154] L. Yu, S. Wang, and K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 1434–1444, Feb. 2008.

[155] B. Liu, S. Wang, R. Long, and K.-C. Chou, "iRSpot-EL: Identify recombination spots with an ensemble learning approach," *Bioinformatics*, vol. 33, no. 1, pp. 35–41, Jan. 2017.

[156] C. Liu, J. Jiang, J. Gu, Z. Yu, T. Wang, and H. Lu, "High-dimensional omics data analysis using a variable screening protocol with prior knowledge integration (SKI)," *BMC Syst. Biol.*, vol. 10, no. S4, p. 118, Dec. 2016.

[157] X. Y. Wang, B. Yu, A. J. Ma, C. Chen, B. Q. Liu, and Q. Ma, "Protein–protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique," *Bioinformatics*, vol. 35, no. 14, pp. 2395–2402, Jul. 2019.

[158] S. M. Hill, L. M. Heiser, T. Cokelaer, M. Unger, and N. K. Nesser, "Inferring causal molecular networks: Empirical assessment through a community-based effort," *Nature Methods*, vol. 13, no. 4, pp. 310–318, Apr. 2016.

[159] C. Liu, X. Wang, G. Z. Genchev, and H. Lu, "Multi-omics facilitated variable selection in Cox-regression model for cancer prognosis prediction," *Methods*, vol. 124, pp. 100–107, Jul. 2017.

[160] M. Torii, Z. Hu, C. H. Wu, and H. Liu, "BioTagger-GM: A gene/protein name recognition system," *J. Amer. Med. Informat. Assoc.*, vol. 16, no. 2, pp. 247–255, 2009.

[161] M. Torii, K. Wagholikar, and H. Liu, "Using machine learning for concept extraction on clinical documents from multiple data sources," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 580–587, Sep. 2011.

[162] S. Doan, N. Collier, H. Xu, P. H. Duy, and T. M. Phuong, "Recognition of medication information from discharge summaries using ensembles of classifiers," *BMC Med. Informat. Decis. Making*, vol. 12, no. 1, p. 36, Dec. 2012.

[163] N. Kang, Z. Afzal, B. Singh, E. M. van Mulligen, and J. A. Kors, "Using an ensemble system to improve concept extraction from clinical records," *J. Biomed. Informat.*, vol. 45, no. 3, pp. 423–428, Jun. 2012.

[164] T.-T. Kuo, P. Rao, C. Maehara, S. Doan, J. D. Chaparro, M. E. Day, C. Farcas, L. Ohno-Machado, and C.-N. Hsu, "Ensembles of NLP tools for data element extraction from clinical notes," in *Proc. AMIA Annu. Symp.*, 2016, p. 1880.

[165] G. Karystianis, A. Dehghan, A. Kovacevic, J. A. Keane, and G. Nenadic, "Using local lexicalized rules to identify heart disease risk factors in clinical notes," *J. Biomed. Informat.*, vol. 58, pp. S183–S188, Dec. 2015.

[166] E. Kogan, K. Twyman, J. Heap, D. Milentijevic, J. H. Lin, and M. Alberts, "Assessing stroke severity using electronic health record data: A machine learning approach," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–8, Dec. 2020.

[167] R. Garg, E. Oh, A. Naidech, K. Kording, and S. Prabhakaran, "Automating ischemic stroke subtype classification using machine learning and natural language processing," *J. Stroke Cerebrovascular Diseases*, vol. 28, no. 7, pp. 2045–2051, Jul. 2019.

[168] C. Kim, V. Zhu, J. Obeid, and L. Lenert, "Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke," *PLoS ONE*, vol. 14, no. 2, Feb. 2019, Art. no. e0212778.

[169] S. Grechishcheva, E. Efimov, and O. Metsker, "Risk markers identification in EHR using natural language processing: Hemorrhagic and ischemic stroke cases," *Proc. Comput. Sci.*, vol. 156, pp. 142–149, Jan. 2019.

[170] N. Afzal, S. Sohn, S. Abram, C. G. Scott, R. Chaudhry, H. Liu, I. J. Kullo, and A. M. Arruda-Olson, "Mining peripheral arterial disease cases from narrative clinical notes using natural language processing," *J. Vascular Surg.*, vol. 65, no. 6, pp. 1753–1761, 2017.

[171] J. Jonnagaddala, S.-T. Liaw, P. Ray, M. Kumar, and H.-J. Dai, "HTNSystem: Hypertension information extraction system for unstructured clinical notes," in *Proc. Int. Conf. Technol. Appl. Artif. Intell.* Springer, 2014, pp. 219–227.

[172] S. Boytcheva, G. Angelova, Z. Angelov, and D. Tcharaktchiev, "Text mining and big data analytics for retrospective analysis of clinical texts from outpatient care," *Cybern. Inf. Technol.*, vol. 15, no. 4, pp. 58–77, Nov. 2015.

[173] X. Liu, Y. Chen, J. Bae, H. Li, J. Johnston, and T. Sanger, "Predicting heart failure readmission from clinical notes using deep learning," 2019, *arXiv:1912.10306*. [Online]. Available: http://arxiv.org/abs/1912.10306

[174] Z. Wang, Y. Zhu, D. Li, Y. Yin, and J. Zhang, "Feature rearrangement based deep learning system for predicting heart failure mortality," *Comput. Methods Programs Biomed.*, vol. 191, Jul. 2020, Art. no. 105383.

[175] M. Topaz, K. Radhakrishnan, S. Blackley, V. Lei, K. Lai, and L. Zhou, "Studying associations between heart failure self-management and rehospitalizations using natural language processing," *Western J. Nursing Res.*, vol. 39, no. 1, pp. 147–165, Jan. 2017.

[176] J. H. Garvin, Y. Kim, G. T. Gobbel, M. E. Matheny, A. Redd, B. E. Bray, P. Heidenreich, D. Bolton, J. Heavirland, N. Kelly, R. Reeves, M. Kalsy, M. K. Goldstein, and S. M. Meystre, "Automating quality measures for heart failure using natural language processing: A descriptive study in the department of veterans affairs," *JMIR Med. Informat.*, vol. 6, no. 1, p. e5, Jan. 2018.

[177] J. D. Osborne, M. Wyatt, A. O. Westfall, J. Willig, S. Bethard, and G. Gordon, "Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 6, pp. 1077–1084, Nov. 2016.

[178] A. K. AAlAbdulsalam, J. H. Garvin, A. Redd, M. E. Carter, C. Sweeny, and S. M. Meystre, "Automated extraction and classification of cancer stage mentions fromunstructured text fields in a central cancer registry," *AMIA Summits Transl. Sci. Proc.*, vol. 2018, p. 16, Jan. 2018.

[179] S. Datta, E. V. Bernstam, and K. Roberts, "A frame semantic overview of NLP-based information extraction for cancer-related EHR notes," *J. Biomed. Informat.*, vol. 100, Dec. 2019, Art. no. 103301.

[180] D. S. Carrell, S. Halgrim, D.-T. Tran, D. S. M. Buist, J. Chubak, W. W. Chapman, and G. Savova, "Using natural language processing to improve efficiency of manual chart abstraction in research: The case of breast cancer recurrence," *Amer. J. Epidemiol.*, vol. 179, no. 6, pp. 749–758, Mar. 2014.

[181] T. Miller, D. Dligach, S. Bethard, C. Lin, and G. Savova, "Towards generalizable entity-centric clinical coreference resolution," *J. Biomed. Informat.*, vol. 69, pp. 251–258, May 2017.

[182] A. Mykowiecka, M. Marciniak, and A. Kupść, "Rule-based information extraction from patients' clinical data," *J. Biomed. Inform.*, vol. 42, no. 5, pp. 923–936, 2009.

[183] S. Bozkurt, J. A. Lipson, U. Senol, and D. L. Rubin, "Automatic abstraction of imaging observations with their characteristics from mammography reports," *J. Amer. Med. Inform. Assoc.*, vol. 22, no. e1, pp. e81–e92, Apr. 2015.

[184] R. Tatman and C. Kasten, "Effects of talker dialect, gender & race on accuracy of bing speech and YouTube automatic captions," in *Proc. Interspeech*, 2017, pp. 934–938. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2017-1746

[185] I. Spasic and G. Nenadic, "Clinical text data in machine learning: Systematic review," *JMIR Med. Informat.*, vol. 8, no. 3, Mar. 2020, Art. no. e17984.

[186] A. Rajkomar, M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin, "Ensuring fairness in machine learning to advance health equity," *Ann. Internal Med.*, vol. 169, no. 12, pp. 866–872, 2018.

[187] J. Gramlich. (2020). *10 Facts About Americans and Facebook. Washington (DC): Pew Research Center*. Accessed: Oct. 12, 2020. [Online]. Available: https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/

[188] (2019). *Treasury Board Secretariat. Directive on Automated Decision-Making. Ottawa (on): Government of Canada*. Accessed: Oct. 12, 2020. [Online]. Available: https://www.tbs-sct.gc.ca/pol/doc-engaspx?id=32592

[189] (2017). *Montreal Declaration for a Responsible Development of Artificial Intelligence. Forum on the Socially Responsible Development of AI: 2017 Nov 2-3: Montreal (QC)*. Accessed: Oct. 12, 2020. [Online]. Available: https://www.montrealdeclaration-responsibleai.com/the-declaration

[190] I. Y. Chen, S. Joshi, and M. Ghassemi, "Treating health disparities with artificial intelligence," *Nature Med.*, vol. 26, no. 1, pp. 16–17, Jan. 2020.

[191] J. Wiens, W. N. Price, and M. W. Sjoding, "Diagnosing bias in data-driven algorithms for healthcare," *Nature Med.*, vol. 26, no. 1, pp. 25–26, Jan. 2020.

[192] K. Holstein, J. Wortman Vaughan, H. Daumé, M. Dudik, and H. Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?" in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–16.

[193] E. Ford, K. Curlewis, A. Wongkoblap, and V. Curcin, "Public opinions on using social media content to identify users with depression and target mental health care advertising: Mixed methods survey," *JMIR Mental Health*, vol. 6, no. 11, Nov. 2019, Art. no. e12942.

[194] U. Erlingsson and C. Radebaugh. (2020). *Introducing Tensorflow Privacy: Learning With Differential Privacy for Training Data. Medium*. Accessed: Oct. 12, 2020. [Online]. Available: https://medium.com/tensorflow/

[195] B. B. P. de Vries, M. van Smeden, F. R. Rosendaal, and R. H. Groenwold, "Abstract, and keyword searching resulted in poor recovery of articles in systematic reviews of epidemiologic practice," *J. Clin. Epidemiol.*, vol. 121, pp. 55–61, Mar. 2020.

[196] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, Oct. 2019.

REHAB E. MOHAMED received the M.Sc. degree from the Computer Science Department, Cairo University, in 2017. She is currently a Teaching Assistant with the Department of Computer Science, Faculty of Computers and Information, Minia University, Egypt. Her research interests include natural language processing, biomedical, machine learning, and computer vision.

ESSAM H. HOUSSEIN received the Ph.D. degree in computer science in 2012. He is currently an Associate Professor with the Faculty of Computers and Information, Minia University, Egypt. He is the Founder and Chair of the Computing & Artificial Intelligence Research Group (CAIRG), Egypt. He has more than 100 scientific research articles published in prestigious international journals in the topics for instance meta-heuristics optimization, artificial intelligence, image processing, the IoT, and its applications. His research interests include WSNs, the IoT, AI, bioinformatics and biomedical, image processing, data mining, and meta-heuristics optimization techniques. He serves as a reviewer for more than 40 journals (Elsevier, Springer, and IEEE).

ABDELMGEID A. ALI is currently a Professor with the Computer Science Department, Minia University, Minya, Egypt. He has published over 80 research papers in prestigious international journals and conference proceedings. He has supervised over 60 Ph.D. and M.Sc. students. His research interests include information retrieval, software engineering, image processing, data security, metaheuristics, the IoT, digital image steganography, and data warehousing. He is a member of the *International Journal of Information Theories and Applications* (ITA).

• • •