Simplifying ICD-10 Coding: Evaluating Pre-trained Language Models and Traditional Methods for Clinical Notes Mapping

Syed Bilal Rizwan[1], Prof Dr. Sajjad Haider[1]*

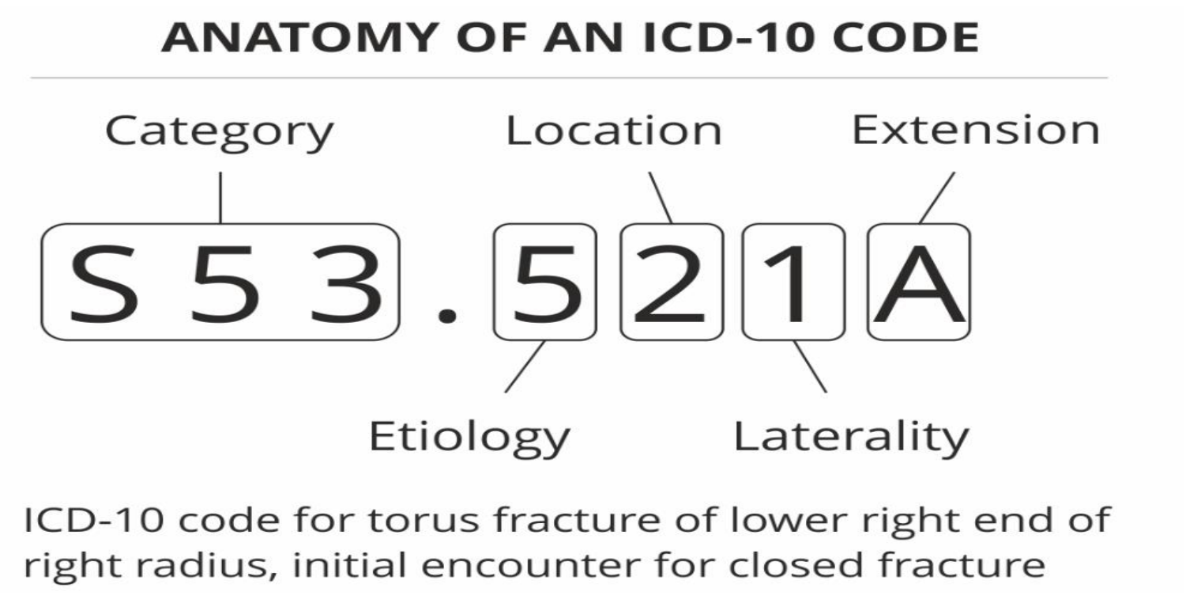
[1]Department of Computer Science, Institute of Business Administration, Karachi, Pakistan

BACKGROUND & MOTIVATION

The healthcare industry is dynamic and data-driven, evolving constantly to improve patient care, operational efficiency, and service quality. The usage of standardized coding systems dates to the 1800s, when the importance of such systems became clear due to nomenclature irregularities and the need for better statistical data.

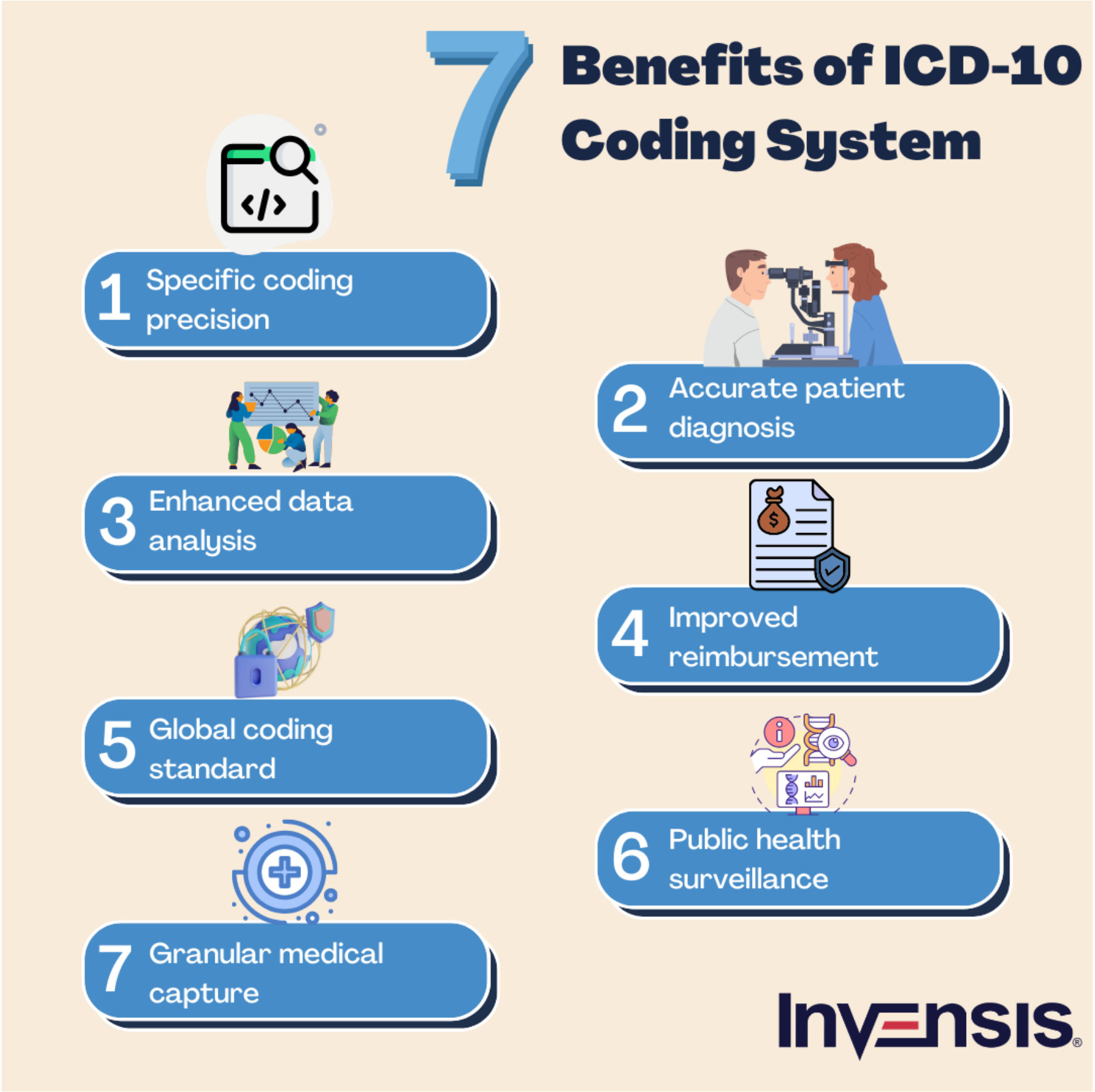
What are ICD-10 Codes?

The International Classification of Diseases, Tenth Revision (ICD-10) is a worldwide recognized and standardized system of alphanumeric numbers used to identify diseases, disorders, and other health-related issues.
The first 3 digits represent the category of disease while the latter digits represents further specifications such as location, etiology


ANATOMY OF AN ICD-10 CODE
Category
Location
Extension
S53.521A
Etiology
Laterality
ICD-10 code for torus fracture of lower right end of right radius, initial encounter for closed fracture

Challenges and Importance in Medical Coding

There are over 73,000 ICD-10 codes which makes manual coding by medical practitioners intricate and time-consuming. This requires healthcare institutes to hire additional experts to carry out the task.
However, despite the challenges, the benefits of medical coding compel healthcare institutes to carry out the tedious task.


7 Benefits of ICD-10 Coding System
1 Specific coding precision
2 Accurate patient diagnosis
3 Enhanced data analysis
4 Improved reimbursement
5 Global coding standard
6 Public health surveillance
7 Granular medical capture
Invensis

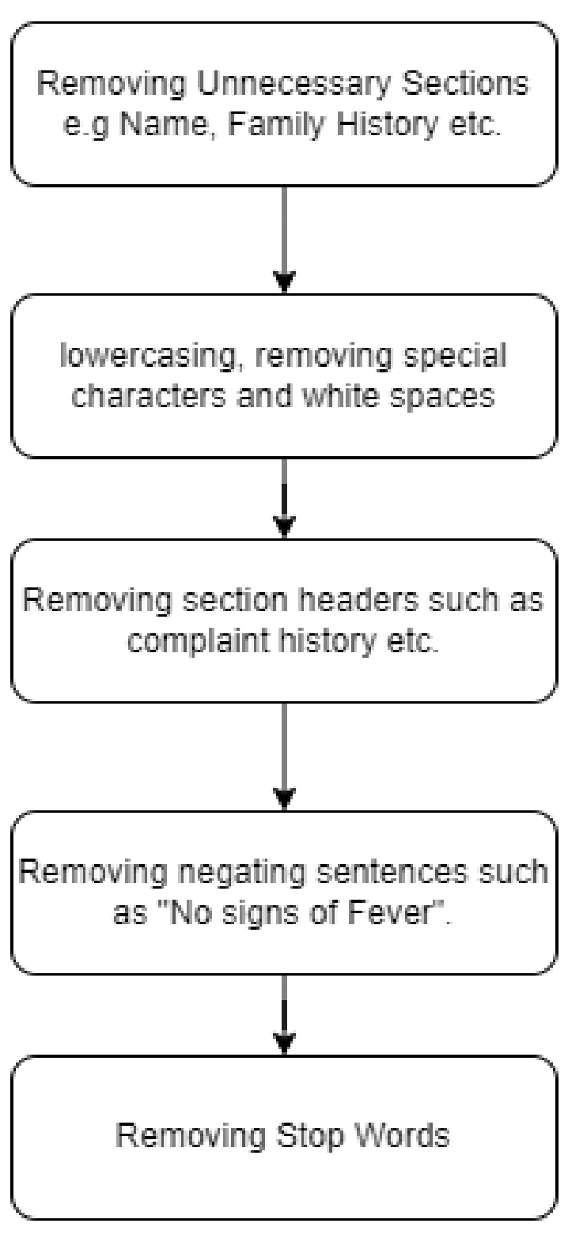
Project Objective

This project aims to improve how clinical notes are mapped to ICD-10 codes by exploring pretrained language models and refining natural language processing techniques. An essential step involves creating a strong preprocessing technique for complex clinical notes. The main goal is to compare these new methods with existing ones like SNOMED CT with MEDCAT library, with the ultimate goal of finding an accurate and efficient existing model for clinical coding

Methodology

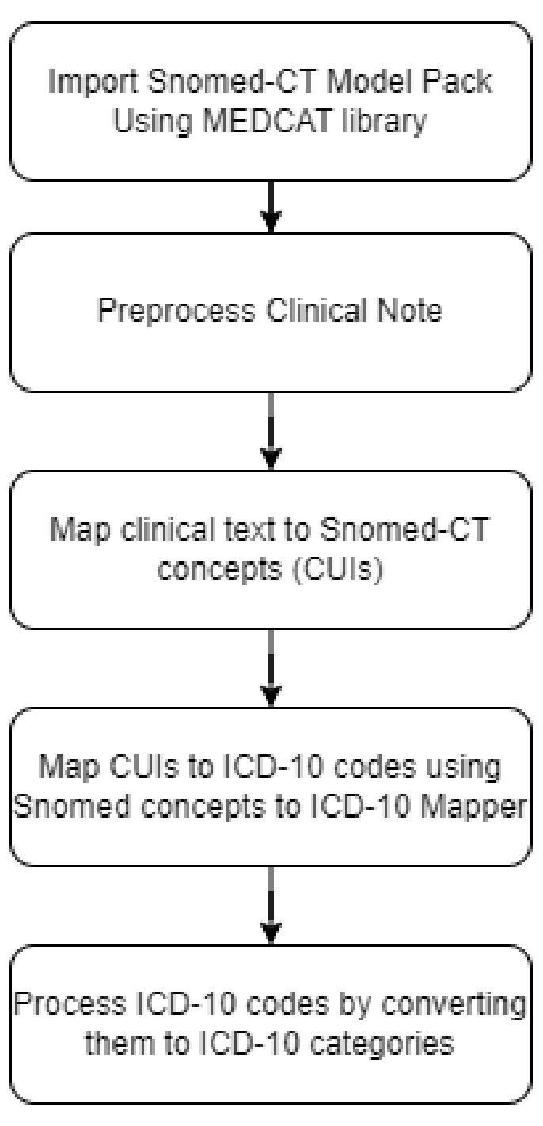
There were 4 methodologies tried in this experiment with a common but extensive preprocessing technique:

Preprocessing Flow



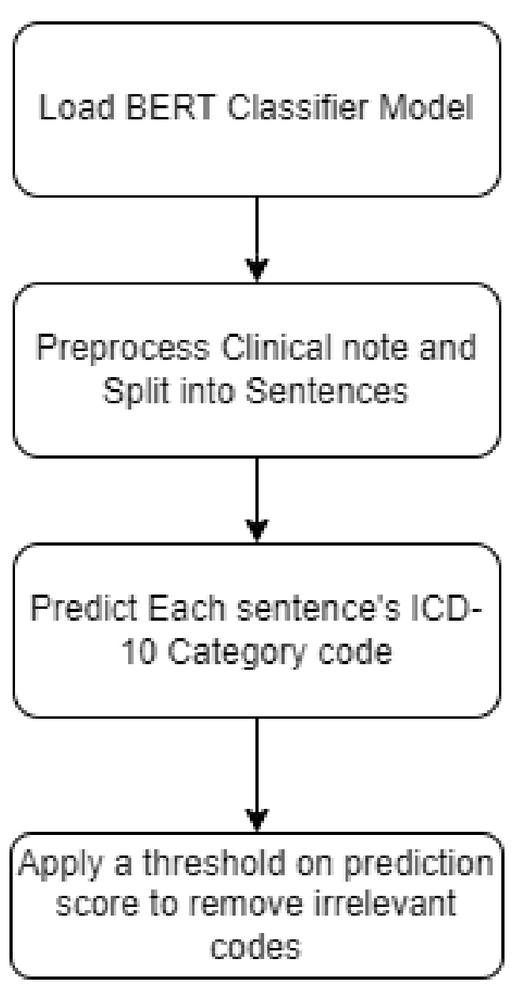
Removing Unnecessary Sections e.g Name, Family History etc.
lowercasing, removing special characters and white spaces
Removing section headers such as complaint history etc.
Removing negating sentences such as "No signs of Fever".
Removing Stop Words

SNOMED-CT (MEDCAT) Flow



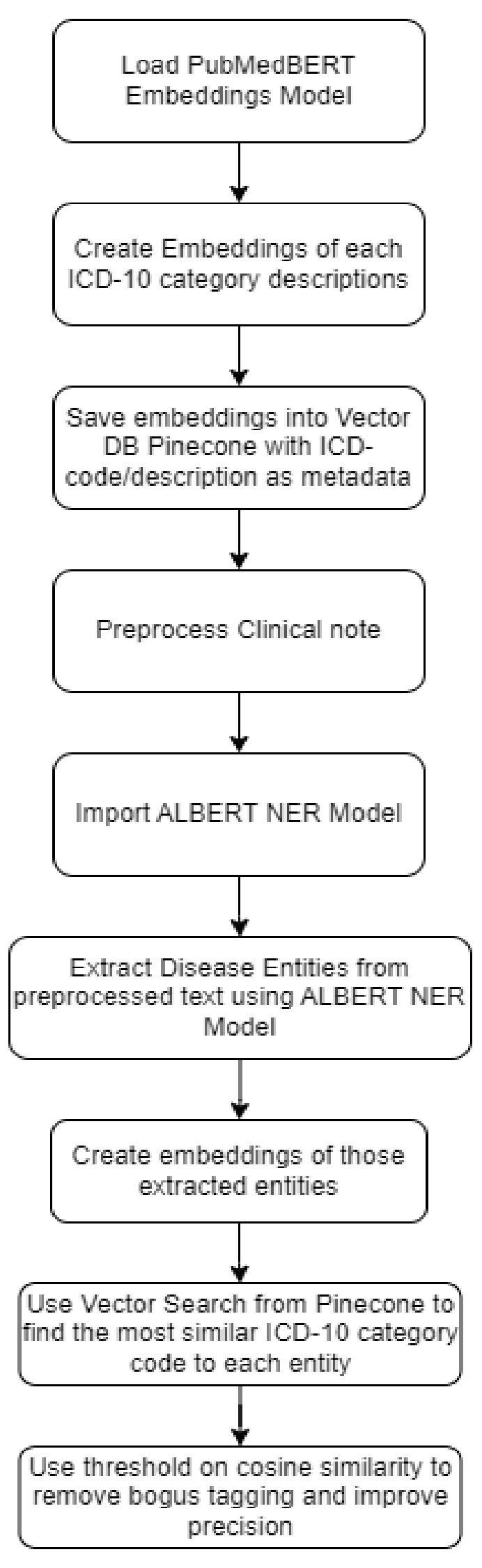
Import Snomed-CT Model Pack Using MEDCAT library
Preprocess Clinical Note
Map clinical text to Snomed-CT concepts (CUIs)
Map CUIs to ICD-10 codes using Snomed concepts to ICD-10 Mapper
Process ICD-10 codes by converting them to ICD-10 categories

BERT Classifier



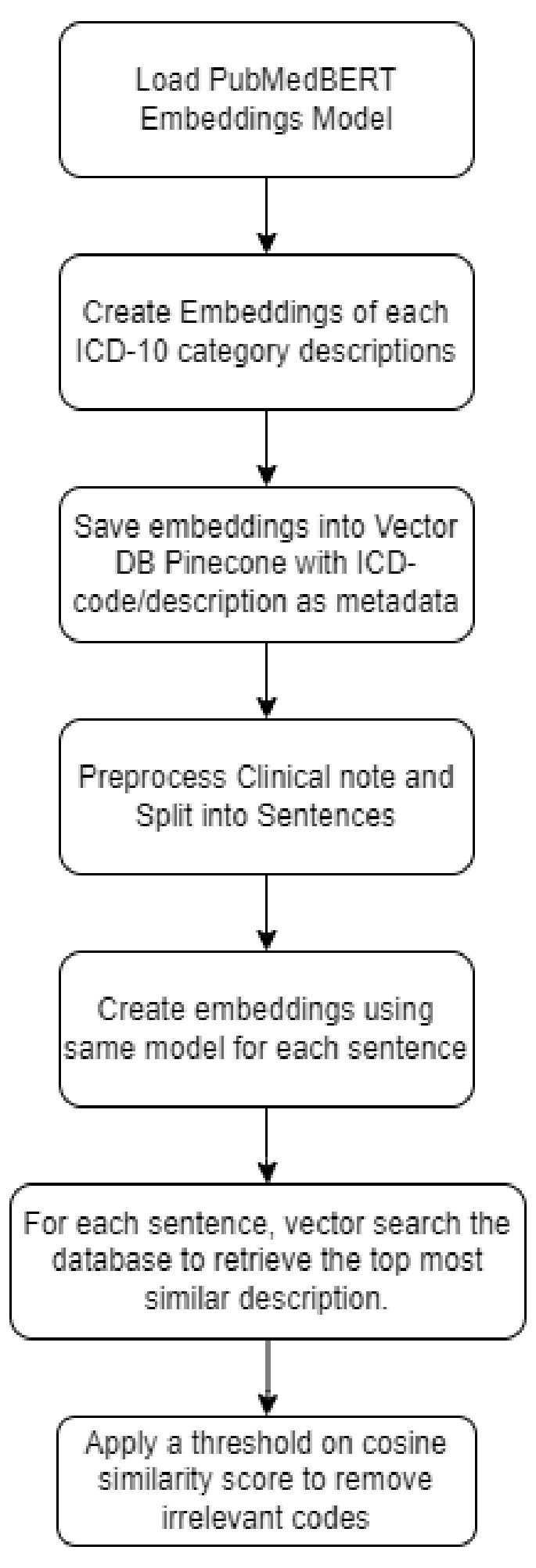
Load BERT Classifier Model
Preprocess Clinical note and Split into Sentences
Predict Each sentence's ICD-10 Category code
Apply a threshold on prediction score to remove irrelevant codes

ALBERT with PubMedBERT



Load PubMedBERT Embeddings Model
Create Embeddings of each ICD-10 category descriptions
Save embeddings into Vector DB Pinecone with ICD-code/description as metadata
Preprocess Clinical note
Import ALBERT NER Model
Extract Disease Entities from preprocessed text using ALBERT NER Model
Create embeddings of those extracted entities
Use Vector Search from Pinecone to find the most similar ICD-10 category code to each entity
Use threshold on cosine similarity to remove bogus tagging and improve precision

PubMedBERT Embeddings Similarity



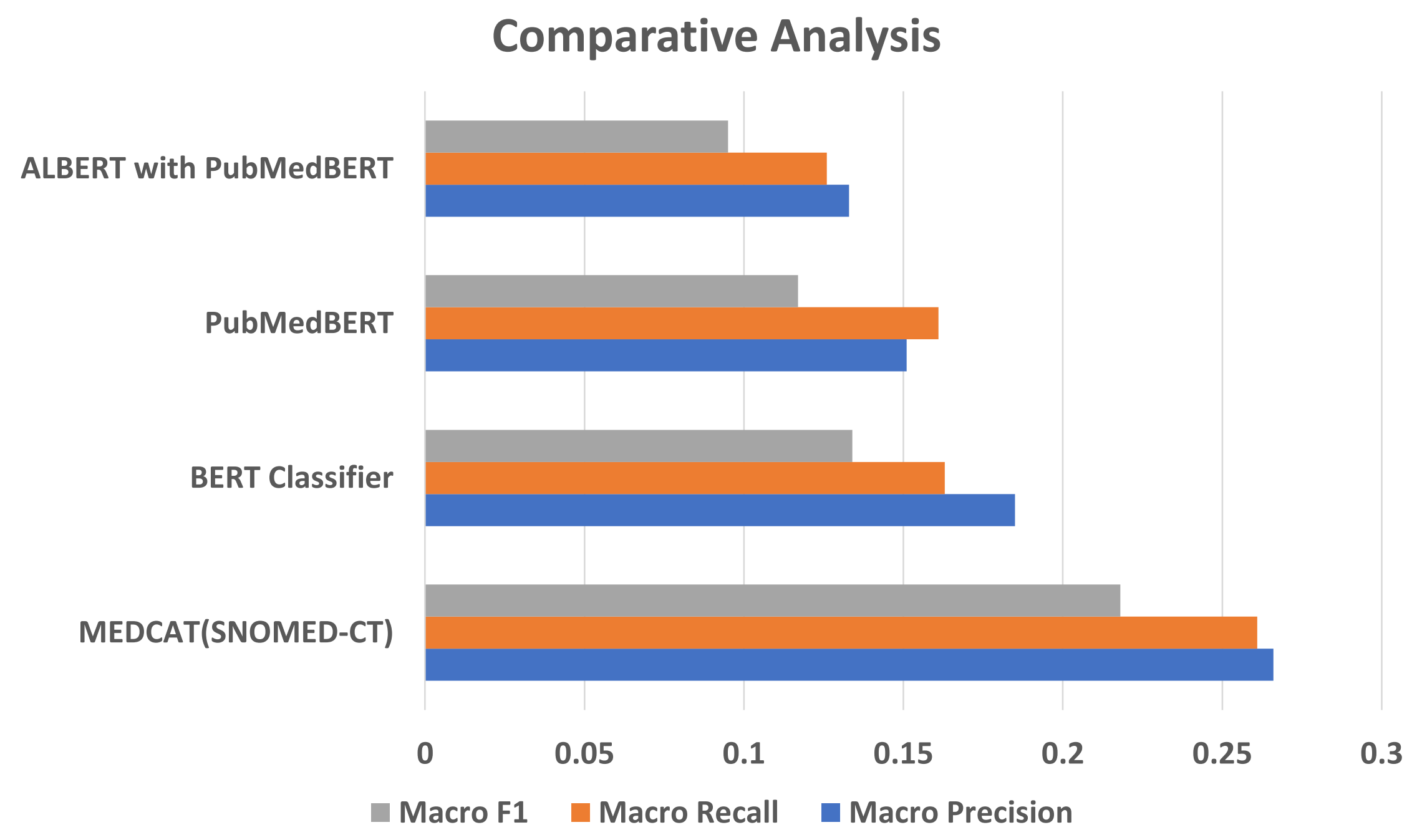
Load PubMedBERT Embeddings Model
Create Embeddings of each ICD-10 category descriptions
Save embeddings into Vector DB Pinecone with ICD-code/description as metadata
Preprocess Clinical note and Split into Sentences
Create embeddings using same model for each sentence
For each sentence, vector search the database to retrieve the top most similar description.
Apply a threshold on cosine similarity score to remove irrelevant codes

RESULTS

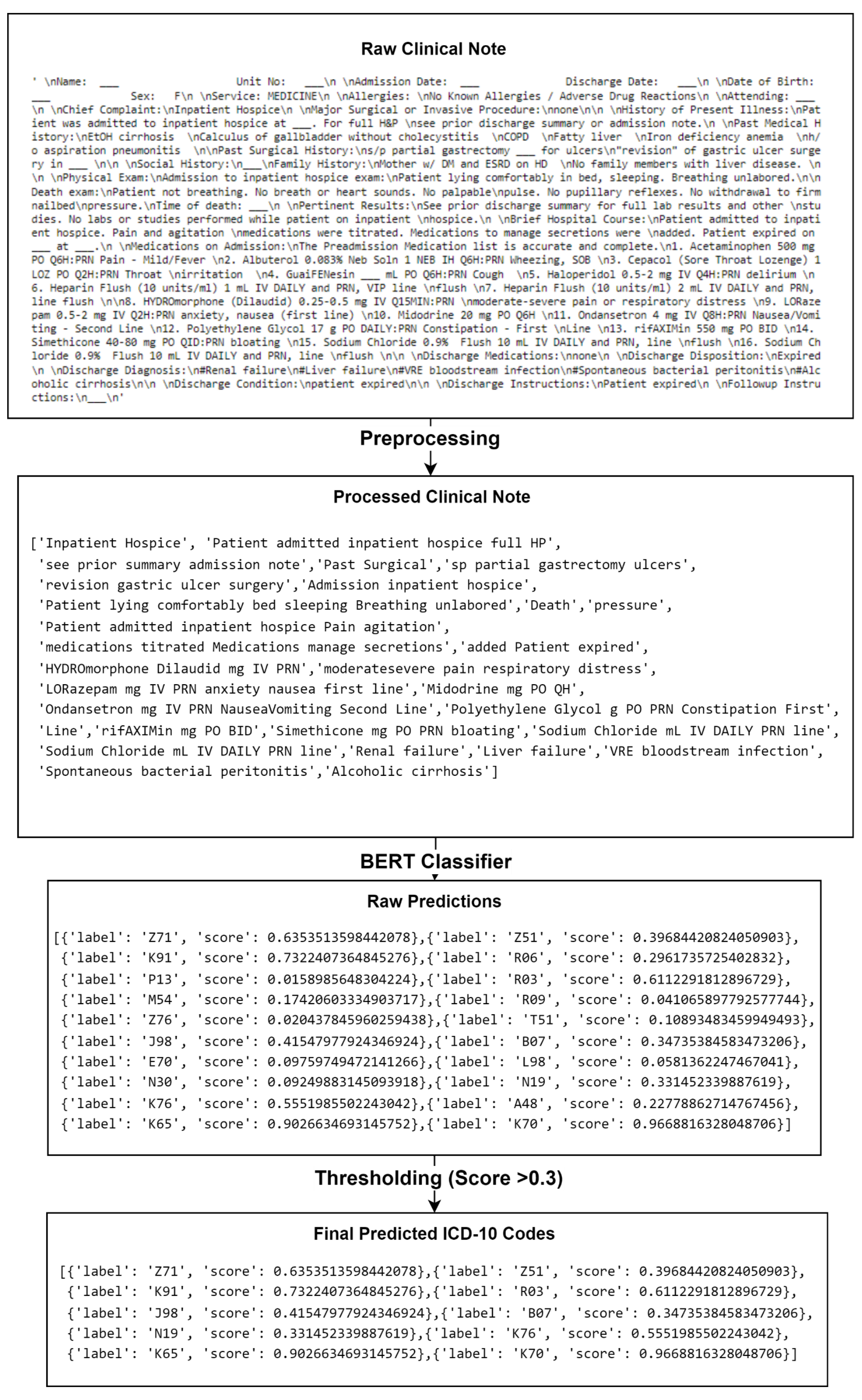
- The dataset chosen for evaluation was the latest version of benchmark clinical notes dataset MIMIC IV.
- 1000 shortest notes were used from the dataset as a testing set.
- the first three digits (category) of the ICD-10 code were predicted for this project. The evaluation metrics chosen inspired by relevant research papers were Macro F1, Precision and Recall.
- Total number of classes (ICD-10 Categories) in testing set were 769.

S.no	Method	Macro Precision	Macro Recall	Macro F1
1	**MEDCAT(SNOMED-CT)**	**0.266**	**0.261**	**0.218**
2	BERT Classifier	0.185	0.163	0.134
3	PubMedBERT	0.151	0.161	0.117
4	ALBERT with PubMedBERT	0.133	0.126	0.095

Table 1: Final Metrics



Walkthrough Example



Raw Clinical Note

```
' \nName:  ___                Unit No:   ___\n \nAdmission Date:  ___              Discharge Date:   ___\n \nDate of Birth:  ___            Sex:   F\n \nService: MEDICINE\n \nAllergies: \nNo Known Allergies / Adverse Drug Reactions\n \nAttending: ___\n \nChief Complaint:\nInpatient Hospice\n \nMajor Surgical or Invasive Procedure:\nnone\n\n \nHistory of Present Illness:\nPatient was admitted to inpatient hospice at ___. For full H&P \nsee prior discharge summary or admission note.\n \nPast Medical History:\nEtOH cirrhosis  \nCalculus of gallbladder without cholecystitis  \nCOPD  \nFatty liver  \nIron deficiency anemia  \nh/o aspiration pneumonitis  \n\nPast Surgical History:\ns/p partial gastrectomy ___ for ulcers\n"revision" of gastric ulcer surgery in ___ \n\n \nSocial History:\n___\nFamily History:\nMother w/ DM and ESRD on HD  \nNo family members with liver disease. \n \n \nPhysical Exam:\nAdmission to inpatient hospice exam:\nPatient lying comfortably in bed, sleeping. Breathing unlabored.\n\nDeath exam:\nPatient not breathing. No breath or heart sounds. No palpable\npulse. No pupillary reflexes. No withdrawal to firm nailbed\npressure.\nTime of death: ___\n \nPertinent Results:\nSee prior discharge summary for full lab results and other \nstudies. No labs or studies performed while patient on inpatient \nhospice.\n \nBrief Hospital Course:\nPatient admitted to inpatient hospice. Pain and agitation \nmedications were titrated. Medications to manage secretions were \nadded. Patient expired on ___ at ___.\n \nMedications on Admission:\nThe Preadmission Medication list is accurate and complete.\n1. Acetaminophen 500 mg PO Q6H:PRN Pain - Mild/Fever \n2. Albuterol 0.083% Neb Soln 1 NEB IH Q6H:PRN Wheezing, SOB \n3. Cepacol (Sore Throat Lozenge) 1 LOZ PO Q2H:PRN Throat \nirritation  \n4. GuaiFENesin ___ mL PO Q6H:PRN Cough  \n5. Haloperidol 0.5-2 mg IV Q4H:PRN delirium \n6. Heparin Flush (10 units/ml) 1 mL IV DAILY and PRN, VIP line \nflush \n7. Heparin Flush (10 units/ml) 2 mL IV DAILY and PRN, line flush \n\n8. HYDROmorphone (Dilaudid) 0.25-0.5 mg IV Q15MIN:PRN \nmoderate-severe pain or respiratory distress \n9. LORazepam 0.5-2 mg IV Q2H:PRN anxiety, nausea (first line) \n10. Midodrine 20 mg PO Q6H \n11. Ondansetron 4 mg IV Q8H:PRN Nausea/Vomiting - Second Line \n12. Polyethylene Glycol 17 g PO DAILY:PRN Constipation - First \nLine \n13. rifAXIMin 550 mg PO BID \n14. Simethicone 40-80 mg PO QID:PRN bloating \n15. Sodium Chloride 0.9%  Flush 10 mL IV DAILY and PRN, line \nflush \n16. Sodium Chloride 0.9%  Flush 10 mL IV DAILY and PRN, line \nflush \n\n \nDischarge Medications:\nnone\n \nDischarge Disposition:\nExpired\n \nDischarge Diagnosis:\n#Renal failure\n#Liver failure\n#VRE bloodstream infection\n#Spontaneous bacterial peritonitis\n#Alcoholic cirrhosis\n\n \nDischarge Condition:\npatient expired\n\n \nDischarge Instructions:\nPatient expired\n \nFollowup Instructions:\n___\n'
```

Preprocessing

Processed Clinical Note

```
['Inpatient Hospice', 'Patient admitted inpatient hospice full HP',
 'see prior summary admission note','Past Surgical','sp partial gastrectomy ulcers',
 'revision gastric ulcer surgery','Admission inpatient hospice',
 'Patient lying comfortably bed sleeping Breathing unlabored','Death','pressure',
 'Patient admitted inpatient hospice Pain agitation',
 'medications titrated Medications manage secretions','added Patient expired',
 'HYDROmorphone Dilaudid mg IV PRN','moderatesevere pain respiratory distress',
 'LORazepam mg IV PRN anxiety nausea first line','Midodrine mg PO QH',
 'Ondansetron mg IV PRN NauseaVomiting Second Line','Polyethylene Glycol g PO PRN Constipation First',
 'Line','rifAXIMin mg PO BID','Simethicone mg PO PRN bloating','Sodium Chloride mL IV DAILY PRN line',
 'Sodium Chloride mL IV DAILY PRN line','Renal failure','Liver failure','VRE bloodstream infection',
 'Spontaneous bacterial peritonitis','Alcoholic cirrhosis']
```

BERT Classifier

Raw Predictions

```
[{'label': 'Z71', 'score': 0.6353513598442078},{'label': 'Z51', 'score': 0.39684420824050903},
 {'label': 'K91', 'score': 0.7322407364845276},{'label': 'R06', 'score': 0.2961735725402832},
 {'label': 'P13', 'score': 0.0158985648304224},{'label': 'R03', 'score': 0.6112291812896729},
 {'label': 'M54', 'score': 0.17420603334903717},{'label': 'R09', 'score': 0.041065897792577744},
 {'label': 'Z76', 'score': 0.020437845960259438},{'label': 'T51', 'score': 0.10893483459949493},
 {'label': 'J98', 'score': 0.41547977924346924},{'label': 'B07', 'score': 0.34735384583473206},
 {'label': 'E70', 'score': 0.09759749472141266},{'label': 'L98', 'score': 0.0581362247467041},
 {'label': 'N30', 'score': 0.09249883145093918},{'label': 'N19', 'score': 0.331452339887619},
 {'label': 'K76', 'score': 0.5551985502243042},{'label': 'A48', 'score': 0.22778862714767456},
 {'label': 'K65', 'score': 0.9026634693145752},{'label': 'K70', 'score': 0.9668816328048706}]
```

Thresholding (Score >0.3)

Final Predicted ICD-10 Codes

```
[{'label': 'Z71', 'score': 0.6353513598442078},{'label': 'Z51', 'score': 0.39684420824050903},
 {'label': 'K91', 'score': 0.7322407364845276},{'label': 'R03', 'score': 0.6112291812896729},
 {'label': 'J98', 'score': 0.41547977924346924},{'label': 'B07', 'score': 0.34735384583473206},
 {'label': 'N19', 'score': 0.331452339887619},{'label': 'K76', 'score': 0.5551985502243042},
 {'label': 'K65', 'score': 0.9026634693145752},{'label': 'K70', 'score': 0.9668816328048706}]
```

Discussion and Final Remarks

Top Performer: SNOMED-CT from MEDCAT library excelled a Macro F1-score of 0.218, establishing a significant project milestone.

BERT Success: Emran's BERT model, coupled with advanced preprocessing, outperformed other transformer-based techniques with a comparable Macro-F1 score of 0.134.

Complexity Acknowledged: Considering the intricate multi-label classification problem involving around 24,000 classes with 769 classes in testing set, achieving such scores reflects commendable progress in addressing the inherent complexity of the task.

Comparable Results: Related works carried out for this problem had very high metrics compared to this project however it was seen that in those research papers the size of a clinical note used were very short which is around 20-50 words while the MIMIC IV dataset had notes ranging between 2000-10000 words. Furthermore, the reference papers mostly focused on a classification of a single ICD-10 code while our methods tried to perform a multi-label classification which came with its complexities.

Contributions to Healthcare: The methodologies tested present tangible solutions to challenges in clinical note mapping, making a noteworthy contribution to the healthcare information systems landscape.

References

Nguyen, Anthony N., Donna L. Truran, Madonna Kemp, Bevan Koopman, David Conlan, John O'Dwyer, Ming Zhang, Sarvnaz Karimi, Hamed Hassanzadeh, Michael Lawley and Damian J. Green. "Computer-Assisted Diagnostic Coding: Effectiveness of an NLP-based approach using SNOMED CT to ICD-10 mappings." AMIA ... Annual Symposium proceedings. AMIA Symposium 2018 (2018): 807-816.

Al-Bashabsheh, Emran, Ahmad Alaiad, Mahmoud Al-Ayyoub, Othman Beni-Yonis, Raed Abu Zitar, and Laith Abualigah. "Improving clinical documentation: automatic inference of ICD-10 codes from patient notes using BERT model." The Journal of Supercomputing (2023): 1-25.

Huang, Chao-Wei, Shang-Chi Tsai, and Yun-Nung Chen. "PLM-ICD: automatic ICD coding with pretrained language models." arXiv preprint arXiv:2207.05289 (2022).

"International Classification of Diseases." Encyclopædia Britannica. Accessed December 29, 2023. https://www.britannica.com/topic/International-Classification-of-Diseases.

Zeljko. "MedCAT: Introduction-Analyzing Electronic Health Records." Medium, April 23, 2020. https://medium.com/@w_is_h/medcat-introduction-analyzing-electronic-health-records-e1c420afa13a.

Acknowledgement

I extend my heartfelt gratitude to all those who contributed to the realization of this project.
My sincere thanks go to Dr Sajjad Haider for being a valuable mentor and supervisor who guided me throughout the project. Your insights have been instrumental in shaping the direction of our work.