SSNSheerinKavitha at SemEval-2023 Task 7: Semantic rule based label prediction using TF-IDF and BM25 techniques

Sheerin Sitara Noor Mohamed and **Kavitha Srinivasan**
Department of Computer Science and Engineering
Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam - 603110, India.
sheerinsitaran@ssn.edu.in, kavithas@ssn.edu.in

Abstract

The advancement in the healthcare sector assures improved diagnosis and supports appropriate decision making in medical domain. The medical domain data can be either radiology images or clinical data. The clinical data plays a major role in the healthcare sector by preventing and treating the health problem based on the evidence learned from the trials. This paper is related to multi-evidence natural language inference for clinical trial data analysis and its solution for the given subtasks (SemEval 2023 Task 7 - NLI4CT). In subtask 1 of NLI4CT, the inference relationship (entailment or contradiction) between the Clinical Trial Reports (CTRs) statement pairs with respect to the Clinical Trial Data (CTD) statement are determined. In subtask 2 of NLI4CT, predicted label (inference relationship) are defined and justified using set of supporting facts extracted from the premises. The objective of this work is to derive the conclusion from premises (CTRs statement pairs) and extracting the supporting premises using proposed Semantic Rule based Clinical Data Analysis (SRCDA) approach. From the results, the proposed model attained an highest F1-score of 0.667 and 0.716 for subtasks 1 and 2 respectively. The novelty of this proposed approach includes, creation of External Knowledge Base (EKB) along with its suitable semantic rules based on the input statements.

1 Introduction

The Natural Language Processing (NLP) is one of the fastest growing sub-domain of Artificial Intelligence (AI) for medical domain (Basu et al., 2020). The contributions of NLP for medical domain are transcribing medical documents, clinical entity resolver and end-to-end drug discovery. To improvise the existing applications, research forums are conducting tasks for current requirements. One among them is SemEval 2023 and the task is multi-evidence Natural Language Inference (NLI) for clinical trial data (Jullien et al., 2023) and analysis. The motivation of this task is to help the clinical practitioner to stay updated on all current literature by providing an opportunity to support the large-scale interpretation and retrieval of medical evidence. To achieve this, NLI is applied on the clinical trial dataset (breast cancer Clinical Trial Reports (CTRs) and Clinical Trial Data (CTD)) to infer the medical evidence based on the relationship between premises and hypothesis. The clinical trial data analysis is divided into two subtasks namely, textual entailment and evidence retrieval.

In textual entailment (subtask 1), the hypothesis is derived from CTRs premises of four sections namely, intervention, eligibility, results and adverse events with respect to the CTD statement. Under each section there are three to eighty-two premises. In evidence retrieval (subtask 2), the anticipated premises which play a significant role in deriving the conclusion are retrieved.

The remaining part spans across the following subsections. The background for the textual entailment and evidence retrieval subtasks are discussed in Section 2. In Section 3, the system overview of the proposed work is explained. The experimental setup and implementation are explained in Section 4. A brief summary about the results, quantitative analysis and error analysis are given in Section 5 and conclusion is summarized at the end.

2 Background

The clinical trial dataset is common to both textual entailment and evidence retrieval subtasks of SemEval 2023 (Task 7). The dataset comprises Clinical Trial Data (CTD) (training set - 1700 samples, development set - 200 samples, test set - 500 samples) and Clinical Trial Reports (CTRs) (CT JSON - 1000 samples). The training and development set of CTD consists of type (single or comparison), section_id (intervention, eligibility, results or adverse events), primary_id (map to CTRs ID), secondary_id (map

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 950–957

to CTRs ID), statement, label (entailment or contradiction), primary_evidence_index and secondary_evidence_index. The test set consists of type, section_id, primary_id, secondary_id and statement. If type is single, then secondary_id and secondary_evidence_index are inappropriate. The dataset structure can be better understood through the following example.

Let the primary_id, secondary_id and section_id of CTD be "NCT00856492", "NCT00009945" and "Eligibility", then the respective CTRs having the CTRs ID as "NCT00856492" and "NCT00009945" are selected from CT JSON folder. Within the selected files, the statements under the eligibility section are considered for the proposed work.

From the dataset analysis, the following key-points are observed, (i). Every statement in the CTRs section need not be required to derive the conclusion. (ii). In case of comparison type, statements from both primary section and secondary section contribute equally in achieving the conclusion. (iii). For better result, (Thamer et al., 2020) stated that semantic level interpretation is required (for example, the clinical trial statement "Most participants in the secondary trial and the primary trial did not suffer from Enterocolitis" is equivalent to the statement "Enterocolitis haemorrhagic 1/167 (0.60%)" and "Enterocolitis haemorrhagic 0/167 (0.00%)" in the adverse event section (iv). External knowledge base (Wu et al., 2020) which maintains synonyms is required to develop better models. For example, the statement "Only patients with HER2 positive breast carcinoma are eligible for the primary trial" is equivalent to "HER-2 overexpressing breast cancer" in the eligibility section because "Carcinoma" is synonymous to "Cancer".

After the analysis of the given clinical trial dataset, the Semantic Rule based Clinical Data Analysis (SRCDA) approach is designed and explained in the following section.

3 System Overview

The proposed SRCDA system for textual entailment and evidence retrieval is divided into three modules namely, SRCDA (Common to subtasks 1 and 2), SRCDA – textual entailment and SRCDA – evidence retrieval. The SRCDA first module consists of five submodules and each of the second and third module consists of two task specific submodules which is shown in Figure 1 for better understanding. The SRCDA module consists of preprocessing, External Knowledge Base (EKB) creation, semantic rules formulation, feature vector generation by Term Frequency Inverse Document Frequency (TF-IDF) technique (Kim and M, 2019) and distance similarity metric (Cha, 2007) by Radial Basis Function Kernel (RBF-Kernel) modules. The two submodules specific to SRCDA – textual entailment are, Clinical Trial Reports (CTRs) section statements filtration and label prediction. The two submodules specific to the SRCDA – evidence retrieval are CTRs section statements ranking by Best Match 25 (BM25) techniques (Kamphuis et al., 2020) and, primary and secondary indices retrieval. These modules are discussed in detail in the following subsections and the process is explained in Algorithm 1.

The proposed SRCDA model uses EKB, semantic rules, TF-IDF techniques, RBF-Kernal distance metrics, BM25 technique because of the following reasons. They are: (i). The EKB collects the knowledge from multiple external sources for model creation; (ii). The semantic rules are defied to understand and analyse the statements by considering the context; (iii). The TF-IDF technique is used since it gives weights based on the uniqueness and relevance rather than frequency and hence more accurate representation among the overall polarity can be achieved; (iv). The RBF-Kernel is selected since it computes the similarity score between two statements based on their distance in the higher dimensional space; (v). The BM25 technique, supports Clinical Trial Report (CTR) statements length and term frequency saturation to rank the statements in the primary and/or secondary indices of CTR statements.

3.1 Semantic Rule based Clinical Data Analysis (SRCDA)

The proposed system design takes CTRs and Clinical Trial Data (CTD) as inputs, processes the data and generates the score value which will be used in subtasks 1 and 2.

3.1.1 Preprocessing

In preprocessing, the statements in the CTRs section and CTD are preprocessed as follows, (i). Removed stop words, punctuation, unnecessary space and preposition using Natural Language Toolkit (NLTK) (ii). Tokenized the statements as individual words (iii). Applied stemming and lemmatization methods on the tokenized input samples to get root words.

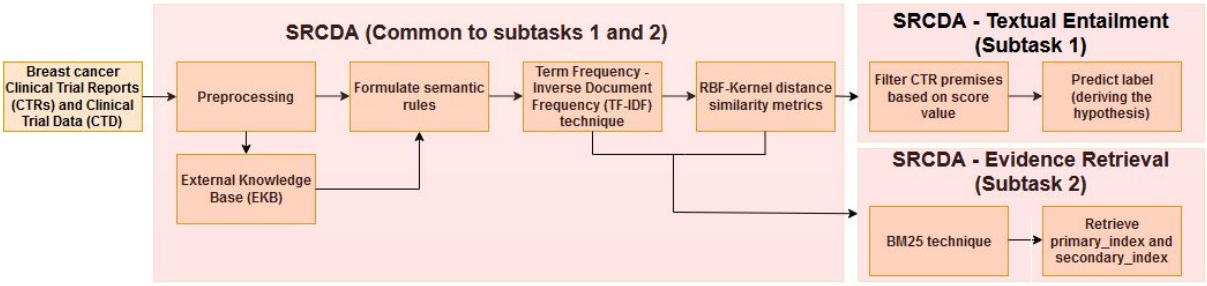

Figure 1: System design of the proposed work

Algorithm 1: Multi-evidence Natural Language Inference for Clinical Trial Data - SRCDA

Input: Breast Cancer Clinical Trial Data (CTD) (I_{in}) and Clinical Trial Reports (CTRs) (I_CTR_{in})

Output (Subtask 1): Determine the inference relation between CTR statement premises as entailment or contradiction (Predict label) (L_{out})

Output (Subtask 2): Retrieve the set of supporting facts (primary and secondary indices) from the CTR premises w.r.t predicted label in subtask 1 (PI_{out} and SI_{out})

function Comm_ST1_ST2 (I_{in}, I_CTR_{in})

- ➔ Procedure common to both textual entailment (subtask 1) and evidence retrieval (subtask 2)
- ➔ N and M be the number of samples in CTD and CTR respectively

```
for i -> 1 to N do
    if I_type_in[i] == "Single"
        for j -> 1 to M do
            if (I_PSID_in[i] == CTR_ID_in[j]) && (I_SID_in[i] == CTR_SID_in[j])
                PP_Sta[i], PP_CTR_Sta[j] -> preprocess (I_ Sta_in[i], CTR_SID_Sta_in[j]) (refer subsection 3.1.1)
                EKB_in -> Compute unique tokens from PP_Sta_in[i] and augment equivalent medical keywords
                          from external sources (refer subsection 3.1.2)
                SEM_Sta[i], SEM_Sel_CTR_Sta[j] -> SEM_Rules (PP_Sta[i], PP_CTR_Sta[j], EKB) (refer subsection 3.1.3)
                FV_Sta[i], FV_CTR_Sta[j] -> TF_IDF (SEM_Sta[i], SEM_Det_CTR_Sta[j], CTR_SID_Sta_in[j]) (refer Equation 1)
                Sco_PI [i, j] = RBF_Kernel (FV_Sta[i], FV_CTR_Sta[j]) (refer Equation 4)
            endif
            L_out -> Tex_Ent (Sco_PI [i, j]) (refer Function 1)
            PI_out -> Evi_Ret (Sco_PI [i, k], CTR_SID_Sta_in, j, k) (refer Function 2)
        endfor
    else if I_type_in[i] == "Comparison"
        Secondary premises also contribute to derive conclusion and hence Sco_SI [i, k] is computed which is similar to Sco_PI [i,
          k] calculation.
            L_out -> Tex_Ent (Sco_PI [i, k], Sco_SI [i, k]) (refer Function 1)
            PI_out, SI_out -> Evi_Ret (Sco_PI [i, k], CTR_SID_Sta_in, j, k, Sco_SI [i, k]) (refer Function 2)
    end for
return L_out, PI_out, SI_out
```

3.1.2 External Knowledge Base (EKB)

From the preprocessed unique tokens, the EKB is developed using external sources like PubMed, ImageCLEF (for medical terms) and two linguistic websites (for non-medical terms). This EKB consists of tokens (with respect to the vocabulary list) and its equivalent synonymous words are used for training the model. During testing phase, for the given CTD statement, tokens are masked with EKB synonymous words and hence equivalent CTR statements followed by conclusion is derived.

3.1.3 Semantic rules formulation

In this paper, four different semantic rules are framed with respect to the dataset. They are: (i). Negation equivalence rule; (ii). Double negation rule; (iii). Deductive reasoning rule; (iv). Condition based equivalence rule. The reason behind using these four rules are: (i). In negation equivalence rule, the complement of CTD statement is equivalent to the negation of CTR statements; (ii). In double negation rule, the double negation of the CTD statement is equivalent to the CTR statements; (iii). In deductive reasoning rule, to deduct the CTR statements, the CTD statement is mapped with the corresponding universal truth statement; (iv). In condition based equivalence rule, based on the condition in CTD statement, the respective CTR statements are retrieved to derive the conclusion.

Negation equivalence rule

Let x be universal quantifier, defined on various premises as $P_1, P_2, ...P_n$ (where n represents the number of premises) on which conjunction and disjunction is applied along with the negation con-

cludes Q.

The CTD statement $\forall x \neg(P_1(x) \wedge \neg(P_2(x)) \vee ...P_n(x)) \Rightarrow \neg Q(x)$ and CTR statement $\forall P_1(x) \wedge \neg(P_2(x)) \vee ...P_n(x) \Rightarrow Q(x)$ are equivalent if the value of x is identical in both case.

For example, the CTD statement "Pre and post menopausal women can enter the primary trial as long as they do not have prior hormone replacement theory" is equivalent to CTR statement "Pre or post menopausal women reporting use of hormone replacement therapy are not eligible for screening".

Let x be an axiom that represents women, $W(x)$ be menopausal women, $P(x)$ be pre-menopausal women, $O(x)$ be post-menopausal women, PHRT(x) be prior hormone replacement therapy and E(x, PT) be the eligible patient for primary trial.

$(1) \quad : \forall x W(x) \wedge (P(x) \vee O(x)) \wedge \neg(PHRT(x)) \Rightarrow E(x, pt)$

$(2) : \forall x W(x) \wedge (P(x) \vee O(x)) \wedge PHRT(x) \Rightarrow \neg E(x, pt)$

The negation of (1) gives (2), hence both statements are equivalent. As a result, this CTR statement is considered for deriving the conclusion.

Double negation rule

Let x be universal quantifier, defined on various premises as P, applying double negation concludes P and it is represented as, $\forall x \neg(\neg P(x)) \equiv P(x)$.

For example, the CTD statement "Patient who is not willing to sign and not willing to give informal consent are not considered for primary trial" is equivalent to the CTR statement "Patient who is willing to sign and give informal consent is considered for primary trial".

Let x be an axiom which represents patient, $W(x)$ be willing patient, $S(x, ic)$ patient willing to sign informal consent.

$\forall x P(x) \wedge \neg W(\neg S(x, ic)) \equiv \forall x P(x) \wedge W(S(x, ic))$

The double negation of CTD statement gives CTR statement and hence this CTR statement is selected for label prediction.

Deductive reasoning rule

Let x be universal quantifier, defined on various premises as P concludes Q when both premises are universal truth statement.

$\forall x P(x) \equiv Q(x) | P(x), Q(x) \epsilon M$ where M is the universal truth statement

For example, let the CTD statement be "Fiona's sister who is 34 years old was diagnosed with a ductual carcinoma and therefore Fiona may be eligible for primary trial" and CTR statement be "A first degree relative with breast cancer under the age of 60 are eligible for primary trial".

Let x be an axiom, $S(x, fiona)$ be Fiona's sister, $HA(34)$ be an Fiona's sister with age 34, $H(x, d)$ be Fiona's sister with disease 'd', $FDR(x, y)$ be an first degree relation, $E(fiona, pt)$ be Fiona is eligible for primary trial, $HA(age, < 60)$ be the FDR relation with age less than 60, $E(y, pt)$ be eligible for primary trial.

$(1) \quad : \forall x S(x, fiona) \wedge HA(34) \wedge H(x, ductualcarcinoma) \Rightarrow E(fiona, pt)$

$(2) : \forall x, y FDR(x, y) \wedge H(x, breastcancer) \wedge HA(age, < 60) \Rightarrow E(y, pt)$

The universal truth statements are

Statement 1: "Sister is an first degree relation"
Statement 2: "Ductual carcinoma is an early stage of breast cancer"

By the universal truth statements, it can been concluded that (2) follows (1) and hence both statements are equivalent. Hence, this CTR statement is also selected for label prediction.

Condition based equivalence rule

Let x be universal quantifier, defined on various premises as P and Q along with some condition which is equivalent to the relevance condition in the CTR statement will give the consequence.

$\forall x P(x) \wedge Q(r, condition)$

where r is a relation. From the deductive reasoning rule example, one of the condition in the CTD statement is $\forall x P(x) \wedge HA(34)$ and in CTR statement is $\forall x P(x) \wedge HA(age, < 60)$. Both these statements are equivalent because, the age 34 is less than age 60, based on the condition it is equivalent and hence these statements fall under condition based equivalence rule. Hence this statement is considered for label prediction otherwise remaining CTR statement is evaluated based on the semantic rules.

3.1.4 Feature vector generation by TF-IDF technique

In Term Frequency Inverse Document Frequency (TF-IDF) computation, the preprocessed CTD statement and preprocessed semantically equivalent CTRs section statements are vectorised. The vectorization finds the relevant section statements based on the given statement, ranks it in the order of relevance and selects the top k section statements. The steps involved in this process are, find the frequency of token in the section statements, perform vectorization based on all possible synonyms words in

the EKB, assign less weight for stop words, find the number of section statements in which the tokens occurred, calculate top k section statements based on the number of tokens in each statement as given in Equations 1 to 3.

$$tf_{idf}(s, c, C) = tf(s, c).idf(s, C) \quad (1)$$

where,

$$tf(s, c) = log(1 + freq(s, C)) \quad (2)$$

$$idf(s, C) = log(\frac{N}{count(c \epsilon C : s \epsilon c)}) \quad (3)$$

In the equation, s represents the processed CTD statement, c represents the processed semantically equivalent CTRs section statement and C is the corpus of all processed semantically equivalent CTRs section statements.

3.1.5 RBF-Kernel distance similarity metric

The RBF-Kernel distance similarity metric computes the score value of tf-idf vector matrices. It is similar to KNN algorithm but it has no space complexity problem because it stores only the support vector during training and not the entire dataset.

$$RBF - Kernel(v_s, v_c) = exp(-\frac{\|v_s, v_c\|^2}{2\sigma^2}) \quad (4)$$

Where v_s, v_c be the vectors computed for each CTD-CTRs statement pair. represents the variance of v_s and v_c. The equation 4 computes the similarity between v_s and v_c and $\|v_s, v_c\|^2$ is the Euclidean (L2 - norm) distance between v_s and v_c. Therefore, RBF-Kernel score is computed for each CTD-CTRs statement pairs.

If there are n CTD-CTRs statement pairs then there will be an n RBF-Kernel score for the particular statement.

3.2 Subtask 1

In subtask 1, the labels are predicted as given in Function 1 based on the critical score value computed from the distance similarity metrics for CTD-CTRs statement pairs,

3.2.1 Filter CTRs section statements

The CTRs section statements are filtered based on the critical score value and it is fixed as 0.80, since the contribution of significant words derives the conclusion. If the score value of the particular statement pairs is greater than the critical score, then the respective CTRs statement is considered.

3.2.2 Predict labels

The label (hypothesis) is derived from the filtered CTRs statements (premises) by applying conjunction and disjunction on the set of premises. The hypothesis may be “contradiction” or “entailment”.

3.3 Subtask 2

In subtask 2, the statements (premises) which contribute to the answer prediction are retrieved by BM25 techniques as shown in Function 2.

3.3.1 BM25 technique

Based on the label (contradiction or entailment), the respective premises which contributes to the label prediction are extracted for the CTRs statements. In BM25 technique, the set of CTRs statements from both primary and secondary indices are ranked based on the CTD statement and labeled as given in Equations 5, 6 and 7.

$$BM25(s_i) = \sum_{i}^{n} IDF(s_i) * S \quad (5)$$

where,

$$S = \frac{f(s_i, c) * (m1 + 1)}{f(s_i, c) * m1 * (1 - b + b * \frac{M}{N})} \quad (6)$$

$$IDF = ln(1 + \frac{CTRstatemnetcount - f(s_i + 0.5)}{f(s_i) + 0.5}) \quad (7)$$

Where s_i is the statement, C is the set of CTRs statements, $m1$ is the term frequency saturation characteristics, b is set as 0.75 (if b is bigger than the effects of the length of the document compared to the average length can be improved), M be each CTR statement length and N be an average of each CTR statement length.

3.3.2 Retrieve primary and secondary indices

Based on the rank value from BM25 technique and the score value from RBF-kernel similarity distance metrics, the primary and the secondary trial statements are extracted with the help of critical score. From the primary and secondary trial statements, the respective indices are retrieved as primary_index and secondary_index.

4 Experimental Setup

The system requirement for the proposed approach includes, (i). Hardware requirement – Intel i5 processor with NVIDIA graphics card, 4800M

Function 1: SRCDA - Textual Entailment

Input: Score values for each CTR statement pairs

Output: Predict the inference relation between CTR statement premises as entailment or contradiction (Predict label) (L_{out})

```
function Tex_Ent (Sco_PI, Sco_SI=0)
    ➔  If premises has no secondary ID then secondary score is assumed to be zero
     score = (Sco_PI + Sco_SI)/2
   if score > 0.72
        L_out = "Contradiction"
   else
        L_out = "Entailment"
return L_out
```

Function 2: SRCDA - Evidence Retrieval

Input: Score values for each CTR statement pairs

Output: Set of supporting facts (primary and secondary indices) from the CTR premises w.r.t predicted label in subtask 1 (PI_{out} and SI_{out})

```
function Evi_Ret (Sco_PI, CTR_SID_Sta_in, j, k,Sco_SI=0)
    ➔  If premises has no secondary ID then secondary score is assumed to be zero
     score = (Sco_PI + Sco_SI)/2
   if score < 0.72
        if Sco_PI < 0.72 && Sco_SI < 0.72
            PI_out = BM25 (CTR_SID_Sta_in[j])) (refer Equation 5)
            SI_out = BM25 (CTR_SID_Sta_in[k])) (refer Equation 5)
        elseif Sco_PI < 0.72
            PI_out = BM25 (CTR_SID_Sta_in[j])) (refer Equation 5)
            SI_out ={}
      elseif Sco_SI < 0.72
            PI_out = {}
            SI_out = BM25 (CTR_SID_Sta_in[k])) (refer Equation 5)
        else
            PI_out = {}
            SI_out ={}
        endif
return PI_out, SI_out
```

at 4.3GHZ clock speed, 16GB RAM, Graphical Processing Unit and 2TB disk space (ii). Software requirement - Ubuntu 20.04 operating system, Python 3.7 package with required libraries like tensorflow, torch, sklearn, json, numpy, nltk, pickle, pandas, rank_bm25, etc.,

The Clinical Trial Data (CTD) in clinical trial dataset is divided into training (training and development) set and test set. Both training and test set uses files in the CT JSON folder for utilizing statements in Clinical Trial Reports (CTRs) section. The proposed model is developed as given in the system design for two subtasks for the training set. In the testing phase, test set is given as input to the proposed model to predict label (subtask 1) and, retrieve primary and secondary section indices (subtask 2). The effectiveness of the proposed model are evaluated by comparing the predicted output with the actual answer using performance metrics. The different performance metrics used for analysing the results are F1-score, recall and precision. In this, the F1-score measures the model accuracy with respect to the dataset, recall measures the number of correctly identified samples by the model to that of labels, and precision identifies the frequency in which a model is correct while predicting the positive class.

5 Results

The performance of the proposed model for different runs are evaluated using suitable quantitative metrics and it is given in Table 1. Even though we have submitted 14 and 38 different combination of runs in two subtasks, some of them are discussed for quantitative and error analysis. In subtask 1, the proposed model is analysed for five different combinations of techniques, such as: (i). TF-IDF followed by Euclidean distance similarity, critical score value is fixed as 0.80 (ii). Same as (i), but the critical score value is reduced as 0.50 (iii). TF-IDF followed by cosine distance similarity (iv). Same as (i), but Radial Basis Function Kernel (RBF-Kernel) is used instead of Euclidean distance similarity (v). Preprocessing, TF-IDF followed by RBF-Kernel for critical score value of 0.72. From the results, it has been inferred that data preprocessing, role of External Knowledge Base (EKB), RBF-Kernel

Table 1: Brief description about sample runs

Task description	Run number	Method and other parameters	Performance metrics F1-Score	Rec	Pre
	1	TF-IDF + Euclidean distance similarity	0.492	0.468	0.520
	3	Run 1 with CS > 0.50	0.262	0.479	0.180
ST 1 - LB	4	TF-IDF + Pairwise distance similarity	0.183	0.500	0.112
	5	TF-IDF + Cosine distance similarity	0.509	0.502	0.516
	10	TF-IDF + RBF-Kernel distance similarity	0.640	0.500	0.888
	14	**Preprocessing + Run 10, CS > 0.72**	**0.667**	**0.500**	**1.000**
	2	BM25	0.350	0.469	0.279
	22	Run 2 with critical score > 0.70	0.287	0.637	0.185
ST 2 - LB	30	Preprocessing + BM25 for ranking	0.479	0.511	0.451
	38	**Run 30 + EKB**	**0.572**	**0.542**	**0.606**
ST 2 - PCLB	**1**	**Run 30 + SR**	**0.716**	**0.558**	**1.000**

*ST 1 - LB - Subtask 1 Leaderboard, ST 2 - LB - Subtask 2 Leaderboard, SR - Semantic Rules, CS - Critical Score, Rec - Recall, Pre - Precision

Table 2: Task 7 (Multi-evidence NLI for Clinical Trial Data) leader board of SemEval 2023

Task description	Rank	Users	Team Name	Performance metrics F1-Score	Rec	Pre
	1	Zhouyx21	THiFLY Research	0.856	0.856	0.856
	2	kamalkraj	Saama AI Research	0.834	0.768	0.912
ST 1 - LB	3	jvladika	TUM-sebis	0.798	0.777	0.820
	14	**SheerinSitara NoorMohamed**	**SSNSheerinKavitha**	**0.667**	**0.500**	**1.000**
	1	Jzy	THiFLY Research	0.853	0.811	0.898
	2	Zhaoxf4	HW-TSC	0.842	0.816	0.871
ST 2 - LB	3	Svassileva	FMI-SU	0.827	0.779	0.881
	19	**SheerinSitara**	**SSNSheerinKavitha**	**0.572**	**0.542**	**0.606**
		NoorMohamed		**0.716**	**0.558**	**1.000**

*ST 1 - LB - Subtask 1 Leaderboard, ST 2 - LB - Subtask 2 Leaderboard, Rec - Recall, Pre - Precision

selection and suitable critical score gives better performance than the other combinations.

For sub-task 2, the proposed model is analysed for five different combination of runs (four runs from prior submission and one runs from post-competition), such as: (i). BM25 technique to retrieve section statement index, critical score value is fixed as 0.80 (ii). Same as (i), but the critical score value is reduced to 0.70 (iii). Preprocessing followed by BM25 technique (iv). Same as (iii), but along with EKB (v). Same as (iii), followed by semantic rule based ranking and retrieval. From the results, it has been inferred that semantic rule based ranking and retrieval gives better performance than the other combinations. Moreover, the performance is still relatively close to the majority class baseline because most of the conclusion are "entailment" and hence the number of false negatives is less which leads the maximum recall value. But the probability of false positive is relatively high which leads to the precision value of 0.558, which is moderate value. The final result of the leaderboard is given in Table 2 where our team achieved 14th and 19th place in the listed ranks.

The quantitative analysis of the subtasks are, (i). Some statements in the section are indirectly deriving the conclusion and it is resolved by incorporating semantic based rules like negation equivalence, double negation, deductive reasoning and condition based equivalence rules (ii). Adopting TF-IDF

technique retrieves the suitable Clinical Trial Report (CTR) statements efficiently among the CTR section statements (usually each section has 4 to 32 statements) (iii). The RBF-Kernel distance similarity performs linear manipulation to map points to higher dimensional space and hence it is better than other distance similarity metrics. (iv). Most of the hypothesis are entailment ("true") and hence computing primary and secondary index is feasible.

The error analysis and the solution for the subtasks are, (i). The pairwise distance similarity (Run 4 of subtask 1) measures distance between two points or matrices and hence it degrades the overall performance. Incorporating the RBF-Kernel distance similarity from Run 10 improves the performance of the system (ii). In Run 3 of subtask 1, the non-optimal critical score reduces nearly 30% of the f1-score as compared to Run 1. This is addressed by fixing the suitable critical score in Run 14 (iii). The EKB is used from 10th run and hence suitable critical score value changes and performance also increased. (iv). Run 38 and Run 1 (post competition) of subtask 2 shows that EKB and semantic based rules improved the overall performance of the system.

6 Conclusion and Future Work

In this paper, the Semantic Rule based Clinical Data Analysis (SRCDA) approach is developed for multi-evidence Natural Language Inference (NLI) for clinical trial data task. In SRCDA approach, negation equivalence, double negation, deductive reasoning and condition based equivalence rules were used along with other textual entailment and evidence retrieval techniques. Hence, the F1-score is increased by 17.5% and 23.7% respectively for subtask 1 and subtask 2 respectively. But the overall performance of the system lacks by 18.9% and 14.0% as compared with the leading team in terms of F1-Score.

In the future work, the performance of the model can be improved by incorporating more semantic based rules like transitive and abductive rules, developing an improved External Knowledge Base (EKB) by including more synonyms words from different open sources, and choosing suitable critical score depending upon the contribution of premises towards deriving the conclusion are the ideas (plans) to be executed for improvement. . By incorporating the above discussed ideas, the researchers can develop an improved system which can be used in other domains like question answering system (chatbot), text summarization, dialogue to text conversion, etc.

References

Kanadpriya Basu, Ritwik Sinha, Aihui Ong, and Treena Basu. 2020. Artificial intelligence: How is it changing medical sciences and its future? *Indian Journal of Dermatology*, 65(5):365–370.

S. H Cha. 2007. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307.

Mael Jullien, Marco Valentino, Hannah Frost, Paul O'Regan, Donal Landers, and André Freitas. 2023. Semeval-2023 task 7: Multi-evidence natural language inference for clinical trial data. In *Proceedings of the 17th International Workshop on Semantic Evaluation*.

Chris Kamphuis, Arjen P. Vries, Leonid Boytsov, and Jimmy Lin. 2020. Which bm25 do you mean? a large-scale reproducibility study of scoring variants. volume 12036 of *Lecture Notes in Computer Science*. Springer.

S. W Kim and Gil J. M. 2019. Research paper classification systems based on tf-idf and lda schemes. *Human-Centric Computing and Information Sciences*, 9(30):1–24.

Thomas Lukasiewicz Oana-Maria Camburu, Tim Rocktschel and Phil Blunsom. 2018. e-snli: Natural language inference with natural language explanations. *Advances in Neural Information Processing Systems*, 31.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter. J. Liu. 2022. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(1):5485–5551.

Mohamed Thamer, Shaker El-Sappagh, and Tarek El-Shishtawy. 2020. A semantic approach for extracting medical association rules. *International Journal of Intelligent Engineering and Systems*, 13(3):280–292.

Wenqing Wu, Zhenfang zhu, Qiang Lu, Dianyuan Zhang, and Qiangqiang Guo. 2020. Introducing external knowledge to answer questions with implicit temporal constraints over knowledge base. *Future Internet*, 12(3):1–14.

https://www.nltk.org/