How to Cite:
Almuhana, H. A.-J., & Abbas, H. H. (2022). Classification of medical specialty for text medical report based on natural language processing and deep learning. *International Journal of Health Sciences*, *6*(S7), 3362–3385. https://doi.org/10.53730/ijhs.v6nS7.12476

Classification of medical specialty for text medical report based on natural language processing and deep learning

Hasanen Abdul-Jawad Almuhana
Student - College of Engineering. University of Karbala, Iraq
Email: hasanen.a@s.uokerbala.edu.iq

Hawraa Hassan Abbas
College of Engineering, University of Karbala, Iraq
E-mail:, Hawraa.h@uokerbala.edu.iq

***Abstract*---**Natural language processing (NLP) is a part of artificial intelligence algorithms that focus on designing and building applications and systems in a way that allows interaction between computers and natural languages developed for human use. NLP has been used in several areas within artificial intelligence and data processing applications such as social media applications and medical applications and translation applications. The medical field is one of the richest in terms of big data, which has not been well invested so far. one of these unused data was the text medical reports. Natural Language Processing analysis is an effective way of examining medical reports and social media data to improve treatments and patient services. by training a computer to make a decision instead of a human, as soon as possible from the decisions of specialist doctors or medical staff working in health institutions to diagnose diseases or Prescribe treatment to the patient. The data of each patient is stored in the form of medical reports mostly as a text document file and kept within the hospital data. Hospitals suffer from a scarcity of medical and nursing staff, in addition to the high cost of hiring medical staff. In this model, we apply a deep learning algorithm on the same textual medical report's dataset used in the past model. Applying NLP to clean data and Convolution Neural Network (CNN) which has five layers. Applying all these layers to our data we classified the medical reports into nine classes that resulted in high accuracy equal to (99.00), F1-Measure (97.82), precession (98.64), and Recall (97.11).

***Keywords*---**Convolution Neural Network (CNN), Natural language processing, health institutions.

Introduction

Natural language processing (NLP) a branch of AI, aims at primarily reducing the distance between the capabilities of a human and a machine and analytic the natural information data such as images, text documents and any unstructured data into structure data able to process. NLP, Artificial intelligence, and Machine learning all have the potential to simplify the use of unstructured data [1].

Traditional analytics typically uses structured data, consisting mainly of claims data and only accounting for approximately 20 percent of all available data. Structured data exists in a specific, consistent format and includes basic information, but not valuable details. Unstructured data include free text data, physician order data, nursing notes, and dictation notes. To access the remaining 80 percent of this data, one must rely on NLP. NLP allows organizations to access the complex, richer data sets that are harder to reach because they require sophisticated technology to derive value from the massive amounts of everyday language sitting in the HER [2].

Unstructured data, residing in EHRs and elsewhere, contains the deeper, more complex information such as a patient's own words to describe symptoms or doctors' notes.

While medical data is growing exponentially, a bulk of it remains unused, mainly because healthcare-related information systems are not equipped to process unstructured data. If the capabilities to interpret, analyze and utilize this unstructured medical data were available, the benefits would be tremendous both for patient treatment, as well as for public health management and medical research.

For clinical research, a huge quantity of detailed patient information, such as disease status, side effects, medication history, treatment outcomes, and lab tests, has been collected in an electronic format called electronic medical record (EMR) and it serves as a valuable data source for further analysis. Therefore, a huge quantity of detailed patient information is present in the medical text, and it is quite a huge challenge to process it efficiently.

1.1 Deep learning.

Deep learning is a set of algorithms and techniques hierarchical machine learning inspired by how the human brain works, called neural networks. Deep learning architectures offer big benefits for text classification because they work at very high accuracy with lower-level of computation [3] .

The concept of 'deep learning' has recently gained a lot of attention. It refers to unsupervised learning algorithms which automatically discover data without the need of supplying specific domain knowledge [4]. This approach usually has higher performance rates than supervised and informed methods when processing large unstructured corpora. However, the ability of these algorithms has to be evaluated to measure their error rate.

Deep learning-based models have surpassed classical machine learning-based approaches in various text classification tasks, including sentiment analysis, news categorization, question answering, and natural language inference[5].

A large amount of biomedical information in databases such as EMRs is a valuable source for information extraction [6]. Information extraction and, in particular, natural language processing methods are required to annotate and process biomedical literature [7].

There are two main architectures for deep learning text classification, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Figure (1) shows simple of deep learning layers.

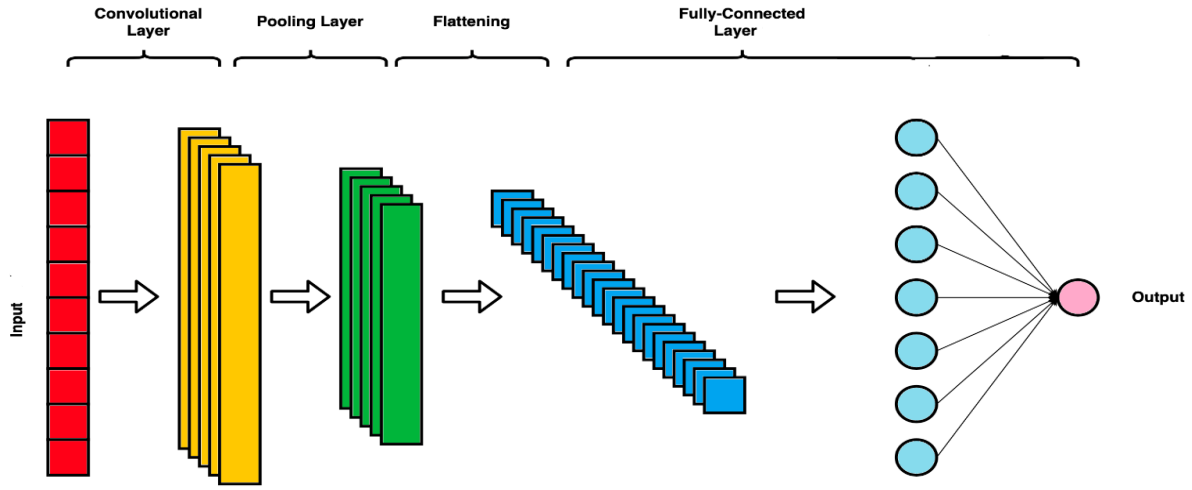
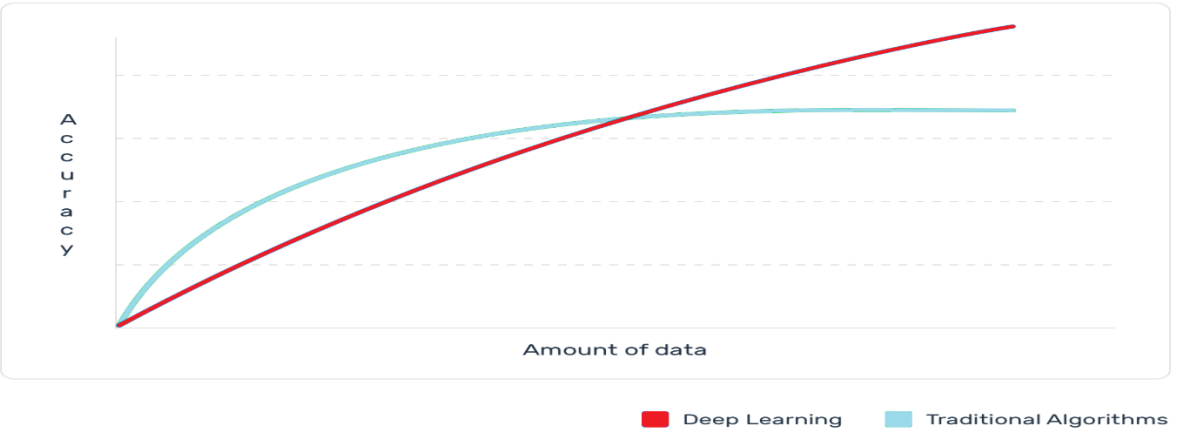


Figure (1) simple of convolution neural networks layers

Deep learning algorithms require big training data (at least millions of tagged) compared with traditional machine learning algorithms. Deep learning algorithms don't have a threshold for learning from training data, like traditional machine learning algorithms, such as Support Vector Machine and DT classifiers continue to get better the more data you feed them with as shown in figure (2)



https://towardsdatascience.com/deep-learning-performance-cheat-sheet-21374b9c4f45

Figure (2) The Relationship Between Accuracy and Amount of Data

1.1.1 One-hot Full Embedding.

This is the most straightforward and commonly used approach to embed categorical features, which assigns each feature value a unique d-dimensional embedding in an embedding table. Specifically, the encoding function E maps a feature value into a unique one-hot vector. the embedding lookup process can be viewed as a 1-layer neural network (without bias terms) based on the one-hot encoding [8].

1.1.2 Word Embedding.

Word embedding is a feature learning technique that is foundational to natural language processing a great asset for a large variety of natural language processing (NLP) tasks, it is a means of turning texts into numbers. We do this because machine learning algorithms can only understand numbers, not plain texts. Word embedding means representing the words in a text in an N-dimensional vector space, thereby enabling the capture of semantics, the semantic similarity between words, and syntactic information for words [9].

Aim of Word embedding mapping words from a vocabulary into vectors of real numbers in a low-dimensional space, by leveraging large corpora of unlabeled text such continuous space representations can be computed for capturing both semantic and syntactic information about words.[10],[3]. Embedding will group commonly co-occurring items together in the representation space.

Embedding is a method that requires large amounts of data and a long training time. The result is a dense vector with an arbitrary number of dimensions. If you have enough training data, enough training time, and the ability to apply the more complex training algorithm (e.g., word2vec or GloVe), go with Embedding.

Embedding is used to represent discrete variables as continuous vectors. It produces a dense vector with a fixed, arbitrary number of dimensions. Word embedding is one of the most popular representations of document vocabulary. The word embedding representation can reveal many hidden relationships between phrases [11].

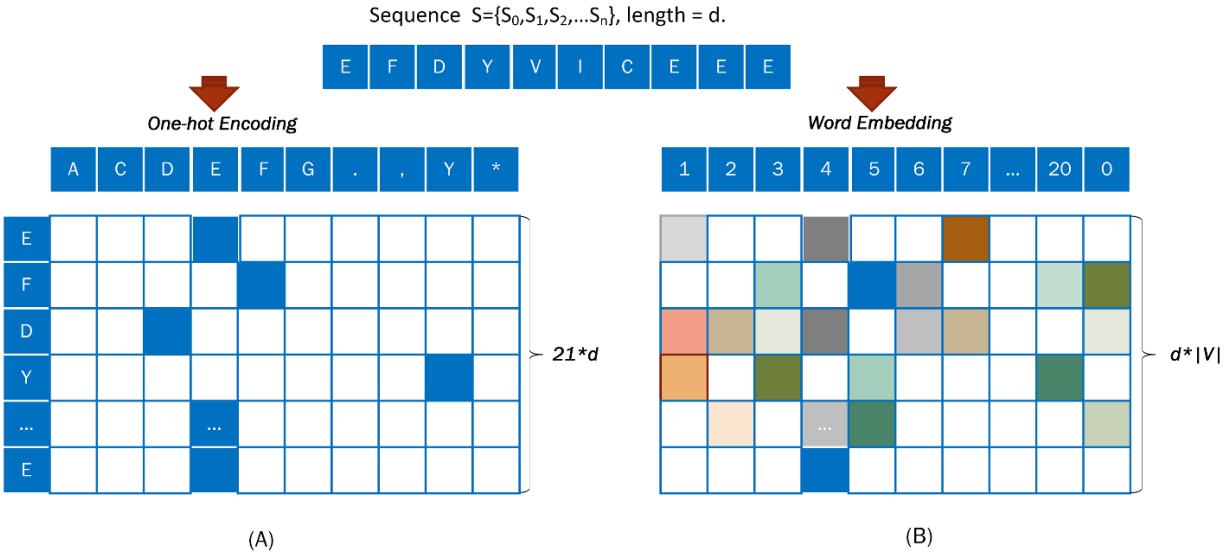


Figure (3) Coding diagram of

(A)One-hot encoding and (B) word embedding encoding

1.1.3 Convolutional neural networks CNNs.

ConvNets are designed to process data that come in the form of multiple arrays. There are four key ideas behind ConvNets that take advantage of the properties of natural signals: local connections, shared weights, pooling and the use of many layers [12]. Figure (4) shows simple of convolution neural networks for text document.

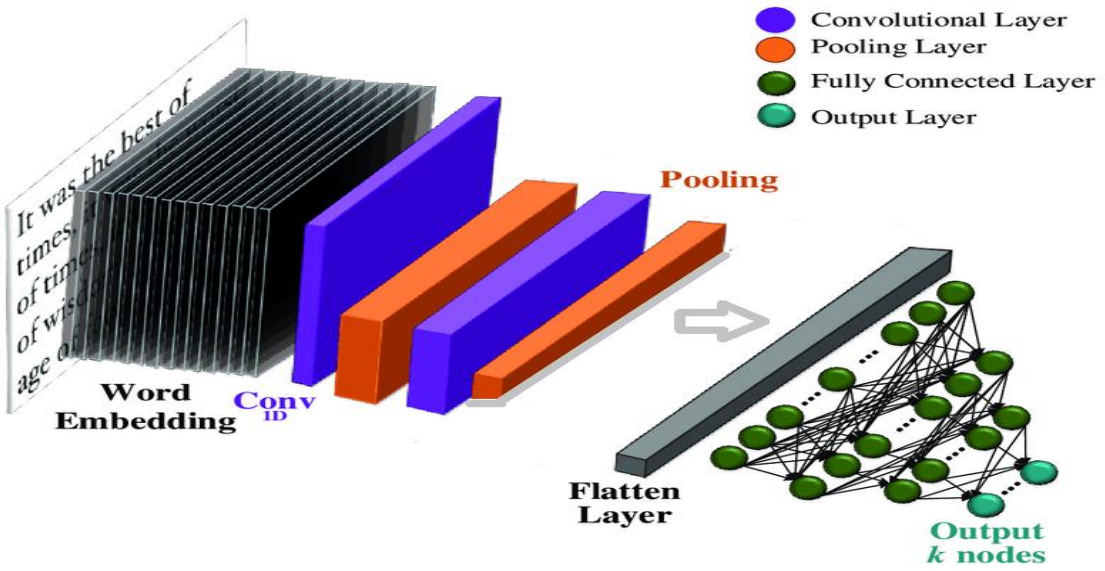


Figure (4) simple of convolution neural networks for text document

The best types of the deep neural networks were convolution neural networks (CNN), and most useful and important, it was usually used in classification for text and segmentation for object in images. Convolution neural networks consists of three main layers: the first layer was convolution, and the second was pooling, and last layer was fully connected. each one of layers does certain spatial operations.

convolution layers convolving the input image or text ,by using different kernels to create the feature maps. the second step was pooling layer. inserted after a convolution layer because the application of this layer reducing the size of network parameters and feature maps.

Flatten layer was the next layer after the pooling layer, there is a followed by some fully connected layers. in this layer, converted the 2D feature maps into 1D feature maps, to be suitable for the following fully connected layers. the flattened vector produced in the previous layer can be used for the text classification.

One advantage of CNNs is the weight sharing mechanism due to the use of the kernels, which results in a smaller number of parameters than a similar fully-connected neural network, making CNNs very easy to train.

1.1.3.1 convolution layer.

There was matching between Convolution and patterns, texts have only one dimension, Therefore, a convolution here is one-dimensional. A typical convolutional model for texts is shown on the figure (5). mostly, a convolutional layer is applied to word embedding, which is followed by a non-linearity (typically ReLU), then a pooling operation was applied. These are the main building blocks of convolutional models: for specific tasks, these blocks are standard.

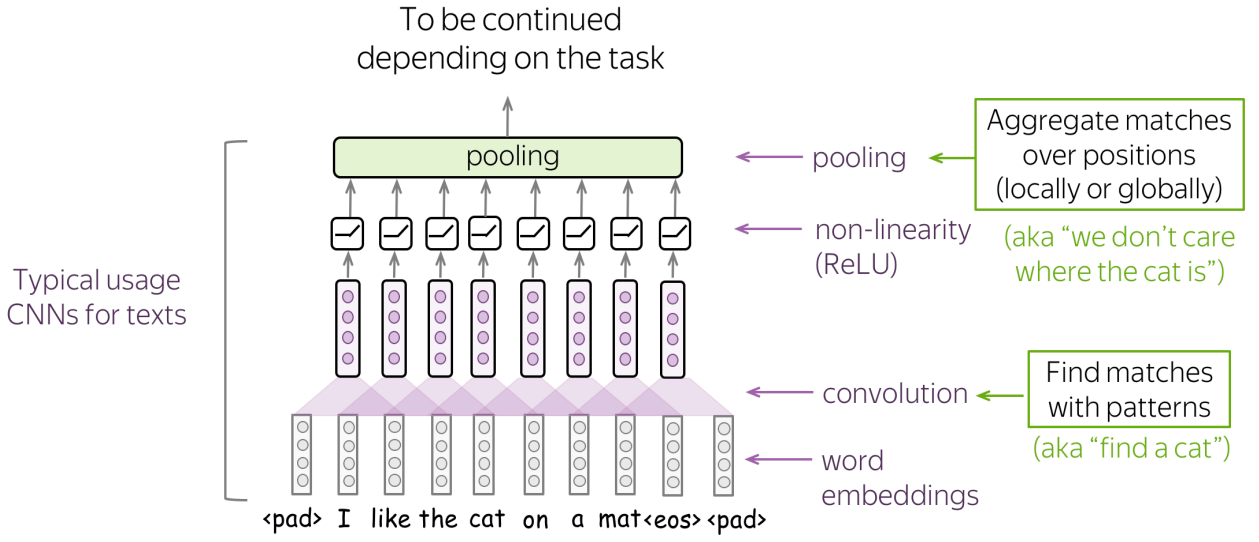


figure (4) typical convolutional model for texts

1.1.3.2 pooling layer.

Pooling layers are used to reduce the number of parameters by reduce the input dimension used by the network. After a convolution extracted features from each window, a pooling layer summarizes the features in some region.[13]

For texts, global pooling is often used to get a single vector representing the whole text; such global pooling is called max-over-time pooling, where the "time" axis goes from the first input token to the last. As shown in figure (5).

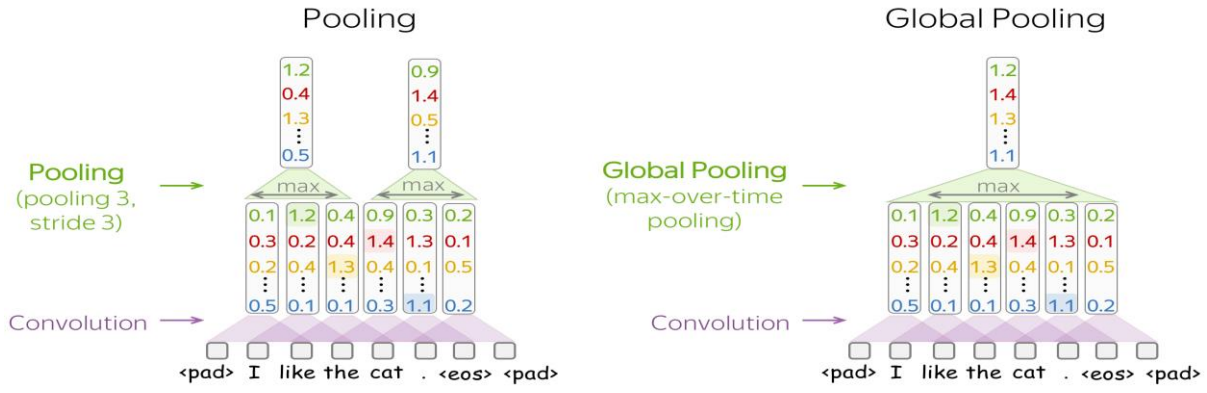


figure (5) simple of pooling layer

max-pooling induces a thresholding behavior. The values below a given threshold are ignored irrelevant to making a prediction. Some time we can that 40% of the pooled ngrams on average can be dropped with no loss of performance.

1.1.3.3 flatten layer and fully connected layer.

Flattening is converting the data into a 1-dimensional array for inputting it to the next layer as shown in figure (6). We flatten the output of the convolutional layers to create a single long feature vector. And it is connected to the final classification model, which is called a fully-connected layer. In other words, we put all the data in one line and make connections with the final layer [14].

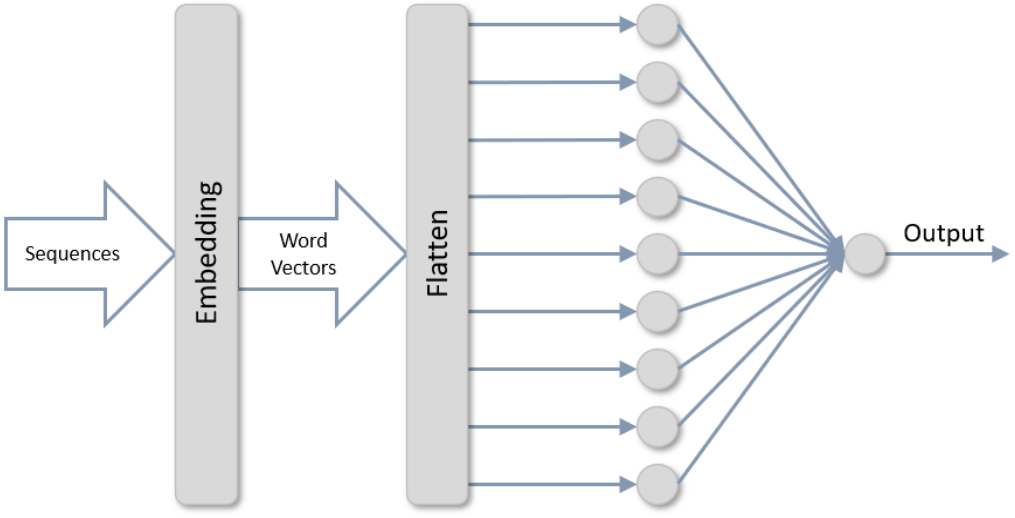


figure (6) Embedding layer and flatten layer

1.1.3.4 Fully-Connected Layer

The fully-connected layer also knows as the dense layer. all the resulting features that selected from the max-pooling layer are combining. the max-pooling layer selects the k-most feature from each convolutional kernel. The fully-connected layer can combine most of the useful assemble and then construct a hierarchical representation for the final stage, the output layer [15]. figure (7) shows the flatten layer and fully connected layer.

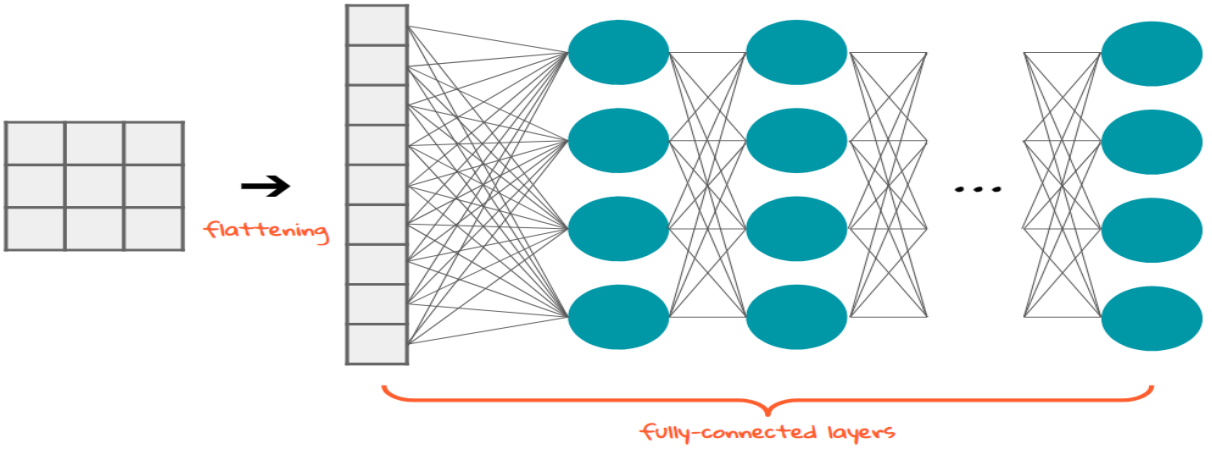


figure (7) flatten layer and fully connected layer

1.1.3.5 Dense layer

This layer is one of the layers that are closely related to the previous layer and it is one of the most used and common layers. Because of this connection, the neurons of the layer are connected to each neuron in the previous layer. The cells of the neuronal layer work to receive the output of each cell from the previous neuronal cells, by multiplying the matrix vector.

The multiplication of the vector requires that the number of layers of the output vector be equal to the number of layers of the column vector in the dense layer.

The general rule of matrix vector multiplication is that the column vector must have as many layers as the row vector.

1.1.4 Parameters.

Kernel size Kernel size means the number of input tokens that the convolution looks at each step.in the text the value of kernel are between 2 and 5. As shown in the figure (3:16).

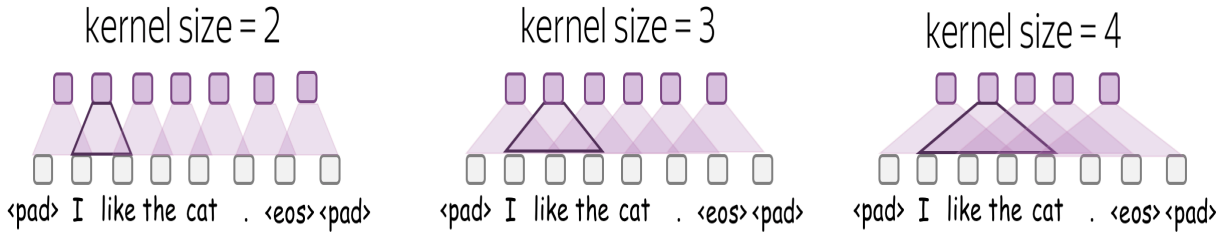


figure (3:16) Kernel

Stride means the step value of move filter between two steps. that was mean, stride equal to 2 that we move the filter by 2 elements from input element (token) for texts at each step. As shown in figure (3:17).

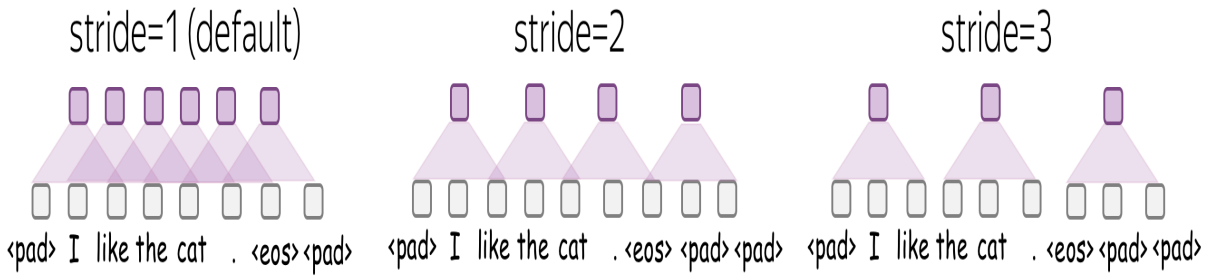


figure (3:17) Stride

Padding: means adding zeros (as a zero vector) to each side of the input. As shown in the figure (3:18).

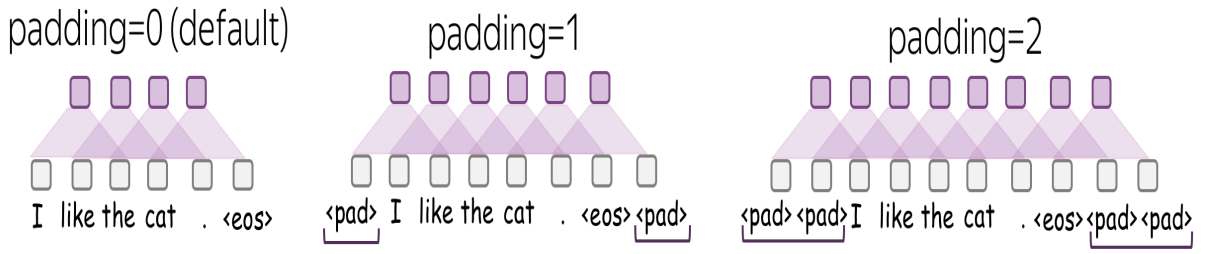


figure (3:18) Padding

Bias: The bias was applied only when multiplying by a matrix in the linear operation in the convolution, as shown in the figure (3:19).

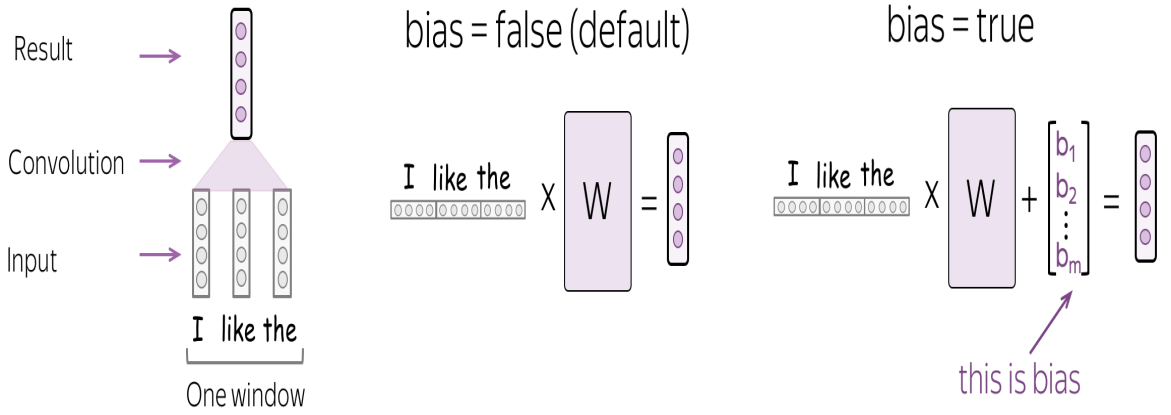


figure (3:19) Bias

Intuition: any filter in the convolution extracts one feature (each filter gives one feature extractor), the filter converts the vector representations in the current window to a single feature as shown in the figure (10).

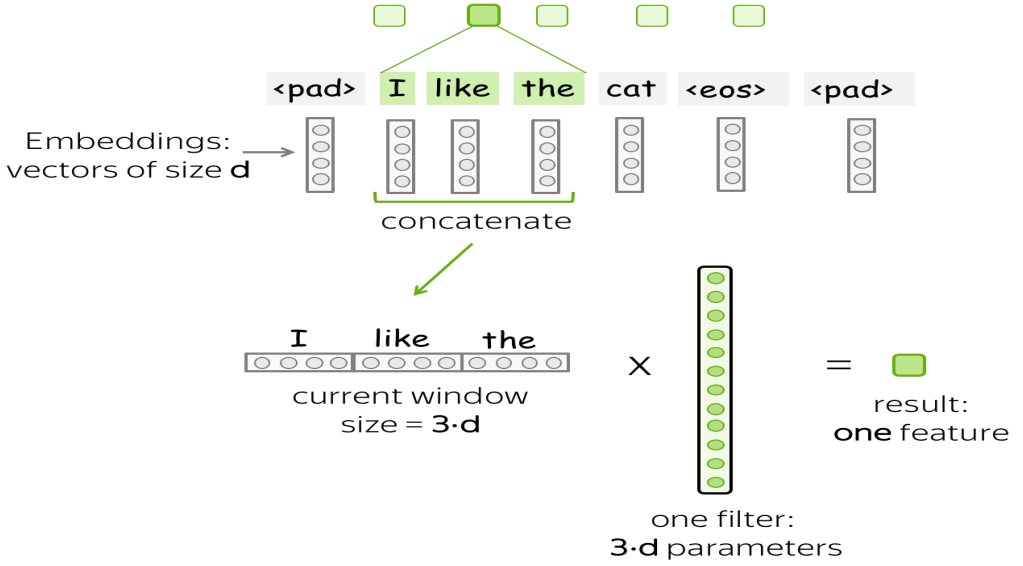


figure (3:20) Intuition

Filters.

As we mentioned earlier that each single filter leads to a single feature, which means that if we want many features we must take several filters. Each filter reads a portion of the input text and extracts a specific feature that is different from the others. [16], so we have the number of filters equal to the number of features of the output you want to have. With M filters instead of one, the size of the convolutional layer becomes**,**

(K . d) × m .

.

k (kernel size)'
d (input channels)

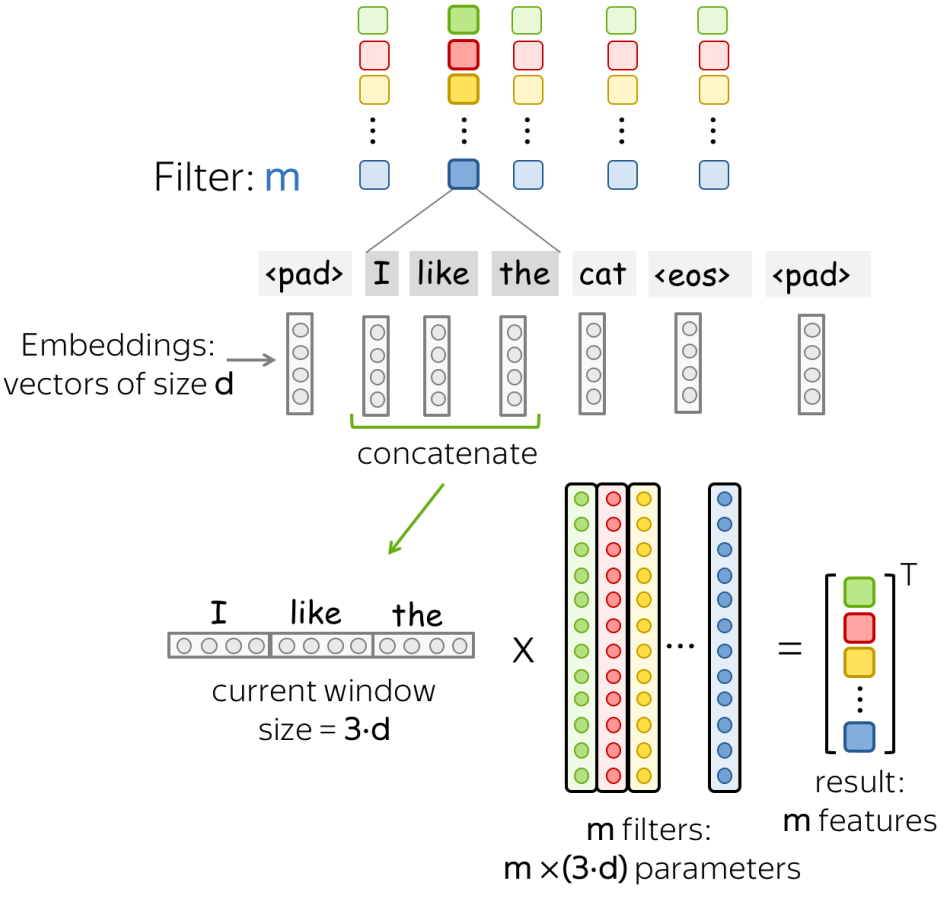


figure (3:21) filters

NLP has been very successful in improving the healthcare process interpreting clinical notes effectively. It extracts details from diagnostic medical reports and doctors' letters and ensures the completeness and accuracy of patient health profiles.

Deep learning is a part of machine learning; it is a neural network with three layers at least. These neural networks attempt to simulate the behaviour of the human brain allowing it to “learn” from large amounts of data.

Deep learning (DL) techniques have begun to dominate because of their simplicity and no need for handcrafted features and efficient processing, meanwhile [17].

Convolutional neural networks deep learning (CNNs) has dramatically improved the approaches to solving many research problems. One of the key differentiators between CNNs and traditional machine learning approaches is the ability of CNNs to learn complex feature representations [18].

1. Related work:

(Fesseha, Awet, et al. [10], 2021), studies convolutional neural networks for Tigrinya which is a Semitic language spoken in Ethiopia and Eritrea by constructing a CNN with a continuous bag-of-words feature extraction method. by developing model that has divided 30,000 text documents into six labels. (categories) include “health”, “politics” “agriculture”, “religion”, “sport” and “education”. CNN's with word2vec method, CNNs without word2vec method. The results found that The CBOW CNN with word2vec achieves the best accuracy with (93.41%).

(Geraci, Joseph, et al.[19], 2017), provided a study of applying machine learning to diagnose youth depression the phenotyping from psychiatric by using 861 labelled documents from electronic medical records (EMRs). The goal was a model to identify individuals who meet inclusion criteria, two psychiatrists who labelled a set of 861 EMRs documents, using a brute force search and training a deep neural network and according to a cross-validation evaluation. The result showed that the model had a specificity of (97%), and a sensitivity of (45%).

(Weng et al. [20], 2017), showed that a supervised learning-based NLP approach is useful for developing medical subdomain classifiers. The medical subdomain, such as cardiology or neurology, of a clinical note is a useful piece of content-derived metadata for developing machine learning downstream applications. NLP was thus used with respect to clinical text analysis and classifying the medical subdomain of a note accurately. In their research, a Unified Medical Language System (UMLS) Metathesaurus, a Semantic Network, and various learning algorithms were used specifically to extract features from a Massachusetts General Hospital data set, with medical subdomain classifiers with different combinations of data representation methods and supervised learning algorithms thus developed.

This led to the construction of a machine learning-based NLP pipeline and the development of medical subdomain classifiers based on the content of each note using a clinical Text Analysis and Knowledge Extraction System (cTAKES). Two datasets, the iDASH data repository (n = 431) and the Massachusetts General Hospital (MGH) database (n = 91,237), were used.

The convolutional recurrent neural network yielded the best F1 scores, with 0.845 for the iDASH dataset and 0.870 for the MGH dataset. This was based on better clinical interpretability from a linear support vector machine-trained medical subdomain classifier using a hybrid bag-of-words and clinically relevant UMLS concepts to develop feature representation.

Using Frequency-inverse document frequency (Tf-IDF) improved the result and proved that weighting outperformed other learning classifier datasets, with an

accuracy of 0.957 and 0.964 and F1 scores of 0.932 and 0.934 for the two datasets, respectively. Classifiers were trained on one dataset and applied to the other dataset, yielding a threshold F1 score of 0.7 with regard to classifiers for half of the medical subdomains.

(Mu, Xiuli, and Hongyan Zhang. [21], 2021), constructed a pure convolutional neural network model (Pure CNN) based on pooled and non-pooled analysis is based on deep learning CNN to mine EMRs text reports.to extract the pregnant women medical information from EMRs text reports. CNN is adopted to classify the Chinese character tags of the neural network model. The model is applied the word segmentation of the actual pregnant women's EMR text. The squired results showed the strong stability of the model when it was conducted on Biomedical Engineering (Peking University) and Microsoft Reserved Partition datasets are between 0.9516 and 0.9684.

(Thomas et al. [22], 2014), performed a study to assess the validity of an NLP program designed to accurately identify patients with prostate cancer: to achieve this, a retrospective review was performed of a prospectively collected database that featured patients from the Southern California Kaiser Permanente. A consecutive series of 18,453 pathology reports were evaluated, and NLP was found to correctly detect 117 out of 118 patients.

This translated to a positive predictive value of 99.1% with 99.1% sensitivity and 99.9% specificity in terms of correctly identify patients with prostatic adenocarcinoma after biopsy. The overall ability of the NLP application to accurately extract variables from the pathology reports was 97.6%.

(Hammoud et al.[23], 2021), presented a new Arabic medical dataset for text classification. The dataset included 2,000 articles invaded into 10 classes (Blood, Bone, Cardiovascular, Ear, Endocrine, Eye, Gastrointestinal, Immune, Liver, and Nephrological) of disease. The model was pre-trained on an Arabic medical reports corpus before fine-tuning on the relevant dataset with the original model produced results of (97.4331) for F1 Validation. and 95.9124 in F1 Testing, with overall SVMs of (89.1308) and (87.3473), respectively.

(Khichdee et al. [24], 2016), introduced an instrument for classification of medical records based on language, based on 24,855 text records. The documents were classified into three groups (endoscopy, ultrasonography, and X-ray), with 13 subgroups, using two methods; these were Support Vector Machine (SVM) and K-Nearest Neighbour (KNN)with feature selection.

At the second stage of classification into subclasses, 23% of all documents could not be linked to only one definite individual subclass (binary system or liver) due to the common features characterising these subclasses.

2. Proposed System.

2.2 pre-processing step.

The methodology proposed for preprocess step in this research is depicted in Figure 1. Many steps have been carried out to achieve the goals of this study:

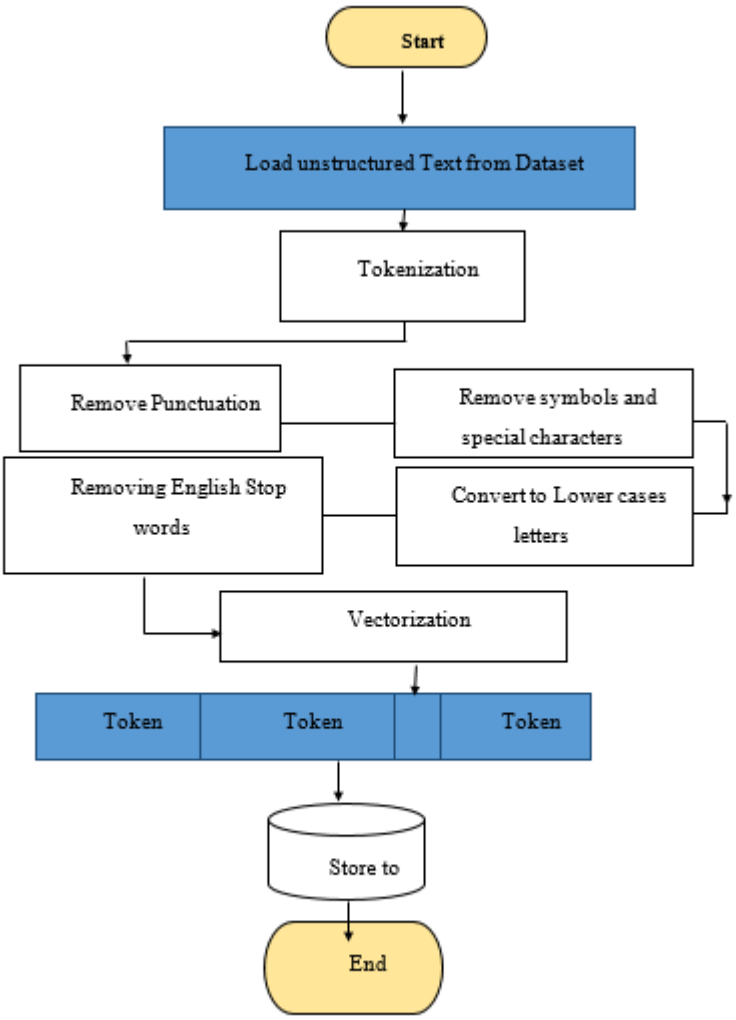


Figure 1 Block diagram of the pre-processing steps

This stage divided into two separate processes, the first one is preprocessing structured data while the second one preprocesses the unstructured data which in turn consists of many steps include (Manual spell Checking, Tokenization, Cleaning Data, removing stop words) to make data suitable for text mining algorithms.

3.1.1 For Unstructured data. pre-processing step includes:

1. **Manual spell Checking:** The spell checker is a software tool that identifies the words that may have been misspelled and also suggest the possible corrections. Every word from the text is looked up in the speller lexicon. when a word is not in the dictionary, it is detected as an error[25] .This step is a sensitive comparison of the word with the closest matched words in dictionary and the result looks different according to the selected word.

   The dataset employed in this study is a clinical reports based multi-class classification which is written by medical stuff and contained a lot of misspelled words, errors which automated spell check packages out there couldn't fix. The manual spell checking Returns a set of possible candidates for the misspelled word and allowing the user selecting the best matching word according to the scope of reports.

2. **Tokenization:** In this study, tokenization is performed by dividing the longer strings of medical reports into smaller pieces, or tokens (words) based on spaces between them.

3. **Clean Data:** which consist of two steps:

   i. **Convert to Lower Case**: This step converts all characters to lowercase to simplify the NLP task.

   ii. **Removing Punctuations**, symbols and special characters**.** This step eliminates unnecessary information such as: '!', '"', '#', '$', '%', '&', "'", '(', ')', '*', '+', ',', '-', '.', '/', ':', ';', '?', '[', '\\', ']', '_', '`', '{', '|', '}', '~' and etc. this process prepare text to be in an appropriate form.

4. **Removing English and domain Stop words:** This stage filters out those words, which contribute little to the overall meaning such as (the, to, this…). For this purpose, the library NLTK, in addition to the set of prespesfied medical words (mg, patient, tb, lt ,..)is used to clean data from stop words.

3.1.2 For structured data preprocessing step include:

1. **Encoding Categorical Data:** in this step all categorical data in the dataset have been encoding to numbers by using integer encoding technique to make data suitable for machine learning models.

2. Handle Missing Values. is an essential part of data cleaning and preparation process because almost all data in real life comes with some missing values. The best method to detect missing values and handle them in a proper.

3. **Data Transformation.** is the process of converting raw data which can be of Text, number, Time series into Vectors (numerical feature). So that we can perform all algebraic operation on it.

Text data usually consists of documents which can represent words, sentences or even paragraphs of free flowing text. These features can then be used in building machine learning or deep learning models easily**.**

3.2 deep learning step.

The methodology proposed of this model was shown in figure (2) below, the first step is the Encoder stage using (one-hot encoding), this stage was in which categorical variables in text reports are converted into a numerical form to enable algorithms to deal with them.

The second stage was CNN deep learning. It contains five Layers, first layer is word embedding, the learning process of Word embedding methods was joint with the neural network model on task, the second layer was Coevolution Neural Network. Then we used the pooling layer, then flatten layer, And the last layer was fully connected layer Figure (2), propose System of Deep Learning Classifier.

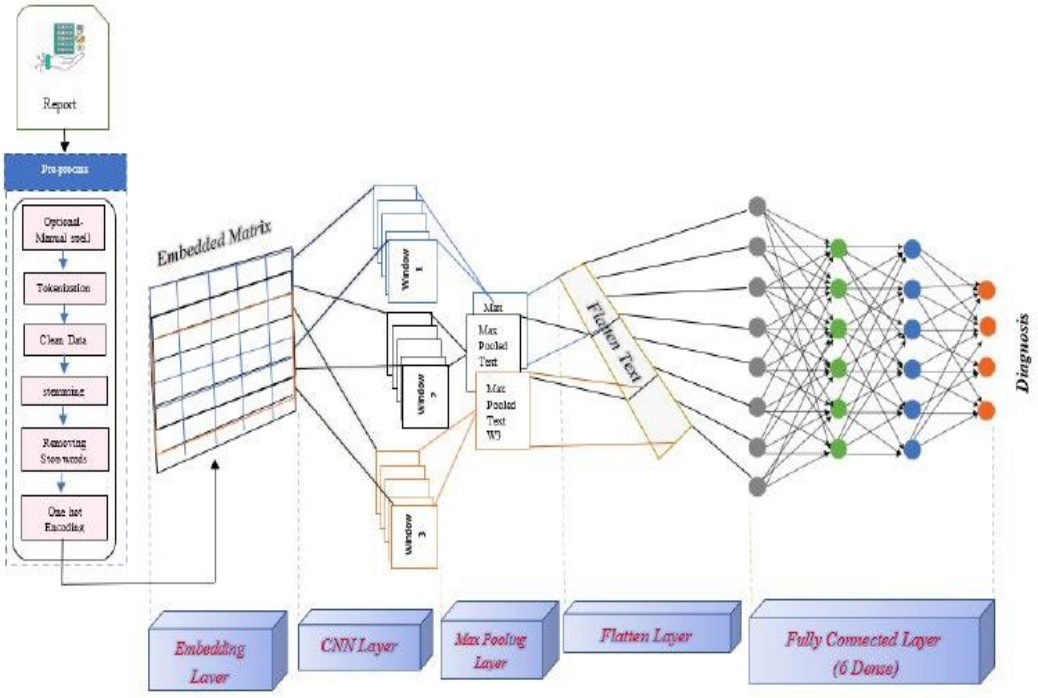


Figure 2 Propose System of Deep Learning Classifier

3.2.1. encoding techniques (Vector Representations of Text).

In order to put the words into the machine learning algorithm the text data should be converted into a vector representations.

in this step, we used a process called one-hot encoding, in which categorical variables in text reports are converted into a numerical form.

The input to the most machine learning algorithms and all deep learning architectures algorithms must be binary values.

every word which are part of the given text data are written in the form of vectors, constituting only of 1 and 0 .each one hot vector being unique, So one hot vector is a vector whose elements are only 1 and 0. Each word is written or encoded as one hot vector, . This allows the word to be identified uniquely by its one hot vector that is no two words will have same one hot vector representation.

3.2.2.Deep learning layers.

In this model we have five layers as shown below,

3.2.2.1 Word Embedded layer.

Embedding layer to transform the “one-hot vectors” into a sequence of dense vectors. Embedding Layer accepts a two-dimensional tensor of size S * L which is the encoded character sequence. Usually, the embedding layer is used for decreasing the dimension of the input tensor.

In this step we used zero-padding tool, to get fixed size by transform the input tensor .and solve the problem of one hot encoding in the previous stage, we can naturally treat the embedding layer as a look-up table. After the processed in embedding layer, the feature extracted by The ConvNets by using the convolutional kernel.

3.2.2.2.Convolution neural network layer.

In convolutional layers, we apply up to three 1D-Convolution layers which have kernel size equal to five and feature map equal to 150 and used No. of filters equal to 128 to extract 128 feature for each enstance . The operation of convolution is widely used in text processing. For the 1D-Convolution we used, there are two signals which are text vector and kernel. After the process, the convolutional operation created a third signal which is the output. The text vector is the output from the embedding layer,vocabolary size equal to 15000 in our setting while kernel has length which is 5.

ReLU was used as our nonlinear activation function, which isalways employed in recent researches. this activation function can better handle the gradient vanishing problem.

ReLU threshold can be best simulate to the brain mechanism of human. L2 regulation is being used in all these layers because it is very effective for solving the overfitting problem.

3.2.2.3.Pooling layer.

It was a necessary layer most followed by the convolutional layer. We used types of pooling called max-pooling layer.

Most important features from the output of 1D-convolution was selected by The pooling layer.

Also, it can accelerate the training speed by diminish the parameters.

In this stage only the most important feature remains because we chose the temporal max-pooling with kernel size equal to the feature maps number.

1.1.4.1 Flatten layer.

In this step we converting the data into a 1-dimensional array for inputting it to the last dence layer. We flatten the output of the convolutional layers to create a single long feature vector. And it is connected to the final classification model, which is called a fully-connected layer. The output of this stage was all the data in one line and make connections with the fully connectionl layer.

The features will be selected from pooling layer sent to the fully-connected layer as a 1D tensor with size (128 * 1).

3.2.2.4.Dense layer.

The most used layer in artificial neural network networks was Dense layer. In any neural network, a dense layer is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer.

3.2.2.5.The fully-connected layer.

Also knows as the dense layer. At this stage, all the resulting features that selected from the max-pooling layer are combining.

the max-pooling layer selects feature from each convolutional kernel. The fully-connected layer can combine most of the useful assemble and then construct a hierarchical representation for the final stage, the output layer.

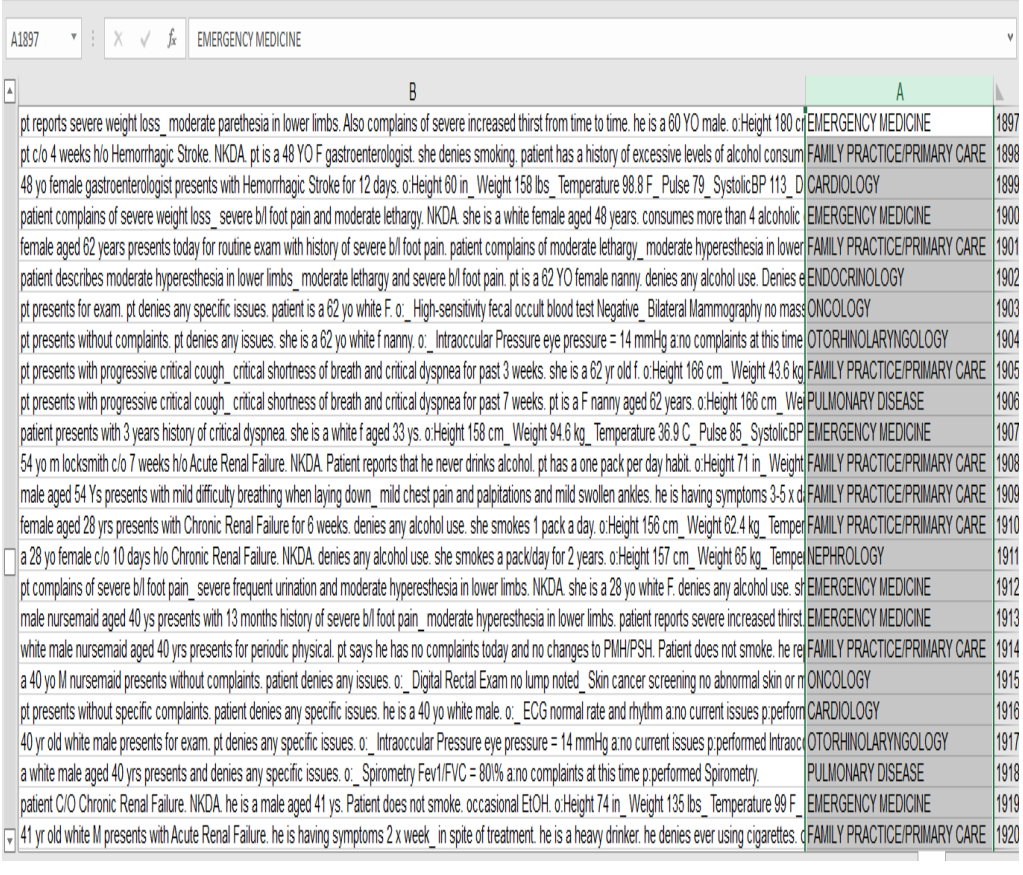
The output layer containe nine neurones because of the number of the target classes. we apply the dropout layer and the dropout rate set to (0.1).

and the dense layer classifies the values emitted from the flatten layer. You can experiment with different dimensions and input lengths in the embedding layer and different numbers of neurons in the dense layer to maximize accuracy.

3.3 dataset.

The data used in this paper was EMRs patient medical reports dataset.

The Encounter dataset records interactions between a patient and healthcare providers for the purpose of providing healthcare services or assessing the health status of the patient. The Date was taken contains medical reports including symptoms and medication notes details of American patients in an emergency hospital collected during ten years. It is available online as a public, free-to-use dataset. It contains data about 3063 medical encounters, classified into 10 classes and 1176 medical fulfilments. Most of the data, however, is contained in the Description Column, free-text natural language description section. This column is thus used for transcription and the assignation of medical specialties. This column was thus assumed to contain the medical specialty for each case, which was then used as a label as shown in figure (3).



A1897 | EMERGENCY MEDICINE

	B	A	
1897	pt reports severe weight loss_ moderate parethesia in lower limbs. Also complains of severe increased thirst from time to time. he is a 60 YO male. o:Height 180 cr	EMERGENCY MEDICINE	
1898	pt c/o 4 weeks h/o Hemorrhagic Stroke. NKDA. pt is a 48 YO F gastroenterologist. she denies smoking. patient has a history of excessive levels of alcohol consum	FAMILY PRACTICE/PRIMARY CARE	
1899	48 yo female gastroenterologist presents with Hemorrhagic Stroke for 12 days. o:Height 60 in_ Weight 158 lbs_ Temperature 98.8 F_ Pulse 79_ SystolicBP 113_ D	CARDIOLOGY	
1900	patient complains of severe weight loss_ severe b/l foot pain and moderate lethargy. NKDA. she is a white female aged 48 years. consumes more than 4 alcoholic	EMERGENCY MEDICINE	
1901	female aged 62 years presents today for routine exam with history of severe b/l foot pain. patient complains of moderate lethargy_ moderate hyperesthesia in lower	FAMILY PRACTICE/PRIMARY CARE	
1902	patient describes moderate hyperesthesia in lower limbs_ moderate lethargy and severe b/l foot pain. pt is a 62 YO female nanny. denies any alcohol use. Denies e	ENDOCRINOLOGY	
1903	pt presents for exam. pt denies any specific issues. patient is a 62 yo white F. o:_ High-sensitivity fecal occult blood test Negative_ Bilateral Mammography no mass	ONCOLOGY	
1904	pt presents without complaints. pt denies any issues. she is a 62 yo white f nanny. o:_ Intraoccular Pressure eye pressure = 14 mmHg a:no complaints at this time	OTORHINOLARYNGOLOGY	
1905	pt presents with progressive critical cough_ critical shortness of breath and critical dyspnea for past 3 weeks. she is a 62 yr old f. o:Height 166 cm_ Weight 43.6 kg	FAMILY PRACTICE/PRIMARY CARE	
1906	pt presents with progressive critical cough_ critical shortness of breath and critical dyspnea for past 7 weeks. pt is a F nanny aged 62 years. o:Height 166 cm_ Wei	PULMONARY DISEASE	
1907	patient presents with 3 years history of critical dyspnea. she is a white f aged 33 ys. o:Height 158 cm_ Weight 94.6 kg_ Temperature 36.9 C_ Pulse 85_ SystolicBP	EMERGENCY MEDICINE	
1908	54 yo m locksmith c/o 7 weeks h/o Acute Renal Failure. NKDA. Patient reports that he never drinks alcohol. pt has a one pack per day habit. o:Height 71 in_ Weight	FAMILY PRACTICE/PRIMARY CARE	
1909	male aged 54 Ys presents with mild difficulty breathing when laying down_ mild chest pain and palpitations and mild swollen ankles. he is having symptoms 3-5 x d	FAMILY PRACTICE/PRIMARY CARE	
1910	female aged 28 yrs presents with Chronic Renal Failure for 6 weeks. denies any alcohol use. she smokes 1 pack a day. o:Height 156 cm_ Weight 62.4 kg_ Temper	FAMILY PRACTICE/PRIMARY CARE	
1911	a 28 yo female c/o 10 days h/o Chronic Renal Failure. NKDA. denies any alcohol use. she smokes a pack/day for 2 years. o:Height 157 cm_ Weight 65 kg_ Tempe	NEPHROLOGY	
1912	pt complains of severe b/l foot pain_ severe frequent urination and moderate hyperesthesia in lower limbs. NKDA. she is a 28 yo white F. denies any alcohol use. sh	EMERGENCY MEDICINE	
1913	male nursemaid aged 40 ys presents with 13 months history of severe b/l foot pain_ moderate hyperesthesia in lower limbs. patient reports severe increased thirst.	EMERGENCY MEDICINE	
1914	white male nursemaid aged 40 yrs presents for periodic physical. pt says he has no complaints today and no changes to PMH/PSH. Patient does not smoke. he re	FAMILY PRACTICE/PRIMARY CARE	
1915	a 40 yo M nursemaid presents without complaints. patient denies any issues. o:_ Digital Rectal Exam no lump noted_ Skin cancer screening no abnormal skin or m	ONCOLOGY	
1916	pt presents without specific complaints. patient denies any specific issues. he is a 40 yo white male. o:_ ECG normal rate and rhythm a:no current issues p:perform	CARDIOLOGY	
1917	40 yr old white male presents for exam. pt denies any specific issues. o:_ Intraoccular Pressure eye pressure = 14 mmHg a:no current issues p:performed Intraoc	OTORHINOLARYNGOLOGY	
1918	a white male aged 40 yrs presents and denies any specific issues. o:_ Spirometry Fev1/FVC = 80\% a:no complaints at this time p:performed Spirometry.	PULMONARY DISEASE	
1919	patient C/O Chronic Renal Failure. NKDA. he is a male aged 41 ys. Patient does not smoke. occasional EtOH. o:Height 74 in_ Weight 135 lbs_ Temperature 99 F_	EMERGENCY MEDICINE	
1920	41 yr old white M presents with Acute Renal Failure. he is having symptoms 2 x week_ in spite of treatment. he is a heavy drinker. he denies ever using cigarettes. c	FAMILY PRACTICE/PRIMARY CARE	

Figure (3) sample of EMR dataset.

4. Experimental Results.

table (1) deep learning results.

Accuracy	*F1-Measure*	*presession*	*Recall*	*time*
99.00	***97.82***	***98.64***	***97.11***	***2m&16seconds***

Figure 5, below shows the confusion matrix for the model explain the result of classifier the reports into nine class.

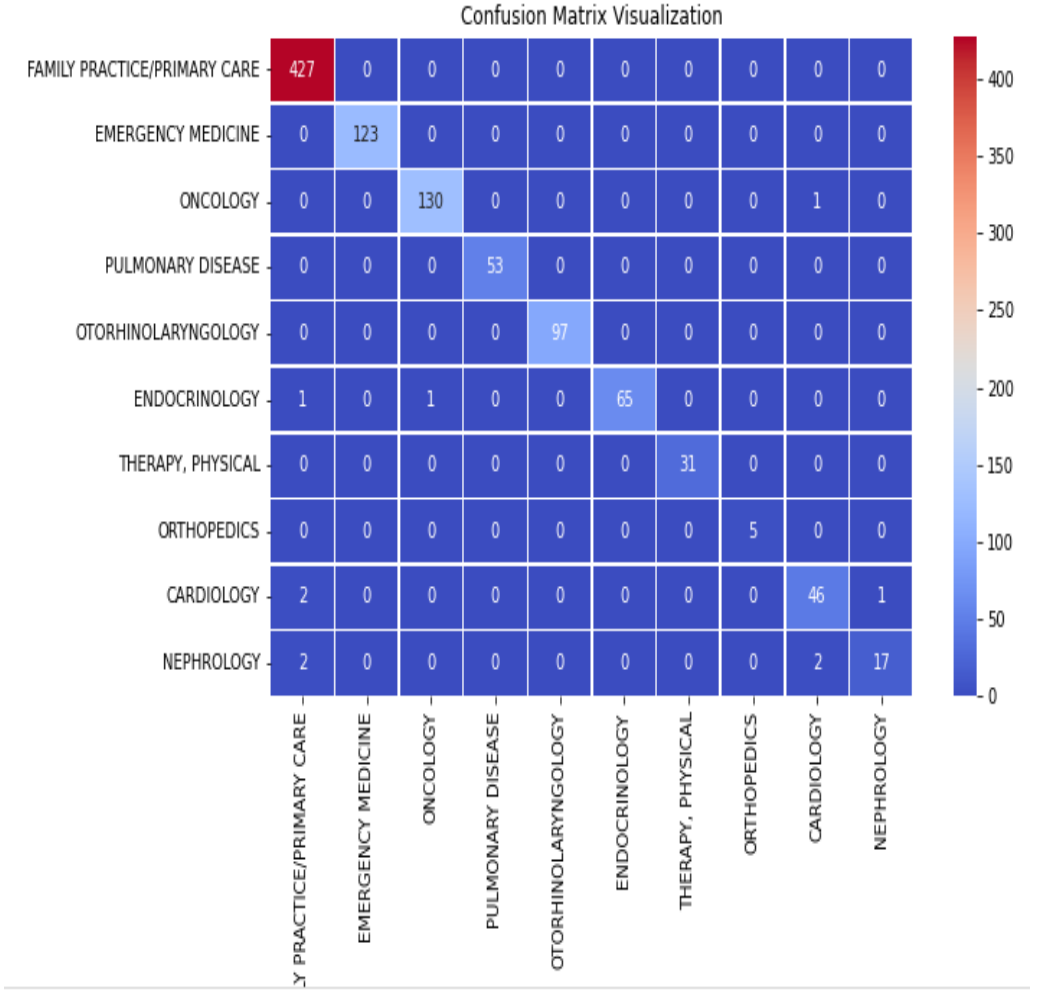


FIGURE (5) ***deep learning CNN results***

Table (2) compare between classical classification and the result and deep learning CNN results.

Table (2) classical classification result and deep learning CNN results

Text Mining Algorithm	***Accuracy Metrics***				
	Accuracy	***F1-Measure***	***precession***	***Recall***	***time***
SVC	***98.22***	***96.86***	***98.68***	***95.39***	***3.19 seconds***
ML-Perceptron	***99.15***	***98.63***	***98.81***	***98.46***	***15.86 seconds***
DT	***98.07***	***96.66***	***96.51***	***96.83***	***0.53 seconds***
CNN DEEP LEARNING	***99.00***	***97.82***	***98.64***	***97.11***	***2m&16seconds***

Discussion the result.

The results show that deep learning can achieve better and more accurate results than the classic algorithms, as shown in Table 2, where the accuracy may reach 99%, but with a difference that the execution time, in this case, was much more than the rest of the algorithms and also needs to advanced processors. In general, promising results emerged for all applied algorithms. The best accuracy achieved was with the ML-Perceptron at 99.39, this also offered a reduction in the amount of time required for training where feature selection was used. Applying all these layers to our data we classified the medical reports into nine classes that resulted in high accuracy equal to (99.00), F1-Measure (97.82), precession (98.64), and Recall (97.11).

Conclusion.

Natural language processing (NLP) is the best way to improve the processing for any medical datasets, It helps to improve the accuracy and efficiency of any process performed on this data, for example extracting patient's information from electronic medical records and determining a patient's medical classification and extract relevant data from electronic records and make a decision about diagnosis

and medication the disease, as this offers high reliability and accuracy, helping create better clinical databases.

A pre-processing step using NLP helps to make the initial unstructured data amenable to processing and mining and rids it of excess and useless attachments in order to allow greater benefit to be derived from the medical information within the initial records.

The results show that deep learning can achieve better and more accurate results than the classic algorithms, Processing tools should be selected according to the characteristics of the data and the principles of dataset design followed.

References

[1] M. Tayefi *et al.*, "Challenges and opportunities beyond structured data in analysis of electronic health records," *Wiley Interdiscip. Rev. Comput. Stat.*, p. e1549, 2021.

[2] R. S. H. Istepanian and T. Al-Anzi, "m-Health 2.0: new perspectives on mobile health, machine learning and big data analytics," *Methods*, vol. 151, pp. 34–40, 2018.

[3] Y. Tedla and K. Yamamoto, "Analyzing word embeddings and improving POS tagger of tigrinya," in *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 115–118.

[4] J. A. Sparks *et al.*, "Rheumatoid arthritis disease activity predicting incident clinically apparent rheumatoid arthritis–associated interstitial lung disease: a prospective cohort study," *Arthritis Rheumatol.*, vol. 71, no. 9, pp. 1472–1482, 2019.

[5] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning--based Text Classification: A Comprehensive Review," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, 2021.

[6] R. J. Roberts, "PubMed Central: The GenBank of the published literature," *Proceedings of the National Academy of Sciences*, vol. 98, no. 2. National Acad Sciences, pp. 381–382, 2001.

[7] J.-D. Kim, T. Ohta, and J. Tsujii, "Corpus annotation for mining biomedical events from literature," *BMC Bioinformatics*, vol. 9, no. 1, pp. 1–25, 2008.

[8] W.-C. Kang *et al.*, "Learning to embed categorical features without embedding tables for recommendation," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 840–850.

[9] S. S. S. KS K, "A Survey of Embeddings in Clinical Natural Language Processing," *arXiv Prepr. arXiv1903.01039*, 2019.

[10] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, "Text Classification Based on Convolutional Neural Networks and Word Embedding for Low-Resource Languages: Tigrinya," *Information*, vol. 12, no. 2, p. 52, 2021.

[11] G. Li, X. Du, X. Li, L. Zou, G. Zhang, and Z. Wu, "Prediction of DNA binding proteins using local features and long-term dependencies with primary sequences based on deep learning," *PeerJ*, vol. 9, p. e11262, 2021.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[13] H. Gholamalinezhad and H. Khosravi, “Pooling methods in deep neural networks, a review,” *arXiv Prepr. arXiv2009.07485*, 2020.
[14] X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.
[15] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018.
[16] R. Ghosh and A. K. Gupta, “Scale steerable filters for locally scale-invariant convolutional neural networks,” *arXiv Prepr. arXiv1906.03861*, 2019.
[17] S. Wu *et al.*, “Deep learning in clinical natural language processing: a methodical review,” *J. Am. Med. Informatics Assoc.*, vol. 27, no. 3, pp. 457–470, 2020.
[18] M. Hughes, I. Li, S. Kotoulas, and T. Suzumura, “Medical text classification using convolutional neural networks,” in *Informatics for Health: Connected Citizen-Led Wellness and Population Health*, IOS Press, 2017, pp. 246–250.
[19] J. Geraci, P. Wilansky, V. de Luca, A. Roy, J. L. Kennedy, and J. Strauss, “Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression,” *Evid. Based. Ment. Health*, vol. 20, no. 3, pp. 83–87, 2017.
[20] W.-H. Weng, K. B. Wagholikar, A. T. McCray, P. Szolovits, and H. C. Chueh, “Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach,” *BMC Med. Inform. Decis. Mak.*, vol. 17, no. 1, pp. 1–13, 2017.
[21] X. Mu and H. Zhang, “Embedded electronic medical record text data mining using neural network association classification algorithm,” *Expert Syst.*, p. e12874, 2021.
[22] A. A. Thomas *et al.*, “Extracting data from electronic medical records: validation of a natural language processing program to assess prostate biopsy results,” *World J. Urol.*, vol. 32, no. 1, pp. 99–103, 2014.
[23] J. Hammoud, A. Vatian, N. Dobrenko, N. Vedernikov, A. Shalyto, and N. Gusarova, “New Arabic Medical Dataset for Diseases Classification,” *arXiv Prepr. arXiv2106.15236*, 2021.
[24] Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2022). Post-pandemic health and its sustainability: Educational situation. *International Journal of Health Sciences*, *6*(1), i-v. https://doi.org/10.53730/ijhs.v6n1.5949
[25] M. Khachidze, M. Tsintsadze, and M. Archuadze, “Natural language processing based instrument for classification of free text medical records,” *Biomed Res. Int.*, vol. 2016, 2016.
[26] G. Neha and P. Mathur, “Spell checking techniques in NLP: a survey,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 2, no. 12, 2012.