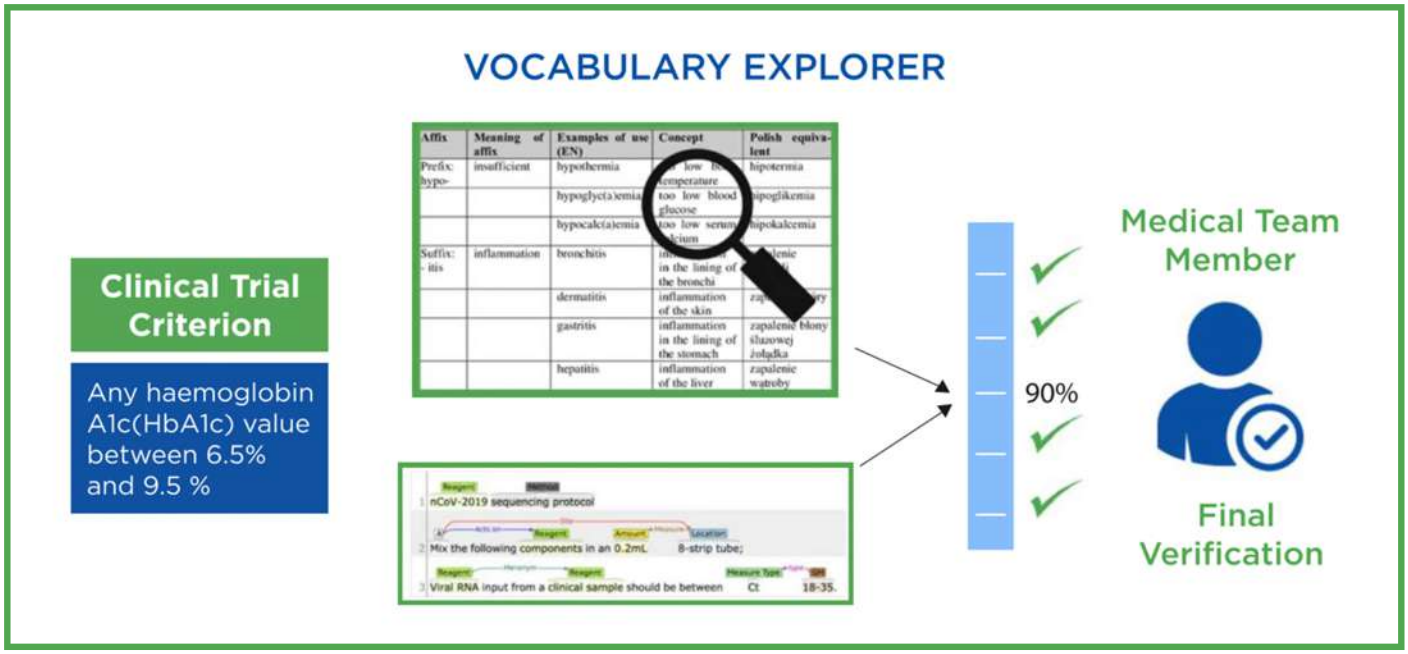
CLINICAL TRIALS COHORT SELECTION NLP MODEL FOR A PHARMACEUTICAL COMPANY

We have an uniquely efficient approach to the difficult task of cohort selection, one which has been thoroughly tested against actual client requests on Crohn's Disease, and Feeding tube detection in radiology reports.

We can generalize the solution to one of three methods.

1. Codification where possible.
2. Collection of concepts, synonyms.
3. Rare relation extraction (RE) by means of soft rules.

Patent matching involves Natural Language Processing (NLP) methods such as concept exploration and relation extraction on a per criterion basis:



BENEFITS OF OUR METHOD

The key technology needed to streamline cohort selection is undoubtedly NLP. The majority of patient data is in free text format, hidden in clinical notes. Social determinants are also becoming increasingly important in matching patients with trials. This type of data is not codified. While the conventional NLP models are well understood, there is a huge hurdle to apply them especially when it comes to domain specific terminology.

- Our solution adopts the most advanced ensemble learning principles which achieves the performance of supervised learning but bypasses the cost of manual labeling.
- The clinical trial organizer benefits from orders of magnitude of reduction in cost and completion time.
- Our innovation eliminates this laborious obstacle and accelerates the traditionally year long process down to a few weeks.

BIGRIO CLIENT PROOF OF CONCEPT

Example for Automating Patient Data Mining for Cohort Selection: ClinicalTrials.gov Identifier: NCT05471492

Illustration of the 3 methods

Exclusive Criteria
Short bowel syndrome

Method 1 - codification
2013 ICD-9-CM Diagnosis Code 579.3
Once the ICD 9 or 10 code has been identified, our algorithm can fire a SQL query against EHR records. This is the most straightforward portion of the 3 approaches.

Inclusive Criteria
Must have inadequate response to, loss of response to, or intolerance to at least one conventional therapy for CD

Method 2 - Concept exploration
Conventional therapy can encompass a long list of treatments ranging from immune system suppressor, biologics, antibiotics, nutritional therapy, to surgery. These need to be collected by an exploration exercise involving keyword search

Method 3 - Relation extraction
This method involves the use of a neural network model trained to recognize the concept of “intolerance”, “resistance”, “rejection”, etc.

POTENTIAL USE CASE FOR A PHARMACEUTICAL COMPANY

This Use Case is a representation of a pharmaceutical company trial where we can see repetition of the same descriptive patterns which would be addressed with our 3 methods mentioned above.

ClinicalTrials.gov Identifier: NCT04381650

Inclusive Criteria
CPI-naïve microsatellite stable-colorectal cancer (MSS-CRC) participants for whom prior standard first-line treatment has failed and who have progressed on no more than 3 chemotherapy regimens. (Approach 2 and 3)

Exclusive Criteria
Has evidence of active, non-infectious pneumonitis.
ICD 10 code: J84.113
(Approach 1)

DATA REQUIREMENTS FROM A PHARMACEUTICAL COMPANY

The application of our cohort selection requires input data which are found in a comprehensive EHR system.

EHR:

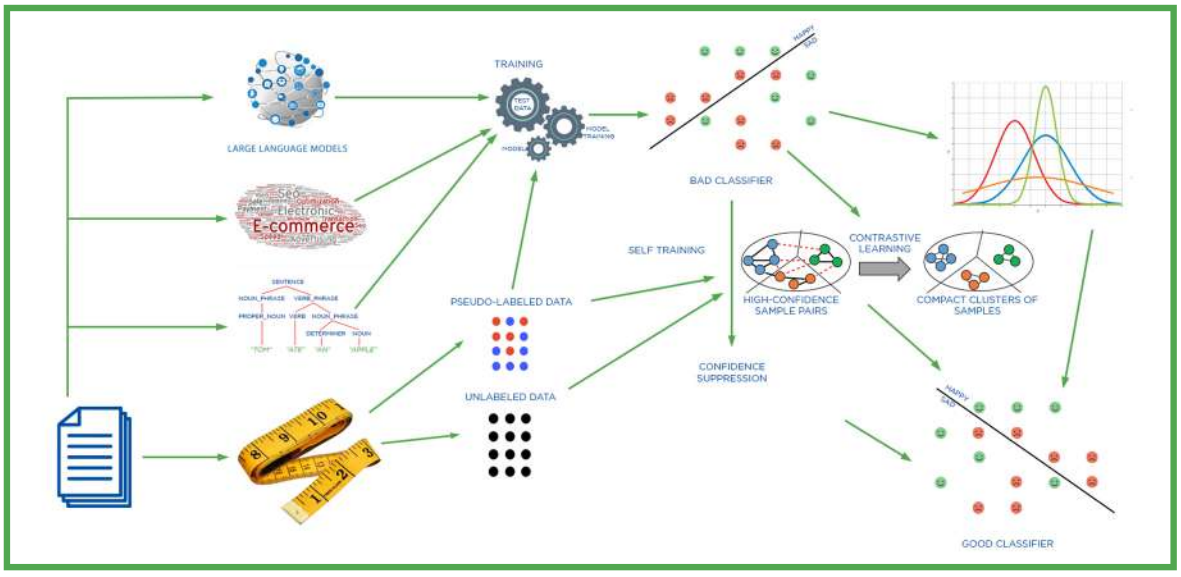
- Admission notes
- Discharge summary
- Nursing notes
- Input / output chart
- Radiology reports
- Lab reports
- ICD 9 / ICD 10 dictionary
- Google Health Knowledge Graph
- Unified Medical Language System (UMLS)

ANNOTATION FREE RELATION EXTRACTION

RE is a well understood neural network technique. However, the conventional RE process required labeled sample count in the thousands. The key challenge is not finding the right model architecture, but rather collecting sufficient labeled sentences which express the target relation. Our innovation leverages the use of “soft rules”, and eliminates the costly process of labeling thousands of sample sentences.

Our method involves the quick creation of loosely constructed rules which do not have to be 100% correct. The ensemble of these rules can be efficiently applied to millions of records at once, resulting in pseudo labels which are used to train a nominally accurate classifier.

We follow this step with a self-training step which incorporates both labeled and unlabeled data and iteratively improves the classification accuracy using techniques such as embedding optimization, contrastive learning, and confidence suppression. The final performance is equivalent to a fully supervised learning algorithm, but without the long and costly process of manual text annotation.



Our application of soft rules in place of manual annotation achieves order of magnitude improvement in efficiency and achieves the same accuracy as supervised learning.

PATIENT DIVERSITY & HEALTH EQUITY IN CLINICAL TRIALS

The conventional approach to identifying patients has been personal contacts to doctors who may refer patients due to this specialty. The chances of finding a match is such a precious occasion, it is often beyond practicality to adjust for population bias in ethnicity, gender, age, social status, and so on.

An effective automation offered by BigRio opens up the capability to efficiently search against a vast database of patient records, making it possible to discard some strictly clinical matches in favor of patient diversity and health equity considerations, thus achieving a balanced study design, which results in greater care for underserved communities.

BIGRIO'S CLINICAL TRIALS COHORT SOLUTION

We have a long-standing consulting practice in application of AI to healthcare. Our technical staff are well versed in patient data and clinical informatics. We achieve the utmost processing efficiency both in vocabulary extension and novel relation extraction, with the emphasis on turn around speed.

Our method starts with intuitively correct linguistic rules which can be quickly applied to a large corpus and feedback the positive candidate percent. Each time a new pattern is discovered, a new rule is generated to capture an unlimited number of additional matching sentences. The rest of the optimization process is all computerized in our machine learning pipeline. Our rigorously tested framework and method brings three orders of magnitude acceleration into clinical trial design efforts.

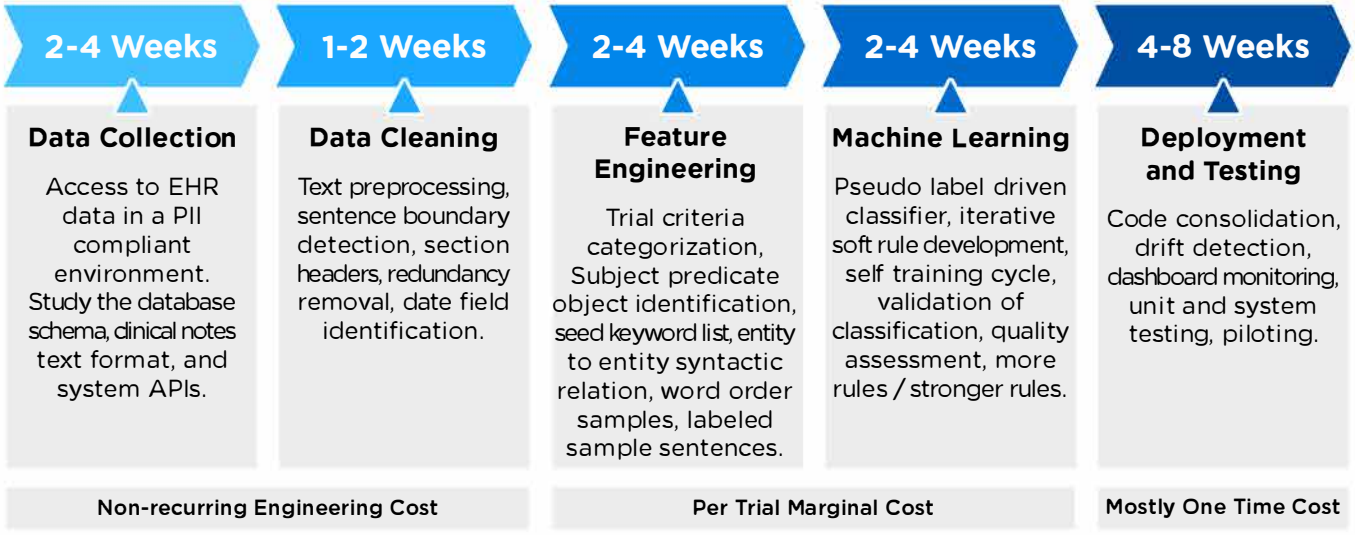
We look forward to delivering our state-of-the-art AI solutions to the most challenging area of patient care and clinical trials.

WHAT'S NEXT

The two key steps in delivering our solution to a pharmaceutical company are:

- Identify a clinical trial which needs help with patient recruitment.
- Access to EMR/EHR data with the pharmaceutical company and their relevant Hospital Partners.

TIMELINE AND PROCESS FLOW



In the overall project flow, there are one-time costs at the beginning and end of this process. Criteria that involve RE model training are the most time consuming. Whereas codification is the quickest when applicable. The per trial work is estimated to be four to eight weeks, depending on the complexity of the particular trial design criteria.

BigRio LLC, Harvard Square, One Mifflin Place, Suite 400, Cambridge, MA 02138
(617) 500-5093 | info@bigr.io | www.BigR.io