Text Summarization and Classification of Clinical Discharge Summaries using Deep Learning

Hardik Gunjal
Dept. of Computer Science
Lakehead University
Thunder Bay, Canada
gunjal1@lakeheadu.ca

KhushalParesh Thaker
Dept. of Computer Science
Lakehead University
Thunder Bay, Canada
kparesh@lakeheadu.ca

Alizar Marchawala
Dept. of Computer Science
Lakehead University
Thunder Bay, Canada
amarchaw@lakeheadu.ca

Preetkumar Patel
Dept. of Computer Science
Lakhead University
Thunder Bay, Canada
ppatel60@lakeheadu.ca

Abhishek Nagrecha
Dept. of Computer Science
Lakhead University
Thunder Bay, Canada
nagrechaa@lakeheadu.ca

Sabah Mohammed
COMP-5112WC Supervisor
Lakhead University
Thunder Bay, Canada

Abstract*—Clinical Discharge Summaries contain a huge amount of data but are difficult to obtain and to be used in analysis because of its unstructured narrative format. Also, they contain spelling mistakes, abbreviations which makes it difficult to summarize. With vast amount of information, there is a need to have a proper segregation of important details like Symptoms, History, Medications to have a better and efficient classification. To save the overall time and energy of the patient and increase the effectiveness of the physician in saving the lives of the sufferers, we suggest a classification of the Clinical Discharge Summaries text, which aims to automatically predict the diagnoses needed for a patient based on clinical notes. The reminiscence is represented by a raw text file with doctor's entnaaries about the patient including his / her age, problems described in normal flow, history of the patient and so on. We are also investigating the automated classification of patient discharge notes into standard disease labels which includes the name of the diagnostic procedure required. In this approach, we use Convolutional Neural Networks to classify and represent complex features from the medical discharge summaries using the MT sample dataset. We make use of GloVE to have a pretrained model learn from it.

***Keywords--* Text Summarization, Natural Language Processing(NLP), Recurrent Neural Network(RNN), Attention, Abstractive, Named Entity Recognition(NER), GloVE**

I. Introduction

With advent rise in medical records, there is an ever-increasing need for automation to classify or summarize them. Trying to understand and to get the grasp of the medical notes prepared by the doctors for future evaluations manually, becomes difficult as the records are lengthy and with a lot of medical terminologies which not everyone is well versed with. The digitization of medical records has uncovered the domain of NLP and deep learning to captivating and novel avenues, which may facilitate in classifying medical documents like clinical narratives, medical transcriptions that successively helps a doctor to search out an identification quicker, that in some cases is also extraordinarily valuable, up to saving lives. Clinical notes, through which the medical reports are basically written in linguistic communication, are perceived as a strong resource to interpret completely different clinical queries by providing elaborated patient conditions, the thinking method of clinical reasoning, and clinical illation, that unremarkably can't be obtained from the other elements of the Clinical Discharge Summaries. Machine-driven document classification is usually useful in any process clinical documents to extract these styles of information. As such, the huge generation of clinical notes and increasing adoption of Clinical Discharge Summaries has caused machine-driven document classification to become a vital analysis in the field of clinical prognosticative analytics, to assist and leverage the utility of narrative clinical notes [1].

Clinical discharge summaries have grown over the years. Summaries incorporate an unprecedented quantity and sort of patient info, together with demographics, sign measurements, laboratory check results, prescriptions, procedures performed, digitized notes, imaging reports, mortality, etc. This vital information needs to be segregated from the entire corpus so that the task of classification becomes easier. They typically contain structured information as well as unstructured information (e.g. medical notes written by specialists)[2].The difficulty in classifying lies in the fact that most records are unstructured and contain spelling or grammatical errors, while many contain abbreviated versions of the contextual words which only the person who wrote it or a medical expert in that domain would be able to understand. This makes it difficult to have a machine learn. One of the difficulties in operating with Clinical discharge summaries is the complex nature by which it is represented, with data types including (1) Date Time object types, (2) numeric data, (3) categorical data, and (4) free-text that include discharge summaries [2].

Until the previous couple of years, most of the strategies for analysing the clinical discharge information were supported by ancient machine learning and applied math techniques like logistic regression, support vector machines (SVM), and random forests. Lately, deep learning techniques have gained nice profit in several regions through deep gradable feature construction and capturing long-range dependencies in information in an efficient manner. Given the increase within the catholicity of deep learning approaches and also the progressively wide quantity of patient information, there has conjointly been a rise within the range of papers applying deep learning to Clinical Discharge Summary information for clinical IP tasks that yield higher performance than ancient strategies and need less pre-processing time and engineering[2].

In this paper, we tend to focus our efforts on the automated labelling of medical notes from the MT Samples Dataset. The dataset is collected from Kaggle and was updated. The dataset contains just about 5000 records of patients later together with their general description that is nothing but a brief description of transcription, a medical specialty that is the classification of the transcription and our target variable, sample name that's transcription title, transcription that are sample medical transcriptions and necessary keywords that are relevant to the classification task and extracted from transcription. Moreover, we have added certain keywords that we extracted from the raw data to enhance the dataset further.

The multi-layered approach enables deep learning models to complete classification tasks such as recognizing subtle anomalies in medical photographs, clustering patients with common features into risk-based cohorts or illustrating associations across vast quantities of unstructured data between symptoms and outcomes. Similar to other forms of artificial learning, deep learning has the additional benefit of being able to make choices with significantly less input from the programmers. Although simple machine learning needs a programmer to decide whether a hypothesis is right or not, due to the complexity of its multi-layered framework, deep learning may gauge the precision of its answers on its own. In addition to that it also takes less pre-processing time when compared to machine learning algorithm. The network itself is responsible for much of the validation and normalization functions that human programmers needs to perform while using certain machine learning techniques. Since neural networks are built for classification, individual linguistic or grammatical components may be classified by "grouping" the related words together and projecting them in relation to each other, this allows the network to understand complex semantic meaning.

The MT discharge summary format in our case will include the below explained topics/features:

1. Sample Type: it includes the type of medical document like patient progress report, discharge summary, clinical narrative etc
2. Sample Name: it includes the medical term/name of the problem patient is suffering
3. Description: it includes brief description of the patient condition at the time of diagnosis
4. Chief Complaint: it includes the details of specific issue that the patient is suffering with
5. History Of Present Illness: it includes the detailed description of the current disease history the patient has gone through over years
6. Allergies: it includes the general allergies the patient has had over these years
7. Past Medical History: it includes the patient medical history with details like a) disease first developed, b) action taken etc.
8. Family History: it includes the detailed information of any disease the patient's family members might have had over the years.
9. Social History: it includes the details of the environment the patient works or stays in, the people who he/she is surrounded with, any other vices the patient might have.
10. Review of Systems: it includes the symptoms the patient has experienced have and other details related to the disease from the patient's point of view
11. Physical Examination: it includes the details of tests done on the patient during initial period in accordance to the symptoms specifies
12. Assessment: it includes the test results of the patients and other observations made by the doctor at time of assessment.

Deep learning covers many approaches under its umbrella. The foremost necessary idea in deep learning is that of illustration. Input options to a machine learning algorithmic rule generally ought to be camp made from information, reckoning on skilled expertise and domain awareness to spot specific patterns of previous interest. The engineering technology to develop, evaluate, decide and assess fittingly. The engineering method of making, analysing, selecting, and evaluating applicable options will be arduous and long and is commonly thought of as a "black art" [3] requiring creativeness, trial-and-error.deep learning strategies learn best options directly from the information itself, with no human steering, letting the automated discovery of latent information relationships which may preferably be unknown or hidden. The overwhelming majority of profound learning algorithms and architectures area unit supported the factitious neural network (ANN) platform. ANNs incorporates multiple interconnected nodes (neurons), sorted in layers. Neurons that don't seem to be within the input or output layers area unit are considered as hidden units. Every hidden unit stores a group of weights W and is regularly updated because the model is being trained [2].

The first step while implementing the proposed model is to fetch the data for the dataset and to preprocess it before we feed it to out model. In our implementation, keywords are extracted for the dataset. Preprocessing steps include cleaning of strings which is basically removal of special characters of English language and all the punctuation. We intend to classify the extracted keywords with the categories of medical specialty they correspond to. The keywords are then tokenized and then the training, testing and validation split is made. The model is trained on text type of input. First layer or the input layer will be the embedding layer which uses the reference for the word vectors from “Glove 6B 100D”. The Model consists of convolution layers, max-pooling layers, flatten layer and dense layer for the final classification.

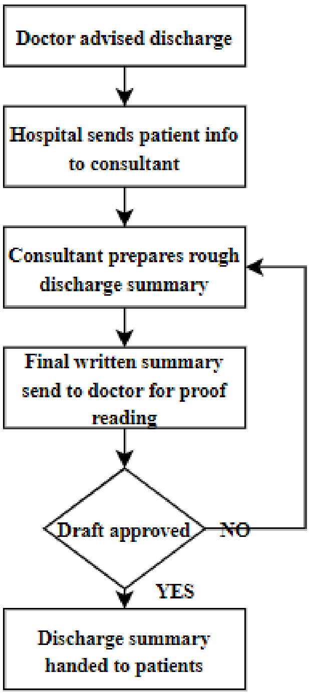


Fig 1: Discharge Summary Generation process flow

There is very little information present in the studies under review about the nature of document development or the precise origins of the discharge summaries. It's sensible to conclude that the more flexibility a clinical author has, the fewer uniform section headings can be anticipated and the more often sections can be identified without heading [20]. Three classes are distinguished by the technology which can be used: rule-based, ML and hybrid. Most rule-based methods include heading pieces, which are interpreted using string and matching patterns. Others have achieved very strong results with the use of probabilistic rules dependent on text functionality.

The ML approaches view section recognition as a classification problem, with the most common paradigms being CRF and SVM. Finally, hybrid techniques, mostly during the development of training data sets, used rule-based methods and then implemented ML methods. Nearly all of the studies focused on a new dictionary; often created by the writers themselves, often reusing material from existing regulated terminologies. SecTag, an excerpt from many terminologies is gaining attention here [19]. Specific degrees of granularity are under investigation. While some research considered traditional dissections of therapeutic records according to formats such as SOAP, others dissected texts according to the temporal shifts alluded to in the text, while others merely separated clinically meaningful material from institutional / administrative details [21].

Review of performance found that much of the published findings are incomparable due to the lack of any reference method for determining section recognition. Hybrid methods produced better results than ML methods, but poorer than those based on rules. Their depth was however more conservative than rule-based experiments in terms of the form of notes they could take. The predictive ability of the features used in some of the ML and hybrid experiments shows extraordinary phenomena. For e.g., features such as token unigrams and n-grams, word / POS pairs, characteristics indicating that the text contained a known section heading, and style features had a major positive effect on classification [21].

II. LITERATURE SURVEY

Health-related text classification is regarded as a special case of text classification. Supervised topic models provide a promising solution for integrating clinical data as features into a predictive problem, provided a record of a patient, and an estimate of a collection of latent factors predicting the output variable. Jason et al., [12] created a prediction focused supervised LDA whose vocabulary selection method enhances the topic coherence of supervised topic models while preserving competitive prediction efficiency [12]. Another important research used by Merck Manual which contains dataset that includes articles of various topics such as brain, cancers etc [4]. The concept of converting each sentence to a word level matrix is proposed where each row in the matrix is a sentence vector extracted from the model using word2vec. This model performed better than Sentence Embedding, mean word Embedding and word embedding with BOW (bag-of-words) in terms of accuracy. A similar approach is used in convolution neural networks for clinical narrative of surgery data categorization which shows significant improvement in terms of the F_1-Score [5].

A Character level language model comprising of CNN (Convolution Neural Network) and bi-RNN is proposed by Qu et al. [10] to have the text classified at the character-level. Unlike word-level model that avoids the problem of unregistered words and improves the robustness of the text representation in character-level model [10]. The proposed language model makes use of the data augmented by different convolution filters of CNN. The contextual information to classify the text is then obtained by bi-RNN. The results show that the model proposed by Qu et al. [10] has a better performance than the combination of CNN and LSTM (Long Short-Term Memory) classification method. Therefore, this approach is relevant to our research topic. Moreover, there are techniques which uses both rule-based features and knowledge guided convolutional neural networks which is used by Yao et al., [11] in clinical text classification [11]. Initially, the trigger phrases are identified using rules. Next, these phrases are used to predict classes with a few examples. Later, a convolutional layer network is trained on the trigger phrases with word embeddings and unified medical language system. Furthermore, a CNN based approach to categorize text fragments from a clinical record at sentence level is implemented in [7]. The concept of converting each sentence to a word level matrix is proposed where each row in the matrix is a sentence vector extracted from the model.

There are different frameworks which are developed in order to conceptualize medical notes [8]. Fodeh et al., [8] proposes a new framework Med Cat. The model is applied to Post traumatic stress disorder clinical notes for its evaluation. The proposed approach includes manual annotation, and metamap for knowledge base expansion. NLP is used to automatically generate automatic annotations which in turn generates a bag of concepts. The research also introduces the process to extract the detailed features from the notes and transform them into less granular set of features by using the concept of specialized concept-category hierarchy [8].

Furthermore, there are various methods which use two design principles for classification i.e. operating at the lowest atomic representation of text (Characters) and using deep stack of local operation, which is convolutions and max pooling of size 3, to learn high level of hierarchical representation of the sentence [6]. This principle is used in the very deep convolutional networks for text classification [6]. This model is able to show the increase in performance by using 29 convolutional neural networks [9]. One such model was proposed in deep pyramid convolutional neural networks for text categorization which also increase the accuracy by increasing the networks depth. It has 15 weight layers and performs 2 strides down sampling which reduces the size of the internal representation of each document by half.

Information on medication is one of the most important forms of clinical data in electronic medical records. The use of electronic medical record data is important for the protection and quality of healthcare, as well as for clinical research. Medication details are, however, also reported as free-text in clinical reports. As such, many computerized applications that rely on coded data do not have access to these. Xu et al., [17] proposed a system called MedEx which is developed in order to extract information from clinical narratives. In this method, a data set of 50 discharge summaries found that it was effective not only in identifying drug names (with F Score of 93.2 %), but also in identifying signature details such as strength, route, and frequency with F Score of 94.5 percent, 93.9 percent, and 96.0 percent. Moreover, in classification of clinical narrative one of the important tasks is clinical name entity recognition. Clinical NER is an essential processing of natural language (NLP) to derive essential concepts (named entities) from the clinical narratives. Scientists have worked extensively on machine learning models for clinical NER. More recent attempts have been made to apply deep learning models to enhance the efficiency of existing clinical NER systems. Using the i2b2 2010 clinical design extraction corpus, Wu et al., [16] compares two deep neural network architectures with three reference Conditional Random Fields (CRFs) models and two state of the art clinical NER systems. The evaluation results showed that for the given clinical NER task, the RNN model trained with the word embedding achieved a new state-of - the-art output (with F1 score of 85.94 per cent), performing the best-reported method using both manually provided and unsupervised learning features. Furthermore, there are many deep learning approaches which show better results than extraction-based method for clinical narrative classification. One of them is discussed by model proposed by Gehrmann et al., [15] wherein, a comparison of concept-based extraction methods with CNNs and other widely used NLP models within ten phenotyping tasks, using 1,610 MIMIC-III database discharge summaries is done. This shows that CNNs outperform concept-based extraction methods in almost all tasks, with an F1-score increase of up to 26 percentage. Other method such as graph convolutional and RNN is also very effective in classification of medical narratives. One such method is applied in the model proposed by Li et al., [18], which gives better results in terms of F1 score. F1 score is 0.692 for the classification of medical therapy problem relationships, 0.827 for medical test problem relationships, and 0.741 for medical-problem-relationships.

III. METHOD

Medical transcriptions are the reports of doctor patient conversations and evaluation of patient’s current medical conditions. These evaluations are highly confidential. To protect the patient’s identity these documents are not shared openly, thus making it very difficult to study them and to develop a model that could help us classify the text of medical transcription. This section explains the details of dataset and the source of its information. This section describes how the dataset was collected and outlines the development steps of our model for predicting category of medical specialty based on the keywords extracted from medical transcription. When large number of documents must be categorized, this task can be identified as a text summarization task. The performance of our model is measured by observing training accuracy, validation accuracy, F1 score, precision and recall.

A. Dataset

The dataset comprises of various medical transcriptions samples collected from the “MTsamples database provided in its link [13]. The dataset includes 4999 medical transcription samples of 40 different medical specialties. The medical categories are as follows: ' Allergy / Immunology', ' Bariatrics', ' Cardiovascular / Pulmonary', ' Chiropractic', ' Consult - History and Phy.', ' Cosmetic / Plastic Surgery', ' Dentistry', ' Dermatology', ' Diets and Nutrition', ' Discharge Summary', ' ENT - Otolaryngology', ' Emergency Room Reports', ' Endocrinology', ' Gastroenterology', ' General Medicine', ' Hematology - Oncology', ' Hospice - Palliative Care', ' IME-QME-Work Comp etc.', ' Lab Medicine - Pathology', ' Letters', ' Nephrology', ' Neurology', ' Neurosurgery', ' Obstetrics / Gynecology', ' Office Notes', ' Ophthalmology', ' Orthopedic', ' Pain Management', ' Pediatrics - Neonatal', ' Physical Medicine - Rehab', ' Podiatry', ' Psychiatry / Psychology', ' Radiology', ' Rheumatology', ' SOAP / Chart / Progress Notes', ' Sleep

Medicine', ' Speech - Language', ' Surgery', ' Urology' [13]. This dataset contains medical transcription samples of above-mentioned categories. Along with transcription, dataset contains sample name, description and keywords. Here we use keywords and corresponding medical specialty. The dataset is free to use and needs no licensing.

B. GloVE Word Vector

Glove: Global vector for word representation, an unsupervised machine learning algorithm which uses global word to word co-occurrence statistics from a word corpus and obtains vector representations of words, was developed by Jeffrey Pennington, Richard Socher, and Christopher D. Manning in 2014 [14]. The word vectors obtained from this model can be further used for training on new dataset and getting the desired predictions. This word vector is available in four different dimensions 50D, 100D, 200D and 300D.

GloVE is used in order for it to learn word vectors so that the product of dots is similar to the log value of the probability of the word's co-occurrence. The ratios extracted can represent the meaning and also provide information that is encoded as differences of vectors. This is the reason why resulting word vectors are functionally well suited for finding word analogies [14].

We use the Glove6B 100D embeddings. And then with other hyper parameters, we compose the out-of-vocabulary word embeddings using n-gram embedding.

C. Deep Learning

The purpose of this project is to use Deep Learning for categorizing medical transcription to their corresponding categories. Our approach makes use of GloVe pre-trained weights of Wikipedia-2014 and Gigaword-5 data. Here we used "glove.6b.100d" for obtaining word vectors of the keywords that are then given as input to the model. Figure 2 represents the flow of our proposed model.

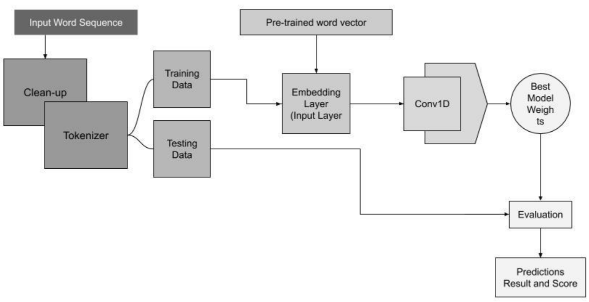


Fig 2: CNN Model Flow-chart

The input sequence shown in the figure 2 are the keywords from the dataset. The keywords go through the cleanup module, which removes unnecessary punctuations and converts all the upper-case characters to lower case so that same word does not produce different set in integer value when converted. They are then tokenized and split into two parts that is training and testing dataset.For validation, 20 percent of training data is used. The next step is to load the pretrained dataset in the input layer, which here is the embedding layer along with a tensor of 1000 dimensional vectors. The output of the embedding layer is fed to 1D convolution layer. The figure 2 shows the model structure.

In the proposed model, we are converting the text into sequences of indexes which is an integer ID for the word. Once this is completed, an embedding matrix is prepared that contains embedding vector at an index. Data dump of pre-trained embeddings are parsed and then known embeddings are mapped to the index mapping words. The embedding matrix is loaded into the Keras Embedding layer which has a 1D convolutional layer built on top of it. CNN has a convolutional hidden layer that runs over a 1D sequence. As shown in the figure 4, we are having an additional convolutional layer over it because the input sequence is very long. The convolutional layer is followed by a max pooling layer that filters the output of the convolutional layer, which in our case is the text (clinical discharge summary) into more important parts. We add a drop out layer so that any overfitting caused by the expansive dataset, is removed. Drop out randomly sets the outwards edges of hidden layers to 0 with every iteration of the training phase. Finally, we have added a flatten layer for the inputs from the drop out layer to be of changed to the shape which is acceptable to the dense layer.

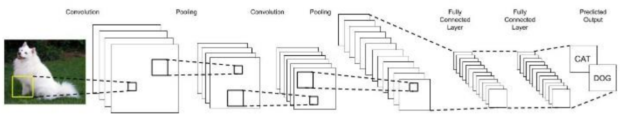


Fig 3: CNN Model Architecture

In Deep Learning, Convolution Neural Network, which is commonly referred to as CNN, is a type of deep neural network. This class of deep neural network method is mostly applied for classifying images. CNN are stacked layers of convolution which are then followed by non-liner activation function. Traditional neural network which are feedforward neural network, each output neurons are connected to each input neuron of the next layer. These kinds of layer are called as fully connected or affine layers. Rather than building the model based on fully connected layers, it uses combination of layers like convolution, max pooling and dense layer (fully connected or affine layer).

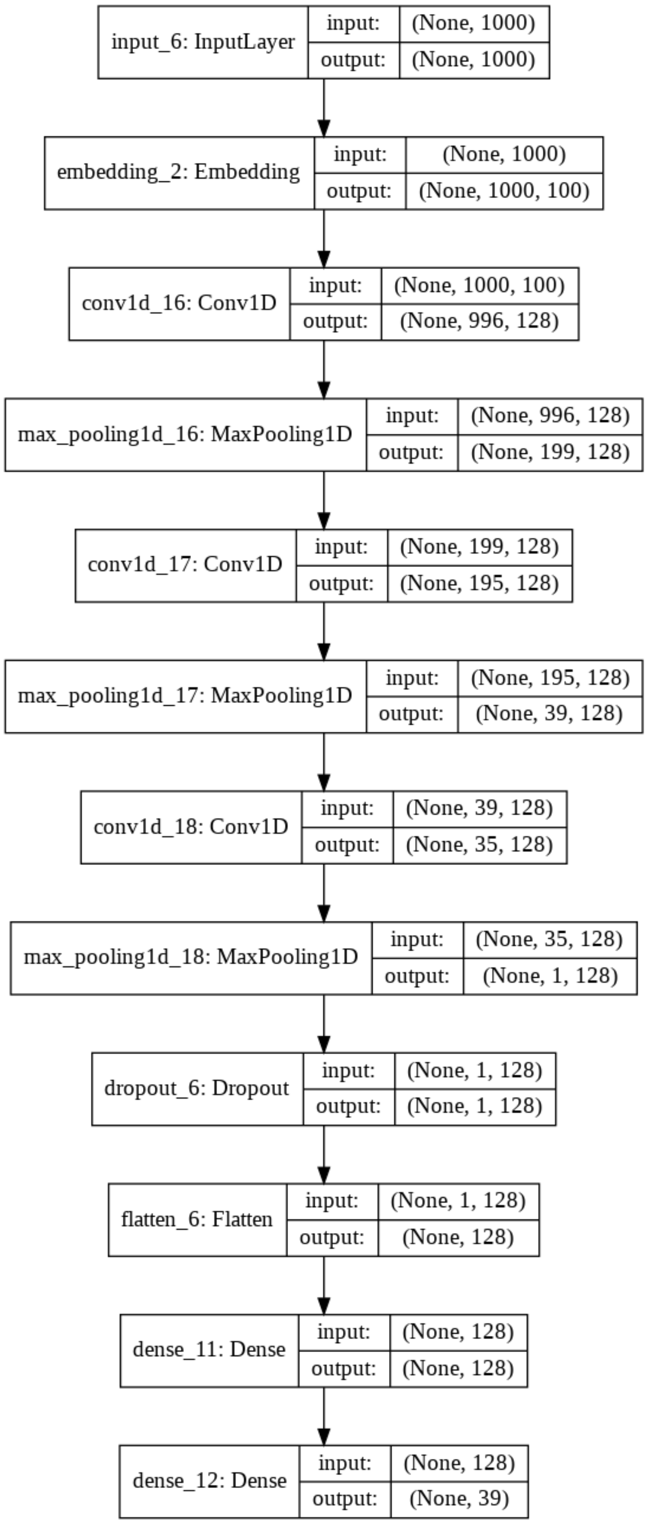


Fig 4: CNN Model Architecture Summary

Convolution is applied over the input to compute output, which makes local connections between each region of input to the output neuron. The convolution layer has many different filters and the result is the combination of all the filters applied to the input. Each filter in a convolution layer gives composition of high-level feature from a part of input. While performing unsupervised learning the CNN automatically learns the type of filter to apply on the basis of the task that needs to be performed by model, which can be classification, regression, or something else. The most common implementation of a CNN model can be seen for image classification as shown in figure 3. An input image goes through different layers. It learns the features present in the image and detects high level feature out of it and those high-level features are then used for classification.

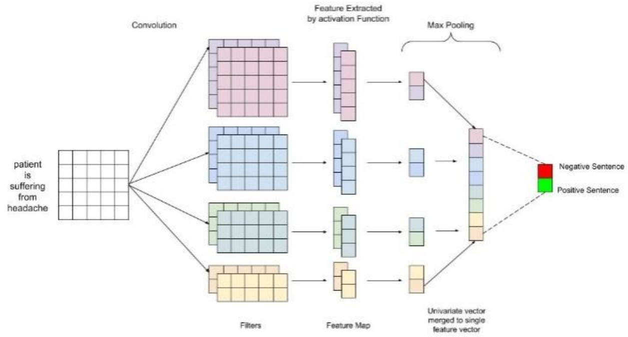


Fig 5: CNN Architecture for sentences classification

The Pooling layer is used for sub sample of the extracted features by the previous layers and those sub samples are fed to the next layer for further deep feature extraction. Pooling is essential in a CNN because they are invariant to rotation, translation, scaling and skewing.

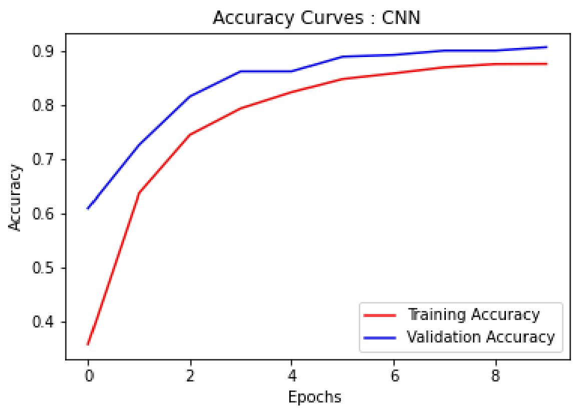
CNN's are popular for handling images and are mostly used in the field of computer vision. When it comes to classification, they can also be used for NLP. NLP problems comprise of text or sentences represented as a matrix, where each row of the matrix represents a token, which are basically words or a single character. These tokens, words or row in case of a matrix are vectors which are word embeddings. They can also be transformed into one-hot encoding where all the elements of the vector are 0 except 1. Word embedding consist of set of feature learning and language techniques for NLP. It represents words with similar meaning through similar representation. Giving text as input to CNN the convolution process remains the same. In image classification the filter size is different as the image is represented as a 2-dimensional matrix with width is different. Whereas the filter width for NLP task is same as the input matrix, region size or height may vary. Figure-5 represents a Convolution Neural Network for NLP.

Previous research has shown that CNN are good for tasks of computer vision because of compositionality and location invariance but in case of NLP that does not apply. Words that are used together in a phrase or a sentence can also be used in different form and with combination with other words and pixel of a single image, close to each other are semantically related. The most important point while using CNN, is its speed because of the hardware acceleration from GPU's.

One of the testing scenarios which has been implemented, takes the whole discharge summaries as input and keyword extraction is performed. From the extracted keywords Glove or TF-IDF can be used to filter out the common words and the processed keywords from which classification can be performed are fed to the RNN model for further computation. There are some of the databases such as MTsamples where the keywords are already extracted hence it can directly be fed to the model. Having said that, our model can also be used for classifying sub specialties of a medical specialty.

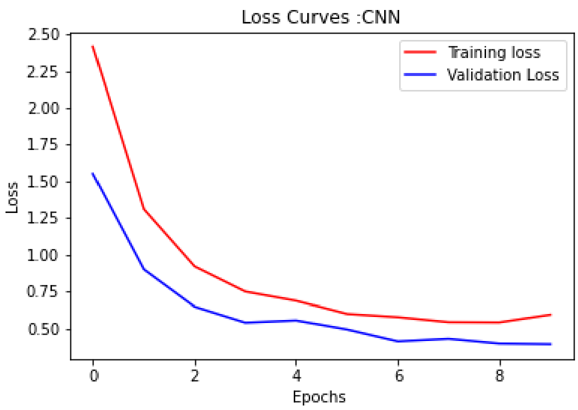
IV. EXPERIMENTAL ANALYSIS

We implemented our CNN model on MT samples dataset with multiple layers being added and removed to improve the efficiency. Our training accuracy increases with every epoch and reaches 0.87 and the validation accuracy is at 0.89 as shown in the graph 1. The model during its validation set also achieved a recall score of 0.888, precision of 0.98 and an F1 score of 0.939.



Graph 1: Accuracy Curve

Graph 2 represents the curve of loss through our model run. Training run started with a higher loss of 2.75 to 0.58 through 10 epochs. While validation did quite well with 0.427 as its loss after 10 epocs.



Graph 2: Loss Curve

The working of the proposed method is similar to that of traditional CNN classifier which are used to classify images. The difference here is the input which is text or keywords and not images. Table 1 shows performance of other models with respect to text classification. Table 1 represents methods that have similar implementation apart from the word embedding. The word embedding used in other models is the wiki-news with dimension of 300 and 1 million-word vectors.

Table 1
Performance of other Models w.r.t. text classification

DCNN - GloVe (ours)	DCNN – Wiki News	CNN - Wiki News	DNN - Wiki News	RNN – GRU - Wiki News	Bidirectional RNN - Wiki News
0.89	0.76	0.78	0.71	0.67	0.68

The use of GloVe word vector embedding greatly improves our model. Clinical discharge summaries are difficult to summarize, as it has medical terminologies which are specific to corresponding medical specialties. As mentioned earlier the GloVe word vector is pre trained on data from Wikipedia pages. Wikipedia holds huge amount of information on almost all kinds topics which includes medical information. This makes the model perform better towards summarizing clinical discharge summary as well general text summarization and categorization, the results of which are shown in the Table 2.

Table 2
Performance of proposed Model

Training Accuracy	Validation Accuracy	Recall	Precision	F1
0.87	0.89	0.88	0.98	0.93

REFERENCES

[1] *Okuyama, Ayako, and Takahiro Higashi. "Usability of Clinical Information in Discharge Summary Data in the Diagnosis Procedure Combination Survey for Cancer Patients." International Journal of Environmental Research and Public Health 17, no. 2 (2020): 521.*

[2] *Murphy, S. F., L. Lenihan, F. Orefuwa, G. Colohan, I. Hynes, and C. G. Collins. "Electronic discharge summary and prescription: improving communication between hospital and primary care." Irish Journal of Medical Science (1971-) 186, no. 2 (2017): 455-459.*

[3] *Domingos P. A few useful things to know about machine learning. Communications of the ACM. 2012;55(no. 10):78–8*

[4] *Hughes, Mark, Irene Li, Spyros Kotoulas and ToyotaroSuzumura. "Medical Text Classification using Convolutional Neural Networks." Studies in health technology and informatics 235 (2017): 246-250 .*

[5] *P. Lauren, G. Qu and P. Watta, "Convolutional neural network for clinical narrative categorization," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 2001-2008.*

[6] *Conneau, Alexis &Schwenk, Holger &Barrault, Loïc&Lecun, Yann. (2017). Very Deep Convolutional Networks for Text Classification. 1107-1116. 10.18653/v1/E17-1104.*

[7] *Hughes, Mark & Li, Irene & Kotoulas, Spyros & Suzumura, Toyotaro. (2017). Medical Text Classification Using Convolutional Neural Networks. Studies in Health Technology and Informatics. 235. 10.3233/978-1-61499-753-5-246. Johnson, Rie& Zhang, Tong. (2017). Deep Pyramid Convolutional Neural Networks for Text Categorization. 562-570. 10.18653/v1/P17-1052.*

[8] *S. J. Fodeh, M. Zirkle, D. Finch, R. Reeves, J. Erdos and C. Brandt, "MedCat: A Framework for High Level Conceptualization of Medical Notes," 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, 2013, pp. 274-280.*

[9] *Johnson, Rie& Zhang, Tong. (2017). Deep Pyramid Convolutional Neural Networks for Text Categorization. 562-570. 10.18653/v1/P17-1052.*

[10] *Hua, Qu, Shi Qundong, Jiang Dingchao, Guo Lei, Zhang Yan-peng and Liu Pengkang. “A Character-Level Method for Text Classification.” 2018 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC) (2018): 402-406.*

[11] *L. Yao, C. Mao and Y. Luo, "Clinical Text Classification with Rule-based Features and Knowledge-guided Convolutional Neural Networks," 2018 IEEE International Conference on Healthcare Informatics Workshop (ICHI-W), New York, NY, 2018, pp. 70-71.*

[12] *Ren, Jason, Russell Kunes, and Finale Doshi-Velez. "Prediction Focused Topic Models for Electronic Health Records." arXiv preprint arXiv:1911.08551 (2019).*

[13] *Mtsamples.com. 2020. Transcribed Medical Transcription Sample Reports And Examples - Mtsamples. [online] Available at: <https://www.mtsamples.com/> [Accessed 22 March 2020].*

[14] *Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.*

[15] *Gehrmann, Sebastian, Franck Dernoncourt, Yeran Li, Eric T. Carlson, Joy T. Wu, Jonathan Welt, John Foote Jr et al. "Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives." PloS one 13, no. 2 (2018).*

[16] *Wu, Yonghui, Min Jiang, Jun Xu, Degui Zhi, and Hua Xu. "Clinical named entity recognition using deep learning models." In AMIA Annual Symposium Proceedings, vol. 2017, p. 1812. American Medical Informatics Association, 2017.*

[17] *Xu, Hua, Shane P. Stenner, Son Doan, Kevin B. Johnson, Lemuel R. Waitman, and Joshua C. Denny. "MedEx: a medication information extraction system for clinical narratives." Journal of the American Medical Informatics Association 17, no. 1 (2010): 19-24.*

[18] *Li, Yifu, Ran Jin, and Yuan Luo. "Classifying relations in clinical narratives using segment graph convolutional and recurrent neural networks (Seg-GCRNs)." Journal of the American Medical Informatics Association 26, no. 3 (2019): 262-268.*

[19] *Denny JC, Miller RA, Johnson KB, Spickard-III A. Development and evaluation of a clinical note section header terminology. In: AMIA Annual Symposium, Washington, DC, USA, November 8-12. USA: American Medical Informatics Association: 2008.*

[20] *Pomares-Quimbaya, A., Kreuzthaler, M. & Schulz, S. Current approaches to identify sections within clinical narratives from electronic health records: a systematic review. BMC Med Res Methodol*

[21] *Dai H-J, Syed-Abdul S, Chen C-W, Wu C-C. Recognition and evaluation of clinical section headings in clinical documents using token-based formulation with conditional random fields. BioMed Res Int. 2015; 2015:10.*