

# Deep CNN with Residual Connections and Range Normalization for Clinical Text Classification

Jonah. K. Kenei<sup>1,\*</sup>, Juliet. C. Moso<sup>2</sup>, Elisha T. Opiyo Omullo<sup>1</sup>, Robert Oboko<sup>1</sup>

<sup>1</sup>School of Computing and Informatics, University of Nairobi, Kenya

<sup>2</sup>Department of Computer Science, Dedan Kimathi University of Technology, Kenya

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

**Abstract** Deep learning has achieved remarkable performance in many classification tasks such as image processing and computer vision. Due to its impressive performance, deep learning techniques have found their way into natural language processing tasks as well. Deep learning methods are based on neural network architectures such as CNN (Convolutional Neural Networks) with many layers. Deep learning methods have shown state-of-the-art performance on many classification tasks through several research works. It has shown great promise in many NLP (Natural language processing) tasks such as learning text representations. In this paper, we study the possibility of using deep learning methods and techniques in clinical documents classification. We review various deep learning-based techniques and their applications in classifying clinical documents. Further, we identify research challenges and describe our proposed convolutional neural network with residual connections and range normalization. Our proposed model automatically learns and classifies clinical sentences into multi-faceted clinical classes, which can help physicians to navigate patients' medical histories easily. Our proposed technique uses sentence embedding and Convolutional Neural Network with residual connections and range normalization. To the best of our knowledge, this is the first time that sentence embedding and deep convolutional neural networks with residual connections and range normalization have been simultaneously applied to text processing. Lastly, this work follows a generalized conclusion on clinical documents classification and references.

**Keywords** Text Classification, Document Classification, Unstructured Text, Deep Learning, Word Embeddings, Sentence Embeddings, Convolutional Neural Network

## 1. Introduction

### 1.1. Background and Context

Information technology is playing an increasingly important role in medical practice. One notable example is the use of Electronic Health Records (EHRs) in healthcare delivery settings which has made it possible to collect and store unprecedented amounts of detail clinical data from patients [1]. The increase availability of electronic health data has led to increased interest in computational techniques [2] to harness information from electronic health data. Unstructured narrative text such as clinical notes, pathology notes, radiology reports, discharge summaries, etc. are important documents stored in Electronic Health Records (EHRs) [3] and contains valuable information [3] which can be useful for patient care [4] and clinical research [5], [6] such as extraction of relations from clinical notes [7]. The medical record serves multiple purposes such as communicating and documenting critical information about patients' medical care among health care professionals [8]. A comprehensive patient's clinical record contributes to quality and efficient healthcare of a patient during hospitalization and also during subsequent follow-up visits, as they can provide a complete and accurate chronology of treatments, patient results and future plans of care [9], [10].

Despite their utility, the unstructured nature of clinical documents makes it challenging to retrieve and extract relevant information about patients during care episodes or retrieve relevant information about a population during clinical research. One such challenge is to get a summarized overview of a given patient medical history. Therefore, the wide adoption of EHR systems does not automatically lead to easy access to EHR data for doctors and researchers. Physicians must read through the same way one would their paper counterparts.

Another reason behind limited access stems from the fact that EHR data are composed of personal identifiers,

which is a combination of potentially sensitive medical information, and it induces privacy concerns

Natural language is widely used to document patients' medical histories, because of its expressivity and efficiency [11]. The resulting clinical documentation is critical in medical practice [12], [13] and it serves a number of purposes such as billing purposes, liability protection, disease research and surveillance, and the patient's future health management[14]. Documenting a patient medical history has traditionally been a paper chart note where physicians write free notes on paper. However, paper based clinical records are giving way to electronic versions using electronic health records (EHRs) [15]. Information in a clinical record conveys meaningful medical knowledge and patients' health information. Extracting information from these clinical texts is therefore an important task of information retrieval in clinical domain, and is the key step to understanding a patient medical history. Text classification of clinical documents would pave the way to categorize the medical sentences into multi-faceted categories of information. This would allow physicians to easily analyze, summarize, and explore content of medical documents related to various patients' diseases and conditions.

## 1.2. SOAP Documentation Framework

The Subjective, Objective, Assessment, Plan (SOAP) is a documentation format used by doctors to write clinical notes [16]. Clinical notes are made in a chart by recording interactions with patients which usually documents what transpired during patient visits. In documenting the clinical aspects of a patient, physicians have traditionally used SOAP documentation format [16] for a long time. SOAP is a documentation format used by physicians to organize and interpret clinical documents with the objective of enhancing understandability of the medical history of a patient's hospital visit as described in a clinical report [17] which is usually a medical text chart. The Subjective, Objective, Assessment, and Plan (SOAP) documentation framework is a high-level model used to structure clinical notes. It organizes clinical documentation in a way that supports assessment, reasoning, and decision making [18] and it is widely used method for documenting clinical notes by healthcare providers [19]. SOAP note was started by Larry Weed [20] and its function is to highlight the importance critical documentation details either in paper records or in electronic health records [21]. SOAP is an acronym which stands for Subjective, Objective, Assessment, and Plan.

**S:** Clinically pertinent things that the patient informs a doctor (except how they respond to today's treatment).

**O:** Clinically pertinent observations that a doctor obtains by clinical examination on that patient visit.

**A:** Assessment of how the patient is progressing (based upon doctor's impression of patient subjective plus

objective data from prior visits).

**P:** Today's treatment, treatment response, and any plans for future care.

In this paper, we borrowed from SOAP documentation format the sentence of structuring clinical notes into multi-faceted clinical information by classifying clinical notes into predefined categories of symptoms, diagnosis, disease and treatment. The idea is to model clinical notes according to SOAP clinical reasoning. This was inspired by the fact that, much of this information within the patient encounter, or treatment instance can be categorized into separate sections.

## 1.3. Challenge of Narrative Clinical Text

Given the abundance of electronic clinical documents and their important role in medical practice, there is need for text classification tools which are able to classify them and get relevant useful information [22]. As the amount of unstructured text data that is produced by electronic health records increases, so does the need to intelligently process it and extract different types of information from it.

Clinical notes in electronic health records (EHRs) usually contain important information for medical investigations. A key challenge facing designers of such systems, however, is that the majority of these documents are free text documents often with no fixed structure or any other explicit semantic information [23]. Therefore, computers can't easily process such unstructured data and get meaningful information. However, most clinical data exists in the form of free text narratives. Patient encounters are often recorded in free text format in form of clinical notes [24]. Notes like "Patient appeared to be extremely tired and to be suffering from high fever." can be written in different ways. While this may be easily understood by human reader, it can be very challenging for a computer to understand it. As a result of this issue, much of the natural language processing research in health care now focuses on analyzing and processing text in clinical documents.

Secondly, these documents often contain personal identifying information and protecting confidentiality and privacy of personal health data is a major ethical issue [25], [26].

## 1.4. Overview Text Classification

Text classification is an active research field in machine learning and information retrieval. Due to the increasing growth of information, text classification has become an important technique for organizing unstructured text [27], [28] in many domains such as in healthcare [31]. In general, text classification plays an important role in information extraction, text summarization, text retrieval, medical diagnosis etc. It has been successfully applied to applications such as spam detection, document routing, categorization of web pages, and medical language

processing [29].

It entails assigning various texts to a particular class label. For example, a clinical document usually has words or sentences which can be assigned to a particular class label (say “symptoms”) among other available class labels like “diagnosis”, “disease” and “treatment”. According to Wilcox and Hripcsak (2003) [29], Text classification is widely recognized as a crucial phase in knowledge discovery process and is often the first step in many data analytics applications [30] and a fundamental task in document processing [31].

### 1.5. Clinical Document Classification

In healthcare, clinical documentation in form of clinical notes, are mainly written in natural language and it is useful in answering clinical questions by giving detailed patient health history, clinical reasoning, and clinical conclusion.

Currently, manual text classification is used where; physicians read through a clinical document in order to interpret the meaning of the underlying text, identify the causal relationships between sentences and categorizing them. While this gives users more control over classification, it is usually expensive and time consuming. Automatic document classification uses machine learning algorithms to classify documents automatically. This has the advantage of being faster, scalable and more objective.

The increase use of health information systems has led to a massive accumulation of electronic clinical documents which are mainly made up of unstructured text. While this data may be useful for physicians and researchers, reading through lots of text (such as clinical notes) to get important information is quite challenging which may negatively impact decision making process. Reading these clinical documents may consume a lot of time at the expense of treating the patient. Despite the substantial research interests and efforts devoted to classification of medical documents, automating this task still remains challenging. Clinicians still rely on manual reading, a prohibitively time-consuming process. Physicians spend a lot of time scanning through clinical documents with a view of identifying key problems and getting an overall patient health profile.

Document classification may therefore be helpful in processing clinical documents by extracting important clinical sentences. The increasing availability of electronic clinical notes resulting from the increasing use of EHR systems in healthcare may be processed using text classification techniques [32]. Several researchers such as in [33], [34] and [35] have attempted to leverage the utility of narrative clinical notes using text classification techniques. Deep neural networks have shown promising results for various clinical text classification tasks. However, training neural networks often requires large labeled training data which requires high computational resources, and hyper parameters optimization which is a

time-consuming process.

Recent advances in deep learning have witnessed unprecedented performance in various tasks such as image processing [36], image recognition [37] and more recently natural language processes [38]. It is possible to achieve higher performance, by training deep learning models with a lot of data though with high computational overhead [39], [40]. Several techniques have been proposed to address this problem such as using distributed deep learning technique that has been investigated extensively in recent years [40]. In the recent past, deep learning has shown promising results in various clinical text classification tasks [41] and is emerging as an efficient machine learning technique for classifying text documents. Although deep learning has achieved state-of-the-art results in many challenging tasks, the adoption of deep learning in clinical NLP has been comparatively slow.

Our research goal in this paper is to propose a deep learning model that can automatically classify representations of unstructured clinical documents.

In this paper we describe deep learning model to identify important medical sentences in a patient medical chart and classify them into symptoms, diagnosis, disease, treatment and personal identifying information (PII). We propose the use of convolutional residual with sentence embeddings and range normalization. The input of our model consists of raw medical text charts.

### 1.6. Clinical Relevance

In this study, the current study is relevant because no previous studies have investigated the possibility of using deep learning to classify clinical documents according to multi-faceted information such as the SOAP documentation framework. As mentioned earlier, SOAP framework is used by clinicians to organize and interpret clinical information within a note with the objective of helping to better understand the course of a patient’s healthcare as described in a clinical report.

Many healthcare providers are increasingly using clinical information systems such as electronic health records (EHRs) in healthcare delivery. Therefore, structured and unstructured patient clinical information are collected during clinical encounters with patients and this creates an opportunity to utilize this information to improve the quality of care and research. However, physicians and researchers have limited time and tools to process all the available data for each patient. Machine learning approaches; have been proven to be effective in extracting information from vast amount of data and generalizing to new cases.

In the recent decades, there has been unprecedented growth in electronic health data due to electronic health records, but has not been matched by an equivalent ability to efficiently leverage this data in medical practice and clinical research. It is therefore challenging for unaided

healthcare professionals to use the available electronic data to deliver patient care efficiently. Successfully classifying clinical documents into multi-faceted semantic classes can help in decision making support based on already existing data. In addition, classifying a monolithic document into multi-faceted components related to patients' medical histories may be both time-saving and efficient.

## 2. Literature Review

In this section, we begin with a review of existing techniques for general text classification tasks and classification of clinical documents. This is followed by a review of recent advances in deep learning. Finally, we summarize specific investigations into using deep convolutional networks.

### 2.1. Overview

Using standard electronic health records is a convenient way of obtaining large amounts of medical data [42]. However, most clinical reports are documented using narrative text [42] and most useful data from individual patients are usually in clinical reports in form of narrative text [43]. Therefore, it is challenging to extract information from this data without appropriate pre-processing [44], [45].

A lot of research work has shown that most electronic health records are documented using narrative text [46], [47] and it is always preferred by clinicians as it is the most natural and expressive way for documenting the clinical patients encounters. However, analyzing and re-using this information requires extraction of important clinical information from clinical text. A variety of technologies have been used to extract important information using techniques such as natural language processing (NLP) [46], data mining [46], and machine learning [47]. In addition, text classification techniques have been used to extract useful information from clinical documents [48] and it is currently an active research field in information retrieval and machine learning.

### 2.2. Text Classification

Researchers have studied and applied a variety of methods for text classification problems. From the literature, a number of statistical classification methods and machine learning techniques have been applied to text classification problems. Common examples includes Naïve Bayes (NB) [49], K-Nearest Neighbor (KNN) [50], [51], Decision Tree (DT) [52], Support Vector Machines (SVM) [53], [54], Centroid Based Classification (CBC) [55], [56], [57], [58], classifier ensemble approaches [59] and maximum entropy [60]. Some of the widely used text classification techniques found in the literature includes Support vector machine, K-nearest neighbor and naive

Bayesian techniques. Trstenjak B, et al. [61] used K-nearest neighbor (KNN) technique with term frequency-inverse document frequency (TF-IDF) method. The use of KNN and TF-IDF gave good performance results.

Recently, deep learning [62] approaches have achieved high performance compared to traditional algorithms [63]. Deep learning has shown major state-of-the-art developments in many domains, such as speech recognition [64] and computer vision [63], [65], image processing [66], [67].

Convolutional neural network (CNN) [68], [69] and recurrent neural network (RNN) [70], [71] are often used and they are powerful in learning representations of texts [72]. From the literature, a number of deep learning techniques have been applied to general text classification problems such as in [68], [69], [73], [74], [75], [76] and [77] and they have proven to be state-of-the-art methods in text classification [78].

### 2.3. Clinical Text Classification Using Deep Learning

In the recent past, research works on text classifications of clinical documents using deep learning such as in [79], [80] and [81] are beginning to emerge. Applying deep learning techniques to classify clinical documents in harnessing the unprecedented amount of unstructured clinical information associated with each patient encounter, such as clinical notes, discharge summaries etc.

Miotto, R et al [82] reviewed deep learning opportunities and challenges in health care domain. They proposed that deep learning techniques could be used for processing big biomedical data which can be used to improve human health [82]. Lee, G.E. (2017) [83] presented a convolutional neural network for classifying medical documents for systematic reviews. The objective was to improve the laborious task of identifying eligible documents for systematic reviews [83]. Hughes, M et al [84] presented a CNN model for classifying medical documents at sentence level. He demonstrated that CNNs can represent the semantics of clinical text enabling semantic classification at a sentence level. Choi, E. et al [85] presented a RNN based model that uses patient history to predict diagnoses and medications during patients' subsequent visits. It assesses used patient's medical history to make multi-label predictions [85]. The state of art development in text classification which is becoming very common is the use of CNNs with word embedding language models for text classification tasks [86], [87].

### 2.4. Challenges in the State of Art

In the last few years, machine learning approaches have been shown to be efficient and effective in clinical text classification tasks [88], [89]. However, successful machine learning model requires a lot of labeled training data which is manually created. This problem is magnified in the healthcare domain, mainly due to:

- i) Lack of publicly available clinical corpora due to patient privacy concerns[90], and
- ii) The requirement of domain medical knowledge to accurately create labeled training data from clinical text.

And therefore, popular methods for creating labeled data, such as crowdsourcing, are not applicable in clinical domain [90].

In the last few years, a combination of deep neural networks and its derivatives (CNN and RNN) and distributional word representations (word embeddings) have been preferred for the clinical documents classifications for the following reasons:

- i) The ability of word embeddings to capture the semantic relationships of words[91]
- ii) The ability of deep learning methods to make use of available data to extract important features without requiring a lot of labeled data (NLP) [92].

The use of traditional machine learning techniques requires a large amount of labeled data to train reliable predictive models [93]. In addition, traditional text representation techniques (Bag of words) are unable to capture the semantic properties and linguistic relationships between words which are very important in clinical documentations. Bag of words word representation is not therefore suitable for modeling clinical sentences because clinical sentences are often expressed using multi-word sentences e.g. “high blood glucose” and “high fever”. Human interpretation of a text is based on the meaning of sentences rather than on the individual words. Therefore, the use of a combination of word embeddings [91] which captures the semantic relationships of words and deep learning methods that make use of available text data have emerged as a state of art solution to the above problems (NLP) [92]. Recently, techniques based on Deep Neural Networks (DNNs) have achieved great performance in classification tasks in a wide range of applications, such as natural language processing [94]. Examples of deep learning applications in healthcare which have shown promising results; include mortality prediction [95], patient note de-identification [96], skin cancer detection [97] and diabetic retinopathy detection [98]. However, apart from the state of art performance of neural networks, it suffers the following problems:

- i) **Overfitting** - Deep learning major problem is overfitting [99] which occur, when a model learns well the training data to the extent that it negatively impacts the performance of the model on new data [100]. Usually, the goal of machine learning algorithm is to perform well on the training data and generalize well to new data. Overfitting negatively affects the generalizability of the model.
- ii) **The problem of covariate shift** - In addition, training deep learning models is difficult and getting them to converge in a reasonable amount of time is still

challenging [101], [102]. A well-known problem of deep neural networks is the problem of covariate shift (Ioffe and Szegedy), which negatively affects the speed of training deep neural networks [103]. This is attributed to the changing nature of layer's inputs during training, as the parameters of the previous layers change. This slows down the training by requiring lower learning rates and careful parameter initialization [101].

- iii) **Vanishing gradient problem** - Deep neural networks have been shown to be more effective in classification tasks. However, increasing the depth of a neural network often makes it more difficult to train, due to the vanishing gradient problem [104].

Thus, there is need to come up with new techniques which solve the above problems in order to come up with models with better training rate (learning rates) and higher classification accuracy to support clinical text classification problems.

## 2.5. Regularization Techniques

The use of convolutional neural network has been successful in many fields such as computer vision [105] and natural language processing [106], [107] and proving to be remarkable in text classification tasks, achieving state-of-the-art results [107],[108], [109], (110). Although, CNNs have shown remarkable performance, training them is difficult due to the problems of overfitting, vanishing gradient and covariate shift as explained above. In the recent past, it has been shown that increasing the depth of CNN architectures further aggravates the problems [111].To address these challenges, regularization have been proposed [112]. Some of the commonly used regulation techniques include dropout [65], batch normalization [102] and kernel regularization [113]. In addition, deep residual learning [104] has been proposed as a technique to overcome vanishing gradient problem in deep networks [111]. However, it has been shown that ResNets does not resolve the vanishing gradient problem [114]. As a result, training deep neural networks still remains an open research problem. Therefore, proposing better techniques to overcome the above challenges is very important. In this paper, we introduce a novel range normalization technique to alleviate training difficulties of deep neural networks, while giving better performance over batch normalization.

In the last few years previous attempts by researchers have addressed the above problems using batch normalization [102] which helps in accelerating deep neural network training and accuracy. It works by normalizing activations in intermediate layers of deep neural networks [115].

It is currently the most widely adopted technique to accelerate and stabilize training of deep neural networks

[116]. However batch normalization has some limitations such as the inability to consider the underlying distribution of the parameters and the nodes [117]. In addition, it depends on batch statistics for layer wise input normalization during training which makes the estimates of mean and standard deviation of input (distribution) to hidden layers inaccurate for validation due to shifting parameter values [118]. Batch normalization has also been shown to require significant computational overheads [119].

A more recent research by Lian and Liu [120] pointed out that the performance of batch normalization depends on how well the batches are constructed during training, and it may not converge to a desired solution if the statistics on the batch are not close to the statistics over the whole dataset [120]. They further demonstrated four issues with batch normalization namely; BN fails/overfits when the mini-batch size is 1, BN's solution is sensitive to the mini-batch size and BN fails if data are with large variation [120].

## 2.6. Overview of Our Contributions

As mentioned earlier, training deep neural networks have been shown to be challenging. Increasing the depth of a neural network often makes it more difficult to train, due to the problems of overfitting, vanishing gradient and covariate shift. In this paper we propose a novel regularization technique called range normalization to help overcome the above problem. We will then apply this technique in classifying clinical documents and comparing with batch normalization which is our baseline. We will then carry out a set of experiments, by comparing the performance of range normalization against batch normalization. To the best of our knowledge, this is the first time that range normalization has been used as a regularization technique. After each convolutional layer the range normalization is applied this maps the data to a desired range (ex. [0, 1] in our case) through a simple linear transformation.

## 3. Description of Our Proposed Model

### 3.1. Overview

In this section, we describe an ongoing project of developing a deep learning model for classifying clinical documents into four clinical sentences namely symptoms, diagnosis, disease/assessment and treatment. These four clinical sentences are based on SOAP documentation format which has four key sections as described above. In addition, we include the personal identifying information such as patient name etc.

We used the generic sentenceualization of Weed's SOAP documentation format to classify clinical sentences according to the Subjective, Objective, Assessment, Plan

(SOAP) headings to classify elements of these notes. We looked especially at symptoms, diagnosis, disease and treatment and how to model these according to SOAP clinical reasoning. This was inspired by the fact that, much of this information within the patient encounter, or treatment instance can be categorized into separate sections. The physicians use a well-known methodology to divide their notes, known as SOAP (Subjective, Objective, Assessment, and Plan) [121].

**Table 1.** Medical Sentences mapping to SOAP standard

Medical Sentence	SOAP Format
Symptoms	S: subjective
Diagnosis	O: objective
Assessments/Conclusions	A: Assessment
Treatments	P: plans

The document is viewed as being made up of several clinical sentences and each sentence belongs to a particular semantic class of sentences such as symptoms, diagnosis, conclusions/disease and treatment. Given a patient medical text chart and a set of semantic medical classes in medical practice, we need to build a classifier that can classify the various medical sentences in the medical chart into the appropriate clinical sentence classes.

We proposed a deep learning algorithm, using convolutional neural network with residual skip connection and range normalization and sentence embeddings for clinical sentences classification. The novelty in the algorithm lies with the skip connections and range normalization.

### 3.2. Formulation of the Problem

We formally restate the clinical document classification problem as follows:

Given a clinical document with a set of clinical sentences,  $S = \{s_1, s_2, s_3, \dots, s_n\}$  and a set of classes,  $C = \{c_1, c_2, c_3, c_4, c_5\}$  and class assignments  $\{s_i, y_i\}$  with  $y_i \in C$  for all  $1 \leq i \leq 5$ , how can we create a classifier  $C$  which correctly predicts the set of classes  $y_i$  for each clinical sentence  $s_j \in S_{\text{test}} \subset S$ , based on class assignments of  $S_{\text{train}} \subset S$ . While considering the fairness condition  $S_{\text{train}} \cap S_{\text{test}} = \emptyset$ .

Where

- $c_1$ : Symptoms
- $c_2$ : Diagnosis
- $c_3$ : Assessment/Disease/Conclusions
- $c_4$ : Treatment
- $c_5$ : PII (Personal identifying information e.g. patient name, date of birth etc.)

We can generalize as follows, given a dataset of pairs  $(s_j, c_j)$  where

- $s_j$  is a sentence of one or more words and
- $c_j$  is one of the possible classes for the sentence i.e.  $c_j = \{\text{Symptoms, Diagnosis, Disease, Treatment and PII(Personal identifying information)}\}$

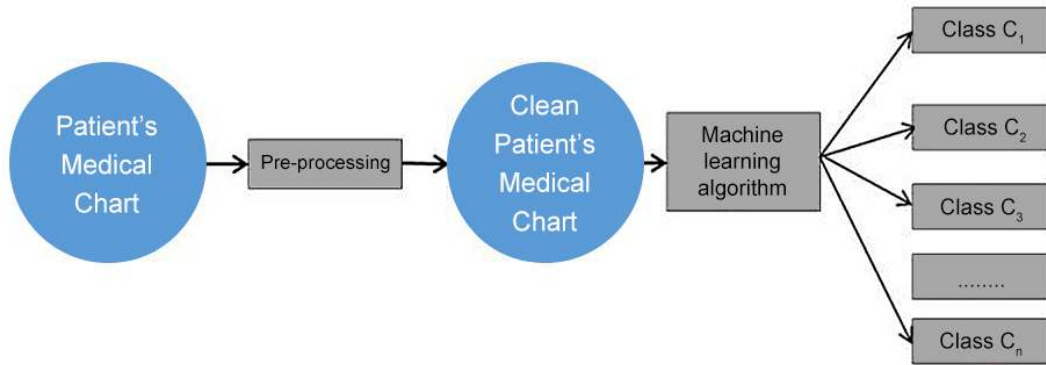


Figure 1. Clinical document visualization architecture

Our objective is to design a function or a classifier  $F$  which maps an arbitrary sentence (clinical sentence) to one of the possible clinical sentence classes.

To begin with, we discuss the general problem, where we have a classification problem with five classes. This means we are given  $n$  samples

$$X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$$

and the classes

$$Y = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$$

Where  $\mathbf{y}_i \in \{1,2,3,4,5\}$ . We use a neural network to model the probability  $P(c_j | \mathbf{x}_i)$  of a class  $c_i$  given sample  $\mathbf{x}_i$ .

We then estimate our prediction as:

$$\mathbf{y}_i = \mathop{\text{argmax}}_{j \in \{1,2,3,4,5\}} P(c_j | \mathbf{x}_i).$$

Now we set up a simple neural net with 5 output nodes, one output node for each possible class.

### 3.3. Architecture of Our Proposed Model

Our objective is to design a function or a classifier  $F$  which takes a patient medical text chart as input and maps an arbitrary sentence (clinical sentence) to one of the possible clinical sentence classes (Symptoms, Diagnosis, Disease, Treatment, PII (Personal identifying information)) as illustrated in table 2 below.

Table 2. Clinical sentences classification

Clinical Sentence	Classification
High fever	Symptom
Fatigue	Symptom
Frequent urination	Symptom
Frequent infections	Symptom
Pressure: 113/79 mmHg	Diagnosis
Temperature:37	Diagnosis
Heart Rate: 57 bpm	Diagnosis
Diabetes	Disease/Conclusion
Insulin therapy started	Treatment
Regular physical exercise	Treatment

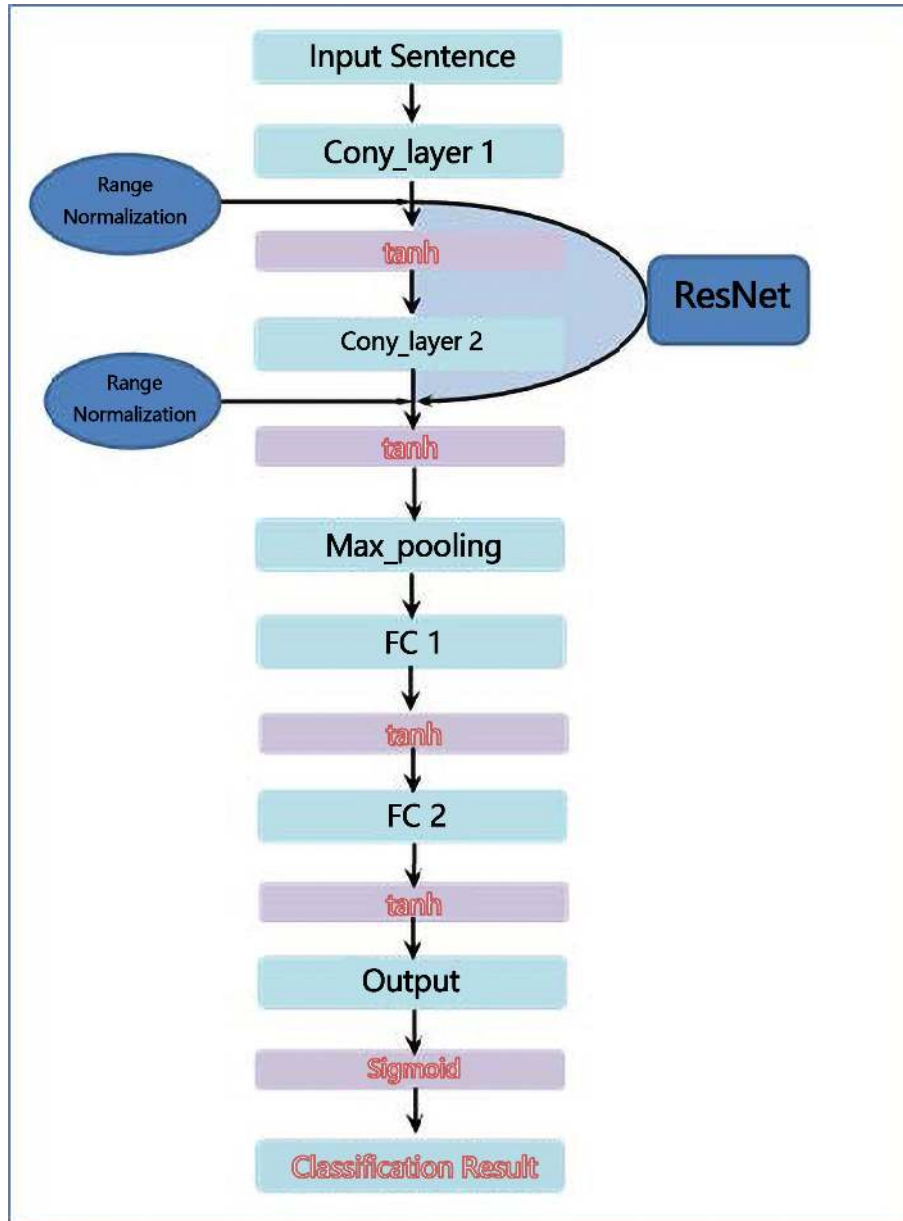
In our proposed model, the clinical sentences (sentences) in a patient medical chart are classified into *symptoms*, *diagnosis*, *conclusion* (disease) and *treatments*, using a convolution deep convolutional neural network (CNN). The input to our model is a patient medical chart which contains narrative text describing patient’s medical history. The narrative text is a doctor’s description, which briefly describes information about the patient health profile in terms of symptoms, diagnosis, disease/condition and treatment according to SOAP documentation format. The outputs (labels) are symptoms, diagnosis, disease/condition and treatment. For example, the clinical text for *symptoms* will be “*Extreme hunger, frequent infections, unexplained weight loss, increased thirst, frequent urination, irritability, slow-healing sores, fatigue, blurred vision*”, which consists of many word-combinations that are separated by comma and have to be classified as symptoms, accordingly. In this case, the clinical text is made of 4 words or less. And for *diagnosis*, the clinical text will be “*Auto-antibodies detected in blood. Patient had an A1C level of 7.0. Ketones detected in urine.*”, and each sentence is separated by “.” has to be classified as diagnosis, accordingly. So is the clinical text for *treatment* such as “*Patient advised to cut weight. Regular monitoring of sugar levels prescribed. Patient advised to center their diet on more fruits, vegetables and whole grains and to cut down on animal products, refined carbohydrates and sweets. Participation in regular physical activities recommended. Insulin therapy started.*” In addition, the clinical chart usually contains personal identifying information (PII) such as patient names, age, sex and date. Similarly, this has to be classified as PII.

### 3.4. Convolutional Neural Network Design

The architecture of our proposed model is summarized in figure 2 below. It has 2 convolutional layers, 1 max pooling layer, 2 fully connected layers (marked as FC), and activation layers after every convolutional and fully connected layer and a softmax layer at the end of network. Our proposed model is a deep convolutional neural network with residual skip connection and range

normalization using sentence embeddings to classify clinical sentences (sentences). The novelty in the proposed model is the idea of using skip connections and range normalization simultaneously. ResNet employs skip connections, or residual layers, which preserve the information from early layers and provide it to the later

layers directly, skipping the intermediary layers. This is done by simple addition of the output of the earlier layer with the output of the layer immediately before the later layer – that is, what would have been the input to the later layer if no skip connections were present.



**Figure 2.** Convolutional neural network model and its components



### 3.4.1. Residual Network (ResNet)

Figure 3 shows the Residual Network used in our model; where the output of the 1<sup>st</sup> convolutional layer  $x_1$  is the input of the 2<sup>nd</sup> convolutional layer via activation layer (tanh),  $x_1$  will be summed directly with the output of the 2<sup>nd</sup> convolutional layer  $x_2$  before activation layer (tanh). The main idea of ResNets is to connect some of the layers with shortcuts, which can avoid vanishing and exploding gradients problems. These problems occur in very deep networks. The shortcut connections have the ability to explicitly let these layers fit a residual mapping with the help of identity transformation. The residual block is defined as:

$$x_2 = F(x_1)$$

$$x_3 = \tanh(F(x_1) + \text{id}(x_1))$$

where

- 1)  $F(\cdot)$  function represents Tanh (after 1<sup>st</sup> convolutional layer) and the 2<sup>nd</sup> convolutional layer;  $x_3$  is the value after activation,
- 2)  $\text{id}(\cdot)$  is an identity mapping function (ResNet);
- 3)  $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$  is the activation function for output of Word Residual block (see Fig. 3).

If the dimensions of  $x_1$  and  $x_2$  are the same, the last equation will simply change as follows:

$$x_3 = \tanh(F(x_1) + x_2)$$

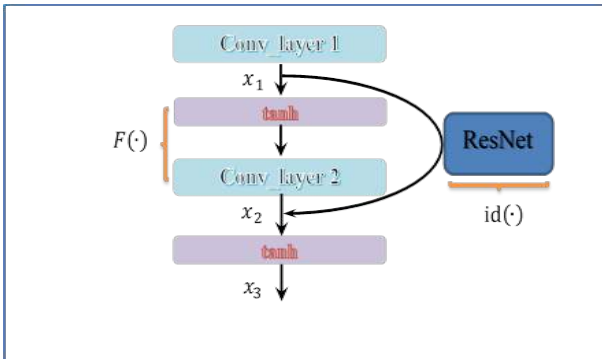


Figure 3. Residual and identity mapping in ResNet

Because of the constant dimension of input ( $n \times m$ ) matrix for our model, the input embedding data contains a lot of zeros. For example, if “unexplained weight loss” (symptoms) is inputted into our model, it only has 3 words; the first 3 row will not be vanishing, but the remaining  $n - 3$  rows must be filled by zeros to let input be  $n \times m$  matrix where  $n$  is the maximum number of words in all the possible input text to be classified. The ResNet in our model help us with speed convergence not only for deepness of our model, but also for sparse input matrix;

### 3.4.2. Range Normalization

After each convolutional layer the range normalization is applied. The objective of applying range normalization is to help map the data in a desired range e.g. [0...1] in our case) through a simple linear transformation. Assume there are  $n$  values  $x_i$  ( $i=1, 2, \dots, n$ ) which have to be mapped in a range [0...1]. Then the result  $y_i$  ( $i=1, 2, \dots, n$ ) can be obtain using the following equation:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

Where  $x_{\min}$  is the minimum value of  $x_i$  and  $x_{\max}$  is the maximum value.

Because this is, of course, the linear function, the minimum value of  $y_i$  is obtained when  $x_i$  is equal to  $x_{\min}$  and so is the maximum value of  $y_i$ . If we substitute the  $x_{\min}$  into  $x_i$ , we obtain  $y_{\min} = 0$ . Similarly, we also obtain  $y_{\max} = 1$ . The new numbers  $y_i$  will therefore be distributed in the range [0, 1] as desired.

Note that there may exist the special case,  $x_{\min} = x_{\max}$ , which means all the values of  $x_i$  are the same ( $x_i \equiv x_{\max}$ ). In this case, the denominator of the above equation is equal to 0, which yields “the division by zero.” In such a case, we can set  $y_i$ ’s to 0.5. Range normalization can simply be implemented and does not increase the complexity of the neural network.

There is more important reason for applying the range normalization. And due to the reason mentioned above, the input matrix is too sparse; it has a lot of elements with vanishing values. The output of each convolutional layer will be very small and extremely vanishes after activation layers (tanh), due to which the relevant final classification results can’t be obtained and the convergence of our CNN model will be not ensured. In order to avoid this point, a relevant treatment on the output of each convolutional layer has to be necessary; i.e. normalization.

There are two types of normalizations; batch normalization and range normalization. The former has been succeeded in image processing field such as classification, face recognition and verification, but it has been shown to exhibit poor performance for the NLP tasks, especially text classification which may result in the sparse input matrix. The latter can help us with the extension of outputs from vanishing region to the relevant region [0, 1] via simple linear transformation; this can solve vanishing problem!

### 3.5. Algorithm Overview

The goal of this phase was to transform clinical narrative documents into quantitative features for training. To do this an algorithm was designed and implemented. The algorithm takes clinical sentences as input and classifies them, outputting a class for each sentence indicating which class it belongs to. The main algorithm can be divided into four modules which can be run independently:

### 3.5.1. Pre-processing

To create word vectors of the input sentences, we first need to do raw document pre-processing. The input sentences are separated into words; the sentence may be composed of various types of punctuation marks such as “,” “.” “:” “/” “(” “)” “-” ... and these marks should be removed; in our code, they are replaced by space “ ”; then the input sentence will be separated into words by space. Sentences are separated using sentence tokens located at the end of each sentence. For each word in the input sentence, checking if it is in the word2vec dataset or not should be performed; if yes, the corresponding word2vec vector will be inputted in the input matrix, but if no, the empty row with zeros will be added. If the number of words is larger than 40, the last words (besides first 40 words) will be ignored.

### 3.5.2. Word Embeddings

The next step after design of a deep learning model is word-embedding which transforms the words into numerical vector representations. Word embeddings which is also called distributional vectors follow the distributional hypothesis, where words with similar meanings tend to occur in similar context. Thus, these vectors always capture the characteristics of the neighbouring words. It has the advantage of capturing similarity between words. The similarity is measured using similarity measures such as cosine similarity.

In this case, pre-trained GloVe (Global Vectors for Word Representation) word vectors are used in this paper for word-embedding. GloVe is an unsupervised machine learning algorithm, the purpose of which is obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations results in interesting linear substructures of the word vector space.

The glove.6B.zip has been chosen; it contains several dimensional word2vec vectors (for example: 50D, 100D, 200D, 300D) and the 50D vector library has been selected for this classification of patient medical chart; i.e. each word is converted into 50D vectors with the positive and negative values; for example, “and” is converted into vector:

### 3.5.3. Sentence Embeddings

As mentioned before, clinical documents can be described by a number of clinical sentences such as symptoms, diagnosis, disease, treatment or personal identifying information such as names. We therefore need a Machine learning (ML) models which can be trained to automatically map documents to these clinical sentences, allowing classifying very large text collections, more than could be processed by a human being.

Before training the neural network and using it for sentence classification tasks, we first generate a numerical representation of the text documents that will serve as an input to the neural classifier. To achieve this, we map each individual word in the document to a vector embedding, and concatenate these embeddings to form a matrix of size the number of words in the document times the dimension of the word embeddings.

Just like word vectors, the same sentence can be applied to sentences where vectors with similar sentences lie close to each other in a high-dimensional space. In this project, we use Convolutional neural networks to extract fix length features from word sequences, which is a patient medical chart in our case. In our approach, we use a pre-trained word embedding model which maps each word to a vector representation. The vectors for multiple word embeddings belonging to the same sentence are then combined to form a single sentence embedding. The patient chart is usually made up of two or more sentences expressing some medical sentences which can be symptoms, diagnosis, disease or treatments. Therefore, we treat each medical expression as a single sentence, separating them using sentence tokens located at the end of each sentence. Similarly, a sentence is made up of two or more words. Therefore, a sentence  $S$  with  $n$  words  $w_1, w_2, \dots, w_{n-1}, w_n$  is mapped into a set of  $n$ -dimensional real-valued vectors  $x_1, x_2, \dots, x_{n-1}, x_n$  using word embedding. Each word is mapped into a single word-vector  $x_i$  where  $i$  ranges from 1, 2 ..... $n$ . To obtain the vector representation of sentence  $S$ , we concatenate the word vectors  $w_1, w_2, \dots, w_{n-1}, w_n$  to obtain a sentence vector  $X$ . This is summarized in figure 4 as shown below.

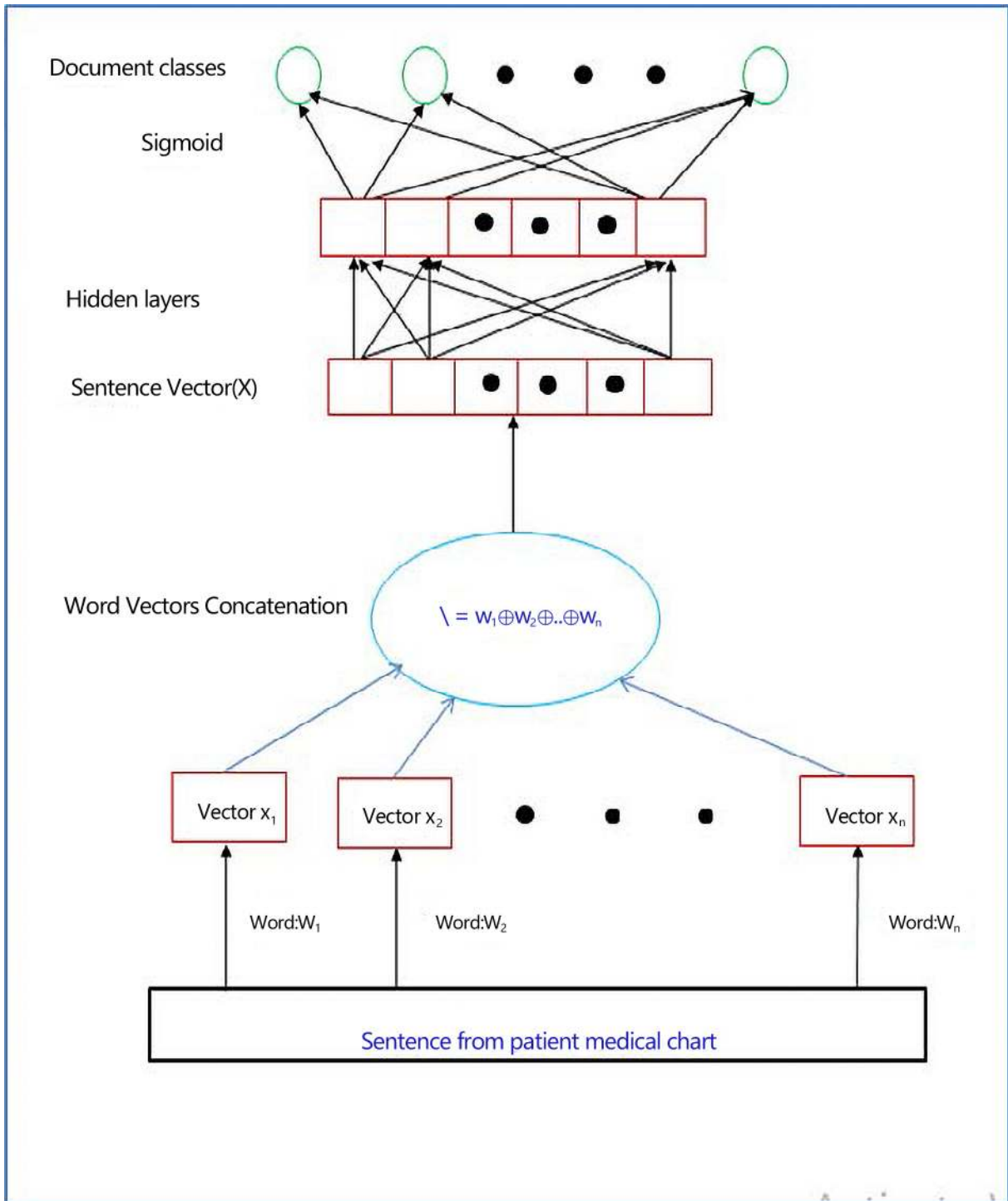


Figure 4. Creating sentence vectors

### 3.5.4. Clinical Sentences Classification

The classification takes the whole dataset of sentence embeddings and classifies them in order to form distinct categories of clinical sentences. The clinical sentence classes are returned as the output of the algorithm. In the following section the proposed algorithm is presented in detail.

Our model receives the clinical text and the corresponding categories; for example, “*unexplained weight loss*” and *symptoms*. The former can be represented via word-embedding procedure as shown below and the latter can be represented by one-hot representation (1, 0, 0, 0, 0), where 1 means the input sentence is classified into *symptoms* and the remaining zeros denotes it is not classified into any sentence class. The former will be fed into our model, while the latter (each element denoted as:  $\mathcal{Y}_1^*, \mathcal{Y}_2^*, \mathcal{Y}_3^*, \mathcal{Y}_4^*, \mathcal{Y}_5^*$ ) will be used for evaluation of loss function and training of our model.

The input sentence e.g. “*unexplained weight loss*” has 3 words *unexplained*, *weight* and *loss* each will be transformed into  $n$ -dimensional (50-dimensional) word2vec vector using word-embedding.

Let  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$  represent the individual words in the sentence. Therefore “*unexplained weight loss*” is converted into  $3 \times 50$  matrix  $X$  as shown below.

$$X = \mathbf{w}_1 \oplus \mathbf{w}_2 \oplus \mathbf{w}_3$$

Where  $\oplus$  denotes concatenation of word2vec vectors and the  $i^{\text{th}}$  column of  $X$  is the embedding vector corresponding to word  $w_i$ . Because of different lengths of input sentences, the resultant matrix  $X$  has different number of rows, which is not valid for CNN model; therefore, the size of input matrix has to be fixed into  $n$  rows and 50 columns. The words in input sentence can't be

omitted; that is,  $n$  must be the largest number of words in all the possible input sentences. For the shorter sentence, the blanks (corresponding to row vector with 50 zeros) can be added to obtained fixed matrix size of input. The input matrix will be too sparse since the input sentence usually has 2-5 words.

For the 1<sup>st</sup> convolutional layer (Conv\_layer 1), the filter  $\Lambda_1 \in \mathbb{R}^{3 \times 50}$  with a window of 3 words is used, which produces feature maps (as shown in Fig. 5):

$$x_1 = X * \Lambda_1 + \epsilon_1$$

where  $x_1 \in \mathbb{R}^{38 \times 1}$  (i.e. row vector) with valid padding [due to valid padding, the number of rows will be  $40$  (original number)  $- 3$  (size of filter window)  $+ 1 = 38$ ],

- $*$  is the convolution operator,
- $\epsilon_1 \in \mathbb{R}^5$  is bias factor which appears due to sparseness of the matrix  $X$ .

The range normalization and activation follow after this convolutional layer and gives  $x_1^*$ . We used 5 filters for the first convolutional layer.

For the 2<sup>nd</sup> convolutional layer (Conv\_layer 2), the filter  $\Lambda_2 \in \mathbb{R}^{3 \times 1}$  with a window of 3 words is used, which gives produces feature maps (as shown in Fig. 6):

$$x_2 = x_1^* * \Lambda_2$$

where  $x_2 \in \mathbb{R}^{38 \times 1}$  with same padding (same padding gives us the same number of rows),  $*$  is the convolution operator, and  $\epsilon_2 \in \mathbb{R}^5$  is bias factor. We also used 5 filters for the first convolutional layer. After the range normalization, the output  $x_1$  of the 1<sup>st</sup> convolutional layer is added before activation layer (Resnet). After Max-pooling layer, all the multiple features are concatenated into one row vector before 2 FC layers.

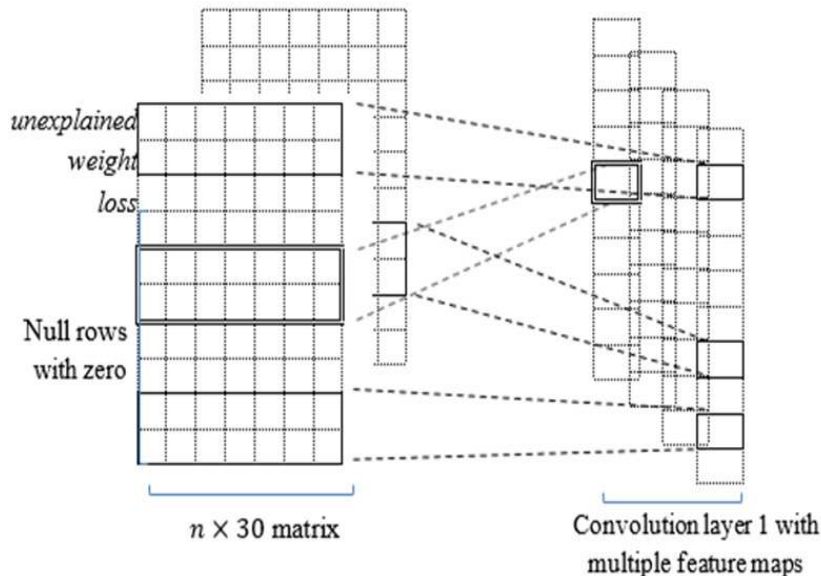


Figure 5. Convolution for input matrix

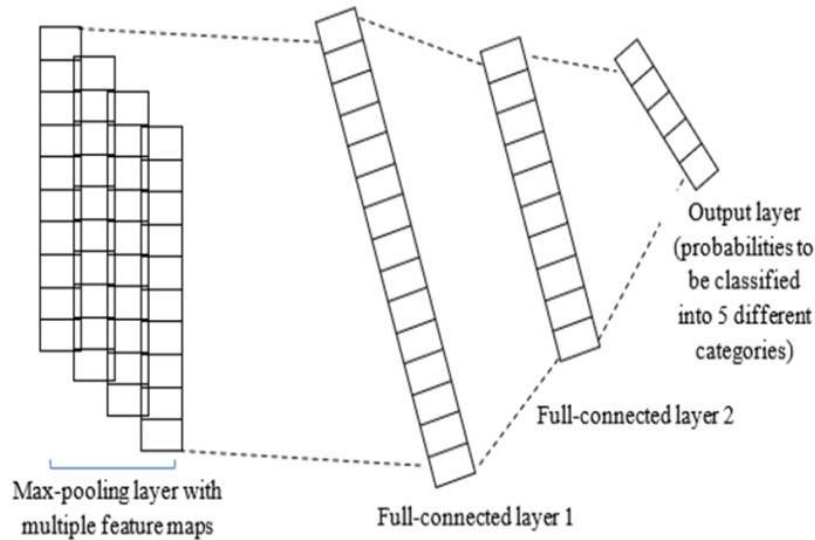


Figure 6. Convolution for output matrix

After 2 full-connected layers, the final activation layer with sigmoid function

$$\sigma(t) = e^t / (1 + e^t)$$

will give the probabilities  $y_1, y_2, y_3, y_4, y_5$  to be classified into 5 types of medical sentences(categories).

#### 4. Proposed Evaluation Experiments

As mentioned earlier, training deep neural networks have been shown to be challenging. Increasing the depth of a neural network often makes it more difficult to train, due to the problems of overfitting, vanishing gradient and covariate shift. In this paper we have proposed the use of range normalization to help overcome the above challenges. In order to determine, the viability of our model, we propose a series of experiments; which compares the performance of range normalization against batch normalization. To evaluate the performance of the two regularization techniques, we propose a series of experiments while comparing cross entropy and classification accuracy for the two techniques.

#### 5. Conclusions and Future Work

In this work, we have presented a deep learning clinical document classification model for improving the laborious task of identifying important clinical sentences in a patient medical chart. To the best of our knowledge, this is the first time that sentence embeddings and deep convolutional neural networks with residual connections and range normalization have been used simultaneously.

Our features are simple: the glove language model is used to generate word embeddings and the word

embeddings are concatenated to produce sentence embeddings. We believe that it is possible to classify patients' clinical charts into multi-faceted clinical information. This research work is a part of on-going work for designing a clinical document classification and visualization model for classifying and visualizing clinical sentences to aid physicians in reviewing clinical notes. Classification and visualisation models play a vital role in identifying facts without wasting time on reading the whole document.

#### Acknowledgements

Our completion of this paper could not have been accomplished without the support of staff members who are working with us. We are very much thankful to them. For the references, we referred many articles and research papers of many authors and institutions which are available in online and offline. We offer our sincere appreciation for the learning opportunities provided by those authors and institutions. Last but not least, our heartfelt thanks to all our family members for caring us and for the huge support.

#### REFERENCES

- [1] Zhang, X., Heno, R., Gan, Z., Li, Y., & Carin, L. (2018). Multi-Label Learning from Medical Plain Text with Convolutional Residual Models. CoRR, abs/1801.05062.
- [2] [6] Choi, E., Biswal, S., Malin, B.A., Duke, J., Stewart, W.F., & Sun, J. (2017). Generating Multi-label Discrete Patient Records using Generative Adversarial Networks. MLHC.

- [3] Wang, S., Fu, L., Yao, J., & Li, Y. (2018). The Application of Deep Learning in Biomedical Informatics. 2018 International Conference on Robots & Intelligent System (ICRIS), 391-394.
- [4] Richard S. Dick, E. (1997). *The Computer-Based Patient Record: Revised Edition: An Essential Technology for Health Care*. National Academies Press. Washington (DC), US
- [5] Moor, G.D., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., Karakoyun, T., Ohmann, C., Lastic, P., Ammour, N., Kush, R., Dupont, D., Cuggia, M., Bozec, C.D., Thienpont, G., & Coorevits, P. (2015). Using electronic health records for clinical research: The case of the EHR4CR project. *Journal of biomedical informatics*, 53, 162-73.
- [6] Bowles, K. H., Potashnik, S., Ratcliffe, S. J., Rosenberg, M., Shih, N. W., Topaz, M., Holmes, J. H., ... Naylor, M. D. (2013). Conducting research using the electronic health record across multi-hospital systems: semantic harmonization implications for administrators. *The Journal of nursing administration*, 43(6), 355-60.
- [7] Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., & Liu, H. (2018). Clinical information extraction applications: A literature review. *Journal of biomedical informatics*, 77, 34-49.
- [8] Wichita.va.gov. (2019). [online] Available at: [https://www.wichita.va.gov/documents/3\\_Documentation\\_of\\_Medical\\_Records.pdf](https://www.wichita.va.gov/documents/3_Documentation_of_Medical_Records.pdf) [Accessed 7 Feb. 2019].
- [9] Biondich, P.G., Anand, V., Downs, S.M., & McDonald, C.J. (2003). Using Adaptive Turnaround Documents to Electronically Acquire Structured Data in Clinical Settings. AMIA ... Annual Symposium proceedings. AMIA Symposium, 86-90.
- [10] Dexter, P.R., Perkins, S., Overhage, J.M., Maharry, K.S., Kohler, R.B., & McDonald, C.J. (2001). A computerized reminder system to increase the use of preventive care for hospitalized patients. *The New England journal of medicine*, 345 13, 965-70.
- [11] Friedman C., Elhadad N. (2014) Natural Language Processing in Health Care and Biomedicine. In: Shortliffe E., Cimino J. (eds) *Biomedical Informatics*. Springer, London
- [12] Mathioudakis, A., Rousalova, I., Gagnat, A. A., Saad, N., & Hardavella, G. (2016). How to keep good clinical records. *Breathe (Sheffield, England)*, 12(4), 369-373.
- [13] Ngo, E., Patel, N.J., Chandrasekaran, K., Tajik, A.J., & Paterick, T.E. (2016). The Importance of the Medical Record: A Critical Professional Responsibility. *The Journal of medical practice management: MPM*, 31 5, 305-8.
- [14] Davidson, S.J., Zwemer, F.L., Nathanson, L.A., Sable, K.N., & Khan, A.N. (2004). Where's the beef? The promise and the reality of clinical documentation. *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine*, 11 11, 1127-1134.
- [15] Roberts, A. (2017). Language, Structure, and Reuse in the Electronic Health Record. *AMA journal of ethics*, 19(3), 281-288.
- [16] Reznich, C.B., Wagner, D.P., & Noel, M.M. (2010). A repurposed tool: the Programme Evaluation SOAP Note. *Medical education*, 44 3, 298-305.
- [17] Mowery, D., Wiebe, J., Visweswaran, S., Harkema, H., & Chapman, W. W. (2011). Building an automated SOAP classifier for emergency department reports. *Journal of biomedical informatics*, 45(1), 71-81.
- [18] Weed L. *Medical records. Medical education and patient care: the problem oriented record as a basic tool*. Year Book Medical Publishers: Press of Case Western Reserve University Cleveland; 1970
- [19] Lew V, Ghassemzadeh S. SOAP Notes. (2018) 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls <https://www.ncbi.nlm.nih.gov/books/NBK482263/>
- [20] Wright, A., Sittig, D. F., McGowan, J., Ash, J. S., & Weed, L. L. (2014). Bringing science to medicine: an interview with Larry Weed, inventor of the problem-oriented medical record. *Journal of the American Medical Informatics Association: JAMIA*, 21(6), 964-968.
- [21] Pearce, P.F., Ferguson, L.A., George, G.V., & Langford, C.A. (2016). The essential SOAP note in an EHR age. *The Nurse practitioner*, 41 2, 29-36.
- [22] Li, Y., Gorman, S.L., & Elhadad, N. (2010). Section classification in clinical notes using supervised hidden markov model. *IHL*.
- [23] Moen, H., Peltonen, L., Heimonen, J., Airola, A., Pahikkala, T., Salakoski, T., & Salanterä, S. (2016). Comparison of automatic summarisation methods for clinical free text notes. *Artificial intelligence in medicine*, 67, 25-37.
- [24] Townsend, H (2013). Natural Language Processing and Clinical Outcomes: The Promise and Progress of NLP for Improved Care. *Journal of AHIMA* 84(2) 44-45.
- [25] Barrows, R.C., & Clayton, P.D. (1996). Privacy, confidentiality, and electronic medical records. *Journal of the American Medical Informatics Association: JAMIA*, 3 2, 139-48.
- [26] Institute of Medicine (U.S.), Donaldson, M. S., & Lohr, K. N. (1994). *Health data in the information age: Use, disclosure, and privacy*. Washington, D.C: National Academy Press.
- [27] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1-47. doi: 10.1145/505282.505283.
- [28] Vasa, K. (2016). *Classification through Statistical and Machine Learning Methods: A Survey*. Marwadi Education Foundations Group of Institutions, Rajkot, India, 4(2).
- [29] Wilcox, A.B., & Hripcsak, G. (2003). Research Paper: The Role of Domain Knowledge in Automating Medical Text Report Classification. *Journal of the American Medical Informatics Association: JAMIA*, 10 4, 330-338.
- [30] Thangaraj, M., & Sivakami, M. (2018). Text classification techniques: A literature review. *Interdisciplinary Journal of Information, Knowledge, and Management*, 13, 117-135. <https://doi.org/10.28945/4066>
- [31] Aldhoayan, M., & Zhou, L. (2016). An accurate and customizable text classification algorithm: Two

- applications in healthcare. 2016 IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences (ICABS), 1-4.
- [32] Henry, J., Pylypchuk, Y., Searcy, T., & Patel, V. (2016). Adoption of Electronic Health Record Systems among U. S. Non-Federal Acute Care Hospitals: 2008-2015.
- [33] Kaurova, O., Alexandrov, M., & Blanco, X. (2011). Classification of Free Text Clinical Narratives (short Review).
- [34] Hiissa, M., Pahikkala, T., Suominen, H., Lehtikunnas, T., Back, B., Karsten, H., Salanterä, S., & Salakoski, T. (2006). Towards Automated Classification of Intensive Care Nursing Narratives. *Studies in health technology and informatics*, 124, 789-794.
- [35] Chapman, W.W., Christensen, L.M., Wagner, M.M., Haug, P.J., Ivanov, O., Dowling, J.N., & Olszewski, R.T. (2005). Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artificial intelligence in medicine*, 33 1, 31-40.
- [36] [41] Deng, L., Dahl, G., Mohamed, A., Jaitly, N., Vanhoucke, V., & Sainath, T.N. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition.
- [37] [42] Chan, T., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing*, 24, 5017-5032.
- [38] [43] LeCun, Y., Bengio, Y., & Hinton, G.E. (2015). Deep Learning. *Nature*, 521, 436-444.
- [39] [44] Chilimbi, T.M., Suzue, Y., Apacible, J., & Kalyanaraman, K. (2014). Project Adam: Building an Efficient and Scalable Deep Learning Training System. OSDI.
- [40] [45] Gupta, S., Zhang, W., & Milthorpe, J. (2016). Model Accuracy and Runtime Tradeoff in Distributed Deep Learning: A Systematic Study. 2016 IEEE 16th International Conference on Data Mining (ICDM), 171-180.
- [41] [46] Lauren, P., Qu, G., & Watta, P. (2017). Convolutional neural network for clinical narrative categorization. 2017 IEEE International Conference on Big Data (Big Data), 2001-2008.
- [42] Fette, G., Ertl, M., Wörner, A., Klügl, P., Störk, S., & Puppe, F. (2012). Information Extraction from Unstructured Electronic Health Records and Integration into a Data Warehouse. GI-Jahrestagung.
- [43] Zheng, S., Lu, J. J., Ghasemzadeh, N., Hayek, S. S., Quyyumi, A. A., & Wang, F. (2017). Effective Information Extraction Framework for Heterogeneous Clinical Reports Using Online Machine Learning and Controlled Vocabularies. *JMIR medical informatics*, 5(2), e12. doi:10.2196/medinform.7235
- [44] Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *Journal of healthcare engineering*.
- [45] Häyrynen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International journal of medical informatics*, 77 5, 291-304.
- [46] Topaz, M., Lai, K.H., Dowding, D.W., Lei, V.J., Zisberg, A., Bowles, K.H., & Zhou, L. (2016). Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application. *International journal of nursing studies*, 64, 25-31.
- [47] Meystre, S.M., Savova, G.K., Kipper-Schuler, K.C., & Hurdle, J.F. (2008). Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of medical informatics*, 128-44.
- [48] Khademi, S., Delir Haghighi, P., Burstein, F., & Palmer, C. (2016). Enhancing rule-based text classification of neurosurgical notes using filtered feature weight vectors. In J. Fisher, & W. Fernandez (Eds.), *Proceedings of the 27th Australasian Conference on Information Systems (ACIS 2016)* (pp. 1-11). Wollongong, NSW, Australia: University of Wollongong.
- [49] Frank, E., & Bouckaert, R.R. (2006). Naive Bayes for Text Classification with Unbalanced Classes. *PKDD*.
- [50] Guo, G., Wang, H., Bell, D.A., Bi, Y., & Greer, K. (2006). Using kNN model for automatic text categorization. *Soft Comput.*, 10, 423-430.
- [51] Yu, Z., Chen, H., Liu, J., You, J., Leung, H.K., & Han, G. (2016). Hybrid k-Nearest Neighbor Classifier. *IEEE transactions on cybernetics*, 46 6, 1263-75.
- [52] Z. Sun, Y. Ye, W. Deng, and Z. Huang. "A cluster tree method for text categorization. *Procedia Engineering*, 2011, vol. 15, pp. 3785-3790.
- [53] Chen, R.C., & Hsieh, C.H. (2006). Web page classification based on a support vector machine using a weighted vote schema. *Expert Syst. Appl.*, 31, 427-435.
- [54] Zhang, B., Su, J., & Xu, X. (2006). A Class-Incremental Learning Method for Multi-Class Support Vector Machines in Text Classification. 2006 International Conference on Machine Learning and Cybernetics, 2581-2585.
- [55] Takçi, H., & Güngör, T. (2012). A high performance centroid-based classification approach for language identification. *Pattern Recognition Letters*, 33, 2077-2084.
- [56] Guan, H., Zhou, J., & Guo, M. (2009). A class-feature-centroid classifier for text categorization. *Proceedings of the 18th International Conference on World Wide Web, ACM* (2009), pp. 201-210
- [57] Lertnattee, V., & Theeramunkong, T. (2006). Class normalization in centroid-based text categorization. *Inf. Sci.*, 176, 1712-1738.
- [58] Han, E., & Karypis, G. (2000). Centroid-Based Document Classification: Analysis and Experimental Results. *PKDD*.
- [59] Yu, Z., Li, L., Liu, J., & Han, G. (2015). Hybrid Adaptive Classifier Ensemble. *IEEE Transactions on Cybernetics*, 45, 177-190.
- [60] Nigam K, Lafferty J, McCallum A (1999) Using maximum entropy for text classification. In *Proceedings: IJCAI-99 Workshop on Machine Learning for Information Filtering* 61-67

- [61] Trstenjak B, Mikac S, Donko D (2014) KNN with TF-IDF based Framework for Text Categorization. *Procedia Engineering* 69: 1356-1364.
- [62] LeCun, Y., Bengio, Y., Hinton, G. (2015) Deep learning. *Nature* 521(7553)
- [63] Voulodimos, A., Doulamis, N.D., Doulamis, A., & Protopapadakis, E. (2018). Deep Learning for Computer Vision: A Brief Review. *Comp. Int. and Neurosc.*
- [64] Hinton, G.E., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Andrew, Senior, Vanhoucke, V., Nguyen, P., Sainath, T.N., & Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition.
- [65] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, pp. 1097–1105.
- [66] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition Boston, MA, USA, 7(12) pp. 1-9
- [67] Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., Bulat, A., Tzimiropoulos, G., Wells, D. M., Murchie, E. H., Pridmore, T. P., ... French, A. P. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, 6(10), 1-10.
- [68] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *EMNLP*.
- [69] Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *ACL*.
- [70] Tai, K.S., Socher, R., & Manning, C.D. (2015). Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. *ACL*.
- [71] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A.J., & Hovy, E.H. (2016). Hierarchical Attention Networks for Document Classification. *HLT-NAACL*.
- [72] Wen, Y., Zhang, W., Luo, R., & Wang, J. (2016). Learning text representation using recurrent convolutional neural network with highway layers. *CoRR*, abs/1606.06905.
- [73] Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent Convolutional Neural Networks for Text Classification. *AAAI*.
- [74] Hassan, A., & Mahmood, A. (2017). Efficient Deep Learning Model for Text Classification Based on Recurrent and Convolutional Layers. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 1108-1113.
- [75] Semberecki, P., & Maciejewski, H. (2017). Deep learning methods for subject text classification of articles. 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), 357-360.
- [76] Kowsari, K., Brown, D.E., Heidarysafa, M., Meimandi, K.J., Gerber, M.S., & Barnes, L.E. (2017). HDLTex: Hierarchical Deep Learning for Text Classification. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 364-371.
- [77] Zhang, X., & LeCun, Y. (2015). Text Understanding from Scratch. *CoRR*, abs/1502.01710.
- [78] Liu, J., Chang, W., Wu, Y., & Yang, Y. (2017). Deep Learning for Extreme Multi-label Text Classification. *SIGIR*.
- [79] Weng, W. H., Waghlikar, K. B., McCray, A. T., Szolovits, P., & Chueh, H. C. (2017). Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC medical informatics and decision making*, 17(1), 155. doi:10.1186/s12911-017-0556-8
- [80] Hughes, M., Li, I., Kotoulas, S., & Suzumura, T. (2017). Medical Text Classification using Convolutional Neural Networks. *Studies in health technology and informatics*, 235, 246-250.
- [81] He, B., Guan, Y., & Dai, R. (2018). Classifying medical relations in clinical text via convolutional neural networks. *Artificial intelligence in medicine*.
- [82] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J.T. (2017). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*.
- [83] Lee, G.E. (2017). A Study of Convolutional Neural Networks for Clinical Document Classification in Systematic Reviews: SysReview at CLEF eHealth 2017. *CLEF*.
- [84] Hughes, M., Li, I., Kotoulas, S., & Suzumura, T. (2017). Medical Text Classification using Convolutional Neural Networks. *Studies in health technology and informatics*, 235, 246-250.
- [85] Choi, E., Bahadori, M.T., & Sun, J. (2016). Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *JMLR workshop and conference proceedings*, 56, 301-318.
- [86] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *EMNLP*.
- [87] Schwenk, H., Barrault, L., Conneau, A., & LeCun, Y. (2017). Very Deep Convolutional Networks for Text Classification. *EACL*.
- [88] Klügl, P., Toepfer, M., Beck, P., Fette, G., & Puppe, F. (2016). UIMA Ruta: Rapid development of rule-based information extraction applications. *Natural Language Engineering*, 22, pp. 1-40.
- [89] Mykowiecka, A., Marciniak, M., & Kupsc, A. (2009). Rule-based information extraction from patients' clinical data. *Journal of biomedical informatics*, 42(5), pp. 923-936.
- [90] Wang, Y., Sohn, S., Liu, S., Shen, F., Wang, L., Atkinson, J.E., Amin, S. and Liu, H. (2019). A clinical text classification paradigm using weak supervision and deep representation. *BMC medical informatics and decision making*.
- [91] Kalyan, K., & Sangeetha, S. (2019). SECNLP: A Survey of Embeddings in Clinical Natural Language Processing. *CoRR*, abs/1903.01039.



- [92] Madhyastha, P. (2019). Exploiting Word Embeddings for Modelling Bilingual Relations. [online] Linguistlist.org. Available at: <https://linguistlist.org/pubs/diss/browse-diss-action.cfm?DissID=45843> [Accessed 12 Apr. 2019].
- [93] Hassanzadeh, H., Kholghi, M., Nguyen, A.N., & Chu, K. (2018). Clinical Document Classification Using Labeled and Unlabeled Data across Hospitals. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2018, pp.545-554.
- [94] Chae, J., Gao, S., Ramathan, A., Steed, C., and Tourassi, G.G. (2017). Visualization for Classification in Deep Neural Networks. Workshop on Visual Analytics for Deep Learning
- [95] Ranganath, R., Perotte, A.J., Elhadad, N., & Blei, D.M. (2016). Deep Survival Analysis. MLHC.
- [96] Deroncourt, F., Lee, J.Y., Uzuner, Ö., & Szolovits, P. (2016). De-identification of patient notes with recurrent neural networks. Journal of the American Medical Informatics Association: JAMIA, 24(3), pp. 596-606.
- [97] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542, pp.115-118.
- [98] Gulshan, V., Peng, L.H., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J.A., Kim, R., Raman, R., Nelson, P.C., Mega, J.L., & Webster, D.R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA, 316(22), pp.2402-2410.
- [99] Tripathi, N., & Jadeja, A. (2015). STOCHASTICALLY REDUCING OVERFITTING IN DEEP NEURAL NETWORK USING DROPOUT.
- [100] Zeeshan, K. (2018). The Impact of Regularization on Convolutional Neural Networks, University of Jyväskylä, Finland
- [101] Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML.
- [102] Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of Statistical Planning and Inference, 90(2), pp.227-244.
- [103] Miličević M., Žubrinić K., Grbavac I., Kešelj A. (2019) Ensemble Transfer Learning Framework for Vessel Size Estimation from 2D Images. In: Rojas I., Joya G., Catala A. (eds) Advances in Computational Intelligence. IWANN 2019. Lecture Notes in Computer Science, vol 11507. Springer, Cham
- [104] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
- [105] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1-9.
- [106] Kalchbrenner, N., Grefenstette, E. & Blunsom, P. (2014). A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188.
- [107] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751.
- [108] Zhang, Y and Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification". In: arXiv preprint arXiv: 1510.03820 (2015).
- [109] Shu, L., Xu, H., and Liu, B. (2017). DOC: Deep open classification of text documents. In: Conference on Empirical Methods in Natural Language Processing, pp. 2901-2906
- [110] LeCun, Y, Bengio, Y & Hinton, G 2015, 'Deep learning', Nature, vol. 521, no. 7553, pp. 436-444. <https://doi.org/10.1038/nature14539>
- [111] Liu, W., Zhang, Y., Li, X., Yu, Z., Dai, B., Zhao, T., & Song, L. (2017). Deep Hyperspherical Learning. NIPS.
- [112] Goodfellow, I.J., Bengio, Y., & Courville, A.C. (2016). Deep Learning. Nature, 521, pp.436-444
- [113] Yu, K., Xu, W., & Gong, Y. (2008). Deep Learning with Kernel Regularization for Visual Recognition. NIPS.
- [114] Veit, A., Wilber, M.J., & Belongie, S.J. (2016). Residual Networks Behave Like Ensembles of Relatively Shallow Networks. NIPS.
- [115] Bjorck, J., Gomes, C.P., & Selman, B. (2018). Understanding Batch Normalization. NeurIPS.
- [116] Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. (2018). How Does Batch Normalization Help Optimization? NeurIPS.
- [117] Salimans, T., & Kingma, D.P. (2016). Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks. NIPS.
- [118] Arpit, D., Zhou, Y., Kota, B.U., & Govindaraju, V. (2016). Normalization Propagation: A Parametric Technique for Removing Internal Covariate Shift in Deep Networks. ICML.
- [119] Gitman, I., & Ginsburg, B. (2017). Comparison of Batch Normalization and Weight Normalization Algorithms for the Large-scale Image Classification. CoRR, abs/1709.08145.
- [120] Lian, X., & Liu, J. (2019). Revisit Batch Normalization: New Understanding and Refinement via Composition Optimization. AISTATS.
- [121] Sorgente, T., Fernández, E. and Petrie, M. (2019). [online] Available at: [https://www.researchgate.net/publication/228707311\\_The\\_SOAP\\_Pattern\\_for\\_Medical\\_Charts](https://www.researchgate.net/publication/228707311_The_SOAP_Pattern_for_Medical_Charts) [Accessed 25 Apr. 2019].