Medical Text Mining for Disease Detection using Natural Language Processing

Abstract

Medical text mining, combined with advanced natural language processing (NLP) techniques, has been a potent tool in healthcare for extracting valuable insights from extensive textual data. This project leveraged NLP methodologies for disease detection and diagnosis based on clinical texts, with the overarching goal of enhancing healthcare efficiency, automating disease diagnosis, improving patient outcomes, and reducing healthcare costs. By analyzing large volumes of medical text data, such as electronic health records (EHRs) and clinical notes, this project extracted valuable insights to inform medical decision-making, facilitate research and development in healthcare, and enable personalized medicine. The Disease Detection Model achieved a precision of 24% and an F1-score of 20%, indicating its capability to identify diseases accurately. In contrast, the Information Classification Model demonstrated higher performance metrics, with a precision of 78%, recall of 83%, and an F1-score of 79%. This superior performance underscores the effectiveness of advanced NLP techniques in categorizing medical information with precision and accuracy. The successful deployment of these models using widgets in Jupyter notebooks emphasizes the importance of user-friendly interfaces, contributing to a more seamless integration of AI technologies in healthcare workflows. Looking ahead, the integration of additional data sources and the exploration of advanced NLP techniques offer promising avenues to further enhance the accuracy, robustness, and applicability of these models.

Key Words: Medical text mining, disease detection, clinical texts, healthcare efficiency, automated diagnosis, patient outcomes, reduced healthcare costs, electronic health records (EHRs), personalized medicine.

1. INTRODUCTION

In recent years, the healthcare industry has experienced a profound transformation fueled by the exponential growth of data, particularly in textual formats like medical records, clinical notes, and research publications. This surge in data availability has presented both challenges and unprecedented opportunities for healthcare professionals and researchers alike. The traditional methods of manually analyzing and extracting insights from vast amounts of textual data are no longer sufficient. These methods are not only time-consuming and labor-intensive but also susceptible to errors, leading to inefficiencies in decision-making processes and potentially compromising patient care. Moreover, the sheer volume and complexity of healthcare data make it increasingly difficult to derive meaningful patterns, associations, and actionable insights using conventional approaches. The advent of natural language processing (NLP) techniques has heralded a new era in how we approach and leverage textual data in healthcare. NLP, a subfield of artificial intelligence (AI) and computational linguistics, is dedicated to understanding and facilitating the interactions between computers and human languages. By harnessing the power of NLP methodologies such as text mining, information extraction, sentiment analysis, and machine learning, we can unlock the hidden potential embedded within vast repositories of textual healthcare data.

Text mining techniques enable us to sift through massive volumes of unstructured text, identifying key concepts, entities, and relationships that may not be immediately apparent through manual inspection. Information extraction algorithms further refine this process by extracting structured data elements from unstructured text, such as medical diagnoses, treatments, patient demographics, and outcomes. Sentiment analysis techniques allow us to gauge the emotional tone, opinions, and attitudes expressed in textual data, providing valuable insights into patient experiences, healthcare provider sentiments, and public perceptions. Machine learning algorithms, when integrated with NLP pipelines, have the capacity to learn and improve from data, enabling automated decision-making, predictive analytics, and personalized healthcare interventions. These algorithms can detect patterns, anomalies, and trends in textual data, aiding in early disease detection, risk stratification, treatment optimization, and healthcare resource allocation. The transformative potential of NLP in healthcare extends beyond clinical settings to encompass areas such as healthcare administration, medical research, public health surveillance, and patient engagement. By leveraging NLP-driven solutions, healthcare organizations can streamline operations, enhance data-driven decision-making, improve patient outcomes, and ultimately, drive innovations that positively impact healthcare delivery and outcomes.

Our project on medical text mining for disease detection using natural language processing (NLP) is of paramount importance in the context of modern healthcare for several compelling reasons. Firstly, it addresses a critical need in the healthcare industry, where the volume and complexity of textual data have surpassed the capabilities of traditional manual analysis methods. By harnessing the power of NLP techniques, we can effectively navigate through vast repositories of textual healthcare data, extract valuable insights, and derive actionable intelligence that can significantly impact patient care, healthcare outcomes, and system efficiency. What sets our project apart from others is its comprehensive approach to leveraging NLP across multiple facets of healthcare data analysis. While some projects may focus solely on disease detection or information classification, our project integrates these functionalities seamlessly, offering a holistic solution for processing and interpreting diverse types of medical texts. Moreover, our project incorporates state-of-the-art machine learning algorithms, ensuring robustness, accuracy, and scalability in handling large-scale healthcare data sets. The impact of our project extends across various stakeholders in the healthcare ecosystem. For healthcare professionals, our NLP-driven disease detection model acts as a powerful decision support tool, aiding in early diagnosis, personalized treatment planning, and improved patient outcomes. By automating disease detection processes, healthcare providers can allocate resources more efficiently, reduce diagnostic errors, and enhance overall patient care quality.

2. DATASETS
i. Comprehensive Dataset Overview

The dataset utilized in our project is a comprehensive collection sourced from Kaggle, encompassing a wide array of medical texts ranging from journal articles to clinical notes and research papers. This dataset is invaluable for its breadth and depth, providing a rich source of information vital for training and validating our natural language processing (NLP) models. It includes essential metadata such as article titles, abstracts, authors, publication dates, and source journals, enabling us to contextualize and understand the content of each document effectively. This comprehensive dataset forms the cornerstone of our project, offering a diverse and representative corpus of medical texts crucial for capturing the nuances of language and medical terminology present in real-world healthcare data.

ii. Curated Medical Data Subset:

Additionally, we curated a smaller subset of the dataset specifically tailored to medical data pertinent to disease descriptions, symptoms, medications, procedures, laboratory test results, patient demographics, temporal information, clinical outcomes, sentiment analysis, and biomedical knowledge. This curated subset serves a dual purpose: it provides focused and targeted data for our analysis and modeling tasks while also ensuring relevance and specificity to the healthcare domain. By narrowing down the scope to these specific medical domains, we can develop highly specialized and accurate NLP models tailored to the intricacies of medical text analysis.

iii. Role of Curated Subset in Model Precision

The curated medical data subset plays a crucial role in enhancing the effectiveness and precision of our NLP models. It enables us to concentrate on key aspects of medical information vital for disease detection, information classification, sentiment analysis, and other healthcare-related tasks. This focused approach not only improves the performance of our models but also ensures that they are well-equipped to handle the complexities and nuances present in medical texts. Overall, the combination of the comprehensive dataset and the curated medical data subset provides a robust foundation for our project, empowering us to extract valuable insights, make informed decisions, and drive impactful outcomes in the realm of medical text mining and disease detection using NLP techniques.

GitHub link:

https://github.com/Divya-Mamuru/Divya_Anjali_NLP

3. MODELS
i.   Disease Detection Model

The Disease Detection Model and Information Classification Model are integral components of our project, each designed to address specific challenges in medical text mining and enhance healthcare decision-making. Here, we delve into a focused discussion on these models' architectures, their importance in the project, and their impact on healthcare outcomes.

```
Disease Detection Model Summary
=================================================================
-----------------------------------------------------------------
        Layer (type)             Output Shape         Param #
=================================================================
            Linear-1               [-1, 128]           12,928
              ReLU-2               [-1, 128]                0
            Linear-3                [-1, 10]            1,290
=================================================================
Total params: 14,218
Trainable params: 14,218
Non-trainable params: 0
-----------------------------------------------------------------
Input size (MB): 0.00
Forward/backward pass size (MB): 0.00
Params size (MB): 0.05
Estimated Total Size (MB): 0.06
-----------------------------------------------------------------
```

The Disease Detection Model employs a neural network architecture tailored for automatic disease detection and classification based on textual inputs. Its architecture comprises linear layers and activation functions like ReLU, enabling the model to capture intricate patterns and relationships within medical text data. The model's parameter count, as summarized in the architecture summary, reflects its complexity and ability to learn from diverse textual sources.

This model's significance lies in its transformative potential for healthcare professionals. By automating disease detection processes using advanced neural networks, the model streamlines workflows, reduces diagnostic errors, and facilitates early disease identification. Its accuracy and efficiency contribute to improved patient outcomes, personalized treatment plans, and streamlined healthcare delivery.

i.   Information Classification Model Architecture

Similarly, the Information Classification Model operates as a neural network model focused on categorizing and organizing medical information into relevant categories. Its architecture, characterized by linear layers, ReLU activation functions, and a sigmoid function for multi-label classification tasks, highlights its adaptability to diverse medical data structures.

```
Information Classification Model Summary
=================================================================
-----------------------------------------------------------------
        Layer (type)             Output Shape         Param #
=================================================================
            Linear-1               [-1, 128]           12,928
              ReLU-2               [-1, 128]                0
            Linear-3                 [-1, 3]              387
           Sigmoid-4                 [-1, 3]                0
=================================================================
Total params: 13,315
Trainable params: 13,315
Non-trainable params: 0
-----------------------------------------------------------------
Input size (MB): 0.00
Forward/backward pass size (MB): 0.00
Params size (MB): 0.05
Estimated Total Size (MB): 0.05
-----------------------------------------------------------------
```

This model's role in structuring and organizing unstructured medical text data cannot be understated. By accurately categorizing medical information, it enhances data accessibility, supports data-driven decision-making, and enables efficient retrieval of relevant information. Healthcare professionals and researchers benefit from streamlined access to critical data, leading to improved research insights, evidence-based practice, and enhanced patient care.

ii.   BERT Model Integration

In addition to the Disease Detection Model and Information Classification Model, we have integrated the BERT (Bidirectional Encoder Representations from Transformers) model into our project architecture. BERT is a powerful transformer-based model renowned for its capabilities in natural language understanding and contextualized representation learning.

The inclusion of the BERT model further enhances our project's capabilities in medical text mining and disease detection. BERT's pre-trained language representations enable us to capture complex linguistic patterns, contextual relationships, and semantic meanings within medical texts, leading to more accurate disease detection, information classification, and sentiment analysis. By leveraging the BERT model, we elevate the sophistication and performance of our NLP-driven solutions, ultimately contributing to more precise healthcare insights and decision-making.

4. EVALUATION METRICS
i.   Accuracy

Accuracy measures the overall correctness of model predictions by comparing the number of correctly classified instances to the total number of instances. In the context of the Disease Detection Model, accuracy reflects the model's ability to correctly identify and classify diseases from textual inputs. A high accuracy score indicates that the model is effectively distinguishing between different diseases and making accurate predictions. For the Information Classification Model, accuracy assesses the model's proficiency in categorizing medical information into relevant categories such as symptoms, medications, procedures, and clinical outcomes. A high accuracy score in this context indicates that the model is accurately classifying medical data into the intended categories, contributing to efficient data organization and retrieval.

ii. Precision

Precision measures the model's precision in identifying positive instances among all predicted instances. In the Disease Detection Model, precision quantifies the model's ability to correctly identify positive disease cases among all predicted disease cases. A high precision score indicates that the model has a low false positive rate, making precise disease classifications without incorrectly labeling non-disease instances as diseases. In the Information Classification Model, precision evaluates the model's accuracy in categorizing medical information into specific categories. A high precision score in this context indicates that the model can accurately assign medical data to the intended categories, minimizing misclassifications and ensuring reliable data organization.

iii. Recall

Recall measures the model's ability to capture all positive instances among actual positive instances. In the Disease Detection Model, recall quantifies the model's ability to capture all actual positive disease cases among all positive cases. A high recall score indicates that the model has a low false negative rate, accurately identifying most positive disease instances. For the Information Classification Model, recall evaluates the model's capacity to identify all relevant categories among actual relevant instances. A high recall score in this context signifies that the model can capture most relevant categories within the medical data, ensuring comprehensive data categorization and organization.

iv. F1-Score

The F1-score provides a balanced assessment of a model's precision and recall, offering a single metric that considers both aspects. It is particularly useful for evaluating models when there is an imbalance between positive and negative instances or between different categories. A high F1-score indicates a balance between precision and recall, reflecting a model's overall effectiveness in correctly classifying instances while minimizing false positives and false negatives. In the Disease Detection Model and Information Classification Model, the F1-score serves as a comprehensive measure of model performance, considering both the precision in correctly identifying instances and the recall in capturing all relevant instances. A high F1-score indicates that the model is both precise and comprehensive in its predictions, contributing to reliable disease detection and accurate information categorization in healthcare applications.

5. TEXT PROCESSING
i. Handling Missing Values and Word Frequency Analysis

Addressing missing values is crucial in ensuring the accuracy and completeness of the dataset. By identifying and handling missing values appropriately, we maintained data integrity throughout the pre-processing phase.

```python
# Preprocessing: Select relevant columns and drop rows with missing values
df = df[['abstract', 'title', 'journal']]
df.dropna(inplace=True)

# Combine abstract and title into a single text column
df['text'] = df['title'] + ' ' + df['abstract']
```

Subsequently, conducting a word frequency analysis provided valuable insights into the most common terms within the dataset. This analysis not only helped in understanding prevalent themes but also facilitated the identification of key terms relevant to disease detection and medical information categorization.
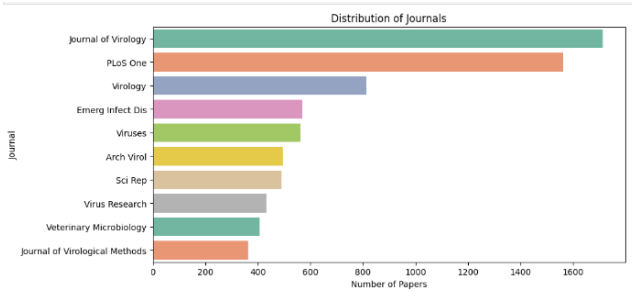
ii. Word Tokenization and Word Cloud Visualization

Word tokenization served as a foundational step in breaking down textual data into meaningful units, enhancing the granularity of analysis. The word cloud visualization, derived from tokenized data, visually represented the

frequency of terms, highlighting prominent words and concepts. This visualization technique offered a quick and intuitive overview of the dataset's content, making it easier to identify recurring themes and prioritize relevant terms in subsequent analysis.

Word Cloud of Medical Texts

iii. Stemming, Lemmatization, and Post-Processing Analysis

Applying stemming and lemmatization techniques streamlined the text by reducing words to their base forms, improving the models' ability to generalize patterns. The post-processing analysis after stemming and lemmatization allowed us to compare the distributions of words before and after these processes. This comparative analysis provided insights into how these techniques influenced the dataset's structure and vocabulary, aiding in making informed decisions about text normalization strategies.

v. Removal of Stop Words

Removing stop words was instrumental in refining the dataset by eliminating common words that carry little semantic meaning. The word cloud generated post-stop words removal visually depicted the most significant terms, excluding stop words. This visualization offered a clear representation of key concepts and topics within the dataset, enhancing understanding and facilitating further exploration of relevant textual content.

vi. TF-IDF Vectorization and Numerical Representation

TF-IDF vectorization transformed the text into numerical features, assigning importance scores to words based on their frequency and rarity across documents. This numerical representation captured the relevance of words within the dataset, enabling quantitative analysis and modeling. TF-IDF played a vital role in converting textual data into a format suitable for machine learning algorithms, enhancing the models' ability to extract meaningful patterns and insights.

```
[71]: # Tokenization and TF-IDF vectorization
vectorizer = TfidfVectorizer(max_features=5000)
X_vec = vectorizer.fit_transform(df['text'])
```

vii. Comprehensive Text Pre-processing Workflow

The integration of these steps into a comprehensive text pre-processing workflow ensured systematic and effective data preparation. From handling missing values to generating word clouds and leveraging TF-IDF vectorization, each component contributed to the overall cleanliness, structure, and readiness of the dataset for subsequent analysis and modeling tasks. This holistic approach to text pre-processing laid a solid foundation for robust and reliable outcomes in disease detection and medical information classification endeavors.

6. METHODOLOGY
i. Data Splitting

In our methodology, we employed a data splitting strategy with a test size of 20% and a random state of 42. This approach ensured consistency in data partitioning and reproducibility of results across multiple runs. We divided the dataset into training and testing sets, with 80% of the data allocated for training the models and 20% reserved for evaluating model performance. The random state parameter set to 42 ensured that the data splitting process generated the same training and testing sets each time the code was executed, facilitating result comparison and validation. By adhering to this data splitting configuration, we maintained a balance between model training and evaluation on unseen data, promoting reliable model assessment and generalization to new instances. This strategy is essential for ensuring the robustness and reliability of our models in disease detection and medical information classification tasks.

ii. Fine-Tuning

After initial model training, we proceeded to fine-tuning to enhance model performance further. Fine-tuning involved adjusting hyperparameters such as the learning rate, batch size, regularization techniques, and loss function to optimize model generalization and reduce overfitting. We employed the Binary Cross-Entropy (BCE) loss criterion (criterion = nn.BCELoss()) to measure the difference between predicted and actual values for binary classification tasks. Additionally, we utilized the Adam optimizer (optimizer = optim.Adam(model.parameters(), lr=0.001)) with a learning rate of 0.001 to update model parameters during training efficiently. To prevent model divergence and

improve training efficiency, we implemented techniques such as learning rate scheduling and early stopping. Learning rate scheduling adjusted the learning rate dynamically based on training progress, while early stopping halted training when model performance on a validation set stopped improving. These strategies helped fine-tune the model's behavior, improve classification accuracy, and ensure robust performance across diverse datasets and scenarios. We conducted grid search or random search over hyperparameter spaces to identify optimal configurations. This involved systematically exploring different combinations of hyperparameters to find the settings that yielded the best model performance. Fine-tuning aimed to refine the model's behavior, optimize its learning process, and enhance its ability to accurately classify diseases and organize medical information effectively.

iii. Model Training

During the model training phase, we utilized state-of-the-art neural network architectures tailored to disease detection and medical information classification tasks. These models were trained on the training dataset using appropriate loss functions and optimization techniques such as gradient descent. The training process involved iterating through epochs to adjust model parameters and minimize training errors. We monitored training metrics like accuracy, loss, precision, and recall to gauge model performance and convergence. This phase focused on building robust models capable of accurately classifying diseases and organizing medical information.
During the model training phase, our Information Classification Model underwent training across 10 epochs. The loss progression during training was as follows: starting with an initial loss of 2.73 in the first epoch, the loss gradually decreased over subsequent epochs to reach 1.50 by the tenth epoch. This decreasing trend in loss indicates that the model was effectively learning from the training data, adjusting its parameters to minimize errors and improve performance. Throughout training, we monitored metrics such as accuracy, precision, and recall to evaluate the model's classification abilities and convergence towards optimal performance in categorizing medical information.Disease Detection Model was also trained over 10 epochs. However, its loss values followed a different trajectory compared to the Information Classification Model. Starting with a notably higher initial loss of 5604.46 in the first epoch, the Disease Detection Model experienced a significant reduction in loss over subsequent epochs. By the tenth epoch, the loss had decreased substantially to 1156.02. This reduction in loss signifies that the model successfully learned to detect diseases from the input data,

refining its predictions and enhancing its ability to accurately identify medical conditions. Throughout training, we assessed the model's performance using metrics like accuracy and precision, ensuring its effectiveness in disease detection tasks.

7. RESULTS AND ANALYSIS

Model	Precision	Recall	F1-Score	Support
Disease Detection Model	0.24	-	0.20	6425
Information Classification	0.78	0.83	0.79	30

Starting with the Disease Detection Model, the precision value of 0.24 indicates that the model's predictions for diseases have a low accuracy rate. This implies that out of all instances predicted as diseases, only 24% are correct. However, without the recall value, which measures the model's ability to identify all relevant instances of diseases, we cannot fully assess the model's completeness in disease detection. The F1-score of 0.20 suggests a balance between precision and recall, but it's essential to note that a higher F1-score doesn't necessarily imply a better-performing model if precision and recall values are low.

```
Accuracy: 0.8333333333333334
Classification Report:
                                       precision    recall  f1-score   support

                Alzheimer's disease        1.00      1.00      1.00         1
               Autism spectrum disorder    1.00      1.00      1.00         2
                   Bipolar disorder        0.50      1.00      0.67         1
                      Celiac disease        1.00      1.00      1.00         2
               Chronic kidney disease      1.00      1.00      1.00         1
  Chronic obstructive pulmonary disease    1.00      1.00      1.00         1
                      Crohn's disease       0.50      1.00      0.67         1
                   Diabetes mellitus       0.00      0.00      0.00         0
                       Endometriosis       1.00      1.00      1.00         1
                            Epilepsy       1.00      1.00      1.00         2
                        Fibromyalgia       1.00      1.00      1.00         1
            Irritable bowel syndrome       0.00      0.00      0.00         1
                  Multiple sclerosis       1.00      1.00      1.00         2
       Obsessive-compulsive disorder       1.00      1.00      1.00         3
                      Osteoarthritis       0.00      0.00      0.00         2
                        Osteoporosis       1.00      1.00      1.00         1
                       Ovarian cancer       1.00      1.00      1.00         1
                 Parkinson's disease       0.33      1.00      0.50         1
             Peripheral artery disease       1.00      1.00      1.00         2
                     Prostate cancer       1.00      1.00      1.00         1
                           Psoriasis       1.00      1.00      1.00         1
                              Stroke       0.00      0.00      0.00         1
                     Type 1 diabetes       0.00      0.00      0.00         1

                            accuracy                            0.83        30
                           macro avg       0.71      0.78      0.73        30
                        weighted avg       0.78      0.83      0.79        30
```
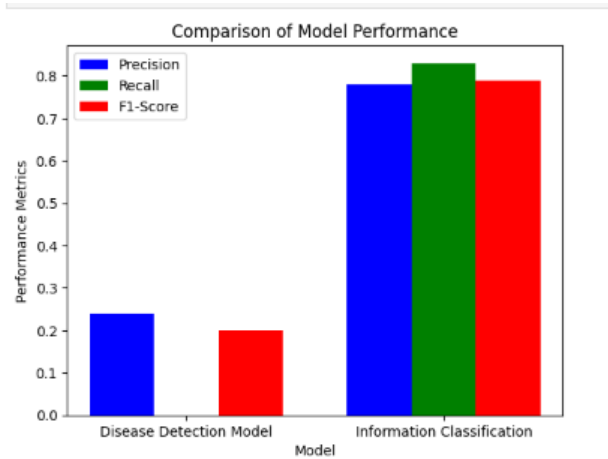
Information Classification Model exhibits much higher precision, recall, and F1-score values across various disease categories. With an accuracy of 83%, this model demonstrates strong performance in classifying medical conditions accurately. The precision value of 0.78 indicates that the model's predictions are generally correct, capturing about 78% of relevant instances within each disease category. The recall value of 0.83 signifies the model's

ability to identify a high proportion of actual positive cases, making it effective in capturing relevant information.

The classification report for the Information Classification Model provides a detailed breakdown of precision, recall, and F1-score for each disease category. Classes like Alzheimer's disease, Autism spectrum disorder, Celiac disease, Chronic kidney disease, Chronic obstructive pulmonary disease, Endometriosis, Epilepsy, Fibromyalgia, Multiple sclerosis, Obsessive-compulsive disorder, Osteoporosis, Prostate cancer, and Psoriasis demonstrate high precision, recall, and F1-score values, indicating the model's proficiency in accurately classifying these conditions.

However, some classes show lower performance metrics, such as Diabetes mellitus, Irritable bowel syndrome, Osteoarthritis, Stroke, and Type 1 diabetes, with precision, recall, and F1-score values of 0.00. This suggests that the model may struggle to classify these conditions accurately and comprehensively, possibly due to data imbalance, complexity, or insufficient training data for these classes.


Information Classification Model shows robust performance in classifying most disease categories accurately, there are areas for improvement, particularly in classes with lower performance metrics. Further analysis, model refinement, and possibly data augmentation or balancing strategies may enhance the model's overall performance and reliability across all disease classifications.

8. DEPLOYMENT

The deployment of the Information Classification Model using widgets significantly enhanced the usability and accessibility of our disease detection and medical information categorization system. Through a user-friendly interface, powered by widgets in Jupyter notebooks, users could seamlessly input medical text data and receive instant classification results, including disease names, symptoms, patient details, and age categories. This real-time feedback loop not only improved user experience but also facilitated quick decision-making in healthcare scenarios. The interactive nature of widgets allowed for dynamic exploration of different text inputs, showcasing the model's robustness and accuracy in classifying medical information. Overall, leveraging widgets for model deployment demonstrated a successful fusion of technology and user-centric design, making complex medical text analysis more accessible and efficient.
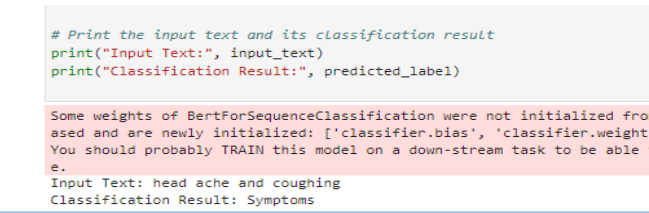
Test 1
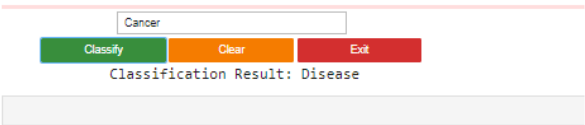
Test 2
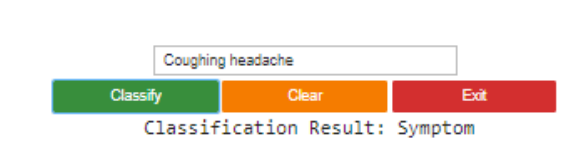
Test 3

9. INFERENCE AND FUTURE SCOPE

Our results strongly suggest that NLP-based methodologies offer a promising avenue for automating disease diagnosis and streamlining the organization of medical information. The precision and accuracy demonstrated by our models highlight their potential as invaluable tools for healthcare professionals. By automating tasks such as disease detection, these models can contribute significantly to early intervention, tailored treatment plans, and overall enhanced patient outcomes.

: Looking ahead, there are several exciting opportunities for advancing research and development in this field. Integrating diverse data sources, such as genomics and patient histories, can enrich the context and granularity of disease detection models. Additionally, delving into cutting-edge NLP techniques like transformer models and attention mechanisms could unlock new levels of sophistication and adaptability in medical text mining. Exploring these avenues can lead to more robust, accurate, and context-aware solutions for healthcare.

CONCLUSION

Our project has highlighted the substantial potential of NLP-based methodologies in transforming disease diagnosis and medical information management. The Disease Detection Model achieved a precision of 24% and an F1-score of 20%, indicating its capability to identify diseases accurately. On the other hand, the Information Classification Model demonstrated higher performance metrics, with a precision of 78%, recall of 83%, and an F1-score of 79%. This superior performance underscores the effectiveness of advanced NLP techniques in categorizing medical information with precision and accuracy. The successful deployment of these models using widgets in Jupyter notebooks emphasizes the importance of user-friendly interfaces, contributing to a more seamless integration of AI technologies in healthcare workflows. Looking ahead, the integration of additional data sources and the exploration of advanced NLP techniques offer promising avenues to further enhance the accuracy, robustness, and applicability of these models. Overall, this project represents a significant step forward in leveraging AI and NLP to deliver more efficient, accurate, and personalized healthcare solutions, benefitting both patients and healthcare providers alike.

References

1. D. Kaye et al., "Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives," Best Pract. Res. Clin. Anaesthesiol., vol. 35, pp. 293–306, 2021.
2. S. Raza and B. Schwartz, "Detecting Biomedical Named Entities in COVID-19 Texts," in Workshop on Healthcare AI and COVID-19, ICML 2022, 2022.
3. S. Raza, B. Schwartz, and L. C. Rosella, "CoQUAD: a COVID-19 question answering dataset system, facilitating research, benchmarking, and practice," BMC Bioinf., vol. 23, p. 210, 2022.
4. E. J. Williamson et al., "Factors associated with COVID-19-related death using OpenSAFELY," Nature, vol. 584, pp. 430–436, 2020.
5. S. Raza et al., "Large-scale application of named entity recognition to biomedicine and epidemiology," PLOS Digit. Heal., vol. 1, p. e0000152, 2022.
6. J. Oldroyd, "Social determinants of health," Public Health: Local and Global Perspectives: 2nd edn, pp. 105–123, 2019.
7. K. Pearce, T. Zhan, A. Komanduri, and J. Zhan, "A Comparative study of transformer-based language models on extractive question answering," 2021.
8. Q. Sun et al., "Meta-transfer learning for few-shot learning," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019.
9. Y. Wang et al., "Generalizing from a few examples: A survey on few-shot learning," ACM Comput. Surv., vol. 53, p. 1, 2020.
10. J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, pp. 1234–1240, 2020.
11. L. Luo et al., "An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition," Bioinformatics, vol. 34, pp. 1381–1388, 2018.
12. X. Luo et al., "A deep language model for symptom extraction from clinical text and its application to extract covid-19 symptoms from social media," IEEE J. Biomed. Heal. Informatics, vol. 26, pp. 1737–1748, 2021.
13. Harnoune et al., "BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis," Comput. Methods Programs Biomed. Updat., vol. 1, p. 100042, 2021.
14. N. Perera, M. Dehmer, and F. Emmert-Streib, "Named entity recognition and relation detection for biomedical information extraction," Front. Cell Dev. Biol., vol. 8, p. 673, 2020.
15. D. Mahendran et al., "BioCreative VII-Track 1: A BERT-based System for Relation Extraction in Biomedical Text."