Journal of Information Systems Engineering and Management
2025, 10(21s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Development of Augmented Deep Learning Approach for Predicting and Diagnosing Generic Health Issues

Manochandar S[1], Mohamed Suhail Mohamed Nabi[2]
[1]*Assistant Professor, Digital Business and Analytics, Bharathidasan Institute of Management-620014, Tamilnadu, INDIA*
manochandar@bim.edu, mano.chandar90@gmail.com
[2]*Assistant Professor, Dept. of Computer Applications, The New College, Chennai, TamilNadu, INDIA*
nifaz33@gmail.com, mmohmedsuhail@thenewcollege.edu.in

ARTICLE INFO

Received: 21 Dec 2024

Revised: 01 Feb 2025

Accepted: 14 Feb 2025

ABSTRACT

In the current context of epidemics, people are increasingly afraid to seek hospitalization. Moreover, dealing with all patients on the physician side is a hugely challenging subject. Hence our research work trying to develop an automated naval framework, which automatically detect abnormal patient from the set of all the patient, and allow those specific patients only to the medical expertise knowledge. Artificial Intelligence based platform is used to accelerate the progress for diagnosing and providing treatment for various diseases. Various deep learning methodologies have been glorified to reach its goal using integrated bio informatics approaches. The data collection is improved with the help of big data analytics which supports cluster computing and automated data processing. So, it becomes easier not only for collecting such kind of data but also for producing the comprehensive healthcare report by converting them into a critical insight. Then Word2vec with Recurrent conditional random filed is utilised in order to extract information. This hybrid data driven findings is interpreted to recognize patterns and analyses the structure to the finest level. So, it helps the doctor to diagnose specific diseases faster and more accurate. From the extracted information, the disease can be easily recognized and diagnosed with the help of combination of NER (Named Entity Recognition) and Bi-directional LSTM (Long short-term memory) algorithms. These algorithms not only identify the keywords but also the context of entities which combines direct matching and stemmed matching. So, it can able to process the health care information based on the lifestyle and environment. Finally, we compare our proposed augmented technique against various state-of-art model, as a result our augmented technique produces better performance against state-of-art models in terms of Accuracy, Recall, Precision and AUC.

Keywords: Big data analytics, Deep Learning, Unigram POS tagger, Word2vec, Recurrent conditional random field, Named Entity Recognition, Bi-directional long short-term memory algorithms.

1. Introduction

In healthcare sector, the role artificial intelligence over the past decade has been significant. It utilises, some specified set of algorithms to mimic the human perception. Furthermore, machine learning and deep learning algorithms are particularly used in terms of understanding and analysing, some difficult processing functionality of data which are existing in different fields like telecommunication, diagnosis, automation, detection and so on. The role of AI towards medical fields has been achieved innumerable result in diversified way also producing very reliable and accurate result that potential of human cannot even handle. By means of handling those tasks, information extraction has to be done which reliably extracting features from the given data up to its finest level. Once processing of information has to be done, analysis of transformation of such data has been happen, that helps to transform analysed data into valuable clinical information in healthcare sector that eventually reach the best result [1]. Although healthcare system, produces feasible outcomes in every aspect but its considerably complex and major challenging task have been associated with those outcomes includes heterogeneous information, complex structure of data and multidimensional feature etc. In modern research of biomedical various different data processing has arises also capability of processing images, sensor data, text, voice etc.

Advancements in information technology assist many industries, including healthcare, especially with the increasing movement of digitalization. Automation of different strategies, cost effective storage mechanisms and effective information and data retrieval which leads to significant increases in efficiency and cost savings [2]. Moreover, the non-expense has been considerable thing which are existing in healthcare industry due to population growth, raising growth which incredibly affects the needful over the people. Therefore, enhanced medical practices, earliest diagnosis of disease which leads to most effective and accurate improvement over the medical filed. All these functionalities have been achieved with the support of AI at the present context. Due to the technological advancement has been assured to provided best deliverable outcomes [3].

Machine learning has grown in popularity in recent years, and it is now used in a wide range of applications, including classification of image, recommendation systems, Analysis of social data, mining of text and so on. Deep learning [DL], also known as language modelling in which feature extraction and algorithm as a single part [4], Also DL involves numerous high-level machine-learning technique in order to process various applications. Recent investigations in deep learning have emerged as a result of the enormous development and availability of the data, as well as the astonishing improvement in hardware technology. Deep learning, which is based on traditional neural networks, outperforms its predecessors greatly. It develops many-layered learning approaches (hidden layer) using graph technology and neuro transformations and this number of hidden layers depends on the types of application. Nowadays, most of the current deep learning algorithms has been presented and proves promising outcomes over numerous applications including processing of speech and audio, processing of visual data and numerous other widely used applications [5, 6].

The digitization of medical data transforms the dimensionality of data while also increasing the volume of information and the usefulness of big data analysis, among other things. Big data seems to have its own set of qualities, including volume, velocity, truthfulness, variability, and monetary worth. Traditional methods of data analysis are not well suited for big data analytics, as previously stated. Big data analytics is a distinct field with its own set of tools and methodologies. Medical data is sensitive information that comes in a variety of ways. Health records, statistical procedures, billing information, person - health details, health surveys, and clinical trials are some of the types of information available. Medical imaging offers a variety of applications in medical research and clinical treatment, each with its own significance. When individuals have anomalies and the illness is clearly manifested, healthcare systems that are now in place can treat them effectively. A timely diagnosis of severe disorders aids in the treatment of the patient and the reduction of risk. Otherwise, it can result in chronic illnesses and even death for the patient. Delay in detection of severe diseases accounts for 59 percent of all deaths each year [7]. A thorough examination of big-data workloads can aid in the comprehension of their consequences for both the design of hardware and software. Mehul Shah et al. [8] try to identify a collection of “data giants" means processing of data kernels— that would facilitate present and, influenced by the 7 dwarfs of numerical computation [9]. They implement a series of categorizing measurements (responsiveness, access structure, operational set, types of data, interpret versus prompt, complexity of processing) based on a detailed collection of workflows and reach the conclusion that, since about 2010, five load - balancing models could satisfactory manner encompass data-centric throughput.

The goal of Named Entity Recognition (NER) is to identify occurrences of specified notations in text that belongs to predetermined set of different semantic kinds which includes humans, region, and institution [10]. NER is used in various natural language processing application, including understanding of text [11, 12], retrieval of information [13, 14], summarization of text in an automated way [15], answering to queries [16], translation of machine [17], and development of knowledgebase [18], and so on. The following are the approaches which are involved in the NER that comprises of approaches based on rule-based, approaches based on unsupervised learning, approaches based on DL, approaches based on features with respective of supervised learning algorithms. Over the past decade, the NER has been successfully utilised in deep learning based on the case study. The phrase or word that exactly similar one object on a set of various objects that has some similar functions [19]. In our study, it particularly works on the recognition of individual patient with their corresponding identification also this helps to map their name with respective medical history.

Named Entity recognition, typically provides three benefit which applied over the deep learning model. From the non-linear transformation, NER facilitates and generates mapping in a non-linear fashion with respect to input-output. Then deep learning models has the capability of reducing the effort and time required to design the features of NER. In terms of traditional based feature techniques necessitate a high level of engineering competence and domain knowledge. Also, DL technique are considerably good at learning usable representation and factors from the data of raw format. On the way DL based NER model can be trained by utilising gradient descent optimiser, hence reduction of error has been minimized at the finest level. All this mentioned characteristic allows us to create potentially complicated Named entity recognition models.

Deep neural networks are comprised of various layers of sophisticated architecture, such as nonlinear and transformation functions. Now, a complex challenge known as deep learning is taking on tasks which could never be completed by other types of AI [20-22]. Deep learning is well-suited to learning heterogeneous information and data, it is an ideal choice for tackling the broad spectrum of unlabelled information, where a model may be trained without the need for extra labelling. Using deep learning for malevolent purposes is probably a possibility, but it is expected to have numerous good outcomes. Deep learning is becoming even more widely applicable as researchers have recently shown it is well-suited to dealing with massive data sets, and future applications of deep learning will benefit from this [23]. There have been a multitude of recent papers demonstrating the powerful potential of powerful computational tools, such as Recognition of image, Categorization of text, and others. Deep learning can help provide better diagnoses for patients [24, 25]. Health informatics, biomedicine, and MRI image analysis (for example) are all included in this description. As you may have guessed, many types of diagnoses are one of these specialised uses of deep learning (ROI). Deep learning is way better than typical machine learning, because it can adapt from original data and includes more than one hidden layer which allows it to understand concepts by themselves. The potential of neural networks to handle big data which also crucial component of deep learning.

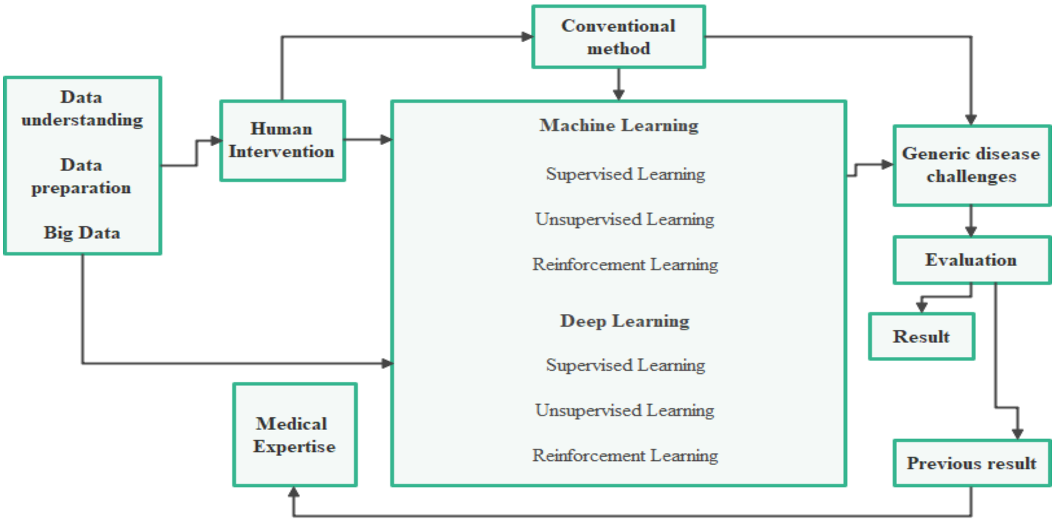


Figure 1: Entire Architecture of ML, DL and BDA

The given figure 1 gives a basic explanation of the architecture of big data disease prediction methods that include machine learning, deep learning, and other similar techniques. Numerous beneficial and life-saving outcomes have come out of various uses of big data analytics in healthcare. In addition, storing large amounts of data and giving attention to the information for longer periods has become quite expensive and requires additional time. The job of extracting structured information from un-structured and semi-structured data using information extraction. Though we did not reveal whether there were any missing data and how they were dealt with in the study, we did say all we discovered. So, there is no evident connection between precision and decision-making. In order to be able to predict and diagnose the specific diseases, such as tumours, eye problems, bone cancer, etc., multiple artificial neural networks and deep learning algorithms, such as FNN, MNN, RNN, and LSTM, are being used independently, but this is not capable of predicting generic diseases and their onset.

2. Related study and literature review

Contextual named entity recognition (CNER) has garnered great attention because to its practical value, and numerous helped to facilitate have indeed been offered in the literature. Several of these previous techniques fall into two general categories which includes, rule based and dictionary-based technique, ML and DL approaches has the capability to intensify the performance over the field.

Rule-based techniques detect items through the use of heuristics and constructed rules [26], [26]. These techniques are generally relied on rule templates developed by speech researchers. These patterns cover a range of aspects, including statistical data, syntax, terms, indications and direction phrases, position syllables (such as termination words), and centre words. They were the prevalent techniques in earlier CNER systems. Nevertheless, listing all the principles for modelling the configuration of healthcare named entities, particularly for diverse medical entities, is impractical. Frequently, such rules are dataset-specific. They may fail to function properly as the dataset evolves, and this type of handmade method will always result in a rather high systems development cost.

Dictionary-based techniques identify attributes by utilising pre-existing clinical terminologies [27] and [28]. They are commonly utilised easy implementation and effectiveness. From given medical documents, a vocabulary CNER system could identify all the relevant entities specified in the dictionary. The majority of these framework. Nevertheless, construction of domain knowledge and vocabularies [29] is a quite frequent in Chinese for a single word to already have multiple meanings depending on the situation. Simple matches between vocabulary items could perhaps ensure that the retrieved words have the desired interpretations. Additionally, it is incapable of dealing with objects that are not included in the dictionary. As a result, this strategy frequently results in low recall.

CNER is typically viewed as a sequence labelling problem in which the objective is to determine the optimal label sequence for an input feature text [30]. Hidden Markov models, probabilistic generative Hidden markov, conditional random fields, and support vector machine (SVM), [31, 32] are all examples of typical techniques. However, because these statistical machine learning algorithms rely on which was before features, their development is prohibitively expensive. The feature model accuracy will be costly in order to identify the optimal set of features that aid in distinguishing entities of a particular type from others. Additionally, it really an art instead of a science, necessitating lengthy trial-and-error investigations.

Deep learning techniques, particularly bidirectional RNN models with a CRF layer as that interface of output, exhibit performance on state-of-art in CNER problems and surpass classic statistical models [33-35]. RNNs with gated recurrent cells, such as long-short term memory (LSTM) and gated recurrent units [36], are capable of recording and recovering extensive global information. On top of the recurrent layers, the sequentially CRF ensures that the ideal arrangement of tags across the entire text is obtained. Additionally, several researchers attempted to increase performance by using additional features such as n-gram features [38] and radical-level features [30]. Zhang and Yang [37] investigated a character-level lattice-structured LSTM model that can directly use word to word sequence information while being free of segmentation errors. Gligic et al. [32] developed a bootstrapped neural network to answer the CNER task via transfer learning. Peters et al. [39] enhanced word representations by pretraining a character language model. Yang et al. [27] employ multi-task learning to leverage cross-domain and cross-lingual expertise. Furthermore, all of the approaches outlined shown are RNN-based. Due to the fact that RNNs are specialized sequence models typically keep a matrix of hidden activation that is propagated across time, learning RNN-based models frequently takes a lengthy period. Devlin et al. [39] recently suggested pretrained bi-directional transformers and significantly increased the performance of several natural language processing tasks. Unfortunately, building this model continues to consume large amounts of computational resources.

3. Methodology

This section encompasses various natural language processing and deep learning approaches that are used for our proposed augmented framework. Initially our framework taken input from user in the form of voice (voice regarding heath issue), with help of bigdata analytics all these, information has to be stored and accessed. Then those input has been processed and transformed into text with help of sequence-to-sequence attention models, hence sequential relation among the information has been preserved. Second, converted text has been further processed with help of natural language processing techniques includes Tokenization, Stemming, Lemmatization, Stop word removal and part-of-speech tagging. Third, detected part-of-speech has been given input to the word2vec with recurrent conditional random field, which is a hybrid data driven approach, in that word2vec particularly convert each and every word present in the corpus to respective vector. Then recurrent conditional random field is used to extract feature from those vectors. Finally, those extracted feature is given input to NER and Bidirectional LSTM which classify those input as normal and abnormal based on the training.

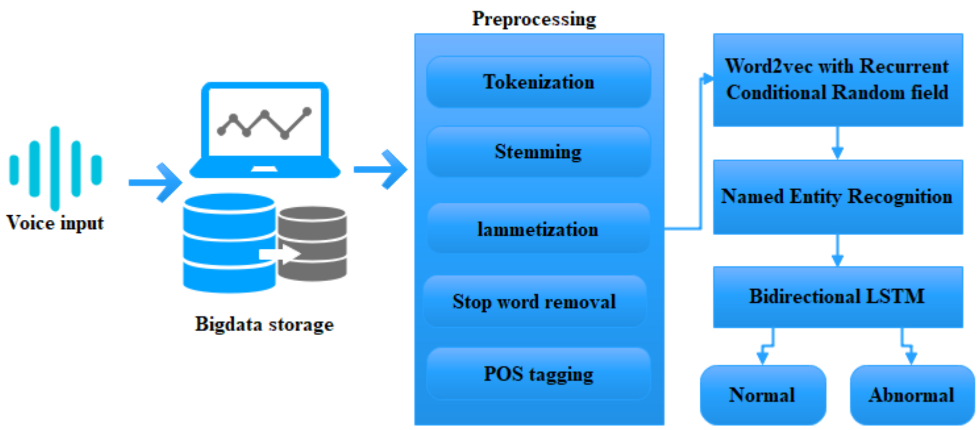


Figure 2: Workflow of proposed augmented framework

A. Big data Analytics

In general, data that arises over the healthcare sector is innumerable and mostly that is heterogeneous in nature. In order to provide efficient storage and processing capacity over healthcare sector, we utilised big data analytics in our work. Big data analytics which are not always analysed in terms of data of size huge, it is a processing technology which holds the capacity of processing numerous kinds of vast data types in an efficient way, which includes streaming data as well [41]. The main reason behind the modification of statistical and data analysis approaches is that, it follows unique approach, so it tends to data of unstructured, heterogeneous, complex, missing value etc. While it may appear that big data enables us to acquire greater data in order to uncover more relevant information, the reality is that too many data does not always equal more usable information. This could hold information that is more confusing or out of the ordinary. For example, a person may have many accounts or a single account could well be exploited by multiple people, lowering the accuracy of acquired data [42].

1. **Input:** Unstructured data (D)
2. Fix candidate solution 's'
3. **For** if the termination technique is not attained
4. f = Scan (D)
5. u = Build (d, s, o)
6. s= update(u)
7. End
8. Output rules 's'

Once data collection is done, analysis of data using KDD which is responsible by means searching the hidden pattern existing in the data. If the dataset which is relatively large in volume, researchers widely utilising technique known as clustering. In healthcare domain, the generated data which is heterogeneous in nature. This state clustering technique has been utilised to understand numerous patterns over those data. The core idea behind this technique is to segregate a sequence of unlabelled data into m different groups. E.g., K-means clustering technique. In our research, we utilised k-means clustering technique by means of segregating the patterns over heterogenous data. The following equation 1 is used to determine cluster over the heterogenous dataset.

$$cluster_i = \frac{1}{k_i}\sum_{j=0}^{k_i} z_{ij} ----\{1\}$$

The occurrence of error has been rectified by the following equation 2.

$$Error = \sum_{i=0}^{n}\sum_{j=0}^{k_i} D\left(z_{ij} - cluster_i\right) ----\{2\}$$

B. Voice to text conversion

Our augmented framework takes input from user in the form of voice. By using big data technologies, the voice signals are stored and organised well efficiently. We used sequence-to-sequence attention mechanism to effectively preserve sequential information among the data.

Cho et al. presented an encoder-decoder paradigm, which is often termed simply preservation of sequential information. In the same paper which established the GRU cell [43]. Although the initial model was developed for a machine translation application, the principle can be used to a variety of different challenges. The sequence-to-sequence paradigm is capable of encoding a variable-length sequences together into fixed representation of vector length and decoding a certain fixed-length representation of vector again into a sequence of variable length. It is referred as a sequential model since the lengths of the outputs and inputs are typically distinct.

Consider the sequence of source can be defined as $x = (x_1, x_2, \dots \dots x_T)$ and the respective sequence of target as $y = (y_1, y_2, \dots \dots y_T)$. The recurrent neural network is responsible for reading every set of input x_t and transformed into desired length of vector c. This vector which is capable of summarizing the entire information sequence. The decoder is responsible for vector c from another RNN, correspondingly produces every time sequence of the output y_T.

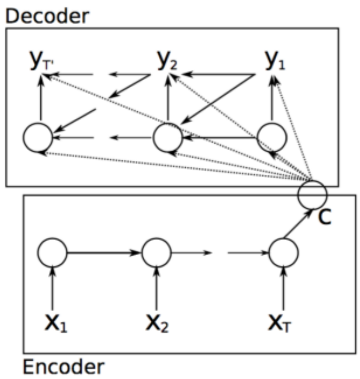


Figure 3: Encoder-Decoder architecture

The objective of attention mechanisms is important for two main reasons: to minimize the computation time needed to analyse high-dimensional data by concentrating on smaller subsets, and to enable the machine to concentrate on diverse features of the input, resulting in more effective extraction of pertinent information. While the first statement could be valid for some systems, this would not indicate that an attention mechanism would always result in a faster running model. Indeed, in some recurrent neural network configurations, this will increase computational effort by requiring the model to assess the entire input at every epoch in order to calculate each step is to take attention. This mechanism is responsible for shows hidden state with respect the set of input parameters.

$$f_{ij} = c(t_{i-1}, h) --- \{3\}$$

$$f_{ij} = w^T \tanh\left(W t_{i-1} + V h_j + b\right) --- \{4\}$$

C. Data Pre-processing

Once voice data is transformed into text, we infuse those data into set of pre-processing steps, which includes tokenization, stemming, lemmatization, stop word removal and POS tagging, which acts a feature extractor. These techniques are widely utilised in natural language processing in terms of doing any pre-processing tasks.

Tokenization: Tokenization is the process of subdividing the larger sentence into the set of tokens by using the punctuation [,]. The technique of mapping paragraphs from characters to string and string to words is the fundamental stage in problem exist in NLP. Since in attempt to comprehend any word or record, we must first interpret the words/sentences contained inside. Tokenization is indeed a necessary component of every Information Retrieval (IR) strategy; it really pre-processes text and also produces tokens for use in the process of indexing.

Stemming and lemmatization:

The expanding volume of data and information on the web has reached an all-time peak in recent years. This massive amount of information as well as data necessitates the development of tools and strategies for easily extracting insights.

Stemming is the process of reverting modulated (or occasionally modified) words to their respective stem or root word – which is typically a written version of a phrase. For instance, stemming eliminates suffixes which are present over the sentence. Thus, after a step stemming, the word “generating” becomes “generate”, just as “liked” becomes “like” etc.

Lemmatization is a term that refers to activities that involve the right use of lexicon and grammatical study of words, with both the goal of removing only interpret the result and returning dictionaries structure of a respective word, referred to as the lemma. In simple terms, lemmatization is the process of lowering the grammatical structure after determining the word's part of speech (POS) present in any article.

Stop word removal:

Once stemming and lemmatization is done over every token, then stop word removal has been actioned. It performs removes unnecessary present in the token. For instance, ‘the’, ‘was’, ‘is’, ‘from’ are some of the words which are sometimes present over the tokens. Stop word removal eliminates all this words with help of NLTK package.

POS tagging:

In general, POS tagging is used to determine part-of-speech for each and every token present in the corpus. A hidden Markov model POS tagger evaluate the highly preferable sequence of POS tag $\hat{p}_1^N = \hat{p}_1, \dots \dots \hat{p}_N$ for given sequence of word w_1^N.

$$\hat{p}_1^N = \arg max_{p_1^N} q(p_1^N, w_1^N) --- \{5\}$$

The two sequence of join probability is termed as produce of lexical and context probability of entire POS tag.

$$P(\hat{p}_1^N, w_1^N) = \pi_{i=1}^N \, p\left(p_i \middle| p_{i-k}^{i-1}\right) p(w_i|t_i) --- \{6\}$$

HMM model is the widely used tagger because of its efficiency. It is successfully performed for a huge amount of training tokens present in the corpus. POS tagger generally, used to train the corpus more than 150 POS tag with help of nltk packages. we used the POS tagger known as hidden Markov model (HMM) tagger that breaks down contextual probabilities into a combination of characteristic probabilities. As explained previously, the likelihood of a feature given the characteristics of the following POS tags and the accompanying features of the anticipated POS tag is assessed using a decision tree. The probability distributions at the decision tree's terminal nodes have been smoothed using the probability distributions at the parent nodes. By means of integrating class weight probability $P_q(c)$ to the frequency $f(c)$ to perform the smoothing operation,

$$P(c) = \frac{g(c) + \beta P_q(c)}{\beta + \sum_c g(c)} ----\{7\}$$

The weight of β has been assigned as 1. In terms of normalizing the probability, the total probability has been divided by overall value of attribute with its corresponding feature. The following formula is used to determine the lexical probability of unknown words: The unidentified words are separated into four distinct classes: words beginning with such an upper-case character, words beginning with a lower-case character, and a separate class for all other terms. The tagger constructs a suffix trie for every unidentified lexical category through using existing types of words within that class. Suffixes can be up to seven characters in length.

$$\frac{g_{\beta\beta}}{|T_{\beta\beta}|} \sum_{t \in T_{\beta\beta}} T_{\beta\beta}(t) \, log \frac{T_{\beta\beta}(t)}{T_\beta(t)} < 1 ----\{8\}$$

D. Word2Vec with Recurrent Conditional Random field:

Once part-of-speech is recognised over all tokens present in the corpus, the we apply Word2Vec with recurrent conditional random field over the tokens. This is the hybrid data driven approach which transform all tokens present into corpus with their respective vector. Then that vector has been fed input the RCRN (Recurrent Conditional Random field).

Word2Vec:

Word2vec decodes and vectorizes the content of unique words on the assumption that combinations of words interpretations in a particular context have proximity distances [44]. The figure 4 illustrates the structured model of Skip-gram and CBOW, two word2vec representation learning introduced by Mikolov. Both algorithms have set of input, output along with respective projections. Their derivation mechanisms of output are distinct. The input layer takes parameters as $W_t = \{W_{(n-2)}, W_{(n-1)}, \dots., W_{(n+1)}, W_{(n+2)}\}$ where W_t signifies words. The set of layers present in the projection is equivalent to the corresponding array of vectors and maintains the summation with numerous vectors. The output layer relates to the layer that delivers the results of projection layer vector. CBOW is comparable to the feed-forward neural network Model [45] in that it identifies the outcome word using adjacent near word representations. The fundamental tenet of CBOW is to forecast the appearance of some word by studying its neighbors. Because layer of projection on CBOW is always projects all items in the same direction, the vectors of all items preserve a mean and share their places. CBOW's structure has the benefit of evenly structuring the data set's information. In contrast, the Skip-gram possesses a structure that enables it to anticipate the vectors of other words given a single word. The fundamental premise of Skip-gram is to predict other keywords near a given word. The Skip-projection gram's layer anticipates words that are adjacent to the term injected directly into layer of input. When new words occur, the Skip-gram layout benefits from vectorization. According to study of Mikolov's, CBOW is smoother and more suited to huge data sets than Skip-gram, although Skipgram outperforms CBOW when attempting to communicate. However, several research comparing the effectiveness of CBOW with Skip-gram indicate that Skip-gram outperforms CBOW [46, 47].

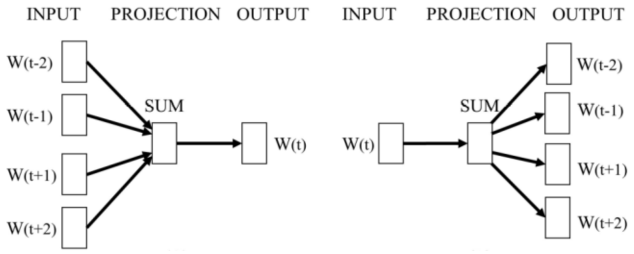


Figure 4: Structure Model for both CBOW and Skip-gram

Data	Size
Collection of data	May 1, 2010 – July 1, 2010 (62 days)
Sequence of News	2,553,234
News magazine	133,543
Medical related words	295,342,943
Healthcare Tweets	432,643
Sequence of medical related twitter	6,432,945
Common twitter words	283,234,742

Table 1: Training dataset for POS

Recurrent Conditional Random field:
The network architecture which comprises of three kinds of a regularization layer and learnable layer, Conditional random filed, recurrent layer integrated GRU memory cell and regularization layer of Bernoulli dropout.

A recurrent neural network (RNN) is a sort of neural network layer wherein the weights are used again across a data sequence through the assumption that the dimension of sequence comprises of relevant signal [48]. The gated recurrent unit (GRU) is a sort of processing unit that is used to calculate the RNN's representation of hidden parameters at each step of time. The GRU makes use of memory cells to enhance the modelling of long-term dependency and convergence rate. The memory units are clearly described by chung et. al.[49].

$$r_n^m = \sigma(h_n^{m-1} W_i^m + h_{n-1}^m U_i^m + c_i^m) ----\{9\}$$
$$y_n^m = \sigma\left(h_n^{m-1} W_y^m + h_{n-1}^m U_y^m + c_y^m\right) ----\{10\}$$
$$\hat{h}_n^m = tanh(h_n^{m-1} W_h^m + (q \odot h_{n-1}^m) U_h^m + c_h^m) ----\{11\}$$

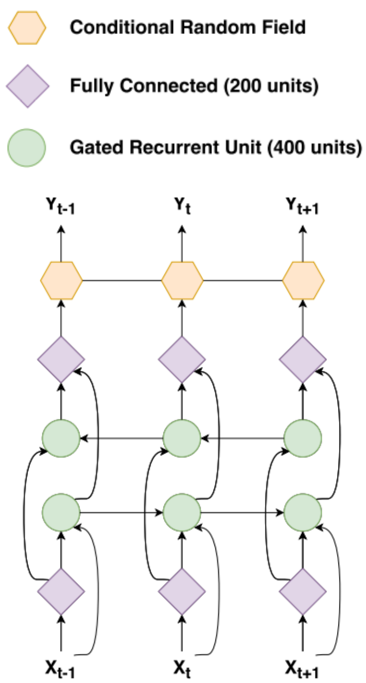


Figure 5: Recurrent Conditional Random field

As mentioned in [50], the hidden states are synchronized with their index, n = [1,2, 3,.....n] in order to integrate the forward and backward passes in both directions.

$$h_n^m = \begin{bmatrix} \overrightarrow{h_n^m} \\ \overleftarrow{h_n^m} \end{bmatrix} - - - \{12\}$$

From the equation forward and backward pass is indicated as $\overrightarrow{h_n^m}$ and $\overleftarrow{h_n^m}$ respectively.

Conditional random field is the joint probability sequence of label $x = x_1, \dots \dots . . x_T$ with corresponding sequnece of input $a = a_1, \dots \dots . . a_T$ follows the some specified form includes,

$$p(x|a) = \frac{1}{x(h)} \prod_{n=1}^{T} exp\varphi x_n(h_n) \prod_{n=1}^{T-1} exp\varphi x_n, x_{n+1}(h_n, h_{n+1}) - - - - \{13\}$$

From the equation, $h = h_1, h_2, \dots \dots, h_n$ represent the hidden layer output of previous node. The linear model $\varphi_n = \varphi_n(h_n)$ works on the basis of input h_n and the respective classes as C corresponding to the output $\phi(h_n, h_{n+1})$ which is a linear model of original valued output as represented by c^2.

$$\varphi_n = \varphi_n h_n + c_n - - - - \{14\}$$

$$\phi_n = W_\phi \begin{bmatrix} h_n \\ h_{n+1} \end{bmatrix} + b_\phi - - - - \{15\}$$

E. Named Entity Recognition

In our proposed augmented framework, named entity acts like a performance intesifier, which helps to actioned on identified every respective name of the patient in an appropriate manner. Multi-task learning [51] is a technique for simultaneously learning a collection of related activities. Multi-task deep learning are intended to perform better than single-task learning algorithms since they take into account the relationship between distinct activities. Collobert et al. [52] developed a network using a sentence strategy to perform POS, Chunk, NER, and SRL tasks concurrently. This multi-task approach enables the classification model to identify word embeddings that are applicable to all relevant tasks. Yang et al. [53] suggested a model of multitask joint for the purpose of learning language-specific regularities. The model was jointly trained on POS, Chunk, and NER tasks. According to Rei [54], the sequence labelling model achieves consistent productivity improvements by integrating an unsupervised language modelling target in the training phase. Lin et al. [55] presented a multi-lingual multi-task architecture for low-resource environments that is capable of effectively transferring various forms of knowledge in terms of improving the primary model. Apart from addressing NER in conjunction with some other sequence labelling tasks, a multi-task based framework can be utilised to identify entities as well as relations concurrently [56], [57], or to represent NER as two distinct subtasks: entity segmentation and categorization of entity prediction [58], [59]. Due to the heterogeneity between datasets in the field of biomedicine, NER on every dataset is regarded a performance in a multi-task environment [60], [61]. A crucial premise is that the various datasets share similar character- and word-level characteristics. Then, multi-task training is used to maximise data efficiency and to motivate models to acquire more broad representations.

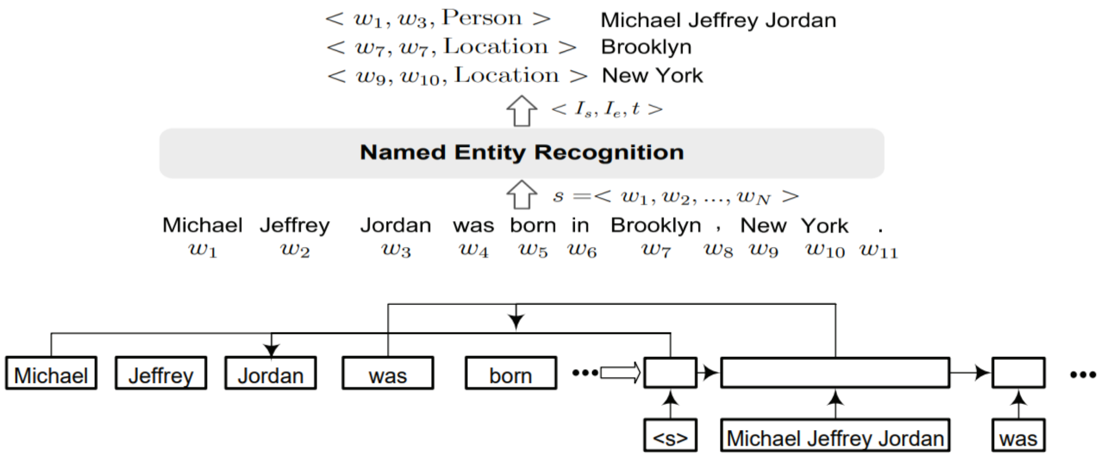


Figure 6: Detailed performance of Named Entity recognition

F. Bi-directional LSTM

Once data has been arranged on the name pair, then by means of classification we utilised bi-directional LSTM, which is responsible for classification the status of incoming patient as normal or abnormal. Normal data is further allow to doctor knowledge and the abnormal data kept stored as log file. Fig. 9 illustrates the sequence based on BLSTM architecture. As can be seen, the proposed framework is divided into four components:

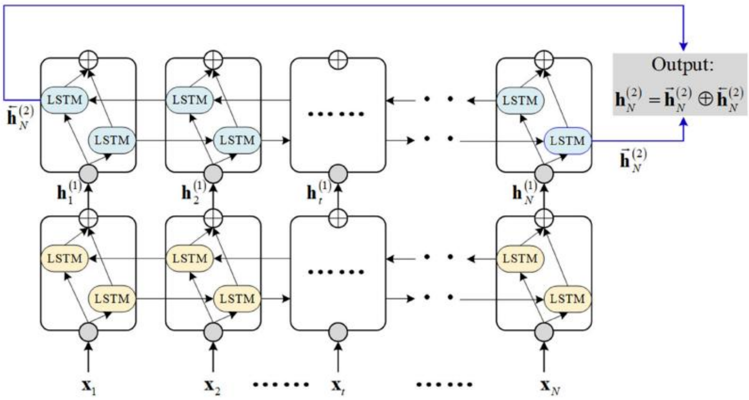


Figure 7: Bi-directional LSTM

The most widely used LSTM network is often succinctly even as vanilla LSTM [62], [63], whose core concept is that multiple non-linear unit gate exist which is input, forget, and output gate. It has the capability to control the information flow around within the cell. A basic LSTM network model can indeed represent the relationship between the current and past states which determines both forward and backward direction exist in the particular context. Providing access towards contexts of past and future. on the other hand, it can be advantageous for a variety of sequential learning tasks, most notably for some data sequence used in actual forecasting applications [64]. As a result, BLSTM networks are often used to extract relevant attributes from the dataset. The fundamental concept underlying BLSTMs seems to be that two distinct hidden layers were used to analyze the data sequence towards different directions in order to grab both the direction of past and future. The continuity equation define the function of the distinct time step t with respect to hidden layer, and two distinct arrow symbols, signify processes of forward and backward (→ and ←), respectively.

$$\vec{h}_n = g(\vec{y}_n, \vec{h}_{n-1}; \vec{\Theta}_{LSTM}) - - - \{16\}$$

$$= \begin{cases} \overrightarrow{X_n} = \tanh\,(\overrightarrow{W_x X_n} + \overrightarrow{R_x\, h_{n-1}} + \overrightarrow{b_x}) \\ \overrightarrow{J_n} = \sigma\,(\overrightarrow{W_J X_J} + \overrightarrow{R_J\, h_{n-1}} + \overrightarrow{b_J}) \\ \overrightarrow{g_n} = \sigma(\overrightarrow{W_g X_n} + \overrightarrow{R_g\, h_{n-1}} + \overrightarrow{b_g}) \\ \overrightarrow{b_n} = \overrightarrow{X_n} \odot \overrightarrow{J_n} + \overrightarrow{J_{n-1}} \odot \overrightarrow{g_n} \end{cases} - - - - \{17\}$$

4. Experimentation and Result discussion

Our augmentated proposed framework has been developed using python, which is sucessfully implemented using anaconda 2020.11. We successfully collected data from Kaggle repository https://www.kaggle.com/shivamb/deep-healthcare-analysis-using-bigquery. We split the dataset for 80:20 ratio for training and testing purposes.

1. Initially, we preprocess the dataset, by using preprocessing techniques of natural langauge processing which inlcudes tokenization, stemming, lammetization, stop word removal and POS tagging.
2. Second, those output has been fed input to word2vec model. Once vectorization has been done those output is infused input as recurrent conditional random field for effective feature extraction. It also helps to maintain sequential information.
3. Third, the resultant is further applied input to named entity recgonition for better recognition of every entity present in the resultant.
4. Finally, Bi-directional LSTM applied, by means of classifying the patient status as normal and abnormal.

Our result comprison comprises of three main folds:

STAGE 1: After preprocessing, we compared the performance of Bi-directional LSTM with state of art models.

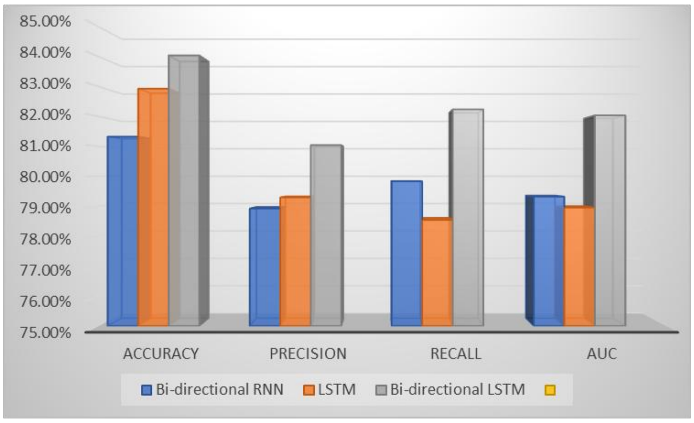
STAGE 2: After the successful execution of word2vec with recurrent conditional random field model, we compared the performance of Bi-directional LSTM with state of art models.

STAGE 3: Our final comparison encampasses the Bi-directional LSTM against state of art model, with the input is processed over Named Entity Recognition.

	Accuracy	Precision	Recall	AUC
Bi-directional RNN	81.32%	78.92%	79.83%	79.32%
LSTM	82.92%	79.29%	78.54%	78.95%
Bi-directional LSTM	**84.03%**	81.04%	82.23%	82.03%

Table 2: Comprison of Bi-directional LSTM against state-of-art Model [Dataset: After processing].

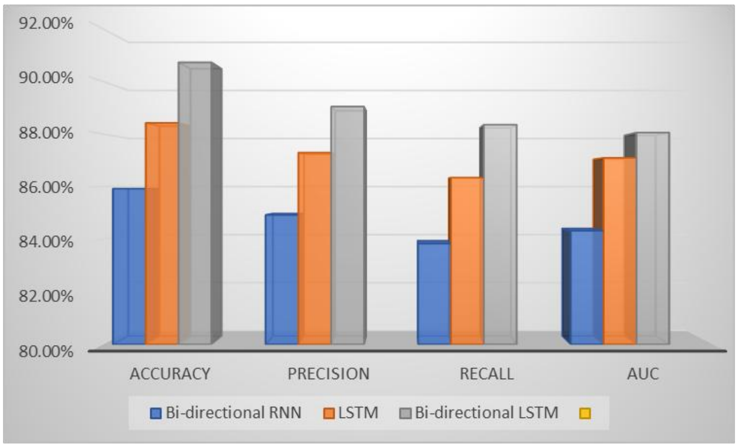
Table 2 encompasses the result comparsion of bi-directional LSTM with other state-of-art models. The dataset which we fed is preprocessed by the techniques tokenization, stemming and lammetization, stop word removal and POS tagging. From the comparsion bi-directional LSTM performs well over the preprocessed dataset and the respective graphs is given as follow:



	Accuracy	Precision	Recall	AUC
Bi-directional RNN	85.92%	84.92%	83.83%	84.32%
LSTM	88.42%	87.29%	86.34%	87.09%
Bi-directional LSTM	**90.72%**	89.04%	88.35%	88.05%

Table 3: Comprison of Bi-directional LSTM against state-of-art Model [Dataset: Preprocessing + word2vec with recurrent conditional random field].

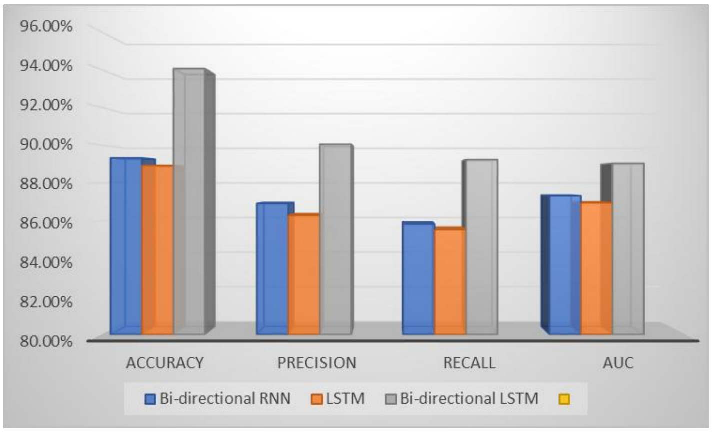
Table 3 encompasses the result comparsion of bi-directional LSTM with other state-of-art models. The dataset which we fed is infused of both preprocessing and word2Vec with recurrent conditional random field. From the comparsion bi-directional LSTM performs well over the data and it achieves the accuracy of 90.72% and the respective graphs is given as follow:



	Accuracy	Precision	Recall	AUC
Bi-directional RNN	89.32%	86.92%	85.83%	87.32%
LSTM	88.92%	86.29%	85.54%	86.95%
Bi-directional LSTM	**94.03%**	90.04%	89.23%	89.03%

Table 4: Comprison of Bi-directional LSTM against state-of-art Model [Dataset: Preprocessing + word2vec with recurrent conditional random field+ Named entity recognition].

Table 4 encompasses the result comparsion of birdirection LSTM with other state-of-art models. Dataset which we have given to the model is completely trained by our proposed augmented technique[Preprocessing + Word2Vec with recurrent condtional random field + Named entity recognition]. As the result shows that our augumented technique classifies the voice of patient as normal and abnormal with the better performance accuracy of **94.03%** and respective graph is given as follow:



5. Conclusion

This study aims at the use of various deep learning approaches along with big data analytics. By improving the resultant quality of guessing the right diagnosis and prescribe the most suitable treatment. The Healthcare Records offer the detailed information about a patient intellectual health, diagnosis system, lab tested results, remedies and so on, such data are stored and analysed using big data analytics. By combining several artificial approaches, like Machine learning and deep learning along with big data analytics, we can diagnose the generic disease of a patient accurately. For example, if the patient having high blood pressure, this model can accurately predict the onset possibilities of diseases like diabetics, heart attacks etc. Thus, we can also prognosticate our biological age and we take further measures to stay healthy. In future we focus on the novel infectious disease prediction using this ensemble of various deep learning approaches.

Reference:

[1] Mohammad Jamshidi; Ali Lalbakhsh; Jakub Talla, “Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment” IEEE Access (Volume: 8), DOI: 10.1109/ACCESS.2020.3001973, Electronic ISSN: 2169-3536, Page(s): 109581 - 109595, Date of Publication: 12 June 2020.

[2] S.H. Koppad, A. Kumar, Application of big data analytics in healthcare system to predict COPD, in: Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference, IEEE, 2016, March, pp. 1–5.

[3] J.A. Patel, P. Sharma, Big data for better health planning, in: Advances in Engineering and Technology Research (ICAETR), 2014 International Conference, IEEE, 2014, August, pp. 1–5.

[4] Li Deng. 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing 3 (2014), 1–29.

[5] Yilin Yan, Min Chen, Saad Sadiq, and Mei-Ling Shyu. 2017. Efficient imbalanced multimedia concept retrieval by deep learning on spark clusters. International Journal of Multimedia Data Engineering and Management 8, 1 (2017), 1–20.

[6] Yilin Yan, Min Chen, Mei-Ling Shyu, and Shu-Ching Chen. 2015. Deep learning for imbalanced multimedia data classification. In The IEEE International Symposium on Multimedia. IEEE, 483–488.

[7] Xiaokang Wang, Laurence T. Yang, Huazhong Liu,M. Jamal Deen, “A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives” in IEEE Transactions on Big Data ,Volume 4 , Issue: 3 , Sept. 1 2018,P 325 – 340.

[8] Mehul Shah, Parthasarathy Ranganathan, Jichuan Chang, Niraj Tolia, David Roberts, Trevor Mudge, Data dwarfs: Motivating a Coverage Set for Future Large Data Center Workloads, Technical Report HPL-2010-115, Hewlett Packard Laboratories, November 8, 2010.

[9] Krste Asanovic, Ras Bodik, Bryan Christopher Catanzaro, Joseph James Gebis, Parry Husbands, Kurt Keutzer, David A. Patterson,William Lester Plishker, John Shalf, Samuel Webb Williams, Katherine A. Yelick, The Landscape of Parallel Computing Research: A View from Berkeley. Technical Report UCB/EECS2006-183, EECS Department, University of California, Berkeley, 2006.

[10] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," Lingvist. Investig., vol. 30, no. 1, pp. 3–26, 2007.

[11] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, "ERNIE: enhanced language representation with informative entities," in ACL, 2019, pp. 1441–1451.

[12] P. Cheng and K. Erk, "Attending to entities for better text understanding," arXiv preprint arXiv:1911.04361, 2019.

[13] J. Guo, G. Xu, X. Cheng, and H. Li, "Named entity recognition in query," in SIGIR, 2009, pp. 267–274.

[14] D. Petkova and W. B. Croft, "Proximity-based document representation for named entity retrieval," in CIKM, 2007, pp. 731–740.

[15] C. Aone, M. E. Okurowski, and J. Gorlinsky, "A trainable summarizer with knowledge acquired from robust nlp techniques," Adv. Autom. Text Summ., vol. 71, 1999.

[16] D. M. Aliod, M. van Zaanen, and D. Smith, "Named entity recognition for question answering," in ALTA, 2006, pp. 51–58.

[17] B. Babych and A. Hartley, "Improving machine translation quality with automatic named entity recognition," in EAMT, 2003, pp. 1–8.

[18] O. Etzioni, M. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates, "Unsupervised named entity extraction from the web: An experimental study," Artif. Intell., vol. 165, no. 1, pp. 91–134, 2005.

[19] R. Sharnagat, "Named entity recognition: A literature survey," Center for Indian Language Technology, 2014.

[20] Szegedy, C.; Wei, L.; Yang, J.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

[21] Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep Learning with Differential Privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 308–318.

[22] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 436–444.

[23] Chen, X.-W.; Lin, X. Big data deep learning: Challenges and perspectives. IEEE Access 2014, 2, 514–525.

[24] Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. Brief Bioinform. 2017.

[25] Wei, J.; He, J.; Chen, K.; Zhou, Y.; Tang, Z. Collaborative filtering and deep learning-based recommendation system for cold start items. Expert Syst. Appl. 2017, 69, 29–39.

[26] C. Friedman, P. O. Alderson, J. H. M. Austin, J. J. Cimino, and S. B. Johnson, "A general natural-language text processor for clinical radiology," J. Amer. Med. Inform. Assoc., vol. 1, no. 2, pp. 161–174, Mar. 1994.

[27] Q. T. Zeng, S. Goryachev, S. Weiss, M. Sordo, S. N. Murphy, and R. Lazarus, "Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system," BMC Med. Informat. Decision Making, vol. 6, no. 1, pp. 1–9, May 2006.

[28] G. K. Savova et al., "Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications," J. Amer. Med. Informat. Assoc. JAMIA, vol. 17, no. 5, pp. 507–513, Sep. 2010.

[29] T. C. Rindflesch, L. Tanabe, J. N. Weinstein, and L. Hunter, "Edgar: extraction of drugs, genes and relations from the biomedical literature," in Biocomputing, vol. 5, pp. 517–528, Oct. 1999.

[30] R. Gaizauskas, G. Demetriou, and K. Humphreys, "Term recognition and classification in biological science journal articles," in Proc. Workshop Comput. Terminol. Med. Biol. Appl., 2000, pp. 37–44

[31] M. Song, H. Yu, and W. Han, "Developing a hybrid dictionary-based bio-entity recognition technique," BMC Med. Inform. Decision Making, vol. 15, no. S1, p. S9, May 2015.

[32] N. Wang, R. Ge, C. Yuan, K. Wong, and W. Li, "Company name identification in chinese financial domain," J. Chin. Inf. Process., to be published.

[33] J. Lei, B. Tang, X. Lu, K. Gao, M. Jiang, and H. Xu, "A comprehensive study of named entity recognition in chinese clinical text," J. Amer. Med. Informat. Assoc., vol. 21, no. 5, pp. 808–814, Dec. 2014.

[34] J. Lei, "Named entity recognition in chinese clinical text," Ph.D. dissertation, School Biomed. Inform., Health Sci. Centre Houston, Univ. Texas, Houston, TX, USA, 2014.

[35] J. Finkel, S. Dingare, H. Nguyen, M. Nissim, C. Manning, and G. Sinclair, "Exploiting context for biomedical entity recognition: from syntax to the web," in Proc. Int. Joint Workshop Natural Lang. Process. Biomed. Appl., 2004, pp. 88–91.

[36] M. Skeppstedt, M. Kvist, G. H. Nilsson, and H. Dalianis, "Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study," J. Biomed. Informat., vol. 49, pp. 148–158, Jun. 2014.

[37] Y. C. Wu, T. K. Fan, Y. S. Lee, and S. J. Yen, "Extracting named entities using support vector machines," in Proc. Int. Workshop Knowl. Discovery Life Sci. Literature, Sep. 2006, pp. 91–103.
[38] Z. Huang, W. Xu, and K. Yu. (2015). "Bidirectional LSTMCRF models for sequence tagging." [Online]. Available: https://arxiv.org/abs/1508.01991
[39] D. Zeng, C. Sun, L. Lin, and B. Liu, "LSTM-CRF for drug-named entity recognition," Entropy, vol. 19, no. 6, p. 283, Jun. 2017.
[40] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Mar. 1997.
[41] Russom P. Big data analytics. TDWI: Tech. Rep ; 2011.
[42] Ma C, Zhang HH, Wang X. Machine learning for big data analytics in plants. Trends Plant Sci. 2014;19(12):798–808
[43] K. Cho, B. van Merrienboer, C¸. Gulc¸ehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. ¨ Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1724–1734, 2014. URL http://aclweb.org/anthology/D/D14/D14-1179.pdf.
[44] Sahlgren M. The distributional hypothesis. Ital J Disabil Stud. 2008; 20: 33–53.
[45] Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. J Mach Learn Res. 2003; 3: 1137–1155.
[46] Lai S, Liu K, He S, Zhao J. How to generate a good word embedding. IEEE Intell Syst. 2016; 31: 5–14.
[47] Chiu B, Crichton G, Korhonen A, Pyysalo S. How to train good word embeddings for biomedical NLP. Proceedings of the 15th Workshop on Biomedical Natural Language Processing. 2016. pp. 166–174.
[48] A. Graves. 2012. Supervised sequence labelling with recurrent neural networks. Springer (2012)
[49] Chung, Ç. Gülçehre, K. Cho, and Y. Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. CoRR abs/1412.3555 (2014). http://arxiv.org/abs/1412.3555.
[50] M. Schuster and K. Paliwal. 1997. Bidirectional Recurrent Neural Networks. Trans. Sig. Proc. 45, 11 (Nov. 1997), 2673–2681. DOI: http://dx.doi.org/10.1109/78.650093.
[51] R. Caruana, "Multitask learning," Mach. learn., vol. 28, no. 1, pp. 41–75, 1997.
[52] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," J. Mach. Learn. Res., vol. 12, no. Aug, pp. 2493–2537, 2011.
[53] Z. Yang, R. Salakhutdinov, and W. Cohen, "Multi-task cross-lingual sequence tagging from scratch," arXiv preprint arXiv:1603.06270, 2016.
[54] M. Rei, "Semi-supervised multitask learning for sequence labeling," in ACL, 2017, pp. 2121–2
[55] Y. Lin, S. Yang, V. Stoyanov, and H. Ji, "A multi-lingual multi-task architecture for low-resource sequence labeling," in ACL, 2018, pp. 799–809.
[56] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu, "Joint extraction of entities and relations based on a novel tagging scheme," in ACL, 2017, pp. 1227–1236.
[57] P. Zhou, S. Zheng, J. Xu, Z. Qi, H. Bao, and B. Xu, "Joint extraction of multiple relations and entities by using a hybrid neural network," in CCL-NLP-NABD. Springer, 2017, pp. 135– 146.
[58] G. Aguilar, S. Maharjan, A. P. L. Monroy, and T. Solorio, "A multitask approach for named entity recognition in social media data," in W-NUT, 2017, pp. 148–153.
[59] N. Peng and M. Dredze, "Multi-task domain adaptation for sequence tagging," in RepL4NLP, 2017, pp. 91–100.
[60] G. Crichton, S. Pyysalo, B. Chiu, and A. Korhonen, "A neural network multi-task learning approach to biomedical named entity recognition," BMC Bioinform., vol. 18, no. 1, p. 368, 2017.
[61] X. Wang, Y. Zhang, X. Ren, Y. Zhang, M. Zitnik, J. Shang, C. Langlotz, and J. Han, "Cross-type biomedical named entity recognition with deep multi-task learning," arXiv preprint arXiv:1801.09851, 2018
[62] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures,"Neural Netw., vol. 18, nos. 5/6, pp. 602–610, 2005.
[63] K. Greff, R. K. Srivastava, J. Koutn´ık, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[64] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," Sensors, vol. 17, no. 2, pp. 273–290, 2017.