ClintriFI

Aiding Recruitment, Empowering Research

Presented by: Surya Tejaswi Yerramsetty

ClintriFI ??

“Recruitment is so difficult and competitive,” "It’s not just finding the right patient; it’s finding the right patient ahead of the next trial or another trial trying to recruit the same population. Being able to do that very quickly and efficiently will be critical in driving the speed everyone is trying to achieve while ensuring the right patient population is included."

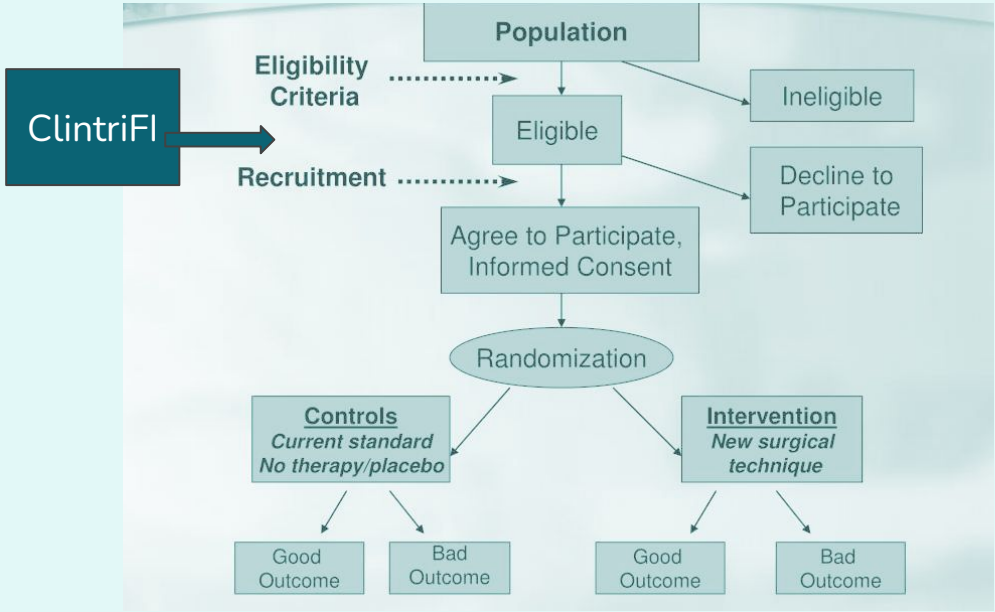
- **Silvia De Carvalho, Clinical Studies Lead, AXON (July 2023).**

- Clinical trials are at the heart of all medical advances as they help us better understand and improve human health and the healthcare system.
- ClintriFi serves as a valuable tool in the initial phase of clinical trials for screening by automating the classification and summarization of eligibility criteria from clinical notes (text). Through its interface, researchers can check for the patient’s eligibility, accompanied by justifying summaries that transparently explain patient eligibility.
- Aims to revolutionize patient recruitment which enhances trial efficiency, and accelerates the development of life-saving treatments and therapies by significantly reducing the time and resources required for trial recruitment.

ClintriFI:

ClintriFi efficiently selects eligible patients from the total population of applicants, screening the clinical notes. ClintriFi ensures that only individuals meeting the specific eligibility criteria** are considered for enrollment.

Once potential candidates are identified, clinicians can further stratify and randomize participants according to their protocol.
This ensures unbiased allocation to treatment groups, enhancing the reliability of trial results.



** limitation:- Criterias(Abdomen, Creatinine, Diabetes, Advanced CAD)

Picture 2

Technical Implementation of ClintriFI:

- Our Main Focus
- Data
- Classification
- Flow of Data in the System
- Summarization
- User Interface
- Future Scope
- Conclusion

Our Main Focus -
MultiLabel Text Classification and Summarization

- Our prototype model leverages the power of NLP and LLM models to perform Multi-Label Clinical Text Eligibility Classification and Summarization.
- Multi-label classification techniques are utilized to categorize patients into different eligibility criteria simultaneously.
- Different classifier models are trained and evaluated for multi-label classification.
- We have performed extractive and abstractive summarization methods to generate a summary to answer “why a patient is eligible for that clinical trial?

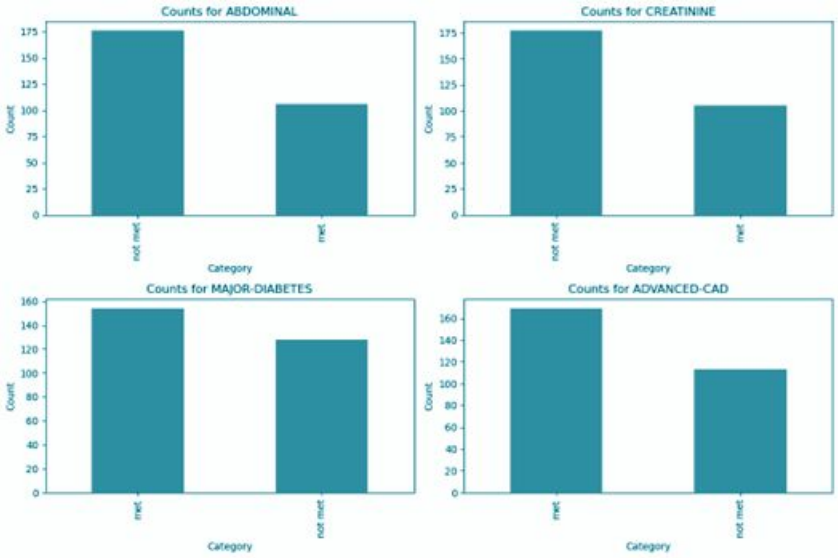
Picture 1

Data

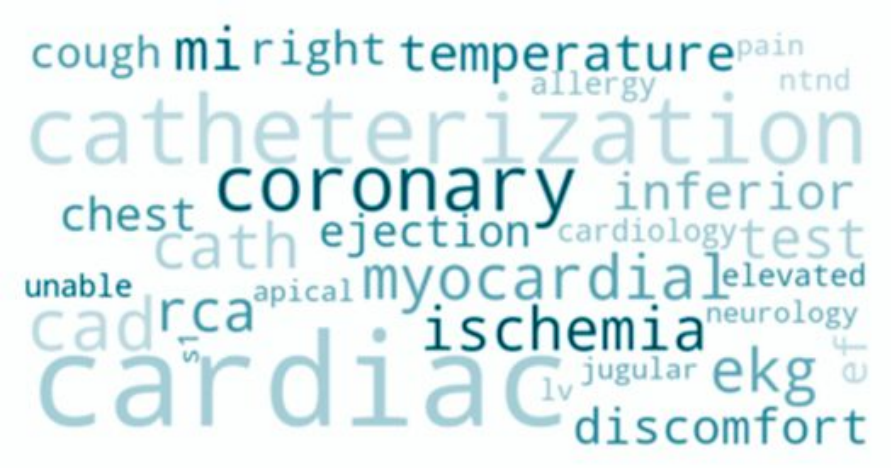
- In this prototype model, we have used an annotated dataset which aimed to answer the question, **“Can NLP systems use narrative medical records to identify which patients meet selection criteria for clinical trials?”**
- The authors of the dataset did the manual annotation procedure for all 288 records; each patient's record indicates whether the patient satisfies a set of selection criteria.
- This dataset comprises 288 patient records, each annotated at the patient level as “met” or “not met” for all the criteria. We have used 280 records throughout the model implementation for training and testing purposes.
- **Data Preprocessing:** We have used Regular expressions to remove unwanted characters, did stop word elimination, lemmatization and tokenization.

Picture 1

Data



Distribution of classes in each target variable



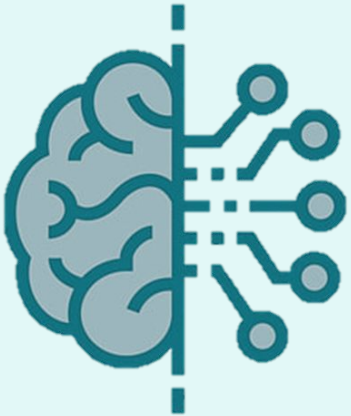
Random Forest Feature importance Word cloud for ADVANCED-CAD criteria

Multi Label Eligibility Classification

- What is Multi label classification?
  - Target variables: ABDOMINAL, ADVANCED-CAD, MAJOR-DIABETES, and CREATININE.
- Approaches used to classify multi-label:
  - Binary Relevance
  - Multi-Output Classifier
  - Classifier chain.
- The multi-label nature of data presented a challenge due to class imbalance. We have performed feature engineering to capture textual data underlying semantics and relationships.

Multi Label Eligibility Classification

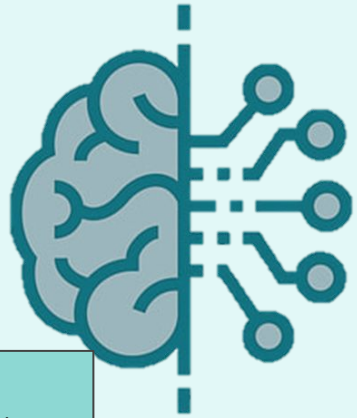
- Feature Engineering:
  - Named Entity Recognition - Clinical-AI-Apollo/Medical-NER. This has successfully identified 18 unique entities in the text corpus, such as Sign_Symptom, Disease_disorder, Diagnostic_procedure. We have calculated the presence/absence of these entities as features.
  - TF-IDF: It is a statistical measure used to score the importance of a word (term) in any content from a collection of documents based on each word's occurrences
  - Count Vectorizer: This efficient method will assign weights to words based on their frequency, ensuring a robust analysis of the text.
  - Word Embeddings(Word2Vec): We have employed the pre-trained Word2Vec model to capture the semantic relationships within words in the text and calculate TFIDF weighted embeddings.



Multi Label Eligibility Classification

Machine Learning Models:

Support Vector Classifier
Random Forest
Multi Layer Neural Network
Multi-Label Zero-Shot GPT Classifier:

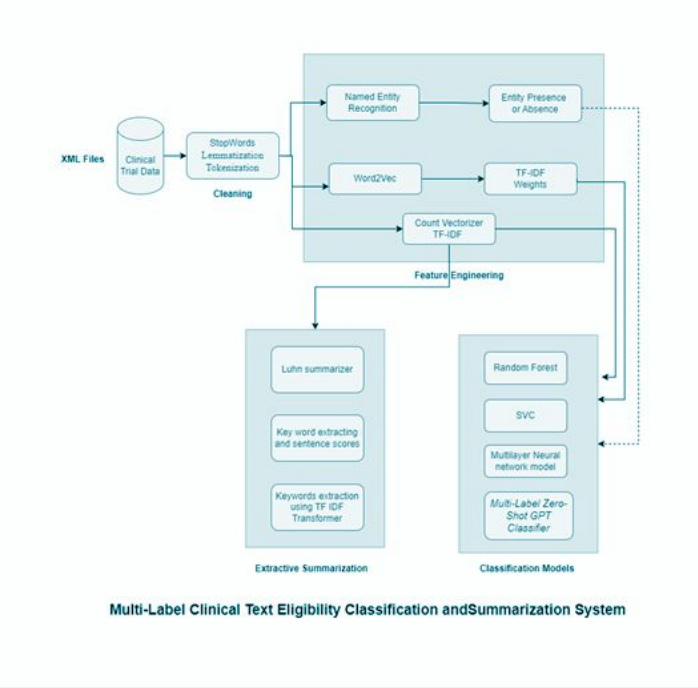


Features: TF IDF weighted embeddings +TF IDF (processed text)

Classifier chains: RF

0.75(Precision) 0.84(Recall)
0.83(F1 Score)

Flow of Data in the System:

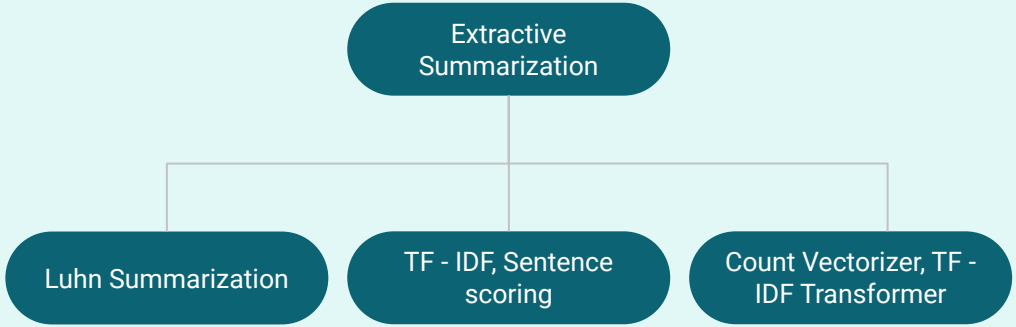


Multi-Label Clinical Text Eligibility Classification andSummarization System

Summarization

The goal of summarization is to reduce the size of the text without losing its key content.

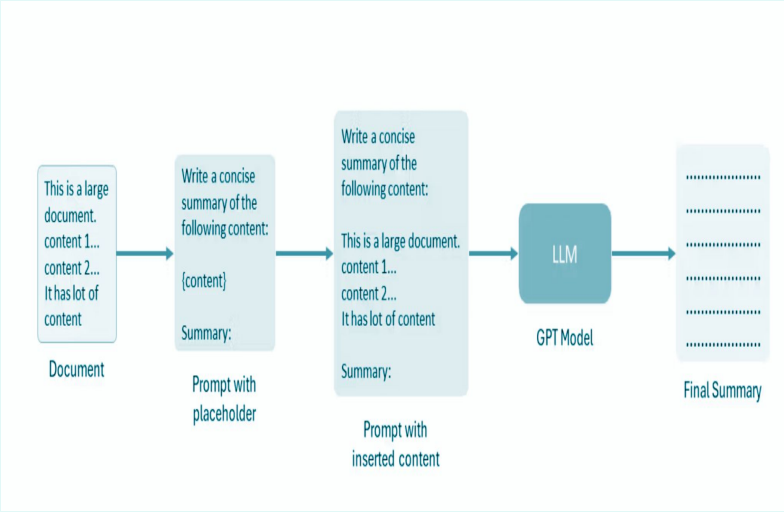
- **Extractive summarization**
  - Extractive summarizations are created by reusing portions of the input text document..
  - Techniques: TF - IDF, word embeddings, count vectorization, Luhn summarization, Text Rank.



Count Vectorizer & TF-IDF Transformer	Rouge-1	1.0 (P)	0.26(R)	0.42(F1)
	Rouge-2	0.96	0.20	0.33
	Rouge-L	1.0	0.26	0.42

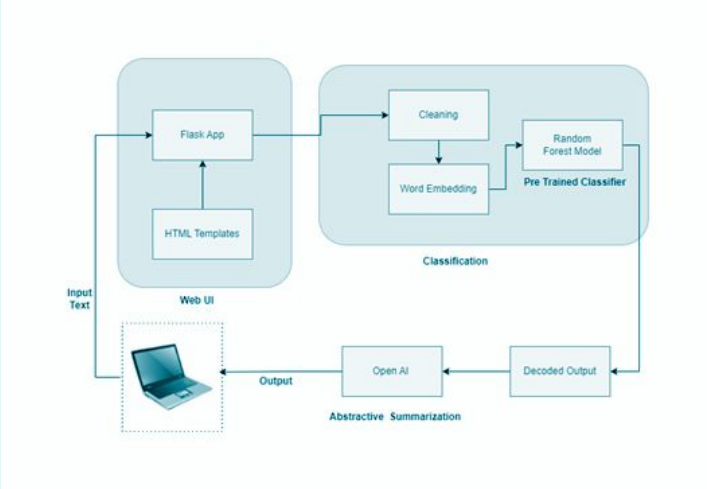
Summarization

- **Abstractive summarization**
  - Express the ideas of source document using different words.
  - Question Prompt: Guide summarization focusing on Chief Complaint, Patient Information, Refine Prompt: Improve summary based on classification process output. \
  - Refine Prompt:This will refine the above-generated summary based on the classifier output.
  - Language Model (GPT3), Framework: LangChain



User Interface

- we created our UI for the user to pass the input text, clinical text data.
- The clinical record is pre processed and is then routed to the backend functions for cleaning and feature extraction. The features are obtained by using techniques like, TF-IDF weighted embeddings and TF IDF vectorizer.
- The data is classified using the imported pre-trained ML model (.pkl file) i.e., our best performing model Classifier chains approach and Random Forest model.
- Summaries are generated using Lang Chain framework with GPT-3 to explain patient eligibility.
- The UI displays classification results and justification summaries for quick assessment of patient eligibility



User Interface

This XML file does not appear to have any style information associated with it. The document tree is shown below.

Settings and more

▼<PatientMatching>
▼<TEXT>
<![CDATA[Record date: 2097-01-27 RENAL FELLOW CONSULTATION NOTE PATIENT: WHITAKER, VINCENT UNIT#: 29964344 ADMIT DATE: 1/26/97 ATTENDING: Dr. Noor Uyeda Source: Notes Reason for consult: Acute on chronic renal failure HPI: 65 y/o man with PMH of stage IV CKD, HTN, CAD, RAS, DM, who was admitted to the CCU on 1/26 after an episode of bradycardia and PEA arrest at home. Pt was found by his wife yesterday to be complaining of severe chest pain and SOB. She called EMS. Upon arrival of EMS, pt was in respiratory distress and bradycardic with a HR in the 20s. He was intubated. He subsequently became puloseless and asystolic and was coded for about 10 minutes with 3 rounds of epinephrine and atropine before he regained his pulse. In the ED, his BP was 230/120 and he was started on a nitro gtt. He was noted to have blood in his OG tube. CXR showed bilateral patchy infiltrates consistent with pulmonary edema. Chest CT showed multiple rib fractures with substernal hematoma and bilateral modest pleural effusions with adjacent atelectasis vs consolidations. On admission, pt’s BUN and creatinine were 43 and 4.35, respectively. He was started on a lasix gtt for diuresis without a good response. His Cr started to rise: 4.6 à 4.9 à 5.4 à 5.33 à 6.3 à 7.0. His urine output continued to be poor, and renal consult was called. MEDICAL HISTORY: Stage IV CKD Right-sided RAS HTN CAD DM Anemia GERD Impotence MEDICATIONS: Versed gtt Propofol gtt Fentanyl gtt Heparin 5000 units Q8h Levofloxacin 250mg IV Q48h Flagyl 250mg IV Q8h ASA 325mg daily Pepcid 20mg IV daily Colace 100mg BID Tylenol PRN Maalox PRN Mucomyst 1200mg BID x 4 doses ALLERGIES: NKDA SOCIAL HISTORY: Tobacco – Denies EtOH – Denies Illicits – Denies Lives with wife and 4 children PHYSICAL EXAM: VITALS: T 99.7, pulse 58, BP 120/60, CVP 9 Vent: AC 14 x 600 / 40% / +5 U/O: 0-30cc/hr GEN: Intubated and sedated HEENT: Atraumatic NECK: Right IJ central line. HEART: RRR, S1S2 normal, no M/R/G LUNG: Coarse vented breath sounds ABD: Soft, NT, ND EXT: Trace edema SKIN: No rashes NEURO: Unable to assess LAB RESULTS: CBC: WBC 9.1, Hgb 7.9, plts 188k BCP: Na 136, K 5.2, Cl 106, HCO3 18, BUN 68, Cr 7.0, glucose 139 Ca 7.9, Mg 2.4 Coags: INR 1.2, PTT 28.6 ABG: 7.33 / 37 / 171 U/A: 1.010, pH 5.5, 2+ prot, 1+ blood, 14 hyaline casts, 6 RBC, 7 WBC IMPRESSION: 65 y/o man with stage IV CKD, DM, HTN, CAD, s/p bradycardia and PEA arrest, now intubated with oligoanuric acute on chronic renal failure. Etiology most likely 2/2 cardiac arrest with hypotension resulting in ischemic ATN. Given his hyperkalemia and acidemia in the context of anuria, we will start the pt on CVVH today. If his BP tolerates it, he may be able to transition over to intermittent HD soon. 1. Will place right femoral vein hemodialysis catheter. 2. Start citrate CVVH 120/2000 with DFL of 50cc/hr as BP tolerates 3. I will examine his urinary sediment. 4. Please check urine Na, urea nitrogen, total protein, creatinine. 5. Please order renal US. 6. Renally dose all medications for CVVH. _____ Travis X. Vogel, MD Longview Hospital Renal Fellow Pager 5-6394 ********** Record date: 2097-03-25 Patient Name: Whitaker, Vincent MRN: 29964344 Admit Date: 3/25/97 Today's Date: 3/25/97 ESRD Attending: Dr. Nancy Gipson Renal Dialysis Fellow Note HPI: 65 yo M h/o HTN, DM, RAS, now ESRD on HD, recent admission in LH CCU 1/26-2/28/97 after presenting with PEA cardiac arrest, course c/b respiratory failure requiring trach and PEG, initiated on HD during that admission, d/c'ed to Homestead Hospital 2/28/97, transferred from HH for fevers (T to 102 on 3/22) and urine and blood cultures positive for gram negative rods for further management. Was given on 3/24/97 linezolid and levaquin, gentamicin at HD. RIJ catheter pulled prior to transfer. + c/o diarrhea for several days. No recent CP, SOB, nausea, vomiting. + productive cough, +back pain. Previous to January admission, patient had CKD stage 4 with Cr in 4s, developed oliguric renal failure and initiated on HD with no evidence of renal recovery. MEDICAL HISTORY: ESRD on HD s/p PEA arrest 1/26/97 Respiratory failure s/p trach PEG Right-sided RAS HTN CAD DM Anemia HCV antibody + MEDICATIONS: Diltiazem 30mg qid Lopressor 25mg tid Insulin Linezolid 600mg bid (started 3/22/97) Levaquin 500mg qod (started 3/24/97) Phoslo 667mg ac meals Colace 100mg bid Heparin sq q8h Ranitidine 150mg qdaily Trazodone 25mg qdaily Coumadin Tylenol 650mg prn Nepro tube feeds ALLERGIES: Statins - rhabdo, Reserpine - anemia, Nifedipine - swelling SOCIAL HISTORY: no alcohol, smoking, drugs. Currently at Homestead FAMILY HISTORY: non-contributory REVIEW OF SYSTEMS: pertinent positives in HPI. Systems otherwise negative in detail. PHYSICAL EXAM: Vitals: T 98.4 HR 118 BP164/82 General: comfortable, awake, able to answer questions appropriately HEENT: neck supple, + trach mask CV: Irregular rhythm, S1, S2 Chest: CTAB Abd: +PEG, soft, non-tender Back: no CVA tenderness Ext: trace edema Access: LIJ tunneled HD catheter LABS: Date/Time NA K CL CO2 03/25/2097 137 4.2 98 27 03/24/2097 138 3.8 (#)[1] 99 27 Date/Time BUN CRE EGFR GLU 03/25/2097 44 (*) 5.31 (*) 11 [2] 114 03/24/2097 34 (*#) 4.10 (*#) 15 [3] 131 (*) Date/Time CA PHOS MG TBILI 03/25/2097 9.6 2.2 (*) 2.4 03/24/2097 9.4 SEE DETAIL[4] 2.1 (#)[5] 0.4 Date/Time TP ALB GLOB 03/24/2097 7.8 3.9 3.9 Date/Time ALT/SGPT AST/SGOT ALKP TBILI 03/24/2097 12 20 66 0.4 Date/Time WBC RBC HGB HCT 03/25/2097 9.08 3.59 (*) 10.7 (*) 32.7 (*) 03/24/2097 9.95 3.69 (*) 10.7 (*) 34.1 (*) Date/Time MCV MCH MCHC PLT 03/25/2097 91.1 29.9 32.8 357 03/24/2097 92.3 28.9 (#) 31.3 (*) 350 (#) Date/Time PT PT-INR PTT 03/25/2097 18.0 (*) 1.4 (*) 30.1 03/24/2097 17.4 (*#) 1.4 (*) 29.1 CXR 3/24/97: no acute cardiopulmonary process Assessment/Recommendations: 65 year old M CAD s/p PEA arrest 1/97, now ESRD on HD, DM, respiratory failure s/p trach/PEG, admitted with gram negative bacteremia. 1. Reason for admission: gram negative bacteremia, fevers -Follow up blood cultures -Call Homestead for speciation and sensitivities of blood cultures -Antibiotics for now: vanco, ceftaz, flagyl -Will check vancomycin level at end of HD tomorrow. -ID consult for assistance with appropriate antibiotic coverage -Check stool studies, urine culture 2. ESRD on HD -Schedule TThS -Access: LIJ tunneled catheter -Will draw blood cultures through tunneled catheter tomorrow at HD-Renal indices currently stable 3. Anemia: Epogen 2000 units with HD tomorrow. Will check iron studies 4. Mineral metabolism: continue phoslo. -Will check PTH level tomorrow at HD 5. Nutrition: continue nepro feeds Heidi Gunn, MD, Renal Fellow Renal Staff Addendum Pt seen and examined and history reviewed in detail. Case discussed at length with Dr. Gunn and I agree with her note from today (above). Will proceed with fever workup and will dialyze tomorrow. _____ Nancy N. Gipson, MD ********** Record date: 2097-08-11 EDVISIT^29964344^WHITAKER, VINCENT^08/11/97^TRAN, FREDDY I confirm that I have interviewed and examined the patient, reviewed the resident's documentation on the patient's chart, and discussed the evaluation, plan of care, and disposition with the patient and the resident, Dr. Nieto. Please see her note for full details. HISTORY OF PRESENT ILLNESS: The patient is a 65-year-old gentleman with significant past medical history of hypertension; diabetes; AFib; CHF; chronic renal insufficiency; end-stage renal disease on hemodialysis Monday, Wednesday and Friday; history of PE and cardiac arrest as well as bacteremia who presents with cough, weakness, chest pain, and palpitations. At dialysis he was found coughing and was sent here via ambulance. He notes decreased appetite. He denies any fevers. He has noted some chills, some sweats, and some shortness of breath. PHYSICAL EXAMINATION: He is slow. Temperature 95.9, pulse 89, respiratory rate 18, blood pressure 180/71, and saturation is 95% on 2 liters. He looks dyspneic, dehydrated with neck that is supple. No rales, rhonchi, or wheezing. Heart is regular without a murmur. Abdomen is soft, nontender, and nondistended. Skin is intact. He has 2+ pedal edema. He appears to be lethargic but no focal

Future Scope

- This project can be used as a prototype and further enhanced to develop a more robust and sophisticated system with both patient-centric and research-centric platforms.
- By implementing machine learning algorithms to enable CLintriFI to continuously learn and improve its accuracy over time. Analyzing past recruitment data can refine selection criteria and identify ideal candidates even more effectively.
- This can be developed in such a way that where a researcher can set his criteria , so that the patients can be selected accordingly. Imagine seamlessly integrating CLintriFI with systems having the information about trial criteria. This would allow CLintriFI to directly extract eligibility criteria from trial descriptions, ensuring tailored selection for each trial.
- Leverage advanced NLP techniques to extract rich information from clinical notes, including medical events, patient sentiment, and medical entities, leading to a more comprehensive understanding of patient suitability for clinical trials.

Conclusion

- Our project introduces a data-driven system leveraging advanced AI technologies to automate the classification of clinical trial eligibility criteria.
- The implementation of this tool has the potential to address biases in participant recruitment processes within pharmaceutical companies, promoting fairness and transparency in clinical trial enrollment.
- As CLintriFI uses sensitive patient data navigating regulatory considerations around data privacy and security will be crucial. Collaboration with regulatory bodies will be essential for ensuring compliance and widespread adoption.
- It is a valuable initial screening tool that saves time for a researcher reviewing lengthy and complicated clinical notes of hundreds of applications to recruit the patients for clinical trails.

References:

1. (2023, July). *Precision in practice: GenAI in clinical trials*. pharmaphorum. https://pharmaphorum.com/digital/precision-practice-genai-clinical-trials
2. Stubbs A, Filannino M, Soysal E, Henry S, Uzuner Ö. Cohort selection for clinical trials: n2c2 2018 shared task track 1. J Am Med Inform Assoc. 2019 Nov 1;26(11):1163-1171. doi: 10.1093/jamia/ocz163. PMID: 31562516; PMCID: PMC6798568.
3. Karystianis G, Florez-Vargas O. Application of a rule-based approach to identify patient eligibility for clinical trials. In: proceedings of the 2018 National NLP Clinical Challenges (n2c2) Workshop Shared Tasks; January 28, 2018; San Francisco, CA.
4. Xiong Y, Shi X, Chen S, Jiang D, Tang B, Wang X, Chen Q, Yan J. Cohort selection for clinical trials using hierarchical neural network. J Am Med Inform Assoc. 2019 Nov 1;26(11):1203-1208. doi: 10.1093/jamia/ocz099. PMID: 31305921; PMCID: PMC7647215.
5. Author, J. A. (2020, November 12). An introduction to multi-label text classification [Blog post]. Analytics Vidhya. [1] https://www.analyticsvidhya.com/blog/2017/08/introduction-to-multi-label-classification/
6. Vashisht, A. (2019, October 27). *Luhn's Heuristic Method for text summarization*. OpenGenus IQ: Computing Expertise & Legacy. https://iq.opengenus.org/luhns-heuristic-method-for-text-summarization.