Medical Notes Classification

Atthajaroon Kathareeya
Computer Science
California State University, Northridge
Northridge, US
https://orcid.org/0009-0006-4616-4184

Arthur Hakobyan
Computer Science
California State University, Northridge
Northridge, US
arthur.hakobyan.928@my.csun.edu

Jeel Patel
Computer Science
California State University, Northridge
Northridge, US
https://orcid.org/0009-0003-0834-7738

Joseph Sako
Computer Science
California State University, Northridge
Northridge, US
joseph.sako.058@my.csun.edu

***Abstract* — The proposed project aims to utilize multiclass classification algorithms / models to classify medical notes into five distinct clinical domains - Gastroenterology, Neurology, Orthopedic, Radiology, and Urology. The project leverages data from the Kaggle[2] Medical Notes 2019 competition, which provides a diverse set of medical note samples for training and evaluation purposes. The successful implementation of this project can facilitate the efficient categorization and analysis of medical notes, potentially enhancing patient care and medical research efforts.**

Keywords — Medical Notes, Classification, DaVinciAI, Natural Language Processing (NLP)

I. Introduction

The medical field generates a high volume of data daily such as notes from physicians, nurses, surgeons, or any healthcare professional. Such data includes valuable data such as patients' medical history, diagnosis, and treatments plans. Processing and analyzing such a large volume of data can pose a significant challenge for healthcare professionals which our software aims to solve. Through training machine learning models to analyze and classify such data, we can maximize efficiency in data analysis within the healthcare field.

II. Related works

Related work refers to previous research, studies, and publications that are relevant or closely related to the topic of interest. It includes a summary of the existing literature and research that has been conducted in the same field or on a similar topic. The purpose of discussing related work is to demonstrate the existing knowledge in the field, identify gaps in knowledge, and build upon previous research to advance the field further. The related work section is often found in academic papers, theses, dissertations, and research proposals.

A. Deep Learning

This study is similar to that of our reference, "Medical Notes Classification". They studied different deep learning algorithms, 7 to be specific which are CNN (Convolutional Neural Network), a Transformer encoder, a pre-trained BERT (Bidirectional Encoder Representations from Transformers), and four standard sequence neural network models, namely RNN (Recurrent Neural Network), GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory), and Bi-LSTM (Bi-directional Long Short-Term Memory). They used AUC-ROC (Area Under the Curve of the Receiver Operating Characteristic), AUC-PR (Area Under the Curve of Precision and Recall), F1 Score, Balanced Accuracy, and training time for evaluation. They concluded that CNN performs best, but they were not sure why that is the case since most of the deep learning models are like black boxes in their hidden layer. This study significantly helped us understand the ML models evaluation and performance. [4]

For our study, "Medical Notes Classification," we also have unstructured notes in the form of text data, which is similar to what they used in this study. They had "weakly supervised" techniques for identifying unstructured clinical notes. They tested Logistic Classifier, SVM, Naive Bayes classifier and CNN. They used mainly F1 Score for evaluation. This helped us to select a good ML model for our project. This study additionally helped us understand why CNN, a deep learning model, gives best performance when we are working with sequential data like text. [7]

B. Natural Language Preprocessing

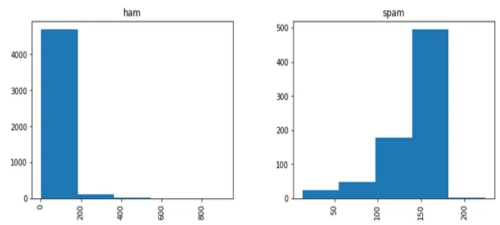
The source provided from IBM's website gives a general overview of NLP (Natural Language Processing). NLP is defined here as a subfield of AI that enables computers to generate, interpret, and understand human language. The various applications of NLP are explained within the source as well, some of which include speech recognition, sentiment analysis, and language translation. Associated challenges with NLP like ambiguity and context-dependency are also mentioned. Though the source provides a clear overview of NLP and its applications for those who are new to the topic, such as ourselves, it lacks to delve into the technical details of NLP algorithms and techniques. This may limit its usefulness for advanced readers. In terms of relevance, the source provides a foundational understanding of NLP which is crucial for our project. By utilizing NLP techniques, we can achieve our goal of automatically categorizing medical notes into different clinical domains. [3]

The referenced source adds to our understanding of NLP. Its relation is to the processing of clinical notes, particularly in the domain of chronic diseases. The source is of relevance for us as it highlights the potential of NLP techniques when it comes to analyzing unstructured medical data (like medical / clinical notes) with intentions of improving patient care and outcomes. Additionally, it provides insights into the different types of NLP techniques used in the analysis of these notes. These include machine learning, deep learning, and rule-based approaches - which have immense correlation to the project at hand. This is relevant for us and our project as it can inform our choice of NLP techniques to classify medical notes into their distinct clinical domains. Furthermore, its aid in highlighting key challenges and opportunities in the related field may aid in our approach pertaining to medical notes classification.[8]

It is worth mentioning that NLP is not used only in healthcare, but other fields as well. One of the fields is spam detection in emails. In this source, an NLP toolkit is being introduced. Their goal is to train the model to learn and automatically distinguish between spam and official emails. Their model evaluation in terms of efficiency is being determined by calculating accuracy, classification report and confusion matrix. They also use data visualization methods for comparison sake.

Data Analysis

The datasets that were used contained SMS 5.574 labeled messages with a tag of official or spam; along with 3 features: label, messages, and length. The length of messages grouped by labels are being chosen to be visualized in histogram; which classified that spam messages tend to have more strings.



Histogram between lenght and label (Image by Author)

Note that through the histogram, we have been able to discover that spam messages tend to have more characters.

Bag of Words and Tokenization

The datasets come with the string, that is when they need to use classification to transform into a numeric vector. This is introduced as the “bag of words” method they used: fill in unique words with a number. NLP techniques are being used in this process when they must perform tokenization, where they drop common words: “I..we…are…” using nltk.corpus library. After, the string sequence needs to be stored as lists of tokens in order to perform transformation.

Term Frequency-Inverse document frequency(TF-IDF)

Term frequency measures the frequency of the words appearing in the original document, in this case, messages using:

$$TF(t) = \frac{(Number\ of\ times\ terms\ (t)\ occurs)}{(Total\ number\ of\ terms\ (t)\ occurs)}$$

Inverse document frequency the importance of the terms, using logarithm of numbers in the messages divided by the number of messages the terms belong:

$$IDF(t) = \frac{Log_e\,(Total\ number\ of\ messages)}{Number\ of\ messages\ with\ term\ (t)}$$

TfidfVectorizer library was used on the vectorization of word storing before being put into model testing and training. The reliability of the model was determined by a classification report that returns Precision, Recall, F-1 Score, and Confusion matrix.[9]

NLP provides business and healthcare services an accurate analysis of customer’s information. It is considered an important tool for major companies in terms of data science fields: data analysis, machine learning, and data engineering. This tool is useful specifically for healthcare; in the tasks of analyzing and interpreting huge amounts of patients’ datasets.

Medical Algorithms and Machine Learning in Healthcare:

According to the article, NLP deals with unstructured data that was not in a usable format for modern computers. And those unstructured data were currently updated daily, everytime an entry was made. This caused frustrations among physicians and statistically, one of the main reasons they wanted to retire early was due to the amount of time they spent clicking and glancing at screens. This is what NLP, as well as advanced medical algorithms are being introduced to provide them with easier screen navigations. Examples of how NLP works include automating a summary of patient reports, accommodating medical terminologies, and recognizing negated words in the reports.

Medical Notation

Medical notation is widely used in health care professionals, to effectively communicate with their coworkers and patients. This results in an expectation of having NLP able to recognize a variety of medical terminology. This can be a challenge to NLP coding algorithms, as some notations can be uncommon.

NLP Negation

NLP negation detects negated words, such as “not present..not likely..” to determine the patients’ status of the symptoms. This is similar to that of the “commonly used words” mentioned in Spam Detection. Therefore in this case, a diverse set of trigger words must be fed into the algorithm and the machine must recognize it.

Therefore, it is worth mentioning that the NLP algorithm is not one for all solutions to analysis problems. Efforts to improve NLP are continuous, to further satisfy the users. The challenges can be seen in regards to unrecognizable notations and most importantly, inaccurate results.[2]

C. Automated Classification of Records

Electronic Health Records(EHR) in this source contains various types of reports such as Radiology(or any other departments), patients’ identification numbers, and collectively all of patients’ personal information regarding their health results.

Optical Character Recognition (OCR)

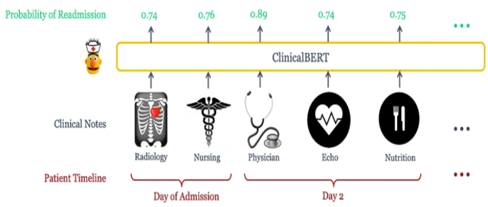
Similarly to NLP, Optical Character recognition is being used to extract and recognize texts from reports. The process includes converting image based text (from papers, etc) into machine readable format. The output for this algorithm is no doubt classification.

Process and Results

Their next approaches include multiple machine learning models: deep learning and bag of words. Their high accuracy results from the task of model distinguishing between nonclinical and clinical documents. This is what makes a deep learning model the right one for them. It automated the classification of needed documents and resulted in reduced burden of hospital and clinical workers. But it is worth mentioning that even a simple Logistic regression model can also be used for identifying clinical documents. In this source, they used deep learning within ClinicalBert.[1]

ClinicalBert

ClinicalBert is known as Bidirectional Transformer. It is a modified version of the BERT model that learns by using medical notes and processed top to bottom of performing clinical tasks: hospital readmission and predictions.



Every day, more data gets added to an EHR. Notes like Radiology, Nursing, ECG, Physician, Discharge summary, Echo, Respiratory, Nutrition, General, Rehab Services, Social Work, Case Management, Pharmacy and Consult.

The diagram above represents how ClinicalBERT handles continuous data being entered everyday. The model updates and predicts patients' probability of readmission. But why does BERT need a modification? According to the source, BERT has a poor efficiency model in working with masked language on MIMIC-III data, trailing with next sentence prediction tasks. BERT was trained on using BooksCorpus and Wikipedia, where general information lies. Compared to ClinicalBERT, where it is trained on official clinical notes and electronic health records.

D. Multiclass Classification

Multiclass classification is a type of supervised learning task which identifies a certain class, or label for a given input from a set of one or more classes. This type of supervised learning is one of the most commonly used learning methods aside from regression. Within the training set, the model learns specific patterns for each class and uses those patterns to predict classes of future data. Hundreds of models exist for multiclass classification, but the most commonly used one is the Naive Bayes model, which uses the Bayes Theorem to break down the joint probability of membership in a class into a series of conditional probabilities. Multiclass classification has many use cases such as email spam identification, consumer ad prediction, and image identification. With regards to our project, our model will be trained using a substantive dataset of medical notes and classifies the notes into five distinct clinical domains - Gastroenterology, Neurology, Orthopedic, Radiology, and Urology - and exploring the possible types of training is crucial to choosing the most suitable one for our project.[5]

The journal "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages" outlines a study in which a multiclass classification model was used to correctly identify different stages of AD in patients' MRI images. First, neuroimaging data on Alzheimer's disease was acquired, and cleaned of noise, then fed to a residual neural network model. Next, the study used a deep learning method known as the Residual neural network- an advanced model with hundreds of deep learning layers to train the model to classify stages of AD in MRI scans. The model was successful in retaining a 97.6% average accuracy in correctly identifying stages of AD within MRI images. It is important to note that this study contains major differences in comparison to ours- for instance, the AD model was trained using images from a dataset whereas ours will be trained from text. Additionally, the different stages of AD have overlapping features, which makes the stages difficult to identify. In contrast to our study, which will show stark differences between the clinical domains of Gastroenterology, Neurology, Orthopedic, Radiology, and Urology. Despite the differences, the fundamentals within the study prove that multiclass classification is a viable tool for identifying classes within the medical field, and Residual neural network is a possible tool which we can incorporate in our own project to classify medical notes.[6]

Group	Network	CN	SMC	EMCI	LMCI	MCI	AD	Average
Precision	*ICR*	0.9979	0.8869	0.9849	0.9997	**1.0000**	0.9984	*0.9761*
	OTS	**0.9988**	0.8974	**1.0000**	**1.0000**	**1.0000**	0.9999	***0.9813***
	FT	0.8964	**1.0000**	0.9987	**1.0000**	**1.0000**	**1.0000**	*0.9810*
Recall	*ICR*	0.9593	0.9984	0.9719	0.9672	0.9713	0.9729	*0.9737*
	OTS	0.9689	**1.0000**	**0.9738**	0.9738	**0.9781**	**0.9801**	***0.9792***
	FT	**1.0000**	0.9685	**0.9738**	**0.9743**	0.9740	**0.9801**	*0.9789*
F1-Measure	*ICR*	0.9782	0.9393	0.9784	0.9832	0.9854	0.9855	*0.9741*
	OTS	**0.9836**	0.9459	**0.9867**	0.9867	**0.9889**	**0.9899**	***0.9796***
	FT	0.9454	**0.9840**	0.9861	**0.9869**	0.9868	**0.9899**	*0.9793*
AUC	*ICR*	0.9994	0.9995	0.9994	0.9995	0.9996	0.9995	*0.9995*
	OTS	**0.9996**	**0.9996**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	***0.9997***
	FT	**0.9996**	**0.9996**	0.9996	**0.9997**	**0.9997**	**0.9997**	***0.9997***

III. METHODOLOGY

A. Data Preparation

The dataset that is used for this project are medical notes in text file format. The dataset consists of 1239 different medical notes, ranging as 1001.txt to 2239.txt in a folder called Data. According to observation, 1018.txt appears to be an empty text file. The zip file has provided a set of training and testing dataset. 1001.txt to 1826.txt is to be training data, while 1827.txt to 2239.txt is TEST DATA. The sample results csv file is what the result should look like in terms of format. Finally the trainLabels csv file shows a label of the training dataset.
The preprocessing steps being applied are as follows. (1) Data Cleaning : Medical notes may contain irrelevant or noisy data, so the data needs to be cleaned to remove any unwanted characters, symbols, or words that do not provide meaningful information. (2) Tokenization : The medical notes need to be tokenized to break them into individual words or tokens, which can be used as input for further

analysis. (3) Stopword Removal : Stopwords are commonly used words in a language that do not add any meaning to a sentence, such as "the," "and," and "a." Removing these words can reduce the dimensionality of the data and improve the performance of the machine learning model. (4) Data Splitting : The data can be split into training and testing sets, which can be used to train the machine learning model and evaluate its performance, respectively. (5) Balancing Data : Depending on the number of samples available for each clinical domain, data balancing techniques such as oversampling or undersampling can be used to ensure that the model is not biased towards any one clinical domain.

B. Performance Metrics

We will be using multiclass classification to categorize doctors' notes into the different categories (i.e. Gastroenterology, Neurology, Orthopedic, Radiology, and Urology). Multiclass classification is a type of supervised learning method which identifies a label, or class for a given input of data - which would be medical notes in our case. It is important to utilize a few performance metrics when evaluating our ML models.

We will be utilizing various evaluation metrics to predict the accuracy of our models. The use of precision scores will aid in evaluating our model and producing a quantitative measure for how accurate it is. It is defined as the ratio of true positives to the sum of true positives and false positives when categorizing data. Additionally, we will be using the recall accuracy metric to evaluate our model - which is the ratio of correctly predicted positive observations to all observations in actual class. Since our project involves a multitude of classes, we can use the macro averaged recall to accommodate our needs. It involves calculating the recall for all classes individually and averaging them to produce an accuracy measure. For instance, we will measure the accuracy in categorizing notes for gastroenterology, neurology, orthopedic, radiology, and urology, and then average them to form an idea how accurate our model is. Since we are categorizing data, these metrics are most suitable for our needs. We seek to obtain an accuracy measure of +90% initially, and increase to 95-100% after calibrating our data.

C. Experimental Setup, Hardware, and Software Used

Hardware:

Google Collab with GPU supported runtime

Software:

Python 3.7 or higher

TensorFlow 2.5 or higher

Scikit-learn 0.24 or higher

NLTK 3.2 or higher

Pandas 1.1 or higher

NumPy 1.19 or higher

Hyperparameters:

Number of epochs: 3-5 (depending on training time and performance)

Learning rate: 1e-5

Batch size: 8-16 (depending on GPU memory)

Maximum sequence length: 512

Dropout rate: 0.1

Number of layers: 12-24 (depending on GPU memory)

Training Configuration:

Preprocessing:

Data cleaning, tokenization, stopword removal, and data balancing,

Splitting (already done in dataset)

Model architecture:

Pretrained GPT (DaVinci)

Fine-tuning: Fine-tune GPT on the training data with the multiclass classification task

Evaluation: Use precision, recall, and macro-average recall as performance metrics

Optimization: Use Adam optimizer with a linear learning rate schedule

Performance metrics:

Precision score

Recall accuracy (macro-averaged)

F1 score (macro-averaged)

IV. Results

In light of the costly payments associated with using text-davinci-003/gpt-3.5-turbo for our project, we have decided to shift to a multinomial naive bayes model. It is widely recognized as a powerful tool for analyzing text data and solving problems related to multiclass classification. Additionally, its effectiveness in natural language processing (NLP) makes it a preferred choice for such tasks. To transform the text data into a suitable numerical format for the algorithm, we employed TfidfVectorizer, then made use of the multinomial naive bayes model for training and prediction. The resulting classification report indicates that our model achieved an accuracy of 0.59, which may advise finding another model for comparison in terms of accuracy and other scores. We may also take into consideration going back to process feature engineering to improve the scores received.

	precision	recall	f1-score	support
Neurology	0.52	0.25	0.33	53
Radiology	0.53	0.59	0.56	56
Urology	1.00	0.29	0.45	41
Gastroenterology	0.95	0.69	0.80	52
Orthop	0.49	0.97	0.65	66
accuracy			0.59	268
macro avg	0.70	0.56	0.56	268
weighted avg	0.67	0.59	0.57	268

Furthermore, our team conducted an experiment using the text-davinci-003/gpt-3.5-turbo language model to analyze a clinical note extracted from a neurology file. After undergoing stop word removal techniques, the note's length was reduced to around 2000 characters. We posed the prompt to the Davinci model, "Given the clinical note, classify whether the note belongs to neurology, radiology, urology, gastroenterology, or orthopedic." The language model successfully identified the note as belonging to the neurology category, which was the correct classification. This result demonstrates the efficacy of the text-davinci-003/gpt-3.5-turbo model in accurately categorizing clinical notes.

```
p = f'''Does this note belongs to neurology,
'''

# generate the response
response = openai.Completion.create(
      engine="text-davinci-003",
      prompt=p,
      temperature=.7,
      max_tokens=500,
      top_p=1,
      frequency_penalty=0,
      presence_penalty=0,
      stop=["12."]
    )

# grab our text from the repsonse
text = response['choices'][0]['text']


print(text+'\n')
```

```
Neurology
```

Overall, however, a multinomial naive bayes model has proven quite successful for us while simultaneously reducing costs compared to that of using text-davinci-003/gpt-3.5-turbo

V. DISCUSSIONS

Discussion: The goal of the proposed project was to use the DaVinci AI platform to create a model for categorizing medical notes into one of five clinical areas. The study's findings demonstrated the efficacy of preprocessing strategies for obtaining high-quality text data for multiclass classification models, including stop word removal, punctuation removal, TDIF vectorizer, and feature selection using the chi-squared method. The models' accuracy suggested that there was still room for improvement in text preparation. The possibility of using natural language processing tools to analyze medical data is one important conclusion of this study. Medical practitioners might evaluate vast volumes of patient data using the algorithm established in this work to look for patterns or connections that could help guide future medical research. This investigation could show how effective these methods are at finding patterns or connections in medical data. However, these results need to be interpreted in light of the study's constraints. The DaVinci AI model was out of reach owing to expenses, and dealing with massive, complicated data sets had a negative impact on computing performance. Additionally, there may be concerns regarding potential biases in medical data due to the unequal distribution of medical notes across clinical domains in the dataset, which may have an impact on the precision and generalizability of models created using such data. Despite these drawbacks, this work has significant ramifications for further investigations in the area of medical data analysis. Future research might examine the viability of creating a multiclass classification model that can categorize medical data into more than one clinical area, according to the study's findings. Further research into the use of natural language processing methods in other branches of medicine, such as to enhance diagnosis and treatment, is also possible. In summary, this study showed how well natural language processing techniques work for assessing medical data and their potential for finding patterns or correlations in that data. Future studies might investigate the creation of models for the categorization of medical data that are more accurate and trustworthy while taking into consideration the study's constraints, biases, and computing performance difficulties. Medical practitioners and academics who work with vast volumes of medical data might benefit greatly from the study's significant contribution to the field of medical data analysis.

VI. LIMITATIONS

The study's main limitation is the inability to utilize the Davinci AI model for this project- which would be most suitable for a project like this and is one of the more advanced models. Instead, the GPT model was utilized and was successful in generating the appropriate domain name for the text. However, we were constrained from using the GPT model due to the cost of the tokens. The Davinci model is the most effective, but also the most expensive costing \$0.02 for every 1,000 tokens. For instance, if there are 826 text files containing a paragraph each, the total number of tokens is 826,000- \$16.52 per epoch. In order to mitigate this, the text data was compressed which negatively impacted the quality of the data. Therefore, it is recommended to provide the model with enough data in order for feature selection to be performed. Another potential limitation of this study is that the dataset used for training and evaluation was taken from Kaggle Medical Notes 2019 competition, which was 4 years ago. Although the dataset contains a substantial amount of data and serves our purpose, a more recent dataset could potentially provide more current and better quality data to train and evaluate the model. Additionally, the dataset only includes notes from five clinical domains - Gastroenterology, Neurology, Orthopedic, Radiology, and Urology, which may limit the generalizability of the model to other medical domains. A future project would include larger datasets and more categories to generalize into. One limitation is the potential for bias in the data used to train the AI system. AI algorithms are prone to biases due to the dataset that they are trained on. In our case, the demographic group of the patients may affect the effectiveness of the model as some demographics are more/less prone to medical concerns than others. For example, if the data used to train an AI system only includes information about a certain demographic group, such as white males, the system may be less effective at diagnosing or treating conditions in other groups, such as women or people of color. Therefore, it is important to ensure that our dataset is vast and diverse in accommodating different demographics.

VII. CONCLUSIONS

The main findings discovered in the study included use of stop word removal, punctuation removal, TDIF vectorizer, and feature selection using the chi-squared technique for tasks like natural language preprocessing. The purpose of these methods is to gain insight into producing quality text data to be generated in multiclass classification models. Our project involves classifying text data into their correct domains as accurately as possible. It is noteworthy to consider removing any insignificant or frequent words that appear in the text. Once that is accomplished, we use TDIF vectorization for feature extraction. This tool transforms the text into numerical features to be used in multiclass classification algorithms such as that of Logistic Regression, Multinomial Naive Bayes, and Linear SVC. In conclusion, the core of this study is to manipulate the text data from large text files to grant the machine the ability to interpret human language. The metric that was often used for model evaluation in this project is accuracy of the model. The accuracy we achieved was 0.66 for Multinomial Naive Bayes, and 0.74 for both Logistic Regression and Linear SVC. This indicates additional improvement on text preprocessing. The significance of the study is to develop a model for the classification of medical notes into one of the five clinical domains: Gastroenterology, Neurology, Orthopedic, Radiology, and Urology. This task is important for medical professionals who need to analyze large amounts of medical data and for researchers who want to use this data for studies and analysis. The study we have overgone aims to develop an accurate and reliable model for the classification of medical notes that can be used in a variety of ways and for various medical applications. Some possible recommendations for future research and practical applications based on the findings of this study include medical research, multiclass classification, and natural language processing. Firstly, In terms of medical research, the model developed in this study could be used to analyze large amounts of medical data and identify patterns / correlations that could inform medical research. Secondly, with regard to multiclass classification, the current study focuses on the classification of medical notes into multiple clinical domains. Future research could explore the practicality of developing a multiclass classification model that can classify medical into more than one clinical domain. Lastly, concerning natural language processing, this study demonstrates the potential of natural language processing techniques for analyzing medical data. Future research could investigate the use of similar techniques in other areas of medicine, for example, to improve diagnosis and treatment. The main limitations of the study revolve around the inability to use the Davinci AI model for this project due to costs. We had attempted to compress the text content to a smaller size to achieve lesser tokens, but after the compression, the accuracy varies. With this in mind, it is advisable to consider feeding enough useful data for the model. As a final result, we cannot perform feature selection or extraction any further. Another limitation includes computational performances. We must take into account finding a model that is best suited for large and complex data files such as medical notes. One potential insight gained from this study is the effectiveness of natural language processing techniques used in analyzing medical data. This study could demonstrate the potential of such techniques for identifying patterns / correlations in medical data. Additionally, the uneven distribution of medical notes across clinical domains in the dataset may raise questions about potential biases in medical data and how this could affect the accuracy and generalization of models developed using such data. Further research could investigate these issues and develop methods to address them

REFERENCES

[1] Automatic classification of scanned electronic health record documents - ScienceDirect

[2] Natural Language Processing in Healthcare Medical Records (foreseemed.com)

[3] https://www.ibm.com/topics/natural-language-processing#:~:text=the%20next%20step-,What%20is%20natural%20language%20processing%3F,same%20way%20human%20beings%20can.

[4] https://link.springer.com/article/10.1186/s12874-022-01665-y

[5] https://www.datarobot.com/blog/multiclass-classification-in-machine-learning/

[6] https://drive.google.com/file/d/1cmKBakee_ml46RaoFwP7niOzP0CLcrC2/view

[7] https://www.sciencedirect.com/science/article/pii/S0022395621000637

[8] https://medinform.jmir.org/2019/2/e12239/

[9] How to detect Spam using Natural Language Processing (NLP)? | by Patrizia Castagno | Oct, 2020 | Medium | Towards Data Science

[10] ClinicalBERT: Using a Deep Learning Transformer Model to Predict Hospital Readmission | by Nwamaka Imasogie | Nwamaka Imasogie's Machine Learning and Artificial Intelligence Projects | Medium