

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO PROGRAMA DE
PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL DOUTORADO

JEZER MACHADO DE OLIVEIRA

**UM MODELO MULTINÍVEL PARA ESTRUTURAÇÃO DE INFORMAÇÕES
CONTIDAS EM EVOLUÇÕES DE PRONTUÁRIOS ESCRITOS EM TEXTO LIVRE**

São Leopoldo
2024

Jezer Machado de Oliveira

**UM MODELO MULTINÍVEL PARA ESTRUTURAÇÃO DE INFORMAÇÕES
CONTIDAS EM EVOLUÇÕES DE PRONTUÁRIOS ESCRITOS EM TEXTO LIVRE**

Proposta de tese apresentada como requisito parcial para obtenção do título de Doutor pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos - UNISINOS.

Orientador:
Prof. Dr. Cristiano André da Costa

Coorientador:
Prof. Dr. Rodolfo Stoffel Antunes

São Leopoldo
2024

CIP- Catalogação na Publicação

O48u Oliveira, Jezer Machado de

Um modelo multinível para estruturação de informações contidas em evoluções de prontuários escritos em texto livre / Jezer Machado de Oliveira. – São Leopoldo : Unisinos, 2024.

123 f., 26 cm.

Tese (Doutorado) – Universidade do Vale dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2024.

Orientador: Prof. Dr. Cristiano André da Costa
Coorientador: Prof. Dr. Rodolfo Stoffel Antunes.

1. Informática Médica. 2. Processamento de Linguagem Natural. 3. Estruturação de informação I. Título.

Bibliotecária Responsável: Gislene Sapata Rodrigues CRB10/1997

“All we have to decide is what to do with the time that is given us.”
— GANDALF THE GREY

AGRADECIMENTOS

Agradeço primeiramente à minha família, principalmente à minha companheira, Sandra Flores Alves, pela cumplicidade e apoio durante todo o nosso relacionamento e que foram essenciais durante todo o meu doutorado.

Agradeço ao meu orientador e coorientador, Cristiano André da Costa e Rodolfo Stoffel Antunes.

Agradeço à Prof.^a Doutora Renata Vieira, Doutora Bruna Donida e ao Prof. Doutor Sandro José Rigo, por todas as contribuições apresentadas na banca de qualificação que foram vitais para a construção final desta tese.

Agradeço ao revisor anônimo Reviewer #2 do periódico Knowledge-Based Systems, que apesar de rejeitar meu artigo para publicação em seu periódico, trouxe uma série de críticas relevantes que contribuíram tanto para a futura publicação do artigo quanto para versão final do meu modelo.

O meu doutorado começou com uma pandemia global e terminou com a maior enchente registrada no meu estado, Rio Grande do Sul. Apesar de ter familiares e parentes que foram extremamente afetados pelos dois eventos, tive a sorte de ter sido pouco afetado por eles em comparação ao caos e destruição trazidos às vidas de muitos. Agradeço a essa sorte, por ter terminado esta tese apesar de tudo isso e principalmente por tê-la concluído antes de uma terceira guerra mundial, uma invasão alienígena ou outro evento inimaginável.

Por fim, agradeço ao Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) pelo apoio financeiro.

RESUMO

A área médica passou por uma série de transformações com a adoção de novas tecnologias. Um dos aspectos que sofreu mudanças significativas é a forma como as informações de um paciente são armazenadas. Os prontuários eletrônicos trouxeram uma série de vantagens, mas ainda apresentam alguns problemas. Um deles é a estruturação das informações contidas na evolução clínica. Essas evoluções podem ser armazenadas em texto livre, isto é, de forma não estruturada; de forma semi-estruturada, contendo uma lista de campos de texto livre para categorizar cada uma das informações; de forma estruturada, onde para cada informação existe uma série de campos específicos; ou uma combinação destas formas. Uma maior estruturação traz uma riqueza de informações e facilidade de consulta automatizada. Por outro lado, possui rigidez e maior complexidade no seu preenchimento. Esse problema se acentua ao realizar uma migração de um prontuário menos estruturado para um prontuário mais estruturado, visto a inviabilidade de uma migração direta. Levando em conta esses aspectos, este estudo surge de uma necessidade concreta, relacionada à migração de um software de prontuário eletrônico de uma empresa que utilizava registros de evolução do paciente não estruturados para um prontuário eletrônico com registros estruturados e semi-estruturados. Para que essa migração seja eficaz, os seguintes requisitos de software são imperativos: que todas as informações relevantes sejam mantidas e que sejam pelo menos semi-estruturadas e, quando possível, sejam completamente estruturadas. Por meio de uma revisão sistemática para encontrar o estado da arte na área, não foi encontrada nenhuma proposta que atenda de forma satisfatória esses requisitos. Considerando esse contexto, este trabalho propõe um modelo multinível para a estruturação das evoluções de prontuários escritos em texto livre na língua portuguesa. As principais premissas do modelo são que, no processo de estruturação, sejam mantidas todas as informações relevantes da evolução clínica, que a informação seja estruturada pelo menos no nível de sua sentença e que, quando possível, cada entidade da sentença também seja estruturada. O modelo é formado por um pipeline com dois componentes principais: o primeiro é responsável pela estruturação da informação no nível de sua sentença, ao dividir o texto e classificar individualmente cada sentença no padrão SOAP. No segundo nível, se possível, é feita uma estruturação no nível de suas entidades, identificando-as e relacionando-as. Para avaliar a viabilidade do modelo, foi implementado um protótipo do pipeline, utilizando técnicas de processamento de linguagem natural e aprendizado de máquina, como modelos BERT, associando seus subcomponentes a tarefas clássicas da PLN, como detecção de limite de sentença, classificação de sentenças, reconhecimento de entidades nomeadas, extração de relação e correspondência de ontologias. Para o treino e avaliação do pipeline e das redes, foi utilizada uma base de dados fornecida pela empresa que motivou o estudo. A base de dados contém 10.000 prontuários e 234.673 evoluções clínicas, das quais 15.883 foram divididas em 100.021 sentenças, classificadas e estruturadas através de uma Interface Gráfica de Usuário (GUI) desenvolvida para essa tarefa, formando o padrão ouro para o treino das redes. Após uma série de treinos e avaliações, as melhores redes foram selecionadas e o pipeline foi implementado. Para a avaliação final desse pipeline, foram utilizados 721 prontuários com um total de 10.013 sentenças, que também foram classificadas usando a GUI, formando o padrão ouro para a avaliação final. Os resultados obtidos foram comparados com os do pipeline, alcançando uma acurácia de 0,8641, precisão de 0,9493 e medida-F de 0,9029 para avaliação do primeiro nível de estruturação e uma acurácia de 0,8354, precisão de 0,9382 e medida-F de 0,8815 para o segundo nível de estruturação.

Palavras-chave: Informática Médica. Processamento de Linguagem Natural. Prontuário Eletrônico. Estruturação de informação.

ABSTRACT

The medical field has undergone a series of transformations with the adoption of new technologies. One aspect that has seen significant changes is how patient information is stored. Electronic health records have brought a series of advantages, but they still present some issues. One of these issues is of structuring of the information contained in clinical notes. These notes can be stored in free text, that is, in an unstructured form; in a semi-structured form, containing a list of free-text fields to categorize each piece of information; in a structured form, where each piece of information has a series of specific fields; or a combination of these forms. Greater structuring brings a wealth of information and ease of automated consultation. On the other hand, the medical staff must dedicate more attention when managing clinical notes due to the strict rules to maintain its structure. This problem is more evident when migrating from a less structured record to a more structured record, given the impracticality of direct migration. Considering these aspects, this study arises from a concrete need, related to the migration of an electronic health record software from a company that used unstructured clinical notes to an electronic health record with structured and semi-structured records. For this migration to be effective, the following software requirements are imperative: that all relevant information is maintained and that it is at least semi-structured and, when possible, fully structured. Through a systematic review to find the state of the art in the field, no proposal was found that satisfactorily meets these requirements. Considering this context, this work proposes a multi-level model for structuring progress notes written in free text in the Portuguese language. The main requirements of the model are that, in the structuring process, all relevant information from the clinical notes is maintained, that the information is structured at least at the sentence level, and that, when possible, each entity in the sentence is also structured. The model consists of a pipeline with two main components: the first is responsible for structuring the information at the sentence level, by dividing the text and individually classifying each sentence in the SOAP notes. At the second level, if possible, structuring is done at the level of its entities, identifying and relating them. To evaluate the viability of the model, a prototype of the pipeline was implemented, using natural language processing and machine learning techniques, such as BERT models, associating its subcomponents with classic NLP tasks such as sentence boundary detection, sentence classification, named entity recognition, relation extraction, and ontology matching. For training and evaluating the pipeline and networks, a database provided by the company that motivated the study was used. The database contains 10,000 records and 234,673 clinical notes, of which 15,883 were divided into 100,021 sentences, classified, and structured through a Graphical User Interface (GUI) developed for this task, forming the gold standard for network training. After a series of training and evaluations, the best networks were selected, and the pipeline was implemented. For the final evaluation of this pipeline, 721 records with a total of 10,013 sentences were used, which were also classified using the GUI, forming the gold standard for the final evaluation. The results obtained were compared with those of the pipeline, achieving an accuracy of 0.8641, precision of 0.9493, and F-score of 0.9029 for the first level of structuring, and an accuracy of 0.8354, precision of 0.9382, and F-score of 0.8815 for the second level of structuring.

Keywords: Medical Informatics. Natural Language Processing. Electronic Health Record. Data Structuring.

LISTA DE FIGURAS

Figura 1 – Estudo de mapeamento sistemático – seleção de artigos.	36
Figura 2 – Evolução cronológica dos estudos por ano de publicação.	37
Figura 3 – Distribuição gráfica dos estudos quanto à forma como as informações estão estruturadas nos registros de saúde.	37
Figura 4 – Distribuição gráfica dos estudos sobre a principal técnica utilizada no desenvolvimento.	42
Figura 5 – Distribuição gráfica de linguagens naturais da base de dados usados para desenvolvimento e avaliação	48
Figura 6 – Distribuição gráfica quanto à origem das bases de dados usados para desenvolvimento e avaliação	48
Figura 7 – Visão geral do modelo proposto com suas entradas e saídas.	53
Figura 8 – Modelo proposto para o componente Estruturador de Sentenças.	54
Figura 9 – Modelo para estruturação de entidade.	56
Figura 10 – Gráfico da medida-F das quatro classes e cinco modelos (escala começando em 0.6).	76
Figura 11 – Matriz de confusão do BioBERTptRT com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.	77
Figura 12 – Curva de Precisão-Recall da BioBERTptRT mostra os compromissos entre precisão e recall para todas as classes.	78
Figura 13 – Versão final da GUI desenvolvida para criação do padrão-ouro.	84
Figura 14 – Matriz de Confusão do componente de reconhecimento de entidades com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.	87
Figura 15 – Exemplo do layout para exportação do Componente de Relacionamento de Entidades.	89
Figura 16 – Matriz de Confusão do componente de relacionamento de entidades com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.	91
Figura 17 – Pseudo-código do componente de Agrupamento de Entidades.	95
Figura 18 – Matriz de confusão do componente Estruturador de Sentenças formado pelo divisor de sentença e o classificação sentença.	97
Figura 19 – Matriz de confusão do componente Estruturador de Entidades com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.	100
Figura 20 – Matriz de confusão final após a execução completa do Pipeline.	102

LISTA DE TABELAS

Tabela 1 – Grupos da CID-10.	28
Tabela 2 – Perguntas gerais e específicas de pesquisa.	34
Tabela 3 – Critérios de avaliação de qualidade.	35
Tabela 4 – Distribuição final dos estudos por ano de publicação.	36
Tabela 5 – Classificação dos estudos quanto à forma como as informações estão estruturadas nos prontuários.	38
Tabela 6 – Distribuição dos estudos sobre o idioma alvo do MER.	40
Tabela 7 – Classificação dos estudos MER quanto ao objetivo proposto.	40
Tabela 8 – Ferramenta NLP/NER usada por algoritmos baseados em regras.	44
Tabela 9 – Arquitetura de redes neurais usada nos estudos	45
Tabela 10 – Distribuição das bases privadas pela quantidade de documentos utilizados.	47
Tabela 11 – Lista de linguagens naturais da base de dados usados para desenvolvimento e avaliação	49
Tabela 12 – Bases de dados pública.	50
Tabela 13 – Grupos de entidades da classe Subjetivo.	57
Tabela 14 – Grupos de entidades da classe Objetivo.	58
Tabela 15 – Grupos de entidades da classe Avaliação.	59
Tabela 16 – Grupos de entidades da classe Plano.	59
Tabela 17 – Distribuição das evoluções dos pacientes entre as especialidades médicas da base de dados utilizada no trabalho.	66
Tabela 18 – Métricas de avaliação.	70
Tabela 19 – Exemplo de evolução do paciente em texto livre.	72
Tabela 20 – Sentenças com maior número de repetições.	72
Tabela 21 – Distribuição de sentenças rotuladas manualmente (padrão-ouro) contendo a classe, o número de sentenças e a proporção em relação ao total.	73
Tabela 22 – Distribuição da base de dados entre as 5 folds.	74
Tabela 23 – Resultados de desempenho para todos os modelos e classes para o componente de Estruturação de Sentença.	75
Tabela 24 – A medida-F foi comparada em pares de modelos sob o teste de soma de postos de Wilcoxon, com valores em negrito indicando significância estatística ($p < 0,05$).	76
Tabela 25 – Comparação da distribuição de sentenças da base de dados desse trabalho e da base de dados de (MOWERY et al., 2012).	79
Tabela 26 – Comparação do ganho da BioBERTpt em relação ao mBERT do modelo treinado e dos modelos de (SCHNEIDER et al., 2020) e (SOUZA et al., 2021).	79
Tabela 27 – Ranking dos modelos em relação a melhor Medida-F balanceada e suas especializações de domínio.	80
Tabela 28 – Distribuição da grupo de entidades após a classificação.	82
Tabela 29 – Resultados da avaliação do componente de reconhecimento de entidade.	85

Tabela 30 – Comparação de Desempenho na tarefa NER com outros trabalhos (valores em porcentagem).	88
Tabela 31 – Exemplo de sentença rotulada para o Componente de Relacionamento de Entidades.	90
Tabela 32 – Resultados de desempenho para o componente de Relacionamento de Entidade.	91
Tabela 33 – Comparação de Desempenho na tarefa de relacionamento de entidade com outros trabalhos (valores em porcentagem).	92
Tabela 34 – Tabelas de contexto para vinculação do componente Estruturar de Sentenças.	94
Tabela 35 – Padrão-ouro para avaliação do pipeline de estruturação.	96
Tabela 36 – Resultados de desempenho do componente de classificação sentença integrado ao divisor de sentenças.	97
Tabela 37 – Vinculação das tabelas de contexto realizadas pelo componente Estruturar de Sentenças	99
Tabela 38 – Resultados de desempenho do componente Estruturador de Entidades. . . .	99
Tabela 39 – Resultados de desempenho após a execução completa do Pipeline.	101
Tabela 40 – Resultados de desempenho após a execução completa do Pipeline, contendo somente as classes com os dois níveis de estruturação.	101

LISTA DE SIGLAS

ANS	Agência Nacional de Saúde Suplementar
Anvisa	Agência Nacional de Vigilância Sanitária
BERT	Bidirectional Encoder Representations from Transformers
Bi-GRU	Bidirectional-Gated Recurrent Unit
Bi-LSTM-CRF	Bidirectional-Long Short-Term Memory-Conditional Random Field
Bi-LSTM	Bidirectional-Long Short-Term Memory
BIO	Begin Inside Outside
BioBERT	Biomedical Bidirectional Encoder Representations from Transformers
CAAE	Certificado de Apresentação para Apreciação Ética
CID	Classificação Internacional de Doenças
CNN	Convolutional Neural Network
CRF	Conditional Random Fields
ECG	Eletrocardiograma
GUI	Graphical User Interface
HAS	Hipertensão Arterial Sistêmica
HDL	High Density Lipoprotein
IA	Inteligência Artificial
LSTM	Long Short-Term Memory
MRE	Medical Relation Extraction
NER	Named Entity Recognition
OMS	Organização Mundial da Saúde
PHR	Personal Health Record
PICO	Population, Intervention/Comparator and Outcomes
PLN	Processamento de Linguagem Natural
RES	Registro Eletrônico de Saúde
RNN	Recurrent neural network
ROC	Receiver Operator Characteristic Curve
SNOMED-CT	Systematized Nomenclature of Medicine-Clinical Terms
SOAP	Subjetivo, Objetivo, Avaliação e Plano
TISS	Troca de Informações na Saúde Suplementar
TUSS	Terminologia Unificada da Saúde Suplementar
UMLS	Unified Medical Language System

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Motivação	21
1.2	Questão de Pesquisa	22
1.3	Objetivo	23
1.4	Contribuições Científicas	23
1.5	Organização do Texto	24
2	REFERENCIAL TEÓRICO	25
2.1	Terminologia e Padrões de Saúde	25
2.1.1	Padrão SOAP	25
2.1.2	TISS e TUSS	26
2.1.3	Classificação Internacional de Doenças (CID)	27
2.2	Processamento de Linguagem Natural (PLN)	27
2.2.1	Classificação de sentenças	29
2.2.2	Reconhecimento de entidades nomeadas	29
2.3	Modelo de Rede Neural BERT	31
2.4	Conclusões Parciais	32
3	TRABALHOS RELACIONADOS	33
3.1	Metodologia de Busca da Literatura	33
3.1.1	Questões de pesquisa	33
3.1.2	Estratégia de Pesquisa	34
3.1.3	Seleção de Artigos	35
3.1.4	Realização da pesquisa	35
3.2	Resultados	36
3.2.1	Formas de estruturação e dados nos registros de saúde	37
3.2.2	Técnicas e ferramentas utilizadas para a estruturação das informações em prontuários de saúde	42
3.2.3	Dados para desenvolvimento e avaliação	46
3.3	Desafios e Oportunidades de Pesquisa	48
3.4	Conclusões Parciais	52
4	MODELO PROPOSTO	53
4.1	Visão Geral	53
4.2	Estruturador de Sentenças	54
4.3	Estruturador de Entidades	55
4.3.1	Reconhecimento de Entidade	59
4.3.2	Relacionamento de Entidade	61
4.3.3	Agrupador de Entidade	61
4.4	Conclusões Parciais	62
5	MATERIAIS E MÉTODOS	65
5.1	Base de Dados	65
5.2	Implementação do Estruturador de Sentenças	66
5.3	Implementação do Estruturador de Entidades	67
5.3.1	Reconhecimento de Entidades	68
5.3.2	Relacionamento de Entidades	68

5.3.3	Agrupamento de Entidades	68
5.4	Métricas de Avaliação	69
5.5	Conclusões Parciais	70
6	RESULTADOS E DISCUSSÃO	71
6.1	Treinamento e Avaliação do Componente Classificador de Sentença	71
6.1.1	Pré-processamento e rotulagem	71
6.1.2	Criação da BioBERTptRT	73
6.1.3	Treinamento e métricas de avaliação	73
6.1.4	Resultados	74
6.1.5	Discussão dos Resultados	81
6.2	Treinamento e Avaliação do Componente de Reconhecimento de Entidades	81
6.2.1	Pré-processamento e rotulagem	82
6.2.2	Treinamento e métricas de avaliação	82
6.2.3	Resultados	83
6.2.4	Discussão dos Resultados	89
6.3	Treinamento e Avaliação do Componente de Relacionamento de Entidades	89
6.3.1	Pré-processamento e rotulagem	89
6.3.2	Treinamento e métricas de avaliação	90
6.3.3	Resultados obtidos	90
6.3.4	Discussão dos Resultados	92
6.4	Implementação e Avaliação do Pipeline de Estruturação	92
6.4.1	Implementação do Estruturador de Sentença	93
6.4.2	Implementação do Estruturador de Entidade	93
6.4.3	Base de dados e métricas de avaliação	94
6.4.4	Resultados Obtidos	96
6.4.5	Discussão	100
7	CONCLUSÃO	103
7.1	Questões de pesquisa	103
7.2	Contribuições	104
7.3	Limitações	105
7.4	Trabalhos Futuros	106
	REFERÊNCIAS	109

1 INTRODUÇÃO

O prontuário eletrônico trouxe uma mudança revolucionária na acessibilidade e utilização das informações dos pacientes por parte dos profissionais de saúde. Essa transformação impactou o processo de tomada de decisões dos profissionais médicos, permitindo que tomem decisões informadas sobre o cuidado dos pacientes (CHUANG et al., 2024).

Um prontuário eletrônico é um sistema de informação usado para manipular e processar as informações de uma pessoa para fornecer serviços relacionados à saúde, oferecendo uma solução digital para armazenar, gerenciar e compartilhar informações dos pacientes. Os prontuários eletrônicos abrangem uma ampla gama de dados dos pacientes, incluindo histórico médico, medicamentos, alergias, resultados de exames laboratoriais e planos de tratamento, entre outros (ADENIYI et al., 2024). Ele é usado para enriquecer as interações médico-paciente e para fornecer serviços de saúde adequados, sendo um elemento importante na prestação de cuidados de saúde universal (CASTILLO et al., 2019).

NEVES et al.(NEVES et al., 2020) destaca como vantagens na utilização de um prontuário eletrônico a melhoria da qualidade do atendimento prestado ao paciente. Todas as informações referentes ao paciente ficam registradas em apenas um instrumento e não são perdidas, ampliando, desta forma, a possibilidade de avaliação da história pregressa e atual do paciente e, ainda, facilitando o diagnóstico precoce. Permite também o compartilhamento de informações entre todos os profissionais que compõem a equipe de saúde e o uso simultâneo por esses profissionais.

Uma das características dos prontuários eletrônicos é armazenar dados sobre a saúde do paciente. Embora existam muitos padrões para isso, não há consenso na adoção de um padrão específico; cada prontuário eletrônico tem sua própria forma de organização de dados. Um dos aspectos dessa organização é o grau de estruturação da informação. Em um registro de uma evolução clínica em um prontuário, por exemplo, a informação de que o paciente tem diabetes pode ser não estruturada, estando no meio do texto livre que representa aquela evolução do paciente. Também pode ser um campo semi-estruturado (ou estruturado no nível da sentença) de nome “histórico de doença”, contendo o texto livre “paciente tem sofrido de diabetes nos últimos 5 anos” ou um campo estruturado contendo uma lista de doenças anteriores, registrando a CID-10 e a data em que a doença começou. Atualmente, aproximadamente 80% desses dados médicos são registrados de forma não estruturada em texto livre (WILLIAMS; PETROVICH, 2023; SEZGIN et al., 2023).

NORGEOT et al.(NORGEOT et al., 2020) cita que a evolução do paciente geralmente contém as informações mais ricas e relevantes disponíveis sobre fenótipos, tratamentos e resultados da doença, bem como o processo de tomada de decisão clínica. Esta narrativa médica escrita frequentemente captura a experiência do paciente e eventos em uma linha de tempo. DE SILVA et al.(DE SILVA et al., 2021) conclui que, com a rápida adoção de sistemas de prontuários eletrônicos, as evoluções do paciente, além de serem uma fonte de dados onipresentes, também podem demonstrar maior qualidade e granularidade, tendo a capacidade de contribuir para o

processo de descoberta de conhecimento na área da saúde.

Dada essa importância, a forma como a evolução clínica é armazenada em um prontuário requer maior atenção. Ainda que esses textos livres não estruturados contenham informações ricas sobre o histórico de saúde do paciente, eles apresentam alguns desafios em seu uso em oposição aos dados estruturados e prontos para uso (KORMILITZIN et al., 2020).

Nesse sentido, muitos estudos têm se concentrado em estruturar essas informações. Eles analisam esses textos livres utilizando uma gama de ferramentas e técnicas de Processamento de Linguagem Natural (PLN) buscando e construindo informações estruturadas a partir deles. Esses estudos, em sua maioria, são focados em elementos específicos da evolução do paciente, como sinais vitais (KAIESKI et al., 2020), sintomas (KOLECK et al., 2019; SHEIKHALISHAHI et al., 2019) ou doenças (COLMENAREJO, 2020) e são desenvolvidos para textos livres escritos na língua inglesa (OLIVEIRA; ANTUNES; COSTA, 2024; OLIVEIRA; COSTA; ANTUNES, 2021; KERSLOOT et al., 2020; FU et al., 2020).

Levando em conta esses aspectos, este estudo surge de uma necessidade concreta, relacionada à migração de um sistema da empresa que forneceu os dados para o desenvolvimento e avaliação do modelo proposto. Esta empresa passou de um sistema de prontuário eletrônico que utilizava registros de evolução do paciente não estruturados para um sistema com registros estruturados e semi-estruturados. Neste novo sistema, algumas informações são organizadas de forma estruturada em nível de sentença, como as queixas principais e histórico familiar, enquanto outras são organizadas em um nível mais avançado, por exemplo, as prescrições, que agora incluem campos para o nome do medicamento, código e posologia, e os diagnósticos, que contêm campos para a descrição, o CID e a parte do corpo afetada. Para que essa migração seja eficaz, permitindo o arquivamento ou descarte do sistema de dados antigo e assegurando que o novo sistema de prontuário eletrônico conserve todas as informações essenciais, os seguintes requisitos de softwares são imperativos: que todas as informações relevantes sejam mantidas e que sejam estruturadas pelo menos no nível de sua sentença e que, quando possível, as informações de interesse como: prescrição e diagnóstico sejam estruturadas em um nível superior, preenchendo as informações do seu grupo de estrutura. A partir dessa necessidade de migração, esta tese visa generalizar essa demanda, sugerindo um modelo que facilite a estruturação de dados, enquanto preserva os requisitos mencionados.

Portanto, o objetivo principal desta tese é criar um modelo de estruturação de informações contidas em evoluções de prontuários escritos em texto livre na língua portuguesa, focado em uma estruturação completa da informação, dividida em dois níveis. No primeiro nível, é feita uma estruturação no nível da sentença, onde as evoluções são divididas em sentenças e cada uma delas é classificada em uma das categorias do padrão SOAP (Subjetiva, Objetivo, Avaliação e Plano). No segundo nível, são extraídas e estruturadas um grupo de estruturas de interesses como: sintomas, contendo descrição, duração e CID; exame físico, com informações como pressão arterial, frequência cardíaca, temperatura axilar; solicitação de exame; prescrição com nome e posologia.

1.1 Motivação

AFSHAR et al.(AFSHAR et al., 2023) avalia que a narrativa não estruturada dos prontuários eletrônicos fornecem uma rica fonte de informações sobre as condições dos pacientes, que podem servir como ferramentas de suporte à decisão clínica. Informações médicas detalhadas são rotineiramente registradas nessas notas. No entanto, essas informações não são organizadas nem priorizadas durante os cuidados de rotina. Além disso, o formato de texto livre das evoluções dificulta os esforços para realizar análises e aproveitar o grande domínio de dados.

ZHANG et al.(ZHANG et al., 2022) ressalta o papel crescente da evolução clínica texto livre nos Registros Eletrônicos de Saúde é tanto uma bênção quanto uma maldição. Permite que informações muito mais detalhadas sobre as condições dos pacientes sejam salvas e documentadas. No entanto, a natureza não estruturada desses dados pode também torná-los indisponíveis para os profissionais de saúde interessados em buscar informações específicas relacionadas aos pacientes.

Já SILVA(SILVA, 2007) analisa que, ao utilizar campos estruturados há uma tendência em se forçar a entrada de dados estruturados, ou seja, algumas informações peculiares e de importância podem deixar de ser registradas enquanto outros campos são deixados sem dados. Sugerindo, a adoção de texto livre pela semelhança aos hábitos de documentação por escrita à mão no prontuário em papel. Segundo essas premissas, alguns prontuários apresentam evoluções clínicas semi-estruturadas com campos de texto livre para preenchimento de informações categorizadas (por exemplo: reclamação principal, exame físico, hipótese diagnóstica, plano, observação, etc) e prontuários que combinam ambas as características mesclando campos semi-estruturados e estruturados.

LAMMERS; ZHENG(LAMMERS; ZHENG, 2011) destaca em seu estudo que, apesar dos desafios e da complexidade, mais de 30% dos hospitais analisados trocaram de software de prontuários eletrônicos em um período de 5 anos, pelos mais variados motivos. Esses motivos vão desde o abandono por parte do fornecedor atual até questões financeiras e competição de mercado. Somando-se a isso, a legislação em vigor exige que os prontuários sejam armazenados por 20 anos após o último atendimento registrado (STUMPF; FREITAS, 1997). Existe, portanto, uma grande chance de que um registro médico acabe passando por uma troca software de prontuário antes de seu descarte.

MIAKE-LYE et al.(MIAKE-LYE et al., 2023) sumariza em seu estudo que a migração pode ser realizada por meio da exportação e importação de arquivos digitais, entrada manual de dados, automação (*scripting*), mapeamento de texto livre para uma terminologia padronizada ou uma combinação destas técnicas.

HUANG et al.(HUANG et al., 2020) afirma que a conversão automatizada apresentando confiabilidade para dados estruturados. Em contra partida, a migração de dados não estruturados, desafia a compatibilidade dos sistemas, com a utilização de PLN representando uma solução potencial ainda pouco explorada. O método ideal para migrar esses dados varia, dependendo do

volume e finalidade, incluindo opções de conversão parcial ou completa, utilização de PLN, ou inserção manual do sistema legado para o novo.

No contexto específico deste estudo, que aborda a migração de uma base de dados composta por mais de 230 mil registros de evolução de pacientes e a transição de registros não estruturados para um formato que integra entradas estruturadas e semi-estruturadas. A migração automatizada revela-se impraticável, uma vez que exigiria uma correspondência direta de campos entre dois sistemas de prontuários eletrônicos com níveis equivalentes de estruturação. Paralelamente, a inserção manual desses dados se mostra excessivamente onerosa e lenta, dada a vasta quantidade de informações a serem processadas. Diante dessas limitações, a implementação de técnicas PLN emerge como a estratégia mais promissora, possibilitando a conversão eficiente das evoluções em texto livre em formatos estruturados e semi-estruturados.

Com base na revisão bibliográfica, foram identificadas muitas propostas para a estruturação automática da evolução dos prontuários eletrônicos, mas, em sua maioria, abrangem subgrupos da informação e não preservam toda a informação relevante. Mesmo os estudos que propõem uma estruturação mais ampla e completa são desenvolvidos em língua inglesa. Isso torna difícil a replicação, visto que as ferramentas e as bases de dados rotuladas utilizadas para treino e validação desses estudos estão em inglês. Além disso, não há garantia de que a aplicação de um modelo que funciona em uma língua terá os mesmos resultados em outra, devido às características de cada língua (OLIVEIRA; ANTUNES; COSTA, 2024; BAMPÁ; DALIANIS, 2020; CAI; ZHOU; ZHENG, 2020; CHENG et al., 2019; LI et al., 2020a; BALABAEVA; KOVALCHUK, 2020). Sendo que nenhum trabalho apresentou um modelo que pode ser aplicado na migração em questão.

Levando em conta essas questões, o que motiva esse trabalho é construir um modelo multi-nível para estruturação de informações contidas em evoluções de prontuários escritos em texto livre na língua portuguesa, onde toda a informação é estruturada em dois níveis. No primeiro nível, a evolução é dividida em sentenças, e cada sentença é classificada individualmente com base no seu contexto geral. No segundo nível, se possível, é feita uma estruturação no nível de suas entidades, identificando e relacionando-as.

1.2 Questão de Pesquisa

Questão de Pesquisa

Como modelar, desenvolver e avaliar um modelo computacional multinível para estruturação das informações contidas nas evoluções em texto livre do prontuário do paciente, sem perda de informações relevantes?

Além dessa pergunta, durante a realização do trabalho pretende-se responder às seguintes questões:

- Quais são as principais técnicas e ferramentas utilizadas para a estruturação da informação?

- Como aplicar essas técnicas e ferramentas para a estruturação multinível do modelo proposto?
- Qual o nível de estruturação que o modelo consegue atingir?

1.3 Objetivo

O objetivo geral do trabalho é modelar, implementar e avaliar um modelo multinível para estruturação de informações contidas em evoluções de prontuários escritos em texto livre na língua portuguesa. Para atingir esse objetivo, foram definidos os seguintes objetivos específicos:

- Fazer uma revisão sistemática do estado da arte da estruturação de informações em prontuários;
- Desenvolver um modelo para estruturação de informações contidas em evoluções de prontuários;
- Identificar os melhores métodos e ferramentas para implementar o modelo proposto;
- Implementar e avaliar um protótipo do modelo proposto.

1.4 Contribuições Científicas

As principais contribuições científicas deste trabalho são:

- Uma revisão sistemática que traz uma série de contribuições para o entendimento do estado da arte da estruturação de informações em prontuários eletrônicos, assim como seus desafios;
- Uma análise do ganho de desempenho com o aumento da especialização de domínios em redes BERT;
- Um modelo multinível para estruturação das informações contidas nas evoluções em texto livre do prontuário do paciente.

Além destas contribuições científicas, este trabalho apresenta algumas contribuições tecnológicas, como:

- Uma rede BERT que representa o estado da arte em redes BERTs para uma base de dados, sendo o modelo mais especializado, possuindo conhecimento específico da língua portuguesa, terminologia clínica e o contexto do grupo médico;
- Uma GUI customizável para rotulação de sentenças e estruturação de entidades relevantes, que permite exportar automaticamente as informações para o treinamento e avaliação das tarefas de classificação de sentenças, reconhecimento e relacionamento de entidades;

- Um componente para divisão e classificação de sentenças de evoluções de prontuários no padrão SOAP;
- Um componente de estruturação de informações relevantes em sentenças de evoluções de prontuários;
- Um pipeline para estruturação de informações relevantes em evoluções de prontuários eletrônicos.

1.5 Organização do Texto

O restante deste documento está estruturado da seguinte forma: o Capítulo 2 apresenta o referencial teórico, incluindo os conceitos fundamentais abordados por esta tese; o Capítulo 3 destaca os trabalhos relacionados, baseados em uma revisão sistemática; o Capítulo 4 descreve o modelo proposto e seus componentes; o Capítulo 5 detalha os materiais e métodos propostos para a implementação do modelo, bem como as métricas para avaliação; o Capítulo 6 discute uma implementação do modelo para sua validação e a respectiva avaliação; e, por fim, o Capítulo 7 apresenta as considerações finais, contribuições, limitações, trabalhos publicados e perspectivas para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta os conceitos essenciais e as terminologias utilizadas neste trabalho. Primeiramente, são descritas as terminologias e os padrões de saúde utilizados no trabalho. Após, é feita a conceituação do processamento de linguagem natural e das tarefas da área aplicadas no trabalho. Por fim, é conceituado o modelo de rede neural BERT e apresentados os modelos de rede pré-treinados utilizados no trabalho.

2.1 Terminologia e Padrões de Saúde

Nesta seção, são apresentados os padrões para a área da saúde utilizados nesse trabalho, com foco no Padrão SOAP, TISS e TUSS, e na Classificação Internacional de Doenças (CID). Cada um desses padrões desempenha um papel crucial na estruturação e codificação de informações médicas, facilitando a interoperabilidade e a análise de dados clínicos.

2.1.1 Padrão SOAP

O padrão SOAP é formado por quatro classes: Subjetivo, Objetivo, Avaliação e Plano. Tendo sido desenvolvido para padronizar a documentação de atendimento ao paciente e facilitar a comunicação entre os profissionais de saúde. O padrão reorganiza a estrutura do prontuário médico de notas médicas pouco documentadas para uma lista bem definida de problemas médicos de um paciente (CAMERON; TURTLE-SONG, 2002; ALVAREZ; REINHART, 2020).

Esse padrão é adotado e implementado por vários profissionais de saúde, incluindo farmacêuticos, como um meio de comunicação de informações sobre os cuidados de um paciente (LISENBY et al., 2018). Cada categoria padrão fornece contexto sobre a nota, o que garante uma melhor compreensão da mesma. Por exemplo, entender o contexto em que o medicamento é descrito no relatório é vital para saber se um medicamento está sendo tomado atualmente ("medicamentos: Prilosec 20 mg 1 comprimido ao dia") ou foi prescrito na consulta ("Prescrito Prilosec 20 mg") (MOWERY et al., 2012).

Cada uma de suas quatro classes representam um aspecto da estrutura do prontuário. A classe Subjetivo representa as queixas do paciente, sintomas e histórico de doenças; o Objetivo representa os dados do exame físico, monitoramento e qualquer investigação; a Avaliação representa a avaliação da condição do paciente, problemas médicos, necessidades e prognóstico; e o Plano representa os planos atuais e futuros para investigações, tratamento, educação ou quaisquer ações futuras (ILAN et al., 2012).

2.1.2 TISS e TUSS

A Troca de Informações na Saúde Suplementar - TISS¹ foi estabelecida pela Agência Nacional de Saúde Suplementar (ANS)² como um padrão obrigatório para as trocas eletrônicas de dados de atenção à saúde dos beneficiários de planos, entre os agentes da Saúde Suplementar. O objetivo é padronizar as ações administrativas, subsidiar as ações de avaliação e acompanhamento econômico, financeiro e assistencial das operadoras de planos privados de assistência à saúde e compor o Registro Eletrônico de Saúde. O padrão TISS tem por diretriz a interoperabilidade entre os sistemas de informação em saúde preconizados pela ANS e pelo Ministério da Saúde, e, ainda, a redução da assimetria de informações para os beneficiários de planos privados de assistência à saúde.

O padrão TISS é mantido pelo Comitê de Padronização das Informações em Saúde Suplementar (COPISS), um grupo de trabalho criado para: propor aprimoramentos do padrão; revisar os termos integrantes do componente de representação de conceitos; promover a divulgação e acompanhar a adoção do padrão; analisar os sistemas de informação da saúde suplementar, coordenados pela ANS, visando a adequação do padrão; fomentar e recomendar estudos relativos à tecnologia de informação e comunicação em saúde.

Dentre os componentes do TISS, se destaca a Terminologia Unificada da Saúde Suplementar (TUSS). Segundo MACIEL; FERREIRA; FÁTIMA MARIN(MACIEL; FERREIRA; FÁTIMA MARIN, 2018), com a implantação do padrão, a ANS identificou que as diferentes terminologias praticadas pelo mercado da saúde suplementar dificultavam a análise dos dados na troca de informações entre os atores envolvidos (prestadores de serviços e operadoras de planos de saúde). Levando isso em conta, a ANS determinou a criação do TUSS, tendo como objetivo consolidar em um único padrão diversas terminologias de saúde.

O TUSS é formado por diversas tabelas que representam os mais diversos padrões de saúde. Em sua versão 202205, é formado por, aproximadamente, 100 tabelas diferentes, que representam os mais variados conceitos médicos como medicamentos, procedimentos, diárias, taxas, gases medicinais, órteses, próteses e materiais especiais. Para este trabalho serão utilizadas as seguintes tabelas:

- **Tabela 22 - Terminologia de procedimentos e eventos em saúde:** Tabela que lista os procedimentos e eventos em saúde unificados, contendo: código do termo (único), termo (nome do procedimento), descrição detalhada, data de início de vigência, data de fim de vigência e data de fim de implantação. Tem em sua versão 202205 aproximadamente 6000 itens;
- **Tabela 20 - Terminologia de Medicamentos:** Tabela que lista os medicamentos, contendo: código do termo, termo, apresentação, laboratório, data de início de vigência, data de fim

¹<https://www.gov.br/ans/pt-br/assuntos/prestadores/padrao-para-troca-de-informacao-de-saude-suplementar-2013-tiss>

²<https://www.ans.gov.br>

de vigência, data de fim de implantação e registro ANVISA. Tem em sua versão 202205 aproximadamente 35000 itens.

2.1.3 Classificação Internacional de Doenças (CID)

Em 1893 a "Classificação das Causas de Morte de Bertillon", a partir da recomendação do Instituto Internacional de Estatística, foi adotada por vários países sendo considerada a primeira classificação internacional de causas de morte. Bertillon era o chefe dos serviços de estatísticas da cidade de Paris e codificou uma tabela com um total de 161 categorias, sendo 150 para doenças e 11 para causas externas. Ele ainda foi responsável por três revisões da mesma. Na sexta revisão que foi assumida pela OMS (Organização Mundial da Saúde), a classificação passou a ser chamada de "classificação de doenças, lesões e de causas de morte da OMS" e passou a incluir o motivo das consultas que não por doenças presentes no momento da consulta. A partir dessa versão e suas subsequentes, suas classificações se voltaram muito mais para o levantamento de estatísticas de morbidade do que para a análise da causa da morte. Tanto que, a partir de sua décima revisão (CID-10) seu nome passou a ser "Classificação Estatística Internacional de Doenças e de Problemas Relacionados à Saúde" (LAURENTI, 1991).

Segundo (CESAR et al., 2001), a CID, inicialmente usada somente para mortalidade, foi expandida a fim de adequar-se à morbidade e, subsequentemente, para incluir achados anormais, queixas e circunstâncias sociais que podem ser encontrados no lugar de diagnósticos em registros relativos à saúde. Pode, portanto, ser usada para classificar registros sob o título de diagnósticos, tanto quanto "razões para admissão" e "razões para consulta" que aparecem em grande variedade de registros dos quais são obtidas as estatísticas de saúde.

A Tabela 1 mostra os grupos existentes na CID-10, sua 10ª revisão. A OMS já divulgou a 11ª revisão³ da CID (CID-11), porém ainda não possui versão em português e sua adoção no Brasil apresenta uma série de desafios (GALVÃO; RICARTE, 2021).

A CID é muito utilizada no PLN na área médica, principalmente nas tarefas de predição de doenças (BLINOV et al., 2020; FENG; SHAIB; RUDZICZ, 2020; SINGH et al., 2020), que, em suma, são tarefas de classificação de sentença/texto. A tarefa de predição tem como entrada de dados uma sentença, texto ou fragmento de texto e tem como saída uma CID ou uma lista delas (*multilabel*) que representam a doença predita a partir da análise da doença.

2.2 Processamento de Linguagem Natural (PLN)

O processamento de linguagem natural (PLN) é um ramo da inteligência artificial (IA) preocupado com a interpretação da linguagem humana. Segundo (GOLDBERG; HIRST, 2017), a linguagem humana é altamente ambígua e variável, também está sempre mudando e evoluindo. As pessoas são ótimas em produzir e compreender linguagens, e são capazes de expressar,

³<https://icd.who.int/en>

Tabela 1 – Grupos da CID-10.

Capítulo	Título
I	Algumas doenças infecciosas e parasitárias.
II	Neoplasmas (tumores).
III	Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários.
IV	Doenças endócrinas, nutricionais e metabólicas.
V	Transtornos mentais e comportamentais.
VI	Doenças do sistema nervoso.
VII	Doenças do olho e anexos.
VIII	Doenças do ouvido e da apófise mastóide.
IX	Doenças do aparelho circulatório.
X	Doenças do aparelho respiratório.
XI	Doenças do aparelho digestivo.
XII	Doenças da pele e do tecido subcutâneo.
XIII	Doenças do sistema osteomuscular e do tecido conjuntivo.
XIV	Doenças do aparelho geniturinário.
XV	Gravidez, parto e puerpério.
XVI	Algumas afecções originadas no período perinatal.
XVII	Malformações congênitas, deformidades e anomalias cromossômicas.
XVIII	Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte
XIX	Lesões, envenenamentos e algumas outras consequências de causas externas.
XX	Causas externas de morbidade e de mortalidade.
XXI	Fatores que influenciam o estado de saúde e o contato com os serviços de saúde.
XXII	Códigos para propósitos especiais.

perceber e interpretar significados muito elaborados e matizados. Ao mesmo tempo, enquanto nós humanos somos grandes usuários de linguagem, também somos muito pobres em entender e descrever formalmente as regras que governam a língua. Desta forma compreender e produzir linguagem usando computadores é altamente desafiador.

Desde o surgimento como uma subdisciplina da ciência da computação na década de 1960, várias abordagens foram aplicadas para lidar com os desafios do processamento de linguagem natural, como análise sintática usando árvores de dependência (DALE; MOISL; SOMERS, 2000), gramáticas transformacionais (MCCAWLEY, 1968), métodos estatísticos (MANNING; SCHUTZE, 1999) e estrutura de distribuição (HARRIS, 1954). Desde os anos 2000, uma nova evolução ocorreu na PLN com a utilização de redes neurais. Por utilizar corpora grandes e abordagens de aprendizado sofisticadas, esses métodos proporcionam bom desempenho em tarefas de natureza estatística, como classificação de texto ou mineração de sentimentos. Além disso, eles têm sido usados com muito mais frequência para aprender estruturas sintáticas ou semânticas superiores (gráficos de sintaxe ou mineração de conceitos, respectivamente) (LE GLAZ et al., 2021).

O PLN é aplicado em uma gama de tarefas como análise de sentimento, classificação de sentenças, tradução automática, reconhecimento de entidades e relacionamento de entidades (DENG; LIU, 2018). Nas próximas subseções serão descritas as tarefas de PLN aplicadas no trabalho.

2.2.1 Classificação de sentenças

A classificação de sentenças no PLN é dividida em duas etapas. Primeiramente, o documento é dividido em frases usando caracteres especiais (símbolos de pontuação especiais, aspas, parênteses, colchetes, etc.) ou técnicas mais complexas, como campos aleatórios condicionais (TOMANEK; WERMTER; HAHN, 2007). Na segunda etapa, cada frase é rotulada com uma classe (*single label*), como em nossa proposta, ou mais rótulos (*multilabel*) (DESHPANDE; PALSHIKAR; ATHIAPPAN, 2010). Se uma tarefa possui exatamente duas classes, é chamada de classificação binária e classificação multiclasse quando apresenta mais de duas classes.

2.2.2 Reconhecimento de entidades nomeadas

O reconhecimento de entidades nomeadas (*Named entity recognition NER*) é um subproblema de extração de informações e envolve o processamento de documentos estruturados e não estruturados em busca de expressões de identificação que se referem a pessoas, lugares, organizações e empresas. Envolve duas tarefas: primeiro, a identificação de nomes próprios no texto e, segundo, a classificação desses nomes em um grupo de categorias pré-definidas de interesses como: pessoas, nomes, organizações (empresas, governos, comitês, etc.), lugares (cidades, países, rios, etc.), expressões de data e hora (MANSOURI; AFFENDEY; MAMAT,

2008). No domínio médico, os sistemas NER também podem ser chamados de Reconhecimento de Entidade Médica Nomeada (*Medical-Named Entity Recognition-MER/MNER*)(WANG et al., 2018a; SORIANO; CASTRO, 2017).

O objetivo do MER é identificar todas as ocorrências de tipos específicos de entidades clinicamente relevantes como doença, medicamento e sintomas no texto médico não estruturado (HAN; RUONAN, 2011; KUNDETI et al., 2016). A maioria deles utiliza um ou mais conceitos UMLS (*Unified Medical Language System*) como identificação dessas entidades, o que permite a interoperabilidade semântica em diversas terminologias biomédicas (SARIPALLE; SOOKHAK; HAGHPARAST, 2020).

A UMLS inclui o Metathesaurus, a Rede Semântica e o *SPECIALIST Lexicon and Lexical Tools*. Metathesaurus é o componente principal UMLS (BROWNE et al., 2003), e é um grande tesouro biomédico organizado sob conceitos ou significados. Ele vincula nomes semelhantes para o mesmo conceito de até 200 vocabulários diferentes, incluindo ICD-10 (ORGANIZATION et al., 1993), SNOMED CT (DONNELLY, 2006) e RxNorm (LIU et al., 2005). Com base nesses padrões, muitas ferramentas de reconhecimento de entidades médicas surgiram para uso geral (ZENG et al., 2018), incluindo cTAKES (SAVOVA et al., 2010), MetaMap (ARONSON; LANG, 2010), MedEx (XU et al., 2010), YTEX (OSBORNE; GYAWALI; SOLORIO, 2014), Bioyodie (GORRELL; SONG; ROBERTS, 2018), CLAMP (SOYSAL et al., 2018), Medlee (FRIEDMAN et al., 2004), SciSpaCy (NEUMANN et al., 2019) e MedTaggerIE (TIBBO et al., 2019). NÉVÉOL et al.(NÉVÉOL et al., 2018) destaca que essas ferramentas não mapeiam todos os elementos UMLS e, em sua maioria, são desenvolvidas para a língua inglesa.

2.2.2.1 Extração de relação

A extração de relações é o processo de detectar e classificar as relações semânticas entre entidades em um determinado pedaço de texto. Do ponto de vista da PLN, a extração de relações do texto tem sido tradicionalmente tratada como um pipeline de duas subtarefas separadas: NER e classificação de relações (ZHANG et al., 2019a; DANDALA et al., 2020; FU; LI; MA, 2019). Como MER é uma aplicação de NER no domínio médico, a extração de relação médica (*medical relation extraction (MRE)*) é uma aplicação de extração de relação no domínio médico. Os tipos de relações podem variar de acordo com o objeto de estudos, como interações, diagnósticos, medicamentos, problemas e tratamentos (YADAV et al., 2020a).

Existe uma gama de propostas para tratar a extração de relação, desde algoritmos baseados em regras até técnicas baseadas em aprendizado de máquina (LI; MAO, 2019). Atualmente, as Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs) são as abordagens comumente usadas para isso (LI et al., 2019a). Tais abordagens podem ter o seu desempenho melhorado ao utilizar técnicas como: *position embeddings* (ZENG et al., 2014), variação de tamanhos de filtro (NGUYEN; GRISHMAN, 2015), funções de perda de classificação (SANTOS; XIANG; ZHOU, 2015) e *embeddings* auxiliares (LEE; DERNONCOURT; SZOLOVITS, 2017).

Outra forma de abordar a tarefa da extração de relação é mapeá-la como uma tarefa de classificação de sentenças e utilizar redes neurais que desempenham essa tarefa, como as redes BERT, para tratar o problema (XUE et al., 2019).

2.3 Modelo de Rede Neural BERT

O BERT (*Bidirectional Encoder Representations from Transformers*) é um modelo de linguagem sensível ao contexto com desempenho de última geração em uma variedade de tarefas de PLN (ALSENTZER et al., 2019). Ele é projetado para pré-treinar representações bidirecionais de textos não rotulados condicionados em ambos os contextos, direita e esquerda em todas as camadas. Como resultado, o modelo BERT pré-treinado pode ser ajustado com apenas uma camada de saída adicional para criar modelos de última geração para uma ampla gama de tarefas, como responder a perguntas e fazer inferência de linguagem, sem a necessidade de modificações substanciais de arquitetura para essas tarefas específicas (DEVLIN et al., 2019).

O treinamento de uma rede BERT é dividido em duas etapas: pré-treinamento e ajuste fino. Durante o pré-treinamento, o modelo é treinado em dados não rotulados, que podem ser um grande corpus de textos genéricos como BooksCorpus (800 milhões de palavras) e Wikipedia em inglês (2,5 bilhões de palavras) (LEE et al., 2019). No ajuste fino, o modelo BERT é inicializado com parâmetros pré-treinados e a rede é treinada com dados rotulados para executar a tarefa específica. Outra vantagem do modelo é que o pré-treinamento pode ser realizado novamente em uma rede BERT pré-treinada. SciBERT (BELTAGY; LO; COHAN, 2019), por exemplo, é um modelo de linguagem pré-treinado especializado em texto científico, a partir da versão genérica em inglês de uma rede BERT pré-treinada que incorpora o texto completo de 1,14 milhão de artigos biomédicos e de ciência da computação da Semantic Corpus acadêmico (AMMAR et al., 2018). As redes BERT mostraram melhor desempenho do que outras abordagens para reconhecimento de entidades médicas (SCHNEIDER et al., 2020; SYMEONIDOU; SAZONAU; GROTH, 2019).

Abaixo são descritos os modelos de redes pré-treinadas utilizados no trabalho:

- **BERT multilíngua (mBERT)**: treinada em 104 idiomas, mostrou desempenho multilíngua surpreendente em várias tarefas de PLN, mesmo sem identificação explícita de idioma. O treinamento da mBERT tentou equalizar os idiomas aumentando a amostragem de palavras de idiomas com baixa frequência e reduzindo a amostragem de palavras de idiomas de alta frequência (DEVLIN et al., 2018; WU; DREDZE, 2020);
- **DistilBERT (Distilled BERT)**: É uma versão pré-treinada do BERT de uso geral, menor e mais rápida, que mantém quase os mesmos recursos de compreensão de linguagem. Em particular, utiliza modelos de linguagem pré-treinados com destilação de conhecimento, uma técnica de compressão na qual um modelo compacto é treinado para reproduzir o comportamento de um modelo maior ou de um conjunto de modelos, resultando em mode-

los mais leves e rápidos no tempo de inferência, além de exigir treinamento computacional menor. Mantém 97% de compreensão da linguagem em seu modelo com aproximadamente 60% de redução no tamanho do modelo, rodando 60% mais rápido (SANTOS. et al., 2021). O DistilBERT pode ser ajustado para várias tarefas, mantendo a flexibilidade de modelos maiores enquanto é pequeno o suficiente para ser executado, por exemplo, em dispositivos móveis (SANH et al., 2019);

- **BioBERT**: Rede especializada para a representação da linguagem biomédica em inglês a partir da versão genérica de uma rede BERT pré-treinada incorporando ao treinamento o corpus PubMed Abstracts (4,5 bilhões de palavras) e o PMC Full-text Articles (13,5 bilhões de palavras) (LEE et al., 2019);
- **BioBERTpt**: Modelo BERT aperfeiçoados e treinados no domínio clínico para a língua portuguesa. Foi especializado para linguagem biomédica em português do BERT multilíngue, narrativas clínicas de hospitais brasileiros (27,7 milhões de palavras) e a parte em português das bases de dados de artigos científicos Scielo e Pubmed (16,4 milhões de palavras) (SCHNEIDER et al., 2020).

2.4 Conclusões Parciais

Este capítulo apresentou uma revisão bibliográfica da literatura sobre os temas estratégicos desta tese, sendo eles: as terminologias e padrões de saúde utilizados no desenvolvimento do modelo; as conceituações e tarefas do PLN necessárias para o entendimento das propostas de modelagem do modelo; e os modelos de redes neurais BERT pré-treinadas utilizados para implementação do protótipo para avaliar o modelo. No próximo capítulo é apresentada a revisão sistemática que identificou o estado da arte na estruturação das informações contidas em prontuários não estruturados onde foram identificados desafios e oportunidades em aberto que orientaram o desenvolvimento do modelo.

3 TRABALHOS RELACIONADOS

Para analisar o estado da arte do assunto, optou-se por uma abordagem de revisão sistemática. Uma revisão sistemática é uma forma de identificar e interpretar a pesquisa relevante disponível para uma questão de pesquisa específica (BUDGEN; BRERETON, 2006). Os estudos individuais que contribuem para uma revisão sistemática são chamados de estudos primários, e uma revisão sistemática é um estudo secundário (KEELE et al., 2007).

A revisão tem como objetivo identificar o estado da arte na estruturação das informações contidas em prontuários não estruturados para responder questões-chave para propor novos estudos na área, tais como: como é feita a estruturação de dados não estruturados nos prontuários, quais técnicas e ferramentas são utilizadas, quais desafios foram enfrentados no processo, como é feita a aquisição e como são organizados os dados utilizados no desenvolvimento e avaliação dos estudos.

O conteúdo deste capítulo foi publicado no periódico *Health and Technology* em outubro de 2021, com título de "*Data structuring of electronic health records: a systematic review*" (OLIVEIRA; COSTA; ANTUNES, 2021).

3.1 Metodologia de Busca da Literatura

Nessa seção é descrito o método de pesquisa adotado na revisão sistemática, apresentando a questão e a estratégia de pesquisa, assim como os critérios para seleção de artigos e a realização da pesquisa.

3.1.1 Questões de pesquisa

As perguntas da questão de pesquisa são baseadas na motivação para conduzir a revisão, fornecendo uma resposta baseada em evidências do estado da arte da estruturação de dados de prontuários eletrônicos.

As questões de pesquisa foram divididas em duas categorias: questões gerais (QG) e questões específicas (QE). A primeira categoria (QG) visa identificar como está estruturado o texto livre no prontuário, quais técnicas e ferramentas são utilizadas e quais desafios foram enfrentados no processo. As questões gerais dizem respeito a uma visão mais ampla da estruturação dos registros de saúde. A segunda categoria (QE) explora algumas questões específicas de pesquisa, com foco na aquisição, forma e organização dos dados utilizados no desenvolvimento e avaliação dos estudos. A tabela 2 mostra as questões gerais e específicas propostas na pesquisa.

Tabela 2 – Perguntas gerais e específicas de pesquisa.

Identificador	Questão
Questões Gerais	
QG1	Quais são as formas de estruturação das informações e quais dados são estruturados no registro de saúde?
QG2	Quais são as principais técnicas e ferramentas utilizadas para a estruturação da informação?
QG3	Quais são os principais desafios?
Questões Específicas	
QE1	Qual é a linguagem natural dos dados usados para o desenvolvimento e a avaliação?
QE2	Qual é a quantidade de dados usados para o desenvolvimento e avaliação?
QE3	Como são adquiridos os dados para o desenvolvimento e avaliação?

3.1.2 Estratégia de Pesquisa

Esse processo envolve a definição das palavras-chave de busca e a definição do escopo da pesquisa. As palavras-chave de busca foram definidas após a leitura dos artigos da área de interesse e a separação dos termos, sinônimos e abreviaturas (GUBERT; COSTA; RIGHI, 2020). A questão de pesquisa é descrita, na língua inglesa, a seguir:

String de Pesquisa

```
("structured" OR "structuring") AND "unstructured" AND ("Health records" OR "medical records")
```

Após a definição da string de pesquisa, ela é executada em uma ou mais bases eletrônicas de artigos ou buscadores científicos. Para esta revisão foi escolhido o buscador científico Google Scholar¹ por agregar muitas bases de dados e o ACM², IEEE³. As bases de dados Science Direct⁴, Springer⁵, Scopus⁶ e PubMed⁷ foram usados para obter ampla cobertura da literatura e aumentar a probabilidade de encontrar artigos de alta relevância. Foi definido como intervalo de busca artigos publicados entre os anos de 2017 e 2020. A escolha por esse intervalo se deve ao surgimento do mecanismo de atenção (*attention*) (VASWANI et al., 2017) em 2017 e os modelos de rede BERT *Bidirectional Encoder Representations from Transformers* (DEVLIN et al., 2018) em 2018, que é o mais recente estado da arte em uma série de tarefas para processamento de linguagem natural e um ponto de virada no campo (QIN et al., 2021; DEVLIN et al., 2019;

¹scholar.google.com

²www.acm.org/dl

³ieeexplore.ieee.org

⁴www.sciencedirect.com

⁵www.springerlink.com

⁶www.scopus.com

⁷pubmed.ncbi.nlm.nih.gov

Tabela 3 – Critérios de avaliação de qualidade.

Critério	Descrição
Critério 1	O artigo é relevante para a questão de pesquisa
Critério 2	O artigo tem uma proposta de pesquisa
Critério 3	O artigo tem uma metodologia nova, clara e replicável
Critério 4	O artigo apresenta a forma de avaliação
Critério 5	O artigo apresenta resultados

ALSENTZER et al., 2019).

3.1.3 Seleção de Artigos

Nessa fase, por meio de alguns critérios, são retirados os artigos que não são relevantes para esta pesquisa, ficando apenas os mais relevantes, onde foram definidos cinco filtros:

Remoção de artigos duplicados: Como a busca é feita em muitas bases de dados, ocorre a duplicação, e o primeiro filtro os remove. Para aumentar a abrangência da busca, optamos por incluir artigos não indexados de bases de dados como arXiv⁸, que contém artigos pré-publicados que também geram uma duplicação após sua publicação. Para esses casos, as versões publicadas são mantidas.

Remoção de impurezas: Nesse filtro, são removidos os não-artigos, como dissertações, teses, patentes e artigos que não estejam escritos no idioma inglês. Não foram removidos artigos que tenham suas bases de desenvolvimento e avaliação em outros idiomas, mas foram excluídos artigos cujos resultados foram apresentados em idiomas não ingleses.

Análise do Título: Nesse filtro, são removidos os artigos que, pelo título, evidentemente não são relevantes para a questão de pesquisa. Os artigos que, pelo título, não são claramente irrelevantes são retidos na busca de filtros posteriores.

Análise do resumo: Assim como no filtro anterior, são removidos os artigos que, pela leitura do resumo, evidentemente não são relevantes para a questão de pesquisa. Artigos que não são claramente irrelevantes são retidos na busca de filtros posteriores.

Análise dos artigos: Por fim, os artigos são lidos e é feita uma análise qualitativa para verificar se eles atendem aos critérios da Tabela 3.

3.1.4 Realização da pesquisa

A Figura 1 resume a seleção e aplicação dos filtros utilizados. Foram encontrados 6.238 resultados em todas as bases de dados, sendo 1.003 desses artigos duplicados. Dos 5.235 restantes, 466 não eram artigos ou não foram escritos em inglês. Restaram 4.769 artigos para o filtro de título, que removeu 3.358 artigos que estavam claramente, por meio do título, não

⁸arxiv.org

Figura 1 – Estudo de mapeamento sistemático – seleção de artigos.

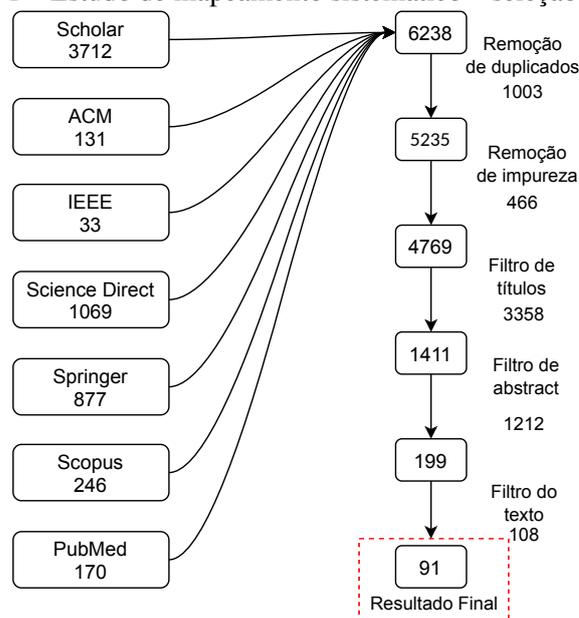


Tabela 4 – Distribuição final dos estudos por ano de publicação.

Ano	Quantidade
2020	33
2019	33
2018	16
2017	9
Total	91

relacionados à questão de pesquisa. Exemplos de artigos removidos incluem "Blockchain-Based Personal Health Records for Patients Empowerment"(EL RIFAI et al., 2020), "Electronic Health Records: Origination, Adoption, and Progression"(REZA; PRIETO; JULIEN, 2020) e "Telemammography: a novel approach for early detection of breast cancer through wavelet-based image processing and machine learning techniques"(SYED; JABEEN; MANIMALA, 2018). Também foram removidos artigos de revisão por não apresentarem um modelo ou técnica nova, restando 1.411 artigos para o filtro de resumos. O objetivo desse filtro foi identificar estudos que estruturam de alguma forma os dados médicos. Por fim, todos os artigos foram lidos, excluindo os artigos que não estavam relacionados ao tema e artigos que não atendiam aos critérios, resultando em 91 artigos. A tabela 4 mostra a distribuição final por ano, e a Figura 2 mostra a evolução dos estudos no período.

3.2 Resultados

Nessa seção, são apresentados os resultados obtidos após a leitura dos artigos para responder às questões de pesquisa levantadas na Tabela 2. A questão QG1 é respondida na subseção 3.2.1,

Figura 2 – Evolução cronológica dos estudos por ano de publicação.

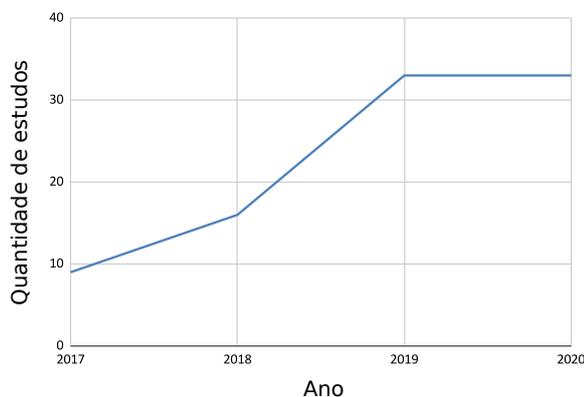
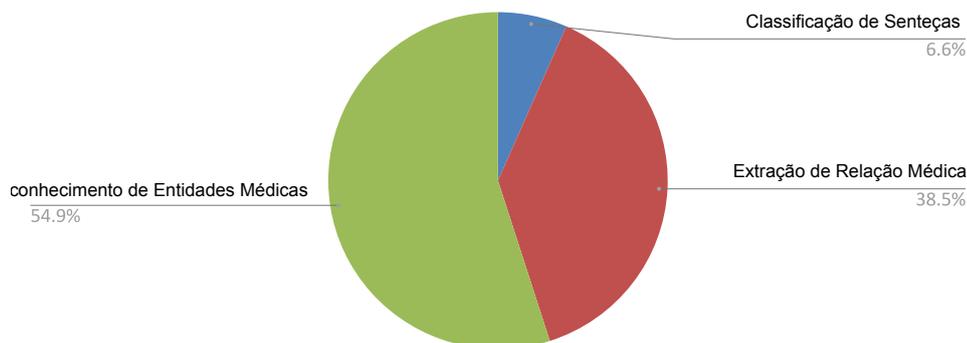


Figura 3 – Distribuição gráfica dos estudos quanto à forma como as informações estão estruturadas nos registros de saúde.



a questão QG2 é respondida na subseção 3.2.2, as questões QE1, QE2 e QE3 são respondidas na subseção 3.2.3 e, finalmente, a questão QG3 é respondida na seção 3.3, seguida de uma avaliação mais ampla dos rumos futuros na área.

3.2.1 Formas de estruturação e dados nos registros de saúde

Após a leitura dos estudos, para responder à QG1, os dividimos em três macro-categorias de acordo com a forma em que as informações foram estruturadas. A Tabela 5 e a Figura 3 mostram a distribuição dos estudos por classificação. As três subseções a seguir descrevem cada macro-categoria. Apresentam também uma análise dos dados estruturados em cada estudo. A quarta subseção analisa a completude das informações estruturadas.

3.2.1.1 Reconhecimento de Entidade Médica Nomeada

Dos estudos analisados, grande parte se concentrou no reconhecimento de entidades médicas em idiomas não ingleses, sendo a maioria no idioma chinês. A tabela 6 mostra esta distribuição. Cinco estudos apresentam novas metodologias para reconhecimento de entidades, como BERT

Tabela 5 – Classificação dos estudos quanto à forma como as informações estão estruturadas nos prontuários.

Classificação	Quantidade	Estudos
Reconhecimento de Entidade Médica Nomeada	50	(KORMILITZIN et al., 2020; WEEKS et al., 2020; DAI; SU; WU, 2019; STEINKAMP et al., 2020; GAVRILOV et al., 2020; HUANG; HONG; TSAI, 2020; LI et al., 2020a; KRALJEVIC et al., 2020; YANG et al., 2020; LOPES; TEIXEIRA; OLIVEIRA, 2020; SCHNEIDER et al., 2020; CAI; ZHOU; ZHENG, 2020; JOUFFROY et al., 2020; LI et al., 2020b; OHNO-MACHADO; SÉROUSSI, 2019; ZHAO, 2019; AMIN et al., 2019; LI et al., 2019b; DONG et al., 2019; NUTHAKKI et al., 2019; YIN et al., 2019; BLANCO et al., 2019; RAISKIN; EICKHOFF; BEELER, 2019; SYMEONIDOU; SAZONAU; GROTH, 2019; MAGUIRE et al., 2019; SEN et al., 2019; LU et al., 2019; LOPES; TEIXEIRA; OLIVEIRA, 2019; ZHANG et al., 2018a; CHOWDHURY et al., 2018; WANG et al., 2018b,a; WEEGAR et al., 2018; LIU et al., 2017; PÉREZ et al., 2017; SORIANO; PEÑA, 2017; BREISCHNEIDER et al., 2017; CHEN et al., 2019a; LEE et al., 2018; LERNER; PARIS; TANNIER, 2020; DU; PANG; XU, 2020; TARCAR et al., 2020; SPANDORFER et al., 2019; WANG et al., 2020; CHEN; HO; LIN, 2020; WUNNAVA et al., 2019; ZHANG et al., 2020; JI; CHEN; JIANG, 2020; SHARMA et al., 2020; Wang; Zeng, 2018)
Extração de Relação Médica	35	(SAGHEB et al., 2020; ALMEIDA; MATOS, 2020; YADAV et al., 2020b; FONFERKO-SHADRACH et al., 2019; CHENG et al., 2019; ALODADI; JANEJA, 2019; SU et al., 2019; ZHANG et al., 2019a; SHI et al., 2019; SHAH et al., 2018; LIU et al., 2018; MUNKHDALAI; LIU; YU, 2018; BOYTCHEVA, 2018; LUO et al., 2018; LUO, 2017; CHEN et al., 2019b; LEITER et al., 2020; LI et al., 2019c; VIANI et al., 2018; SINGH et al., 2017; DANDALA et al., 2020; BALABAEVA; KOVALCHUK, 2020; MAXIMILIAN; J.; MICHAEL, 2019; Song et al., 2017; IQBAL et al., 2017; YANG et al., 2019; NATARAJAN et al., 2016; LI et al., 2020c; CAI et al., 2019; KERSLOOT et al., 2019; LAMY et al., 2018a; PETERSON; LIU, 2020; YEHA et al., 2019; TOGNOLA; MURRI; CUDA, 2018; STEINKAMP et al., 2019)
Classificação de Sentença	6	(MOEN et al., 2020, 2019; KENEI et al., 2019; ZHANG et al., 2018b; MOEN et al., 2018; WU et al., 2020)

(SYMEONIDOU; SAZONAU; GROTH, 2019). DAI; SU; WU(DAI; SU; WU, 2019) propõe um modelo de reconhecimento de entidades médicas, incluindo nomes de medicamentos, efeitos colaterais de medicamentos e outras entidades que se referem ao medicamento, mas deixam como trabalho futuro o mapeamento dessas relações. Os demais estudos focam no reconhecimento de entidades para fins específicos, como entidades relacionadas ao câncer de mama (BREISCHNEIDER et al., 2017) e medidas de qualidade (OHNO-MACHADO; SÉROUSSI, 2019). A tabela 7 mostra a classificação completa dos estudos MER.

3.2.1.2 Extração de relação médica

Dos estudos de extração de relação médica, dezenove dos estudos analisados focam em associações binárias, como eventos clínicos e temporalidade (ALODADI; JANEJA, 2019), doença e sintomas (SHAH et al., 2018), condição de órgão ou sistema anatômico (BOYTCHEVA, 2018), um problema médico com seu tratamento, um problema médico com um teste, ou um problema médico com problemas médicos coexistentes (LUO et al., 2018; LUO, 2017) e parentes doentes (BALABAEVA; KOVALCHUK, 2020).

Três estudos (DANDALA et al., 2020; IQBAL et al., 2017; MUNKHDALAI; LIU; YU, 2018) estão focados em um desafio proposto no Workshop on Natural Language Processing Challenges for Clinical Records (UZUNER et al., 2010) para associar um medicamento com a dosagem, um medicamento com os efeitos colaterais e gravidade com os efeitos colaterais.

Três estudos focam na associação da doença e suas características e sintomas. YADAV et al.(YADAV et al., 2020b) associa a doença, sinal vital ou sintoma mencionado no documento com sua gravidade, curso e temporalidade. FONFERKO-SHADRACH et al.(FONFERKO-SHADRACH et al., 2019) extraiu dados estruturados de epilepsia de anotações clínicas contendo o tipo de epilepsia, frequência das crises e investigação neurológica; CHEN et al.(CHEN et al., 2019b) extraiu e estruturou informações sobre tumores como tamanho, textura e cor do prontuário do paciente.

Seis estudos têm uma abordagem mais ampla para extrair um grupo maior de relações. VIANI et al.(VIANI et al., 2018) desenvolveu uma ontologia contendo 11 eventos e 61 atributos, sendo 44 numéricos e outras 8 categorias para relações entre entidades. No entanto, avaliou o sistema apenas nos cinco eventos mais frequentes: diagnóstico principal, medicamentos prescritos e três exames diagnósticos (ECG, Holter ECG, teste de esforço ergométrico); SINGH et al.(SINGH et al., 2017) extraiu UMLS correspondente a PICO (*Population, Intervention/Comparator and Outcomes*); Song et al.(Song et al., 2017) dividiu as informações que extraem os prontuários em 5 grupos: queixa principal, sintomas agravados, informações sobre complicações, informações médicas e intervalo desde a última internação; PETERSON; LIU(PETERSON; LIU, 2020) propôs a estruturação completa do texto não estruturado da lista de problemas dos registros convertendo-os em Expressões SNOMED CT, que são uma combinação estruturada de um ou mais identificadores de conceito usados para representar uma ideia clínica de forma

Tabela 6 – Distribuição dos estudos sobre o idioma alvo do MER.

Idioma	Quantidade	Estudos
Chinês	16	(LI et al., 2020a; YANG et al., 2020; CAI; ZHOU; ZHENG, 2020; LI et al., 2020b; CHENG et al., 2019; LI et al., 2019b; DONG et al., 2019; SU et al., 2019; ZHANG et al., 2019a; YIN et al., 2019; LU et al., 2019; ZHANG et al., 2018a,b; CHOWDHURY et al., 2018; LIU et al., 2018; WANG et al., 2018b,a; LIU et al., 2017; CHEN et al., 2019a,b; DU; PANG; XU, 2020; WANG et al., 2020; ZHANG et al., 2020; Wang; Zeng, 2018)
Inglês	3	(LOPES; TEIXEIRA; OLIVEIRA, 2020; SCHNEIDER et al., 2020; LOPES; TEIXEIRA; OLIVEIRA, 2019)
Francês	2	(JOUFFROY et al., 2020; LERNER; PARIS; TANNIER, 2020)
Espanhol e Sueco	2	(WEEGAR et al., 2018; PÉREZ et al., 2017)
Alemão	1	(AMIN et al., 2019)
Búlgaro	1	(ZHAO, 2019)
Coreano	1	(LEE et al., 2018)
Espanhol	1	(BLANCO et al., 2019)
Russo	1	(GAVRILOV et al., 2020; BALABAEVA; KOVALCHUK, 2020)

Tabela 7 – Classificação dos estudos MER quanto ao objetivo proposto.

Classificação	Qtd	Estudos
Idioma não inglês	28	(GAVRILOV et al., 2020; LI et al., 2020a; YANG et al., 2020; LOPES; TEIXEIRA; OLIVEIRA, 2020; SCHNEIDER et al., 2020; CAI; ZHOU; ZHENG, 2020; JOUFFROY et al., 2020; LI et al., 2020b; ZHAO, 2019; AMIN et al., 2019; LI et al., 2019b; DONG et al., 2019; YIN et al., 2019; BLANCO et al., 2019; LU et al., 2019; LOPES; TEIXEIRA; OLIVEIRA, 2019; ZHANG et al., 2018a; CHOWDHURY et al., 2018; WANG et al., 2018b,a; WEEGAR et al., 2018; LIU et al., 2017; PÉREZ et al., 2017; CHEN et al., 2019a; LEE et al., 2018; LERNER; PARIS; TANNIER, 2020; ZHANG et al., 2020; Wang; Zeng, 2018)
Objetivo específico	15	(STEINKAMP et al., 2020; HUANG; HONG; TSAI, 2020; OHNO-MACHADO; SÉROUSSI, 2019; NUTHAKKI et al., 2019; RAISKIN; EICKHOFF; BEELER, 2019; MAGUIRE et al., 2019; SEN et al., 2019; BREISCHNEIDER et al., 2017; DU; PANG; XU, 2020; SPANDORFER et al., 2019; WANG et al., 2020; CHEN; HO; LIN, 2020; WUNNAVA et al., 2019; JI; CHEN; JIANG, 2020; SHARMA et al., 2020)
Nova Metodologia	6	(KORMILITZIN et al., 2020; WEEKS et al., 2020; KRALJEVIC et al., 2020; SYMEONIDOU; SAZONAU; GROTH, 2019; SORIANO; PEÑA, 2017; TARCAR et al., 2020)
Relação futura	1	(DAI; SU; WU, 2019)

lógica; TOGNOLA; MURRI; CUDA(TOGNOLA; MURRI; CUDA, 2018) também converteu os registros de saúde em Expressões SNOMED CT; YEHIA et al.(YEHIA et al., 2019) estruturou os registros de saúde para sua própria ontologia do domínio clínico já proposta em estudo anterior (BOSHNAKA; ABDELGABERB; AMANYABDOC, 2018).

Nos demais estudos, SAGHEB et al.(SAGHEB et al., 2020) extraiu informações de prontuários médicos cirúrgicos para cirurgias de joelho e procedimentos associados: joelho envolvido (direito ou esquerdo), lateralidade cirúrgica (direito, esquerdo, ambos), tipo de restrição, presença de recapeamento patelar, e números de modelo de implante; STEINKAMP et al.(STEINKAMP et al., 2019) relata os dados da radiologia estruturada transformando as frases em fatos onde cada fato representa uma afirmação, possuindo uma âncora, que é o elemento mais relevante dos achados, e um conjunto de informações adicionais (modificantes), que contextualizam ou modificam o fato; CAI et al.(CAI et al., 2019) estrutura as informações dos sinais vitais: pressão arterial, frequência cardíaca, frequência respiratória, temperatura, níveis de saturação de oxigênio; SHI et al.(SHI et al., 2019) extraiu a história familiar e identificou parentes, lado da família, status de vida e doença.

3.2.1.3 Classificação de sentença

Dos seis estudos analisados, quatro estudos (MOEN et al., 2019, 2020, 2018) apresentaram abordagens diferentes para o mesmo problema de classificação de sentenças para anotações clínicas de enfermagem. KENEI et al.(KENEI et al., 2019) classificou as sentenças das notas clínicas utilizando o padrão SOAP (MOWERY et al., 2012). ZHANG et al.(ZHANG et al., 2018b) classificou as sentenças em estado de saúde, diagnóstico hospitalar, processo de tratamento, tratamento médico e consulta de retorno de acompanhamento clínico e criou uma segunda camada de classificação extraindo a temporalidade da sentença, ou seja, quando a sentença ocorreu para estruturar as informações em uma linha do tempo. WU et al.(WU et al., 2020) classificou as sentenças, identificou as entidades e as relacionou. Porém, teve, como resultado final, apenas a sentença reescrita em seu estudo sobre o processamento de itens de laudos de exames patológicos de doenças renais para gerar formulários estruturados padronizados a partir de um dicionário médico.

3.2.1.4 Completude da estruturação

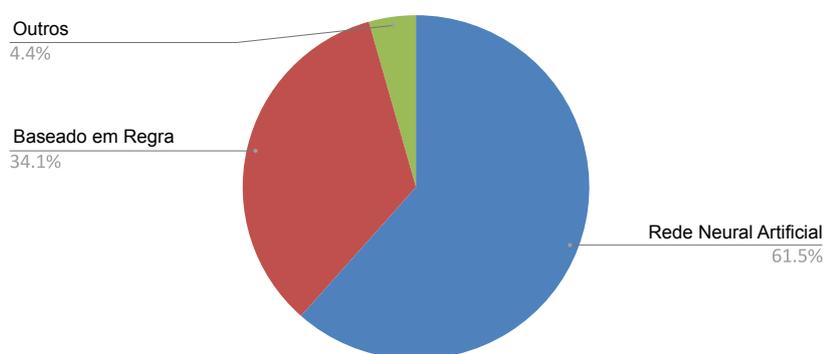
Outro aspecto abordado na revisão é a completude da estruturação. Em outras palavras, o estudo analisado estrutura apenas parte das informações, deixando parte das informações sem classificação, rótulo ou relação. Desta forma, perdendo partes das informações e, conseqüentemente, tornando necessário manter o prontuário não estruturado, ou o estudo analisado estrutura todas as informações relevantes, permitindo o descarte ou arquivamento das informações não estruturadas, sem perda relevante de informações.

Dos 91 estudos analisados, apenas dez apresentaram esse nível de completude. Os seis estudos, que são de classificação de sentenças, abrangem todas as sentenças de um documento classificando cada uma delas. Três estudos de MER apresentam como resultado final todas as entidades relevantes identificadas e relacionadas, porém limitam-se a um elemento específico do prontuário ou especialidade médica. STEINKAMP et al.(STEINKAMP et al., 2019) estrutura laudos de radiologia, PETERSON; LIU(PETERSON; LIU, 2020) propôs a estruturação completa do texto não estruturado de uma lista de problemas, mas não contemplava outros elementos dos registros de saúde como o plano de diagnóstico; TOGNOLA; MURRI; CUDA(TOGNOLA; MURRI; CUDA, 2018) converteu os registros de saúde em SNOMED CT Expressions, mas apenas para registros de saúde de audiometria e otorrinolaringologia. Apenas um estudo classifica, identifica e relaciona as entidades de registro de saúde como um todo. YEHIA et al.(YEHIA et al., 2019) estruturou os registros de saúde para sua própria ontologia.

3.2.2 Técnicas e ferramentas utilizadas para a estruturação das informações em prontuários de saúde

Para responder a questão QG2, os estudos foram divididos em dois macrogrupos de técnicas baseadas na técnica mais relevante utilizada no processo de estruturação da seguinte forma: redes neurais artificiais (GALLANT; GALLANT, 1993) com 56 estudos e algoritmos baseados em regras (DASH et al., 2017) com 30 estudos. Cinco estudos não se enquadraram nessas duas categorias e foram classificados separadamente. A Figura 4 mostra a distribuição das técnicas, e as subseções mostram em detalhes cada uma delas. Alguns estudos utilizam mais de uma técnica para comparar o desempenho e, nesses casos, foram consideradas as que apresentaram melhor desempenho.

Figura 4 – Distribuição gráfica dos estudos sobre a principal técnica utilizada no desenvolvimento.



3.2.2.1 Redes neurais artificiais

Uma gama de propostas de arquiteturas de redes neurais foi encontrada nos estudos analisados, e o processo foi semelhante em quase todos eles: adquirir um corpus de dados de tamanho

adequado com os elementos relevantes para o estudo já rotulados ou rotulá-los, dividir os dados em uma base de treinamento e uma base de teste, definir a arquitetura da rede neural, treiná-la e avaliar os resultados. A Tabela 9 lista todas as redes. As arquiteturas de redes neurais mais relevantes encontradas foram memória bidirecional de longo prazo (*Bidirectional long-short term memory Bi-LSTM*) com ou sem campos aleatórios condicionais (*Conditional random fields CRF*) e BERT. Estes são descritos abaixo.

Bi-LSTM com ou sem CRF. Uma rede LSTM, é uma arquitetura de rede neural recorrente (RNN) capaz de processar dados sequenciais em que cada decisão é influenciada pela entrada anterior. No processamento de texto, uma sequência é uma frase entregue à rede, uma palavra de cada vez (MOEN et al., 2019). Internamente, os LSTMs mantêm um status que a rede pode modificar, esquecendo partes das informações antigas e acrescentando algo novo à entrada atual. Assim, na leitura das palavras de entrada, uma rede LSTM tem a capacidade de manter a informação semântica de cada palavra relevante. Um BI-LSTM é composto por duas redes LSTM conectadas, uma rede que recebe a sequência na ordem regular e outra que a recebe na ordem inversa. Assim, é possível observar não apenas a posição anterior, mas também a posição futura (LU et al., 2019). As redes LSTM produzem uma lista de classificação/rotulagem com probabilidades, mas ao processar essa saída em um CRF, elas podem inferir o resultado com mais precisão, analisando as probabilidades em um grupo dentro do contexto da sentença (LI et al., 2019b). CRFs são um tipo de modelo gráfico probabilístico que leva em consideração o contexto da amostra vizinha para tarefas como classificação. A previsão é modelada como um modelo gráfico que implementa dependências entre as previsões (LAFFERTY; MCCALLUM; PEREIRA, 2001). Com a adição dessa camada nos LSTMs, identificou-se uma melhora nos resultados dos estudos que fizeram essa comparação (CHENG et al., 2019; LOPES; TEIXEIRA; OLIVEIRA, 2020).

3.2.2.2 Algoritmo Baseado em Regras

Dos 30 estudos analisados que utilizam um algoritmo baseado em regras como método principal, 17 também combinam alguma ferramenta de PLN ou, mais especificamente, uma ferramenta de NER como as descritas na subseção 3.2.1.1, onde a ferramenta é usada para identificar as entidades e o sistema de regras para tratar o resultado para seu objetivo. A tabela 8 lista todas as ferramentas utilizadas, e abaixo estão dois exemplos.

SAGHEB et al.(SAGHEB et al., 2020) usou MedTaggerIE (TIBBO et al., 2019) para nomear as entidades e um algoritmo baseado em regras para associar o procedimento ao joelho, lateralidade e outros elementos em seu estudo para extração e estruturação de cirurgias de joelho em notas clínicas. FONFERKO-SHADRACH et al.(FONFERKO-SHADRACH et al., 2019) usou Bio-YODIEs (GORRELL; SONG; ROBERTS, 2018) para identificar conceitos de UMLS e uma combinação de algoritmos e estatísticas baseados em regras para extrair e estruturar nove categorias de informação sobre epilepsia na extração de texto clínico de epilepsia (ExECT). Os demais

Tabela 8 – Ferramenta NLP/NER usada por algoritmos baseados em regras.

Ferramenta	qtd	Referência
SciSpaCy	3	(KORMILITZIN et al., 2020; GAVRILOV et al., 2020; KRALJEVIC et al., 2020)
MetaMap	2	(SHAH et al., 2018; TOGNOLA; MURRI; CUDA, 2018)
cTAKES	2	(BREISCHNEIDER et al., 2017; KERSLOOT et al., 2019)
UIMA framework	1	(VIANI et al., 2018)
OB-CIE	1	(YEHIA et al., 2019)
NLTK	1	(LIU et al., 2018)
MedTaggerIE	1	(SAGHEB et al., 2020)
Leo	1	(OHNO-MACHADO; SÉROUSSI, 2019)
HanLP e NLTK	1	(HUANG; HONG; TSAI, 2020)
GraphIE	1	(LEITER et al., 2020)
GATENLP	1	(IQBAL et al., 2017)
CLAMP	1	(ALODADI; JANEJA, 2019)
Bio-YODIE	1	(FONFERKO-SHADRACH et al., 2019)

Tabela 9 – Arquitetura de redes neurais usada nos estudos

Técnica	Qtd	Referência
Bi-LSTM-CRF	11	(DAI; SU; WU, 2019; YANG et al., 2020; LOPES; TEIXEIRA; OLIVEIRA, 2020; CHENG et al., 2019; LU et al., 2019; ZHANG et al., 2018a; WEEGAR et al., 2018; CHEN et al., 2019a; LI et al., 2019c; WUNNAVA et al., 2019; Wang; Zeng, 2018)
Bi-LSTM	9	(YADAV et al., 2020b; MOEN et al., 2020; JOUFFROY et al., 2020; MOEN et al., 2019; LI et al., 2020b; SU et al., 2019; SHI et al., 2019; YIN et al., 2019; MOEN et al., 2018)
Convolutional neural network (CNN)	5	(ZHAO, 2019; KENEI et al., 2019; SINGH et al., 2017; SPANDORFER et al., 2019; SHARMA et al., 2020)
CRF	4	(LOPES; TEIXEIRA; OLIVEIRA, 2019; LIU et al., 2017; PÉREZ et al., 2017; LEE et al., 2018)
LSTM	3	(NUTHAKKI et al., 2019; LUO, 2017; TARCAR et al., 2020)
Bi-LSTM-CRF + Attention	3	(LI et al., 2019b; DU; PANG; XU, 2020; WANG et al., 2020)
BERT	3	(STEINKAMP et al., 2020; SCHNEIDER et al., 2020; SYMEONIDOU; SAZONAU; GROTH, 2019)
Bi-GRU	2	(BLANCO et al., 2019; STEINKAMP et al., 2019)
bi-directional RNN	2	(DONG et al., 2019; CHOWDHURY et al., 2018)
support vector machines (SVM)	1	(MUNKHDALAI; LIU; YU, 2018)
Segment convolutional neural networks (Seg-CNNs)	1	(LUO et al., 2018)
RNN + Attention	1	(LI et al., 2020c)
Residual Network (ResNet) + Attention	1	(ZHANG et al., 2019a)
Neural network probabilistic language model (NNLM)	1	(WANG et al., 2018a)
LSTM-CRF + Rule-based	1	(YANG et al., 2019)
LSTM-CRF	1	(JI; CHEN; JIANG, 2020)
HAC-RNN	1	(SEN et al., 2019)
GRU	1	(RAISKIN; EICKHOFF; BEELER, 2019)
ELMo-lattice-LSTM-CRF	1	(LI et al., 2020a)
CRF-segmentation	1	(CHEN et al., 2019b)
Bi-LSTM-CRF + Rule-based	1	(DANDALA et al., 2020)
Bi-GRU-CRF	1	(LERNER; PARIS; TANNIER, 2020)
BERT-Bi-LSTM	1	(PETERSON; LIU, 2020)

não utilizaram nenhum tipo de rede neural ou ferramenta de PLN, ao invés disso, percorrem a string utilizando dicionários ou expressões regulares para identificar as entidades, tratando-as posteriormente para seus objetivos. O MedExtractR (WEEKS et al., 2020) é um exemplo dessa abordagem. Ele usa uma combinação de dicionários léxicos e padrões de expressão regular para identificar informações relevantes sobre drogas. Ele, primeiramente, identifica o medicamento e cria uma janela de busca para outras informações, como potência, quantidade da dose, número de doses, tempo de ingestão, frequência e hora da última dose.

3.2.2.3 Outros estudos

Em outros estudos, CAI; ZHOU; ZHENG(CAI; ZHOU; ZHENG, 2020) usou o aprendizado ativo para desenvolver NER em chinês. NATARAJAN et al.(NATARAJAN et al., 2016) usou redes lógicas de Markov para fazer relações entre drogas e condições da literatura médica. AMIN et al.(AMIN et al., 2019) e LAMY et al.(LAMY et al., 2018a) propuseram a tradução de seu idioma original (alemão e português) para o inglês como método principal, tendo assim acesso a recursos, ferramentas e redes pré-treinadas indisponíveis para seus idiomas originais.

3.2.3 Dados para desenvolvimento e avaliação

Quanto à linguagem natural dos dados utilizados para o desenvolvimento e avaliação, observamos que os estudos, em sua maioria, desenvolveram, treinaram e testaram seus modelos em documentos clínicos redigidos na língua inglesa. O chinês é a segunda língua mais comum. A soma dos demais idiomas não chega a 20% dos estudos analisados. Percebemos que a linguagem natural afeta diretamente o desenvolvimento de ferramentas em vários aspectos, principalmente em estudos em outras línguas que não o inglês. Esses aspectos são discutidos na Seção 4. A Figura 5 mostra um gráfico com a distribuição dos idiomas, e a Tabela 11 mostra os artigos.

Foi identificada uma grande variação na quantidade de dados utilizados para o desenvolvimento, treinamento e teste dos estudos. As bases de dados podem ser divididas em dois grupos: bases de dados públicos e bases de dados privados. Dois estudos não esclareceram a origem ou quantidade de dados utilizados (KENEI et al., 2019; ZHANG et al., 2018b). A Figura 6 mostra esta distribuição.

As bases de dados públicas são geradas por meio da anonimização e anotação de dados fornecidos por uma ou mais instituições de saúde e estão disponíveis para uso no desenvolvimento de estudos na área ou desafios em concursos. A Tabela 12 lista essas bases de dados com uma breve descrição de cada um e dos estudos que os utilizam.

As bases de dados privadas são concedidas por meio de convênios entre os pesquisadores do estudo e instituições de saúde como hospitais, clínicas e laboratórios. A maioria dos estudos que utilizam bases de dados privadas explicita a anonimização de suas bases de dados. A Tabela 10 apresenta a distribuição das bases de dados privados pelo número de documentos utilizados.

Tabela 10 – Distribuição das bases privadas pela quantidade de documentos utilizados.

Quantidade de Documentos	Qtd	Referências
1 a 1000	24	(WEEKS et al., 2020; HUANG; HONG; TSAI, 2020; LI et al., 2020a; LOPES; TEIXEIRA; OLIVEIRA, 2020; ALMEIDA; MATOS, 2020; JOUFFROY et al., 2020; FONFERKO-SHADRACH et al., 2019; LI et al., 2019b; YEHA et al., 2019; DONG et al., 2019; LOPES; TEIXEIRA; OLIVEIRA, 2019; TOGNOLA; MURRI; CUDA, 2018; MUNKHDALAI; LIU; YU, 2018; CHEN et al., 2019b; LEE et al., 2018; LEITER et al., 2020; BALABAEVA; KOVALCHUK, 2020; MAXIMILIAN; J.; MICHAEL, 2019; STEINKAMP et al., 2019; Song et al., 2017; WUNNAVA et al., 2019; WU et al., 2020; NATARAJAN et al., 2016; KERSLOOT et al., 2019)
1001 a 10000	27	(STEINKAMP et al., 2020; GAVRILOV et al., 2020; SAGHEB et al., 2020; YANG et al., 2020; CAI; ZHOU; ZHENG, 2020; OHNO-MACHADO; SÉROUSSI, 2019; ZHAO, 2019; CHENG et al., 2019; AMIN et al., 2019; SU et al., 2019; YIN et al., 2019; SYMEONIDOU; SAZONAU; GROTH, 2019; LIU et al., 2018; WANG et al., 2018b,a; WEEGAR et al., 2018; BREISCHNEIDER et al., 2017; CHEN et al., 2019a; LERNER; PARIS; TANNIER, 2020; VIANI et al., 2018; TARCAR et al., 2020; SPANDORFER et al., 2019; IQBAL et al., 2017; WANG et al., 2020; CHEN; HO; LIN, 2020; YANG et al., 2019; Wang; Zeng, 2018)
10001 a 10000	9	(BLANCO et al., 2019; RAISKIN; EICKHOFF; BEE-LER, 2019; MAGUIRE et al., 2019; LAMY et al., 2018a; CHOWDHURY et al., 2018; DU; PANG; XU, 2020; ZHANG et al., 2020; CAI et al., 2019; SHARMA et al., 2020)
100001 a 100000	9	(MOEN et al., 2020, 2019; SHAH et al., 2018; SEN et al., 2019; BOYTCHEVA, 2018; MOEN et al., 2018; LIU et al., 2017; PÉREZ et al., 2017; SORIANO; PEÑA, 2017)
acima 100000	3	(SINGH et al., 2017; LI et al., 2020c; PETERSON; LIU, 2020)

Figura 5 – Distribuição gráfica de linguagens naturais da base de dados usados para desenvolvimento e avaliação

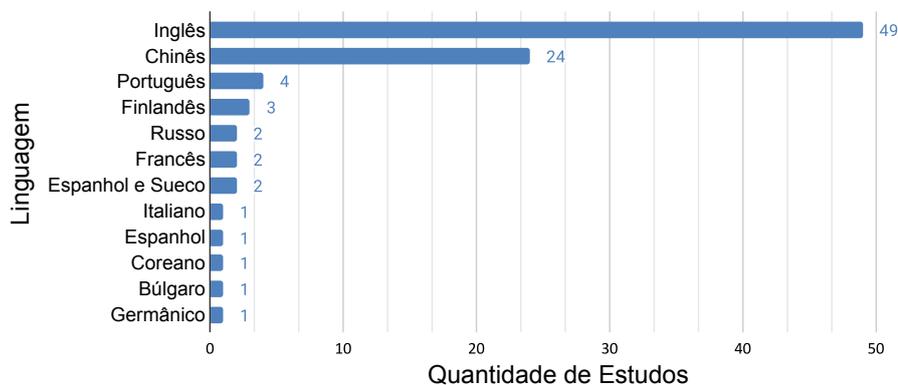
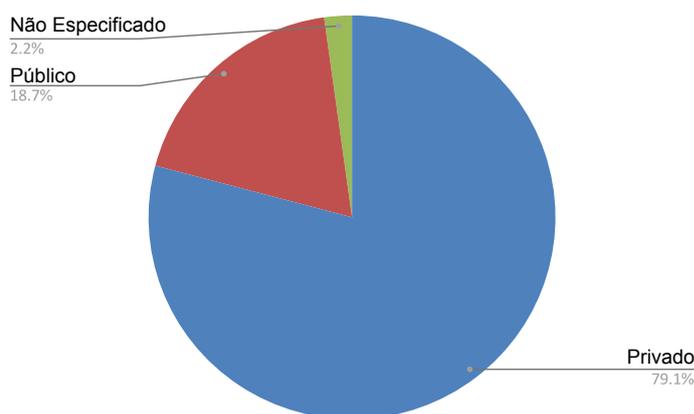


Figura 6 – Distribuição gráfica quanto à origem das bases de dados usados para desenvolvimento e avaliação



3.3 Desafios e Oportunidades de Pesquisa

Muitos estudos mencionam, como principal desafio, a dificuldade na aquisição de bases de dados rotuladas (padrão ouro) adequadas para treinamento e avaliação (RAISKIN; EICKHOFF; BEELER, 2019; KORMILITZIN et al., 2020; ZHANG et al., 2019b; SHAH et al., 2018; FONFERKO-SHADRACH et al., 2019; PETERSON; LIU, 2020; VIANI et al., 2018). Em contraste com outros problemas de PLN, nos problemas médicos, há sérias questões de privacidade, que criam uma série de dificuldades para a criação de um corpus público (LAMY et al., 2018a). Métodos populares de criação de dados rotulados, como *crowdsourcing*, por exemplo, não são aplicáveis no domínio clínico (KENEI et al., 2019) devido a questões de privacidade e à necessidade de conhecimento médico para tornar a rotulagem dos dados precisa. Erros nessas bases de dados podem causar problemas no treinamento e avaliação dos modelos (MOEN et al., 2018). Além disso, o uso de dados de apenas uma instituição pode afetar o desempenho quando aplicado a outras instituições. (SAGHEB et al., 2020; CHEN et al., 2019b). LI et al. (LI et al., 2019c) também observou que algumas categorias de problemas, como doenças e sintomas raros

Tabela 11 – Lista de linguagens naturais da base de dados usados para desenvolvimento e avaliação

Idioma	Qtd	Referências
Inglês	49	(KORMILITZIN et al., 2020; WEEKS et al., 2020; DAI; SU; WU, 2019; STEINKAMP et al., 2020; HUANG; HONG; TSAI, 2020; SAGHEB et al., 2020; PETERSON; LIU, 2020; KRALJEVIC et al., 2020; ALMEIDA; MATOS, 2020; YADAV et al., 2020b; FONFERKO-SHADRACH et al., 2019; OHNO-MACHADO; SÉROUSSI, 2019; YEHA et al., 2019; ALODADI; JANEJA, 2019; NUTHAKKI et al., 2019; SHI et al., 2019; SHAH et al., 2018; RAISKIN; EICKHOFF; BEELER, 2019; SYMEONIDOU; SAZONAU; GROTH, 2019; MAGUIRE et al., 2019; SEN et al., 2019; KENEI et al., 2019; TOGNOLA; MURRI; CUDA, 2018; MUNKHDALAI; LIU; YU, 2018; BOYTCHEVA, 2018; LUO et al., 2018; LUO, 2017; SORIANO; PEÑA, 2017; BREISCHNEIDER et al., 2017; LEITER et al., 2020; LI et al., 2019c; SINGH et al., 2017; DANDALA et al., 2020; TARCAR et al., 2020; SPANDORFER et al., 2019; MAXIMILIAN; J.; MICHAEL, 2019; STEINKAMP et al., 2019; Song et al., 2017; IQBAL et al., 2017; CHEN; HO; LIN, 2020; WUNNAVA et al., 2019; WU et al., 2020; YANG et al., 2019; NATARAJAN et al., 2016; LI et al., 2020c; JI; CHEN; JIANG, 2020; CAI et al., 2019; SHARMA et al., 2020; KERSLOOT et al., 2019)
Chinês	24	(LI et al., 2020a; YANG et al., 2020; CAI; ZHOU; ZHENG, 2020; LI et al., 2020b; CHENG et al., 2019; LI et al., 2019b; DONG et al., 2019; SU et al., 2019; ZHANG et al., 2019a; YIN et al., 2019; LU et al., 2019; ZHANG et al., 2018a,b; CHOWDHURY et al., 2018; LIU et al., 2018; WANG et al., 2018b,a; LIU et al., 2017; CHEN et al., 2019a,b; DU; PANG; XU, 2020; WANG et al., 2020; ZHANG et al., 2020; Wang; Zeng, 2018)
Português	4	(LOPES; TEIXEIRA; OLIVEIRA, 2020; SCHNEIDER et al., 2020; LOPES; TEIXEIRA; OLIVEIRA, 2019; LAMY et al., 2018a)
Finlandês	3	(MOEN et al., 2020, 2019, 2018)
Russo	2	(GAVRILOV et al., 2020; BALABAEVA; KOVALCHUK, 2020)
Francês	2	(JOUFFROY et al., 2020; LERNER; PARIS; TANNIER, 2020)
Espanhol e Sueco	2	(WEEGAR et al., 2018; PÉREZ et al., 2017)
Italiano	1	(VIANI et al., 2018)
Espanhol	1	(BLANCO et al., 2019)
Coreano	1	(LEE et al., 2018)
Búlgaro	1	(ZHAO, 2019)
Germânico	1	(AMIN et al., 2019)

Tabela 12 – Bases de dados pública.

Base de dados	Qtd	Referências	Descrição
MIMIC-III	6	(KORMILITZIN et al., 2020; DAI; SU; WU, 2019; KRALJEVIC et al., 2020; ALODADI; JANEJA, 2019; NUTHAKKI et al., 2019; DANDALA et al., 2020)	O MIMIC é uma base de dados aberta desenvolvido pelo Laboratório de Fisiologia Computacional do MIT, contendo dados de saúde anônimos associados a aproximadamente 60.000 internações em unidades de terapia intensiva. Esses incluem dados demográficos, sinais vitais, relatórios laboratoriais, medicamentos e muito mais.
i2b2/n2c2	5	(ZHANG et al., 2019a; LUO et al., 2018; LUO, 2017; LI et al., 2019c; JI; CHEN; JIANG, 2020)	I2b2(<i>Informatics for Integrating Biology & the Bedside</i>) é uma organização que patrocinou uma série de desafios em aprendizado de máquina com foco em texto de 2006 a 2018, usando essa base como referência para eles. Atualmente, está armazenado no Departamento de Informática Biomédica (DBMI) da <i>Harvard Medical School</i> , sendo posteriormente renomeado para n2c2: <i>National NLP Clinical Challenges</i> .
CCKS 2017	3	(LI et al., 2020b; LU et al., 2019; ZHANG et al., 2018a)	Base de dados publicado para a <i>Chine Conference on Knowledge Graph and Semantic Computing</i> 2017.
CLEF-2014	1	(YADAV et al., 2020b)	Base de dados publicada para a <i>Conference and Labs of the Evaluation Forum</i> 2014.
OHNL2018-FH	1	(SHI et al., 2019)	Base de dados publicada para a tarefa de extração de história da família para desafio <i>BioCreative/OHNL2018</i> de 2018.
SemClinBr	1	(SCHNEIDER et al., 2020)	SemClinBr é semanticamente anotado para tarefas clínicas de PLN em português (Brasil) para fornecer um recurso comum de desenvolvimento e avaliação para pesquisadores biomédicos de PLN, sendo também o primeiro corpus clínico disponível em português (Brasil).

com grupos de dados, acabam sendo muito pequenas, o que dificulta a compreensão semântica do modelo.

No campo linguístico e semântico, problemas simples de escrita natural, como erros ortográficos ou numéricos (GAVRILOV et al., 2020; KORMILITZIN et al., 2020) ou ambiguidades de expressão (YADAV et al., 2020b) afetam a assertividade das ferramentas. PETERSON; LIU(PETERSON; LIU, 2020) identificou que as relações conceituais para a descrição de problemas clínicos no mundo real são muito mais complexas do que as propostas pelo modelo SNOMED CT, e devido a esse fato, a construção do padrão ouro é complexa, pois pode haver mais de uma forma sintaticamente correta para representar as mesmas expressões conceituais.

Os estudos que operam sobre línguas não inglesas apresentam desafios maiores, principalmente pela falta de ferramentas e estudos prévios e pela escassez ou inexistência de dados rotulados para línguas nativas. Além disso, embora existam muito mais bases de dados disponíveis em inglês, ainda é um desafio mesmo nesse idioma. Como exemplo, BAMPA; DALIANIS(BAMPA; DALIANIS, 2020) afirma que a língua sueca é diferente do inglês, então as técnicas de PLN usadas também são diferentes. A língua sueca é uma língua altamente flexível e composta, semelhante ao alemão. CAI; ZHOU; ZHENG(CAI; ZHOU; ZHENG, 2020), CHENG et al.(CHENG et al., 2019) e LI et al.(LI et al., 2020a) mencionam que as estruturas gramaticais e ortográficas do chinês são mais complexas que o inglês, o que pode causar segmentação de palavras, principalmente, na tokenização, o que dificulta o reconhecimento da entidade. LOPES; TEIXEIRA; OLIVEIRA(LOPES; TEIXEIRA; OLIVEIRA, 2019), ZHANG et al.(ZHANG et al., 2019a), ZHANG et al.(ZHANG et al., 2018a) e ZHAO(ZHAO, 2019) afirmam que a maioria das pesquisas na área é voltada para a língua inglesa, sendo escassa ou inexistente em outras línguas. LAMY et al.(LAMY et al., 2018b) nos lembra que já existem muitos sistemas de PLN capazes de extrair informações clínicas de documentos médicos, mas grande parte deles funciona apenas para o idioma inglês. BALABAEVA; KOVALCHUK(BALABAEVA; KOVALCHUK, 2020) aponta a dificuldade em encontrar um analisador de árvore sintática para o russo; JOUFFROY et al.(JOUFFROY et al., 2020),AMIN et al.(AMIN et al., 2019),VIANI et al.(VIANI et al., 2018) e SCHNEIDER et al.(SCHNEIDER et al., 2020) observam que o processamento de linguagem natural na área médica em seu idioma é um desafio, principalmente porque não há grandes corpora médicos em seus idiomas e a maioria deles está em inglês.

Na estruturação das informações, poucos estudos processam todas as informações do prontuário. A maioria deles se concentra em extrair um grupo específico de informações, como sinais vitais, sintomas, drogas e suas relações, deixando outras informações relevantes desses grupos desestruturadas. Dos estudos que estruturam toda a informação, a maioria estrutura apenas elementos específicos do prontuário, como lista de problemas, laudos ou grupos de especialidades. Apenas um estudo realizou essa estruturação de forma completa e genérica, mas sua avaliação foi feita com 150 registros de saúde com um pequeno grupo de especialidades.

As redes BERT e o mecanismo de atenção não apresentaram grande adesão; apenas 3 estudos usaram redes BERT em comparação com Bi-LSTM-CRF, que ainda domina com 11 estudos.

Embora menos estudos as utilizem, as redes BERT apresentam um resultado melhor que o BI-LSTM-CRF para as mesmas tarefas nos estudos analisados. Nas BERTs, embora seja uma rede recente, o pré-treinamento com um grande corpus de textos genéricos não rotulados ou de um domínio específico como biomédico antes da especialização para uma tarefa pode transformá-lo na ferramenta ideal para superar a tarefa atual da PLN no domínio médico. Apenas um estudo apresenta redes BERT em língua não inglesa, a BioBERT_{pt}, rede BERT pré-treinada para o domínio biomédico em português. Embora vários estudos (BLINOV et al., 2020; FENG; SHAIB; RUDZICZ, 2020; SINGH et al., 2020) que utilizam redes BERT para diagnóstico ou previsão na área médica a partir de textos não estruturados tenham sido identificados durante a busca na literatura, eles estavam fora do escopo deste estudo e foram removidos durante o processo de filtragem.

3.4 Conclusões Parciais

Esse capítulo apresentou uma revisão sistemática para encontrar o estado da arte da estruturação de informações contidas em registros de saúde não estruturados. Procurou responder a questões-chave para propor novos estudos na área, levantar questões de pesquisa relevantes, respondê-las e fazer uma análise transversal dos achados.

A maioria dos estudos concentra-se no reconhecimento de entidades na área médica. Embora bem estabelecidos na língua inglesa, novos estudos buscam alcançar um novo estado da arte usando novas abordagens ou resolver problemas específicos que as soluções atuais não cobrem satisfatoriamente. No entanto, em outras línguas, os estudos ainda não atingiram esse nível de maturidade e ainda enfrentam grandes desafios que têm despertado o interesse da comunidade científica, como ilustra o número de estudos encontrados em outras línguas, principalmente o chinês.

As redes neurais apresentam-se como uma solução robusta, não só para estruturação mas também para reconhecimento de entidades. Embora existam dificuldades apresentadas para a aquisição de dados rotulados, essas redes neurais têm se mostrado uma boa solução para adaptar as características complexas do processamento de linguagem natural. Embora as redes BERT tenham apresentado melhor desempenho nos estudos que as utilizaram, sua utilização foi menor do que o esperado pelos autores do estudo.

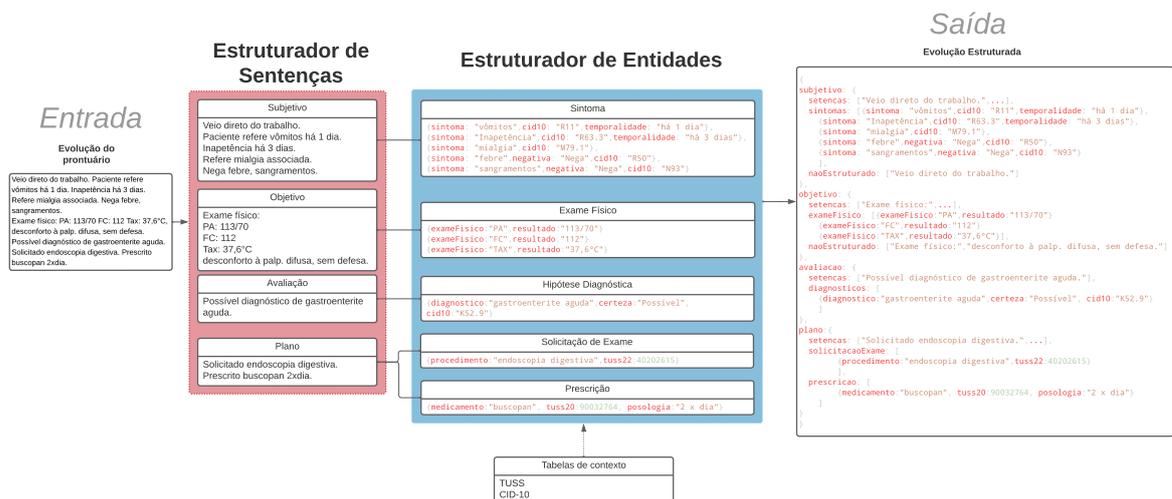
As principais contribuições da revisão são listadas a seguir: revisão sistemática do estado da arte da estruturação de prontuários eletrônicos; uma taxonomia para a classificação das formas de estruturação das informações não estruturadas do prontuário eletrônico; identificação de tendências na utilização de ferramentas, técnicas de estruturação e algoritmos; um inventário para entender como é feita a aquisição de dados para o desenvolvimento e avaliação; e um mapeamento dos desafios na estruturação do prontuário eletrônico.

4 MODELO PROPOSTO

Este capítulo apresenta um detalhamento do modelo proposto, detalhando seus componentes e subcomponentes, com ênfase na funcionalidade de cada componente, suas entradas de dados e saídas esperadas. Inicialmente, é fornecida uma visão geral do modelo; nas seções seguintes, cada componente é descrito detalhadamente.

4.1 Visão Geral

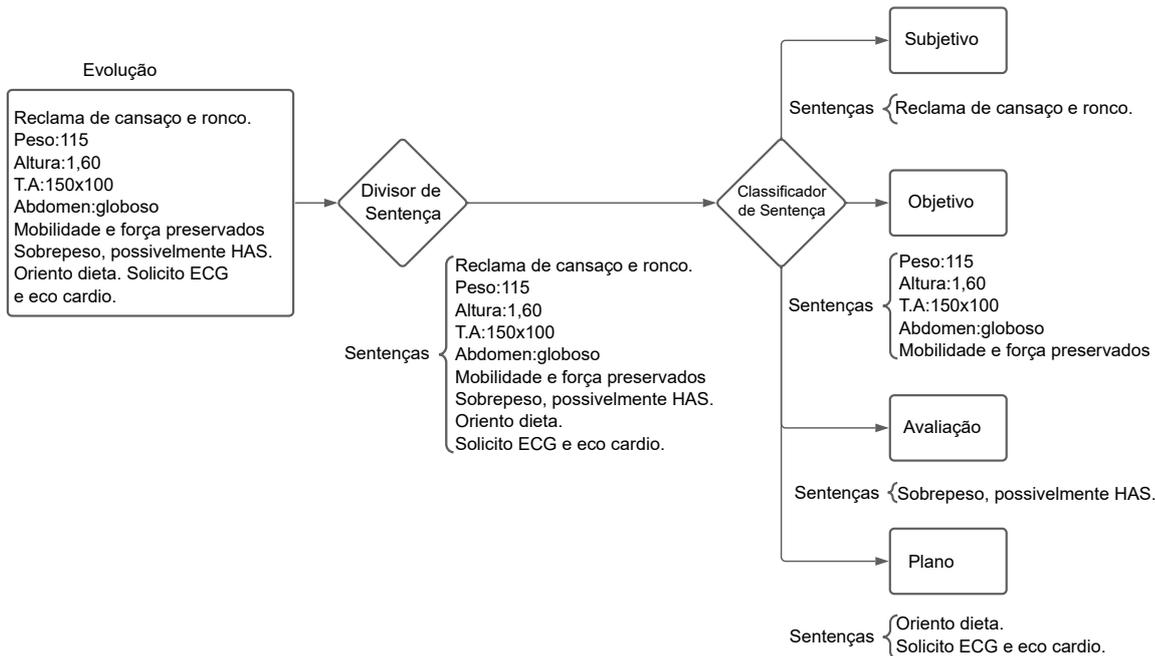
Figura 7 – Visão geral do modelo proposto com suas entradas e saídas.



A Figura 7 apresenta a visão geral do modelo proposto, sendo um pipeline que tem como entrada a evolução do paciente em texto e como saída as sentenças da evolução classificadas no padrão SOAP e quando possível estruturadas no nível de suas entidades.

O processamento do pipeline é formado por dois componentes principais: o Estruturador de Sentença e o Estruturador de Entidade. O Estruturador de Sentença recebe a evolução do paciente e a divide em sentenças, classifica-as seguindo o padrão SOAP. O Estruturador de Entidades recebe essas sentenças classificadas, busca as entidades relevantes, relaciona essas entidades, e vincula seus termos às tabelas de terminologia médica. O resultado final do modelo é a evolução estruturada em formato de árvore, onde os nodos mais próximos da raiz representam a estruturação de mais alto nível e as folhas representam as de mais baixo nível. Desta forma, nenhuma informação é perdida no processo. O processo de estruturação é baseado no melhor esforço, onde o modelo vai tentar estruturar a sentença no nível mais baixo possível.

Figura 8 – Modelo proposto para o componente Estruturador de Sentenças.



4.2 Estruturador de Sentenças

Este componente visa estabelecer um primeiro nível de estruturação da evolução. Ele divide a evolução do paciente em uma lista de sentenças e classifica cada uma delas com uma classe do padrão SOAP, garantindo assim, que todas as sentenças tenham somente uma classe e nenhuma sentença fique sem uma classificação. Dessa forma, o componente foi modelado como uma tarefa de classificação de sentenças do PLN, mais especificamente, uma tarefa de classificação de sentenças multi-classe de rótulo único, sendo dividido em dois subcomponentes: Divisor de Sentença e Classificador de Sentença. A Figura 8 mostra o modelo proposto para o componente.

O componente Divisor de Sentença tem como objetivo identificar os limites das sentenças e dividir o texto em uma lista de sentenças. A detecção de limite de sentença (*sentence boundaries*) ou desambiguação de limite de sentença (*sentence boundary disambiguation*) também é um subproblema do PLN. Esse problema consiste em identificar os limites das frases em um texto, o que inclui a pontuação, reduzindo-se efetivamente à decisão sobre quais dos símbolos que potencialmente indicam limites de frases (., !, ?) são de fato delimitadores. Por exemplo, na frase "Dr. Jones palestra na U.C.L.A.", apenas o ponto final denota o fim de uma frase (STEVENSON; GAIZAUSKAS, 2000).

O Classificador de Sentença recebe cada uma destas sentenças e classifica com uma das quatro classes do padrão SOAP. A escolha do padrão SOAP leva em conta uma série de fatores. Além de ser um modelo internacional, bem documentado e consolidado no meio médico, ele é utilizado por diversos outros profissionais na área da saúde. Suas classes são uma representação de alta ordem bem definida da evolução do paciente, o que reduz a ambiguidade e traz contexto

às sentenças quando compartilhadas com outros profissionais de saúde ou em consultas futuras.

A decisão de design pela rotulação única em contrapartida a um modelo multi-rótulo, onde a sentença poderia receber mais de um rótulo da classe SOAP, se deve à própria natureza do padrão SOAP aplicado na área médica. Quando o profissional de saúde preenche uma evolução estruturada no padrão, há um campo para cada classe, não sendo possível essa múltipla rotulação. Como contraponto, MOWERY et al. (MOWERY et al., 2012) usa uma abordagem multi-rótulo para modelar o problema, mas apenas 4,0% das sentenças da base de dados possuem dois rótulos e somente 0,02% possuem três. Levando em conta essa baixa incidência, optou-se por manter a abordagem de rotulação única.

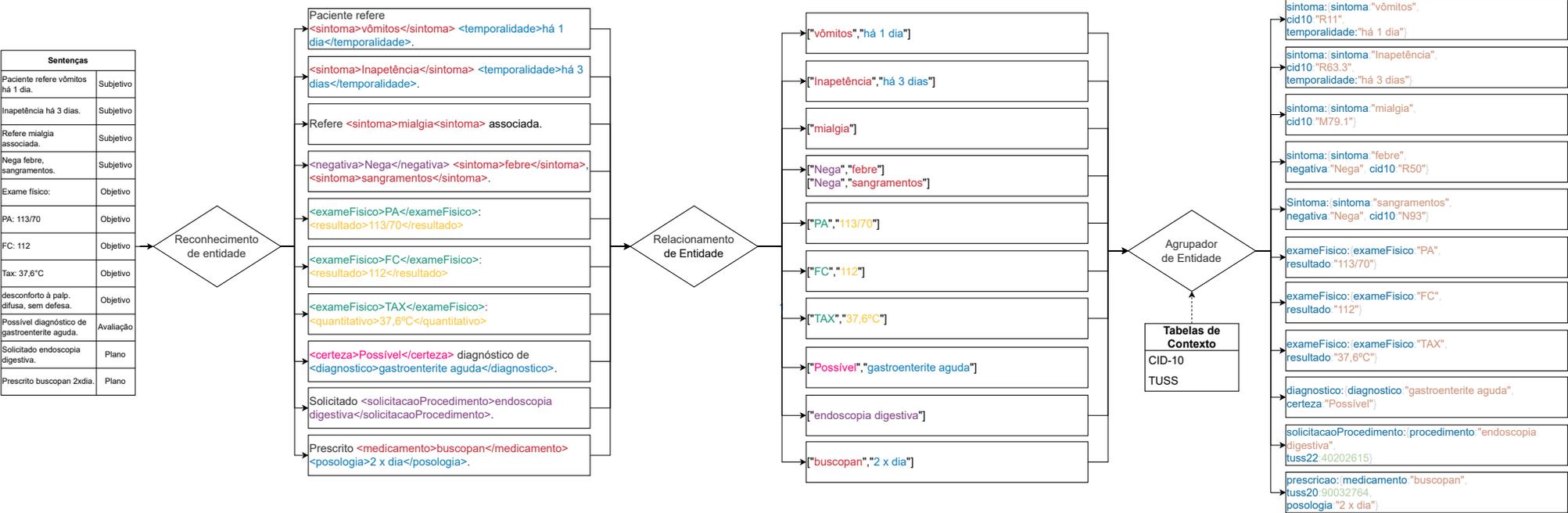
4.3 Estruturador de Entidades

O segundo nível faz uma estruturação mais complexa. A Figura 9 mostra o modelo proposto, e é dividida em 3 subcomponentes. No primeiro subcomponente, é feito o reconhecimento das entidades relevantes nas sentenças que serão utilizadas nas próximas fases do processamento. No segundo subcomponente é feito o relacionamento dessas entidades, onde são formadas várias tuplas de interesse a partir das entidades identificadas. Por fim, no último subcomponente, são montados os grupos de estruturas relevantes das entidades, com base nas relações e contextos das sentenças. O modelo não determina de forma rígida quais grupos de estrutura devem existir e quais suas entidades, sendo que o mesmo deve ser determinado a partir das suas necessidades de aplicação. Levando isso em conta, nesse trabalho são propostos 12 grupos de entidades, que foram baseados no problema prático e na base de dados disponível para teste e avaliação. Duas entidades comuns na maioria dos grupos de entidade, são Negativa e Observação, a Negativa serve para identificar a negativa ou inexistência do evento ou fato, por exemplo: 'Nega problema cardíaco na família.' A entidade negativa 'Nega' representa essa inexistência, já a entidade Observação representa uma informação semi-estruturada que é relevante para o grupo de entidade mas acaba não se encaixando em nenhuma outra entidade relevante. Por exemplo, o sintoma 'Dor na perna esquerda quando corre.' A entidade 'quando corre' faz parte e é relevante para o grupo de entidade Sintoma, mas não se enquadra em nenhuma de suas entidades. As entidades: Código TUSS 20, Código TUSS 22 e CID-10, não são extraídas diretamente da sentença, mas sim da vinculação das tabelas de contexto.

A Tabela 13 apresenta quatro grupos de entidades associadas à classe Subjetivo, cada contendo suas entidades, uma sentença de exemplo e o resultado de sua estruturação. As informações desse grupo, por se tratar da classe SOAP subjetiva advém de informação fornecidas pelo próprio paciente. Abaixo é feita a descrição de cada um deles:

- **Histórico Familiar:** Representa a identificação de doenças hereditárias ou condições prevalentes na família do paciente. Este grupo abrange entidades como doença específica, relação familiar, e a codificação da doença conforme o CID-10;

Figura 9 – Modelo para estruturação de entidade.



- **Doença pré-existente:** Grupo de entidades que representam as informações fornecidas pelo paciente sobre sua condição de saúde preexistente, detalhando a doença, a parte do corpo afetada, a duração da condição e o código CID-10;
- **Medicação em uso:** Contém a informação farmacológico de uso atual ou anterior fornecida pelo paciente, incluindo o nome do medicamento, posologia, período de uso e o código TUSS correspondente ao medicamento;
- **Sintoma:** Registra as manifestações sintomáticas relatadas pelo paciente, abrangendo o tipo de sintoma, localização, duração e o código CID-10.

Tabela 13 – Grupos de entidades da classe Subjetivo.

Grupo	Entidades	Sentença Exemplo	Estruturação
Histórico Familiar	Doença Familiar Negativa Observação CID-10	Mãe tem histórico de diabetes.	Doença: diabetes Familiar: Mãe CID-10: E14
Doença pré-existente	Doença Temporalidade Parte do corpo Negativa Observação CID-10	Artrose no joelho esquerdo há cinco anos.	Doença: Artrose Temporalidade: cinco anos Parte do corpo: joelho esquerdo CID-10: M19
Medição em uso	Medicamento Posologia Parte do corpo Temporalidade Negativa Observação Código TUSS 20	Refere que faz uso de Carbamazepina 200 mg/dia (7 anos).	Medicamento: Carbamazepina Posologia: 200 mg/dia Temporalidade: 7 anos Código TUSS 20: 90348346
Sintoma	Sintoma Parte do corpo Temporalidade Negativa Observação Cid-10	Dor na nuca intensa e constante nas últimas 24 horas.	Sintoma: Dor Parte do corpo: nuca Temporalidade: últimas 24 horas Observação: intensa e constante Cid-10: R52

A Tabela 14 apresenta os três grupos de entidades associadas à classe Objetivo, cada um contendo suas entidades, uma sentença de exemplo e o resultado de sua estruturação. A seguir, é feita a descrição de cada um deles:

- **Medição:** Representam as medições feitas pelo profissional de saúde no processo de atendimento, incluindo medidas físicas, como altura e peso, ou sinais vitais, incluindo temperatura, frequência cardíaca, frequência respiratória e pressão arterial. Sendo formado, principalmente pelas entidades de nome da medição e resultado;

- **Exame físico:** Inclui procedimentos ou avaliações realizados no momento da consulta, como verificação de parte do corpo à procura de problemas ou outras características. As entidades principais são: nome do exame físico, parte do corpo e resultado. Nem sempre o exame físico apresenta um nome, muitas vezes é descrito somente a parte do corpo examinado e o resultado, por exemplo: 'Olho direito: edema das pálpebras, conjuntiva', 'abdômen: globoso' e 'membros inferiores sem varizes';
- **Resultado de Exame:** Representam resultados de exames clínicos realizados fora do atendimento, normalmente trazidos pelo paciente para avaliação e diagnóstico do profissional de saúde. Tem como principais entidades: nome do exame, parte do corpo, resultado e data em que foi realizado.

Tabela 14 – Grupos de entidades da classe Objetivo.

Grupo	Entidades	Sentença Exemplo	Estruturação
Medição	Nome da medição Resultado Observação	Peso:115kg.	Nome da medição: Peso Resultado: 115kg
Exame físico	Nome do exame físico Parte do corpo Resultado Observação Negativa	oroscopia: amígdala esquerda com edema.	Nome do exame físico: oroscopia Parte do corpo: amígdala esquerda Resultado: com edema
Resultado de Exame	Nome do exame Parte do corpo Resultado Data Observação Negativa Código TUSS 22	glicemia de jejum (02/05/19): 126 (106).	Nome do exame: glicemia de jejum Resultado: 126 (106) Data: 02/05/19 Código TUSS 22: 40302040

A Tabela 15 apresenta a Hipótese Diagnóstica, que representa o grupo de entidades proposto para a estruturação da classe Avaliação, que é a postulação do profissional de saúde sobre uma possível condição médica que pode estar afetando o paciente. As principais entidades incluem: o diagnóstico, o grau de certeza, a parte do corpo e a temporalidade.

A Tabela 16 descreve os quatro grupos da classe Plano, que detalham as propostas de resolução para o atendimento. Abaixo, cada grupo é descrito detalhadamente:

- **Solicitação de Procedimento:** Representa a requisição de um procedimento médico, que abrange uma série de tipos de procedimentos, como clínicos, cirúrgicos, diagnósticos e terapêuticos. As principais entidades incluem: nome do procedimento, parte do corpo e data;
- **Encaminhamento:** O grupo representa o direcionamento de um paciente a um especialista ou a um serviço médico específico. A principal entidade que compõe este grupo é a

Tabela 15 – Grupos de entidades da classe Avaliação.

Grupo	Entidades	Sentença Exemplo	Estruturação
Hipótese diagnóstica	Diagnóstico Grau de certeza Parte do corpo Observação Temporalidade Negativa CID-10	Provavelmente angina por esforço.	Diagnóstico: angina Grau de certeza: Provavelmente Observação: por esforço CID-10: I20

especialidade, que identifica o tipo de especialista ou departamento ao qual o paciente deve ser encaminhado;

- **Prescrição:** Este grupo aborda a prescrição de medicamentos feita durante o atendimento. As principais entidades incluem: medicamento, posologia e parte do corpo;
- **Atestado:** Representa a emissão de atestados feita no atendimento médico. A principal entidade é a temporalidade, que define a duração do período de afastamento.

Tabela 16 – Grupos de entidades da classe Plano.

Grupo	Entidades	Sentença Exemplo	Estruturação
Solicitação de Procedimento	Nome do Procedimento Parte do corpo Observação Data	Refazer ressonância magnética do joelho direito daqui há 3 semanas.	Nome do Procedimento: Ressonância magnética Parte do corpo: joelho direito Data: daqui há 3 semanas Observação: Refazer
Encaminhamento	Especialidade Observação	Encaminhamento para ortopedista, verificar joelho.	Especialidade: ortopedista Observação: verificar joelho
Prescrição	Medicamento Posologia Parte do corpo Observação Negativa	Amoxicilina, 500mg três vezes ao dia.	Medicamento: Amoxicilina Posologia: 500mg três vezes ao dia
Atestado	Temporalidade Observação	Atestado de três dias por gripe.	Temporalidade: três dias Observação: por gripe

4.3.1 Reconhecimento de Entidade

Esse subcomponente faz o primeiro processamento da sentença e tem como objetivo identificar e anotar as entidades relevantes na sentença. Cada início e término de frase ou fragmento de texto relevante para o modelo é marcado na sentença e ganha um rótulo que represente sua classificação como entidade. As entidades do reconhecimento de entidade são diretamente

relacionadas aos grupos de entidades previstas no modelo. Abaixo são descritas as entidades previstas para este componente:

- **Doença:** Condição particular anormal, que possui sintomas específicos e causas conhecidas, que afeta negativamente o organismo e a estrutura ou função de parte ou de todo um organismo. Exemplo: diabetes, câncer e alzheimer;
- **Familiar:** Refere-se a qualquer membro da família com vínculo biológico ou social do paciente. Exemplo: mãe, pai, marido;
- **Negativa:** Usada para indicar a ausência de uma condição, sintoma ou resultado específico. Exemplo: nega, não apresenta, sem indicativo e descarto;
- **Observação:** Parte semi-estruturada de um grupo de entidades que é relevante mas não enquadra em nenhuma das demais entidades do grupo. Exemplos: Ansiedade - *come sem fome* e encaminhado para cardiologista *caso não melhore com a medicação*;
- **Temporalidade:** Refere-se ao aspecto temporal associado aos eventos, como a duração dos sintomas ou tempo de uso de medicamento. Exemplos: há 5 meses, desde que nasceu, quando jovem;
- **Parte do corpo:** Especifica uma região ou órgão do corpo, além do membro pode indicar a região. Exemplo: braço esquerdo, região lombar e ponta dos dedos;
- **Medicamento:** Nomenclatura de medicamento. Exemplo: Buscapan, Alegria D e dorflex;
- **Posologia:** Instruções sobre como administrar medicamentos, incluindo dosagem, frequência e duração do tratamento. Exemplos: 02 cp 12/12h, 75mg-100 mg dia e 1 cp quando dor;
- **Sintoma:** Queixas do paciente em relação ao que ele está sentindo ou manifestações notadas na pessoa. Exemplo: tosse, rouquidão e tremor;
- **Nome da medição:** O nome específico da medição feita pelo profissional de saúde no processo de atendimento. Exemplos: altura, frequência respiratória e pressão arterial;
- **Resultado:** Os dados quantitativos ou qualitativos obtidos a partir de exames médicos ou testes laboratoriais. Exemplos: 10.30, não reagente e acima do padrão;
- **Nome do exame físico:** Refere-se a nomenclatura de um exame físico. Exemplo: romberg, rinoscopia e otoscopia;
- **Nome do exame:** O título específico de um teste ou procedimento diagnóstico. exemplo: ácido úrico, HDL e Mapa;

- **Data:** A data e/ou hora, ou estimativa do evento médicos. Exemplos: em três semanas, ago/20 e 10/01/19;
- **Diagnóstico:** A determinação sobre a natureza e a causa de uma doença do paciente. Exemplo: carcinoma, lúpus e transtorno depressivo;
- **Grau de certeza:** Expressa o nível de confiança sobre a precisão do diagnóstico. Exemplo: provável, possível e certamente;
- **Nome do Procedimento:** Nomenclatura de procedimento ou exame médico. Exemplos: Raio X, audiometria, biópsia óssea ECG e cirurgia de polidactilia;
- **Especialidade:** Nomenclatura de especialidade médica ou serviço médico específico. Exemplos: Fisioterapeuta, ginecologista, pediatra ou pronto socorro.

Caso não seja reconhecida nenhuma entidade na sentença, baseado na política do melhor esforço, o modelo considera que a sentença atingiu o grau máximo de estruturação da informação. Sendo considerado, para essa sentença, o nível máximo de estruturação a classificação SOAP do componente de estruturação de sentença.

4.3.2 Relacionamento de Entidade

O objetivo deste subcomponente é determinar se duas entidades, identificadas no subcomponente anterior, pertencem ao mesmo grupo de entidades. Para isso, propõe-se modelar o componente como uma tarefa de extração de relações de entidades de forma binária (sim ou não), com o intuito de determinar se duas entidades compartilham o mesmo grupo. A tarefa de extração de relações visa identificar relações semânticas entre as entidades. Formalmente, dada uma sentença s contendo duas entidades anotadas e_1 e e_2 , o objetivo é prever as relações entre e_1 e e_2 , formando uma tupla que inclui ambas as entidades e a relação entre elas (FENG et al., 2017). Normalmente, a extração de relações é modelada como uma tarefa multiclasse, onde, além de determinar a relação entre duas entidades, essa relação é qualificada. Por exemplo, uma relação pode indicar que uma entidade é parte de outra, está localizada em outra, ou pertence a outra. No entanto, como no modelo proposto a relação das entidades já está determinada pela classificação SOAP e pela estrutura do seu grupo de entidades, a qualificação torna-se desnecessária. Portanto, o principal objetivo desta extração de relações é decidir se essas duas entidades compartilham o mesmo grupo de entidades, resultando numa decisão binária de pertencimento (sim ou não).

4.3.3 Agrupador de Entidade

Para realizar a última parte do processamento, é necessário associar as tuplas produzidas no subcomponente anterior aos grupos de entidades propostos. Desta forma, o componente pode

ser mapeado como um problema de correspondência de ontologias de informação estrutural. As tuplas são uma representação da informação semântica da ontologia, incluindo relações como *parte de*, *é-um*, *subclasse* e *subpropriedade*, e um algoritmo realiza o mapeamento dessas informações nos elementos das ontologias (grupos de entidades) (LIU et al., 2021).

Além disso, nesta etapa, é realizada a vinculação das entidades às terminologias médicas. O campo “cid10” faz a associação ao item da CID-10, o campo “tuss22” relaciona o procedimento à terminologia de procedimentos e eventos em saúde da TUSS (tabela 22) e o campo “tuss20” relaciona o medicamento à terminologia de medicamentos (tabela 20).

O modelo não prevê que todas as sentenças sejam estruturadas nesse nível. Apenas as sentenças que se enquadram em algum grupo de estrutura são estruturadas, sendo possível que algumas sentenças produzam mais de um grupo, por exemplo, uma lista de sintomas.

4.4 Conclusões Parciais

Este capítulo apresentou uma descrição detalhada do modelo proposto, abordando seus componentes e subcomponentes com ênfase na funcionalidade de cada parte, nas entradas de dados necessárias e nas saídas esperadas. Inicialmente, forneceu-se uma visão geral do modelo, destacando a relação entre os componentes principais e suas respectivas funções em cada etapa do pipeline.

O Estruturador de Sentenças realiza a primeira fase de processamento ao dividir a evolução do paciente em sentenças e classificá-las segundo o padrão SOAP. Este componente é composto pelo Divisor de Sentença, que identifica os limites das sentenças, e pelo Classificador de Sentença, que atribui uma classe do padrão SOAP a cada sentença.

O Estruturador de Entidades, por sua vez, realiza um processamento mais complexo ao reconhecer entidades relevantes nas sentenças classificadas, relacioná-las e vinculá-las a termos das tabelas de terminologia médica. Este componente é dividido em três subcomponentes: Reconhecimento de Entidade, Relacionamento de Entidade e Agrupador de Entidade. Juntos, eles formam uma estrutura hierárquica onde os níveis mais altos representam a estruturação global e os níveis mais baixos detalham as entidades específicas.

Os grupos de entidades são definidos conforme a necessidade prática e a base de dados disponível para teste e avaliação. Exemplos de grupos de entidades incluem Histórico Familiar, Doença Pré-existente, Medicação em Uso, Sintoma, Medição, Exame Físico, Resultado de Exame, Hipótese Diagnóstica, Solicitação de Procedimento, Encaminhamento, Prescrição e Atestado. Cada grupo é associado a classes específicas do padrão SOAP, como Subjetivo, Objetivo, Avaliação e Plano, garantindo uma estruturação abrangente e contextualizada da evolução do paciente.

Ao reforçar a conexão entre os componentes e suas respectivas funções, fica clara a importância de cada parte no processo global de estruturação da evolução do paciente, promovendo uma visão integrada e funcional do modelo proposto. O próximo capítulo (Capítulo 5) se concentrará

na implementação desses componentes, detalhando as tecnologias e ferramentas que serão avaliadas, bem como as métricas e a base de dados utilizadas para essa avaliação.

5 MATERIAIS E MÉTODOS

Neste capítulo, são descritos os materiais e métodos utilizados para o desenvolvimento, teste e validação da implementação dos componentes e do pipeline desenvolvido para a validação do modelo. Inicialmente, é apresentada a base de dados, que consiste em prontuários médicos coletados ao longo de um período de dez anos em uma clínica de médio porte. Em seguida, é abordada a implementação do Estruturador de Sentenças, que envolve a utilização de ferramentas para divisão de sentenças e uma rede BERT para a classificação de sentenças no padrão SOAP. Posteriormente, é detalhada a implementação do Estruturador de Entidades, que inclui as técnicas de reconhecimento, relacionamento e agrupamento de entidades para realizar o agrupamento de entidades de interesse. Finalmente, são descritas as métricas de avaliação aplicadas para validar e medir a eficácia da implementação do modelo proposto.

5.1 Base de Dados

A base de dados utilizada neste estudo é privada, fornecida por uma clínica de médio porte com 10.000 prontuários pré-selecionados aleatoriamente entre os pacientes da instituição no período de 18-05-2010 a 14-04-2020. Os mesmos foram entregues com identificação do paciente e do médico já anonimizados. Este processo foi aprovada pelo Comitê de Ética em Pesquisa da Unisinos sob o número CAAE (Certificado de Apresentação para Apreciação Ética) 61086722.0.0000.5344.

Os prontuários são constituídos por dois componentes textuais não estruturados: o prontuário permanente do paciente e uma série de registros denominados anotações clínicas. Cada prontuário médico possui apenas um registro permanente que é editado por vários médicos, sem histórico ou data, contendo, principalmente, informações de pouca relevância ou duplicadas das anotações médicas. Por outro lado, cada evolução clínica representa um único atendimento clínico realizado por um dos 83 médicos credenciados em uma das dez especialidades disponibilizadas pela instituição. A Tabela 17 mostra a divisão das evoluções por especialidade. A base de dados possui 234.673 evoluções, com média de 23,4673 por paciente. Para este estudo, optou-se por trabalhar apenas com as evoluções dos pacientes, por conter uma grande quantidade de informações e maior relevância.

Os prontuários permanentes são formados por:

- Código do paciente: identificação numérica do paciente;
- Observação permanente: texto de observação genérica.

As evoluções são formadas por:

- Código: identificador sequencial de lançamento no prontuário;
- Data da evolução: data do atendimento;

Tabela 17 – Distribuição das evoluções dos pacientes entre as especialidades médicas da base de dados utilizada no trabalho.

Especialidade	Evoluções
Clínica Geral	133973
Ginecologia	20848
Dermatologia	14453
Psiquiatria	12707
Oftalmologia	12530
Cardiologia	12508
Nutricionista	10468
Urologia	6570
Psicologia	5633
Endocrinologia	4983

- Evolução: texto contendo a evolução do paciente no atendimento;
- Código do médico: identificação numérica do médico que realizou o atendimento;
- Especialidade: especialidade do médico;
- Código do paciente: identificação numérica do paciente.

As identificações numéricas do paciente e do médico têm o objetivo de garantir o agrupamento de múltiplas evoluções/prontuários pertencentes a um mesmo paciente ou médico, não sendo possível a engenharia reversa da identificação. As evoluções foram realizadas por 83 médicos diferentes de 10 especialidades diferentes. A Tabela 17 apresenta essa distribuição.

5.2 Implementação do Estruturador de Sentenças

Esse componente é modelado como uma tarefa de classificação de sentenças multi-classe de rótulo único, formado por dois subcomponentes: divisor de sentenças e classificador de sentenças. Para implementação do divisor de sentença é proposta a utilização de uma ferramenta já desenvolvida para esse propósito que opere na língua portuguesa. Entre elas se destacam:

- **iSentenizer**: Modelo de detecção de limite de sentença multilíngue (WONG; CHAO; ZENG, 2014);
- **PySBD**: Desambiguador pragmática de limite de sentença (SADVILKAR; NEUMANN, 2020);
- **Syntok**: Segmentador de frases e tokenizador de palavras¹;
- **Stanza**: Coleção de ferramentas para análise e processamento linguístico².

¹<https://github.com/fnl/syntok>

²<https://stanfordnlp.github.io/stanza/performance.html>

Entretanto, para a classificação de sentença no padrão SOAP, na revisão bibliográfica, não foi encontrada nenhuma ferramenta que a faça em língua portuguesa. Existe uma gama de ferramentas, propostas e modelos para tratar a classificação de sentença na área médica (KIM et al., 2020), porém a maioria deles está focada na língua inglesa. Além disso, trabalhos que operam fora da língua inglesa apresentam desafios mais significativos, justamente por essa falta de ferramentas e trabalhos anteriores (OLIVEIRA; COSTA; ANTUNES, 2021). Levando isso em consideração, foi escolhido um modelo BERT por, atualmente, representar o estado da arte na maioria dos desafios de PLN (DEVLIN et al., 2019; QIN et al., 2021), incluindo tarefas na área da medicina (ALSENTZER et al., 2019). O modelo tem ainda a vantagem de já incluir no seu modelo de rede pré-treinado *word embeddings* e possuir modelos pré-treinados na língua portuguesa, incluindo um modelo de apoio a tarefas de PLN clínica e biomédica. Desta forma, foram escolhidos quatro modelos de redes públicas BERT pré-treinadas para avaliar qual apresenta o melhor resultado para a solução do problema. A seguir são descritos cada um deles e o motivo para sua escolha.

- **Multilingual BERT:** Versão genérica para múltiplos propósitos da BERT, será usada como referência de comparação para as demais redes;
- **DISTILBERT (base-pt):** Versão menor e mais rápida da BERT genérica, porém possui uma base de conhecimento menor. Caso tenha um desempenho similar às demais, se torna mais vantajosa;
- **BioBERT:** Versão especializada para a representação de conceitos biomédicos em inglês. Tem como objetivo analisar o que traz mais ganho, o conhecimento do idioma ou domínio do campo biomédico mesmo em outro idioma;
- **BioBERTpt:** Versão especializada para a representação de conceitos biomédicos em português. É esperado que apresente o melhor desempenho, visto que possui tanto o domínio do campo biomédico quando o idioma adequado.

Além das quatro redes públicas pré-treinadas, uma quinta rede foi avaliada. Observando-se a melhoria no desempenho do BioBERTpt em relação ao mBERT e ao BioBERT, sugere-se que personalizar o pré-treinamento para um domínio biomédico específico, como um determinado grupo médico ou uma rede hospitalar, pode levar a uma performance ainda superior. Com o objetivo de testar essa hipótese, foi feito o processo de retreinamento da BioBERTpt com todas as 2.979.987 sentenças disponíveis na base de dados. O modelo resultante desse re-treinamento foi nomeado **BioBERTptRT**.

5.3 Implementação do Estruturador de Entidades

Para a implementação dos subcomponentes do Estruturador de Entidades são propostas algumas abordagens e técnicas que serão descritas nas próximas subseções.

5.3.1 Reconhecimento de Entidades

Existem propostas de ferramentas desenvolvidas para o reconhecimento de entidades em tarefas médicas, como as propostas por ROCHA et al.(ROCHA et al., 2022), SILVA(SILVA, 2023) e SCHNEIDER et al.(SCHNEIDER et al., 2020). No entanto, torna-se inviável a sua utilização, pois utilizam listas de entidades reconhecidas diferentes das propostas neste trabalho. Desta forma, é necessária o treino de uma nova rede para essa tarefa. Uma das vantagens das redes BERT pré-treinadas é a possibilidade de re-treiná-las para outras tarefas de PLN, incluindo o reconhecimento de entidades. Assim, propõe-se o re-treinamento da rede que apresentou o melhor desempenho na classificação de sentenças para a tarefa de reconhecimento de entidades.

5.3.2 Relacionamento de Entidades

Redes Neurais Convolucionais (CNNs) e Redes Neurais Recorrentes (RNNs) são abordagens comumente usadas para a tarefa de extração de relações de entidade (LI et al., 2019a). Contudo, têm como contraponto a necessidade de *word embeddings* adequados para o domínio médico na língua portuguesa (SCHNEIDER et al., 2020). Levando isso em conta, ainda que essa tarefa não seja uma das tarefas tratadas pelas redes BERT, este problema pode ser modelado como um problema de classificação de sentenças (RINK; HARABAGIU; ROBERTS, 2011; KIM; CHOE; MUELLER, 2015). Utilizando o *token* especial ‘SEP’ da BERT, que permite a separação de contexto da entrada, a rede teria como entrada duas entidades e_1 e e_2 , a sentença com todas as entidades anotadas e, como saída, uma classificação que representa o relacionamento das duas entidades.

Dessa forma, também é previsto o retreino da mesma rede utilizada nas duas tarefas anteriores para essa tarefa, seguindo essas adaptações. Caso o desempenho não seja adequado para a tarefa, serão avaliadas outras propostas apresentadas.

5.3.3 Agrupamento de Entidades

Inicialmente, é previsto a utilização de algoritmos baseados em regras para fazer esse mapeamento, visto que os grupos de entidades são limitados, as entidades já estão mapeadas e a classificação da sentença reduz consideravelmente a ambiguidade entre os grupos. Caso essa abordagem apresente um desempenho insatisfatório, serão abordadas outras técnicas para solucionar o problema, como algoritmos meméticos (*Memetic algorithms*) (XUE; WANG, 2015), hiper-grafo (ZHU; HUA; GAO, 2020) e agregação de semelhança (HOUSHMAND; KHORRAM, 2011).

Para a associação das entidades de procedimentos, doenças/sintomas e medicamentos é proposta a utilização da distância de Levenshtein (YUJIAN; BO, 2007) ou tabela de tradução.

5.4 Métricas de Avaliação

Para fazer as avaliações do modelo proposto, serão utilizadas métricas consolidadas para avaliação de modelos de PLN na área da saúde, sendo elas: acurácia (*accuracy*), precisão (*precision*), revocação (*recall*) e medida-F (*F1-score*) (MUJTABA et al., 2019; KOLECK et al., 2019; KERSLOOT et al., 2020). A Tabela 18 descreve cada uma dessas métricas.

Para avaliar individualmente os componentes do modelo que requerem treinamento, será aplicada a validação cruzada k-fold (ZHANG; LIU, 2023) com $k = 5$, sempre que possível, de forma estratificada. Nesta abordagem, a base de dados é dividida em cinco partes iguais, garantindo-se que a proporção de cada classe seja preservada em cada *fold*. Esta precaução é importante, especialmente em bases de dados onde a distribuição das classes não é uniforme, assegurando que cada *fold* seja representativo do conjunto total.

Em cada uma das cinco iterações, quatro dos cinco *folds* são usados para treinar o modelo, enquanto o *fold* restante é usado como conjunto de teste. Esta estratégia garante que cada parte dos dados seja utilizada tanto para treino quanto para teste exatamente uma vez, permitindo uma avaliação abrangente e justa do desempenho do modelo.

A avaliação é realizada calculando as métricas definidas para cada iteração, usando os resultados do modelo no conjunto de teste. Após completar todas as cinco iterações, as métricas de desempenho são agregadas, para fornecer uma visão mais equilibrada do desempenho do modelo.

Um problema comum nas tarefas multi-classe de PLN da área médica é o desbalanceamento de classe (TAFT et al., 2009), onde algumas classes, entidades ou vínculos possuem uma frequência menor que as demais. Essa situação, além de trazer dificuldade para o treino dos modelos, pode afetar as métricas de avaliação (TANHA et al., 2020). É particularmente danoso em modelos onde uma classe é mais relevante que outra, por exemplo, um sistema de predição de doença que possui 5 classes, onde 4 classes representam uma escala de propensão à aquisição da doença, e a 5ª classe representa que o paciente realmente possui a doença. Caso essa quinta classe seja mais relevante que as demais, ao utilizar uma base de dados de avaliação desbalanceado, onde existam somente 5% de itens rotulados nessa classe, o modelo pode errar todas as predições da 5ª classe e ainda ter uma acurácia de 95%. Ainda que o modelo proposto não tenha classes ou entidades com relevância distinta, pela análise dos resultados preliminares, que serão apresentados no próximo capítulo, a base de dados é fortemente desbalanceada.

Levando isso em conta para os componentes multi-classe, as médias da precisão serão calculadas usando uma média simples e uma média balanceada. A média simples calcula a métrica independentemente para cada classe e, em seguida, obtém a média. Essa métrica trata todas as classes igualmente e não leva em consideração o desequilíbrio entre elas. Por outro lado, a média balanceada calcula as métricas de cada classe, mas a contribuição de cada classe para a média é ponderada pelo número de exemplos disponíveis. Desta forma, algumas classes contribuem mais que outras (THORAT; TONGAONKAR; JAGTAP, 2020; CHAKRAVARTHI

Tabela 18 – Métricas de avaliação.

Acurácia <i>accuracy</i>	Proporção de previsões corretas (verdadeiros positivos e verdadeiros negativos) para o número total de previsões.	$\frac{TP+TN}{TP+TN+FP+FN}$
Precisão <i>precision</i>	Proporção de corretas, das classificações positivas do modelo, quantas foram acertadas.	$\frac{TP}{TP+FP}$
Revocação <i>recall</i>	Das amostras positivas existentes, quantas o modelo conseguiu classificar corretamente.	$\frac{TP}{TP+FN}$
Medida-F <i>F1-score</i>	Média harmônica de precisão e revocação.	$\frac{2*Precision*Recall}{Precision+Recall}$

et al., 2020; FORMAN et al., 2003).

Para a avaliação final da implementação do pipeline do modelo, será construída uma nova base rotulada (padrão ouro), composta de aproximadamente 10.000 sentenças selecionadas a partir das evoluções dos 234.673 registros da base de dados original, garantindo que essas evoluções não tenham sido utilizadas em nenhuma outra avaliação. Após a rotulação, as mesmas evoluções serão estruturadas pela implementação do pipeline e seus resultados serão comparados. Esta avaliação tem como objetivo avaliar individualmente os dois componentes principais do modelo, identificando se o componente Estruturador de Sentenças dividiu e classificou corretamente cada sentença, e se o Estruturador de Entidades identificou e estruturou de forma adequada cada grupo de entidades. Além disso, visa avaliar a integração dos componentes e o resultado geral do pipeline em relação às métricas propostas.

5.5 Conclusões Parciais

Neste capítulo, foram apresentadas as estratégias de implementação dos componentes e subcomponentes do modelo proposto, detalhando as tecnologias e ferramentas selecionadas para cada etapa do processo, incluindo a base de dados utilizada.

A utilização de diferentes variantes do modelo BERT, incluindo Multilingual BERT, DISTILBERT, BioBERT, BioBERTpt e o treinamento da BioBERTptRT, busca encontrar a melhor performance para a tarefa de classificação de sentenças. A partir dos resultados obtidos, propõe-se selecionar a melhor rede resultante da tarefa de classificação de sentenças e retreiná-la para a tarefa de reconhecimento e relacionamento de entidades do componente Estruturador de Entidades. Para a implementação do Agrupamento de Entidades, é proposta a utilização de algoritmos baseados em regras ou técnicas avançadas, caso necessário.

As métricas de avaliação, bem como a estratégia de validação cruzada k-fold, foram definidas para assegurar a robustez e a precisão na avaliação dos componentes. A avaliação final, com a construção de uma base rotulada padrão-ouro, permitirá uma análise mais rigorosa da eficácia do pipeline desenvolvido.

6 RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados o desenvolvimento e avaliação do pipeline de estruturação, projetado para validar o modelo proposto. Detalhando os dois componentes principais: o Estruturador de Sentenças e o Estruturador de Entidades.

Inicialmente, é mostrado o treinamento e a avaliação dos subcomponentes: Classificador de Sentenças, Reconhecimento de Entidades e Relacionamento de Entidades, que exigem treinamento em rede. Em seguida, é descrita a implementação e integração dos demais componentes para criação do Pipeline de Estruturação. Estes incluem a implementação do Divisor de Sentenças e sua integração com o Classificador de Sentenças para formar o Estruturador de Sentenças, além da implementação do componente de Agrupamento de Entidades e sua integração com os componentes de Reconhecimento de Entidades e Relacionamento de Entidades.

Por fim, é explicada a metodologia empregada na avaliação do pipeline, utilizando uma nova base de dados de teste composto por registros de evoluções de pacientes não previamente utilizados. Os resultados desta avaliação são apresentados em detalhes.

6.1 Treinamento e Avaliação do Componente Classificador de Sentença

Essa seção apresenta os resultados e a discussão da avaliação do componente classificador de sentenças do modelo proposto. O conteúdo desta seção foi publicado no periódico *Expert Systems with Applications* em 1º de Julho de 2024 com título de *SOAP classifier for free-text clinical notes with domain-specific pre-trained language models* (OLIVEIRA; ANTUNES; COSTA, 2024).

A avaliação tem como objetivo determinar a viabilidade de treinar uma rede BERT capaz de classificar sentenças em português utilizando o padrão SOAP. Para atingir esse objetivo, foram avaliados quatro modelos de redes pré-treinadas, propostos na seção 5.2, e uma quinta rede, a BioBERTptRT, através do re-treinamento no BioBERTpt usando a base de dados privada apresentada na seção 5.4, que compreende 2.979.987 sentenças. Essa análise visa identificar, dentre os modelos avaliados, aquele que apresenta o melhor desempenho, por fim sendo feita uma comparação com resultados similares em outros trabalhos.

6.1.1 Pré-processamento e rotulagem

O primeiro processamento aplicado a base de dados foi a divisão das evoluções em sentenças. Cada evolução do paciente foi dividida em uma lista de sentenças. Gerando 2.979.987 sentenças, uma média de aproximadamente 12,69 sentença por evolução e 297 sentenças por paciente.

Após a divisão das sentenças, dois aspectos chamaram a atenção: o número de sentenças curtas e o número de sentenças duplicadas. Com base nas evoluções analisadas, isso se deve à semelhança das informações dos exames físico e médico e à natureza resumida dos preenchimentos

dos prontuários pelos médicos, que utilizam muitas abreviações e frases curtas.

Tabela 19 – Exemplo de evolução do paciente em texto livre.

Sentença	SOAP
Reclama de cansaço e ronco.	Subjetivo
Peso:115	Objetivo
Altura:1,60	Objetivo
T.A:150x100	Objetivo
Abdomen:globoso	Objetivo
Mobilidade e força preservados.	Objetivo
Sobrepeso, possivelmente HAS.	Avaliação
Oriento dieta.	Plano
Solicito ECG e eco cardio.	Plano

A Tabela 19 mostra um exemplo dessas características: as queixas do paciente (Subjetivo) se resumem a uma frase curta com dois substantivos. O Objetivo é dividido em cinco sentenças contendo peso e altura, pressão arterial (T.A), resultados de exames abdominais e descrição física com pouco contexto. A Avaliação é representada por uma frase indicando excesso de peso e possível hipertensão arterial sistêmica representada por uma sigla médica HAS. O Plano se resume a uma frase prescrevendo uma dieta e a segunda frase solicitando dois exames médicos de formas abreviadas, Eletrocardiograma (ECG) e Ecocardiograma (eco cardio).

A tabela 20 mostra as sentenças que mais se repetem, sendo em sua maioria, resultados de exames físicos de rotina com valores normais. De 2.979.987 sentenças, 1.796.642 são duplicadas, resultando em 1.183.345 sentenças únicas. O contexto das sentenças não foi levado em consideração, pois cada sentença foi classificada apenas uma vez, tendo a mesma classificação para suas outras aparições.

Para criar um padrão ouro, é necessário rotular manualmente as sentenças. Devido ao número limitado de pesquisadores autorizados a acessar o banco de dados, estabeleceu-se a meta de rotular aproximadamente 100.000 sentenças únicas, uma GUI (*graphical user interface*) foi desenvolvida para facilitar a rotulagem. O processo aplicado foi o seguinte: seleciona-se

Tabela 20 – Sentenças com maior número de repetições.

Sentença	Qtd
Abdomen:SP	31,893
Pele e mucosa:SP	29,017
T.A:120x80	19,979
HAS	17,435
T.A:130X80	11,675
T.A:140x80	8,519
DISLIPIDEMIA	8,005
Pele e mucosa:ÍNTEGRAS	7,879
FC:74	7,789

Tabela 21 – Distribuição de sentenças rotuladas manualmente (padrão-ouro) contendo a classe, o número de sentenças e a proporção em relação ao total.

Rótulo	Sentença	%
Subjetivo	48784	48,77%
Objetivo	34596	34,59%
Avaliação	2751	2,75%
Plano	13890	13,89%
Total	100021	

aleatoriamente uma evolução não rotulada, cujas sentenças são classificadas como Subjetivo, Objetivo, Avaliação e Plano. Após classificar todas as sentenças, seleciona-se uma nova evolução. Após esse processo, 100.021 sentenças únicas foram rotuladas e, no total, foram rotuladas 15.883 evoluções clínicas. A Tabela 21 mostra a distribuição entre as classes. A escassez de sentenças da classe Avaliação pode ser atribuída a dois fatores principais: a falta de informações sobre o diagnóstico do problema médico e a repetição de sentenças. Das 15.883 evoluções clínicas analisadas, apenas 4.791 apresentam sentenças da classe Avaliação. Dentre estas, 2.040 são duplicatas, resultando em 2.751 sentenças únicas.

6.1.2 Criação da BioBERTptRT

Para criar o BioBERTptRT, foi re-treinado o modelo BioBERTpt usando a base de dados privado não rotulado de 2.979.987 sentenças. As sentenças foram tokenizadas usando o tokenizador do BioBERTpt. O modelo foi treinado por quinze épocas, o mesmo número usado no treinamento original do BioBERTpt. O treinamento foi conduzido em uma GPU Quadro RTX 6000 de 24GB, usando a biblioteca PyTorch Transformers desenvolvida pela Hugging Face¹. Os hiperparâmetros foram configurados com:

- *Maximum token size*: 512;
- *Batch Size*: 2;
- *Learning rate*: $2e-5$.

6.1.3 Treinamento e métricas de avaliação

Para avaliação, utilizou-se a validação cruzada estratificada k-fold (k=5), e a divisão das classes é apresentada na Tabela 22. Para a função de ativação da rede, adotou-se a função Softmax (HU et al., 2018), típica da regressão logística multinomial, ideal para cenários multiclasse. A função Softmax é uma função de ativação empregada principalmente no final das redes neurais para classificação, especialmente em problemas de classificação multiclasse. Ela converte um

¹<https://github.com/huggingface/transformers>

Tabela 22 – Distribuição da base de dados entre as 5 folds.

Classe	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Subjetivo	9757	9757	9757	9757	9756
Objetivo	6919	6919	6919	6919	6920
Avaliação	551	550	550	550	550
Plano	2778	2778	2778	2778	2778
Total	20005	20004	20004	20004	20004

vetor de valores reais não normalizados em um vetor de valores reais que somam 1, representando efetivamente uma distribuição de probabilidade sobre as possíveis classes de saída. Isso torna a Softmax particularmente útil para a interpretação dos resultados em tarefas de classificação, pois seus valores de saída podem ser compreendidos como probabilidades de cada classe.

Todos os modelos foram treinados usando a implementação BERT do PyTorch proposta por Hugging Face executando em uma GPU GeForce, RTX 3070 SG, 8GB, GDDR6 com os seguintes hiperparâmetros de treinamento:

- *Maximum token size:* 256;
- *Batch Size:* 32;
- *Number of train epochs:* 4;
- *Learning rate:* 2e-5.

Experimentos foram conduzidos usando mais épocas, mas não foram observados ganhos ou perdas substanciais.

Foram consideradas as seguintes métricas de desempenho: acurácia, precisão, revocação e Medida-F, calculados individualmente para cada classe e suas médias simples e ponderadas.

6.1.4 Resultados

Os resultados da avaliação de cada *fold* na sua interação de teste foram utilizados para calcular as métricas de avaliação. A Tabela 23 apresenta um resumo de todos os resultados obtidos da validação cruzada de cinco modelos, exibindo para cada modelo os valores de positivo verdadeiro, negativo verdadeiro, falso positivo, falso negativo, acurácia, precisão, revocação e medida-F de cada classe, além das médias simples e balanceadas de cada modelo. Os valores em negrito indicam o melhor resultado na categoria entre os modelos.

BioBERTptRT (medida-F 94.63%) alcançou os melhores resultados em todas as métricas avaliadas. Por outro lado, o modelo mBERT teve o pior desempenho, com uma medida-F de 93.51%. A maior diferença entre as classes foi observada na classe Avaliação, com um aumento de 5.33% na medida-F ao comparar mBERT (68.95%) com BioBERTptRT (74.28%). A Figura 10 mostra um gráfico comparativo da medida-F com cada classe e modelo avaliado.

Tabela 23 – Resultados de desempenho para todos os modelos e classes para o componente de Estruturação de Sentença.

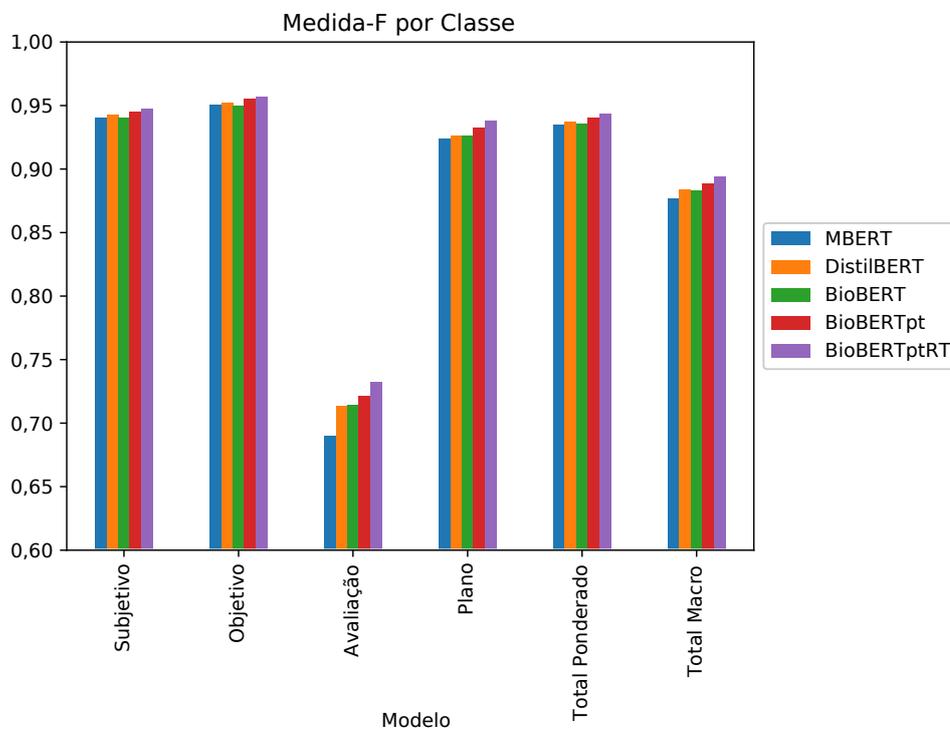
Modelo/Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação	Medida-F
mBERT								
Subjetivo	45849	48389	2848	2935	0.9398	0.9415	0.9398	0.9407
Objetivo	33022	63601	1824	1574	0.9545	0.9477	0.9545	0.9511
Avaliação	1793	96613	657	958	0.6518	0.7318	0.6518	0.6895
Plano	12903	85006	1125	987	0.9289	0.9198	0.9289	0.9243
Média (Balanceada)					0.9355	0.9349	0.9355	0.9351
Média (Simples)					0.8688	0.8852	0.8688	0.8764
DistilBERT								
Subjetivo	46034	48418	2819	2750	0.9436	0.9423	0.9436	0.9430
Objetivo	33017	63689	1736	1579	0.9544	0.9500	0.9544	0.9522
Avaliação	1883	96624	646	868	0.6845	0.7446	0.6845	0.7133
Plano	12864	85109	1022	1026	0.9261	0.9264	0.9261	0.9263
Média (Balanceada)					0.9378	0.9373	0.9378	0.9375
Média (Simples)					0.8772	0.8908	0.8772	0.8837
BioBERT								
Subjetivo	46127	48076	3161	2657	0.9455	0.9359	0.9455	0.9407
Objetivo	32744	63846	1579	1852	0.9465	0.9540	0.9465	0.9502
Avaliação	1880	96634	636	871	0.6834	0.7472	0.6834	0.7139
Plano	12869	85106	1025	1021	0.9265	0.9262	0.9265	0.9264
Média (Balanceada)					0.9360	0.9356	0.9360	0.9358
Média (Simples)					0.8755	0.8908	0.8755	0.8828
BioBERTpt								
Subjetivo	46057	48646	2591	2727	0.9441	0.9467	0.9441	0.9454
Objetivo	33060	63845	1580	1536	0.9556	0.9544	0.9556	0.9550
Avaliação	1893	96663	607	858	0.6881	0.7572	0.6881	0.7210
Plano	13117	85015	1116	773	0.9443	0.9216	0.9443	0.9328
Média (Balanceada)					0.9411	0.9407	0.9411	0.9408
Média (Simples)					0.8830	0.8950	0.8830	0.8886
BioBERTptRT								
Subjetivo	46263	48903	2334	2521	0.9483	0.9520	0.9483	0.9501
Objetivo	33220	63976	1449	1376	0.9602	0.9582	0.9602	0.9592
Avaliação	1991	96651	619	760	0.7237	0.7628	0.7237	0.7428
Plano	13189	85175	956	701	0.9495	0.9324	0.9495	0.9409
Média (Balanceada)					0.9464	0.9462	0.9464	0.9463
Média (Simples)					0.8955	0.9014	0.8955	0.8983

Para medir se a melhoria de desempenho na medida-F é estatisticamente significativa, o teste de soma de postos de Wilcoxon (WILCOXON et al., 1970) foi utilizado nos resultados da medida-F de cada k-fold. O teste de soma de postos de Wilcoxon é um método não paramétrico que determina se as médias são estatisticamente diferentes uma da outra com base em postos em vez das medida-F originais dos experimentos, sendo usado para comparar modelos de PLN (LI et al., 2022). Os resultados são exibidos na Tabela 24. Apenas BioBERTptRT exibe um ganho estatisticamente significativo em comparação a todos os outros modelos, com um $p < 0,05$ no teste de Wilcoxon. BioBERTpt mostra um ganho em relação aos outros modelos, mas uma perda estatisticamente significativa em comparação ao BioBERTptRT. Os outros redes não mostram variações estatisticamente significativas quando comparadas.

Tabela 24 – A medida-F foi comparada em pares de modelos sob o teste de soma de postos de Wilcoxon, com valores em negrito indicando significância estatística ($p < 0,05$).

p	BioBERTptRT	BioBERTpt	BioBERT	DistilBERT	mBERT
BioBERTptRT	-	0.0079	0.0079	0.0079	0.0079
BioBERTpt	0.0079	-	0.0159	0.0159	0.0079
BioBERT	0.0079	0.0159	-	0.1508	0.6905
DistilBERT	0.0079	0.0159	0.1508	-	0.0556
mBERT	0.0079	0.0079	0.6905	0.0556	-

Figura 10 – Gráfico da medida-F das quatro classes e cinco modelos (escala começando em 0.6).



6.1.4.1 Análise do melhor modelo (BioBERTptRT)

Analisando cada classe no modelo individualmente, as classes Subjetiva, Objetiva e Plano apresentam resultados similares, sendo a Objetiva a que tem o melhor medida-F, com 95,92%. A classe Avaliação apresenta o pior resultado. Na Figura 11, a matriz de confusão mostra os valores reais e previstos para as quatro classes. As células ao longo da diagonal correspondem aos valores positivos verdadeiros para cada classe. Os valores falsos negativos são mostrados nas linhas fora da diagonal, enquanto os valores falsos positivos são exibidos nas colunas fora da diagonal.

O tipo mais comum de confusão ocorre entre Subjetivo e Objetivo, resultando em um total de 2378 falsos positivos. No entanto, deve-se notar que essas duas classes compõem a maioria da base de dados (83,36%). Considerando o desequilíbrio da base de dados, uma curva de precisão-recall e sua área para cada classe também foram geradas para fazer uma análise mais precisa da qualidade de previsão do modelo. A Figura 12 mostra essa informação. As classes de Subjetivo (área = 0,9850), Objetivo (área = 0,9885) e Plano (área = 0,9674) mostraram valores AUC comparativamente altos, enquanto a classe Avaliação (área = 0,8081) exibiu um valor mais baixo. Isso ocorre porque tanto as medidas de precisão quanto de recall são menores na classe Avaliação do que nas outras classes.

Figura 11 – Matriz de confusão do BioBERTptRT com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.

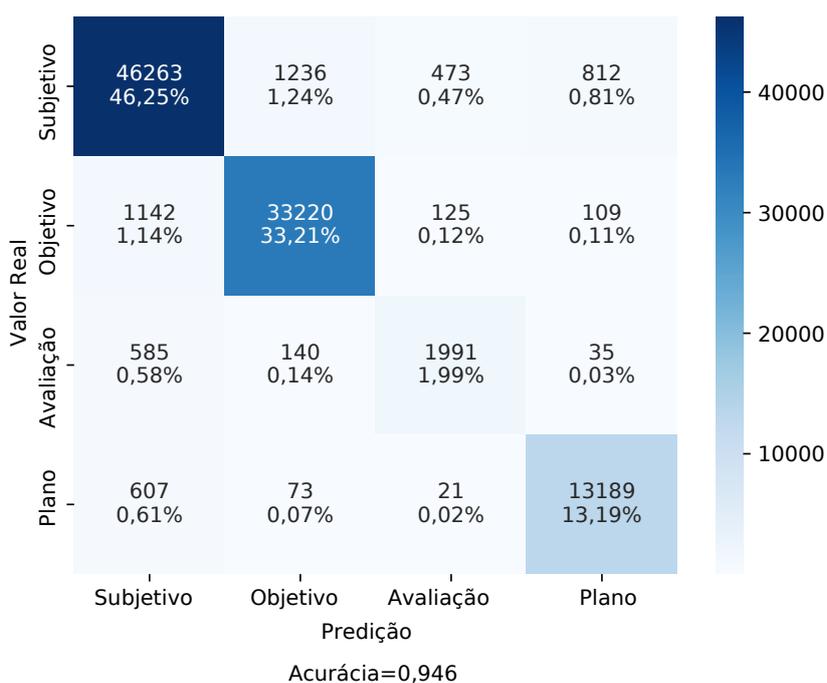
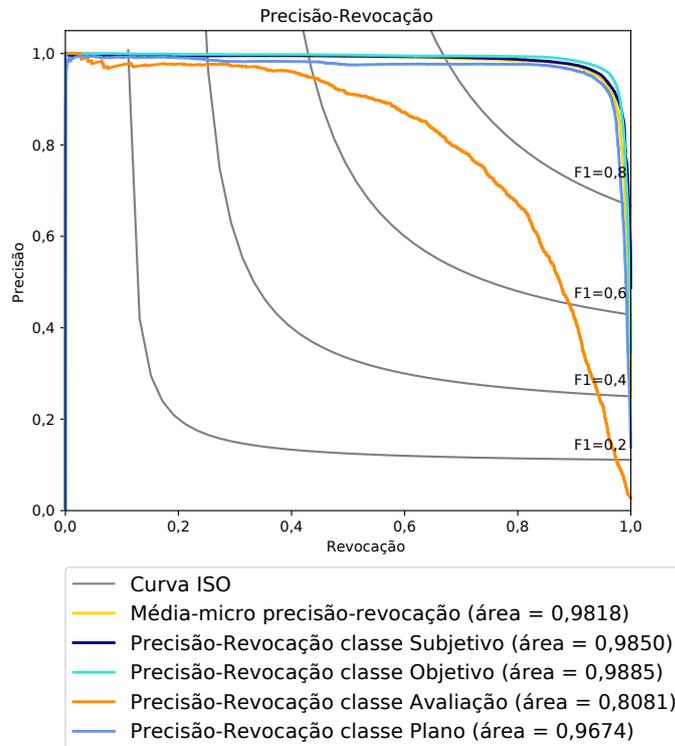


Figura 12 – Curva de Precisão-Recall da BioBERTptRT mostra os compromissos entre precisão e recall para todas as classes.



6.1.4.2 Análise do desempenho na classe Avaliação

Avaliação é a classe que apresenta o pior desempenho em todas as métricas quando analisamos a base de dados, a matriz de confusão e uma análise amostral de falsos positivos e falsos negativos resultantes da classe. Duas principais causas podem ser listadas: a baixa amostragem da classe em comparação com as outras, o que pode causar um pior desempenho na classe em problemas multiclasse com uma base de dados desbalanceado (ALDRAIMLI et al., 2021; SREEJITH; Khanna Nehemiah; KANNAN, 2020; WANG et al., 2021) e a dificuldade de avaliar certas sentenças sem analisar o contexto da evolução clínica. Isso ocorre principalmente em casos de falsos positivos entre as classes Avaliação e Subjetivo. Por exemplo, a frase "pressão arterial elevada devido a prostatite aguda" foi classificada como Avaliação. No entanto, no contexto da evolução clínica precedida pela frase "Paciente foi encaminhado por um clínico geral", torna-se uma sentença Subjetiva, pois representa uma queixa trazida pelo paciente encaminhado. Sem a análise do contexto da evolução clínica, a frase torna-se ambígua, dificultando a classificação pelo modelo. Esses casos também foram identificados em outras classes, mas são mais recorrentes na classe Avaliação.

Tabela 25 – Comparação da distribuição de sentenças da base de dados desse trabalho e da base de dados de (MOWERY et al., 2012).

Rótulo	(MOWERY et al., 2012)			Modelo Proposto		
	Sentenças	%	Medida-F	Sentenças	%	Medida-F
Subjetivo	1468	35,5%	0,939	48784	48,77%	0,9501
Objetivo	1818	44,0%	0,945	34596	34,59%	0,9592
Avaliação	227	5,5%	0,757	2751	2,75%	0,7428
Plano	465	11,3%	0,770	13890	13,89%	0,9409
Não Aplicável	335	8,1%				
Total	4130			100021		

Tabela 26 – Comparação do ganho da BioBERTpt em relação ao mBERT do modelo treinado e dos modelos de (SCHNEIDER et al., 2020) e (SOUZA et al., 2021).

	(SCHNEIDER et al., 2020)			(SOUZA et al., 2021)			Modelo Proposto		
	Precisão	Recall	Medida-F	Precisão	Recall	Medida-F	Precisão	Recall	Medida-F
mBERT	0,912	0,931	0,921	0,518	0,538	0,527	0,9349	0,9355	0,9351
BioBERTpt	0,917	0,935	0,926	0,566	0,557	0,561	0,9407	0,9411	0,9408
Ganho	0,005	0,004	0,005	0,048	0,019	0,034	0,006	0,005	0,006

6.1.4.3 Comparação dos resultados obtidos com um trabalho similar

Embora não seja possível comparar diretamente o modelo com o de (MOWERY et al., 2012), uma vez que eles usam bases de dados e idiomas diferentes, é possível identificar algumas semelhanças. A Tabela 25 mostra a quantidade e a porcentagem de sentenças na base de dados e o melhor medida-F de cada classe de cada trabalho. O trabalho de (MOWERY et al., 2012) utilizou uma abordagem multirrótulo para modelar o problema, mas apenas 165 (4,0%) sentenças possuem duas classes, e 9 (0,02%) possuem três classes. Ambos as bases de dados apresentam algumas semelhanças, onde as classes Subjetivo e Objetivo têm o maior número de sentenças, sendo o Objetivo maior em uma base de dados e o Subjetivo maior em outra. A Avaliação apresenta a menor quantidade em ambos as bases de dados, indicando que uma falta de informação sobre diagnósticos de problemas médicos, mostrada na subseção 4.2, é recorrente em outras bases de dados. A Avaliação apresenta o pior desempenho em ambos. Isso pode implicar um maior nível de complexidade na classificação da classe discutida no parágrafo anterior. Alternativamente, pode sugerir que não há sentenças suficientes disponíveis para treinar adequadamente a rede. Isso é particularmente evidente na classe Plano, onde o trabalho proposto alcançou resultados favoráveis e possuía mais sentenças na base de dados.

6.1.4.4 Especialização do BioBERTpt

Através da especialização de domínio, foi possível identificar um ganho na tarefa proposta. A Tabela 27 apresenta a classificação do modelo em relação ao melhor Medida-F (ponderado). O BioBERTptRT alcançou o melhor resultado e também é o mais especializado, com três domínios

Tabela 27 – Ranking dos modelos em relação a melhor Medida-F balanceada e suas especializações de domínio.

Modelo	Domínio Português	Domínio Biomédico	Domínio Local	Medida-F Ponderado	Posição
BioBERTptRT	X	X	X	0.9463	1°
BioBERTpt	X	X		0.9408	2°
DistilBERT	X			0.9375	3°
BioBERT		X		0.9358	4°
mBERT	X			0.9351	5°

(português, biomédico e local), seguido pelo BioBERTpt com duas especializações (português e biomédico). Com apenas uma especialização, os outros modelos ocupam as três últimas posições.

É relevante notar que o DistilBERT teve um desempenho melhor que o mBERT, apesar de ter o mesmo domínio. Uma possível explicação é que o mBERT é multilíngue, enquanto o DistilBERT usado no estudo foi treinado exclusivamente para o português. Analisando o resultado do BioBERT, pode-se concluir que o domínio biomédico sozinho só traz ganhos significativos com o domínio da língua.

Especializar para grupos médicos ou redes hospitalares resultou em um aumento de 0,55% na medida-F em comparação com o BioBERTpt sem esta especialização. Apesar deste ganho real, é importante considerar algumas questões ao avaliar a viabilidade deste processo. O pré-treinamento requer uma base de dados substancialmente maior em comparação com o ajuste fino. Para o pré-treinamento, foram usadas 2.979.987 frases não rotuladas, em comparação com 100.021 frases rotuladas para o ajuste fino. Além disso, não foi possível realizar o pré-treinamento usando o mesmo hardware, exigindo recursos computacionais mais avançados, como uma GPU Quadro RTX 6000 de 24GB, com um tamanho de lote máximo de 2, em contraste com o ajuste fino realizado em uma GPU RTX 3070 de 8GB, com um tamanho de lote de 32. Além disso, espera-se que este modelo tenha um desempenho menor quando testado em bases de dados de outros grupos médicos ou redes hospitalares.

6.1.4.5 Comparação da especialização de domínio em outros trabalhos

Foi possível identificar que a especialização de domínio, além da especialização na língua portuguesa, contribui para melhoria no desempenho. Embora esses ganhos sejam relativamente pequeno, eles se mostram consistentes ao compararmos esse trabalho com os estudos propostos por (SCHNEIDER et al., 2020) e (SOUZA et al., 2021). Esses trabalhos também realizaram comparações entre o mBERT e o BioBERTpt, utilizando bases de dados e objetivos distintos. Os resultados obtidos são similarmente positivos.

A Tabela 26 destaca o ganho de desempenho em termos de precisão, revocação e medida-F nos três estudos analisados. O estudo realizado por (SOUZA et al., 2021) demonstra um aumento

mais acentuado na medida-F (0.034), porém o valor absoluto dessa medida-F é de apenas 0.561. Por outro lado, o modelo proposto neste estudo e o modelo de (SCHNEIDER et al., 2020) apresentam aumentos na medida-F (0.004 e 0.005, respectivamente) com valores absolutos mais elevados, explicando assim a menor variação observada.

Com base nesta análise, podemos afirmar que a especialização de domínio no campo clínico demonstra um desempenho superior em tarefas de PLN em domínios médicos em português em comparação com modelos que se especializam apenas no português.

6.1.5 Discussão dos Resultados

Dentre os cinco modelos BERT pré-treinados testados, a BioBERTptRT alcançou os melhores resultados. Obtendo uma precisão de 0,9462, uma acurácia de 0,9464, uma Revocação de 0,9464 e Medida-F de 0,9463. O BioBERTptRT é o modelo mais especializado, possuindo conhecimento específico da língua portuguesa, terminologia clínica e o contexto do grupo médico. Comparado ao segundo modelo mais especializado, o BioBERTptRT obteve um aumento de 0,55% no Medida-F. A classe que teve os piores resultados foi avaliação com 0,7428. Isso se deve a dois fatores principais: o menor número de sentenças em relação às demais classes e a falta de contexto em relação às demais sentenças do prontuário. Após a análise do desempenho, a comparação dos resultados com um trabalho similar em língua inglesa e a análise da consistência do ganho, é possível afirmar que a BioBERTptRT é adequada para classificação de sentenças da classe SOAP em português.

6.2 Treinamento e Avaliação do Componente de Reconhecimento de Entidades

Para implementar o Componente de Reconhecimento de Entidades, é necessário treinar uma rede específica para a tarefa de NER, uma vez que essa tarefa é comumente associada às redes BERT, que se destacam pela sua capacidade de serem re-treinadas. Dessa forma, foi selecionada a rede BioBERTptRT, pré-treinada para tarefa anterior de classificação de sentença, para a implementação desse componente. A escolha da BioBERTptRT deve-se à sua especialização (foco no domínio biomédico e em português, além de ser localizada) e ao fato de ter apresentado o melhor desempenho em comparação a outras redes avaliadas no processo de implementação do componente de classificação de sentenças.

Nas subseções seguintes, são detalhados os processos de treinamento e avaliação deste componente, incluindo o pré-processamento e a rotulagem da base de dados, o treinamento propriamente dito, as métricas de avaliação utilizadas, os resultados alcançados, a análise dos achados e, por fim, as conclusões preliminares do processo de treinamento e avaliação do componente.

6.2.1 Pré-processamento e rotulagem

A rotulagem para criar um padrão-ouro foi realizada a partir das sentenças rotuladas para o componente de classificação de sentenças. Para isso, a GUI desenvolvida para o componente anterior foi estendida para realizar além da anotação NER, a estruturação das entidades nos grupos propostos nas Tabelas 13, 14, 15 e 16 para o posterior treino e avaliação dos demais componentes. Nos casos que não seja possível elevar o nível da estruturação na sentença, a sentença é classificada como "sem itens para classificar", representando o limite da estruturação da mesma. A Figura 13 mostra a versão final da GUI. Devido ao número limitado de pesquisadores com autorização para acessar o banco de dados, optou-se por rotular somente as sentenças das classes Objetivo, Avaliação e Plano, somando um total de 51.237 sentenças para classificação.

Ao término da rotulação das 51.237 sentenças, 10.580 não apresentaram nenhum grupo de entidades e 40.657 apresentaram um ou mais grupos, conforme detalhado na Tabela 28.

Tabela 28 – Distribuição da grupo de entidades após a classificação.

SOAP	Grupo de Entidades	Quantidade
Objetivo	Resultado de Exame	29640
	Exame Físico	9315
	Medição	2270
Avaliação	Diagnóstico	3219
Plano	Encaminhamento	1721
	Prescrição	5921
	Solicitação de Procedimento	4390
	Atestado	127

Para o treinamento e avaliação do componente de reconhecimento de entidades, as sentenças e suas notações foram exportadas utilizando a notação BIO (*Begin Inside Outside*) (JIANG, 2012), com essa notação, para cada tipo de entidade T, dois rótulos são criados, nomeadamente, B-T e I-T. Um token rotulado com B-T é o início de uma entidade nomeada do tipo T, enquanto um token rotulado com I-T está dentro (mas não é o início de) uma entidade nomeada do tipo T. Além disso, existe um rótulo O para tokens fora de qualquer entidade nomeada.

6.2.2 Treinamento e métricas de avaliação

Também foi usada a validação cruzada k-fold (k=5) para o treinamento e a coleta de métricas de desempenho da rede, não sendo utilizado o recurso de estratificada, pois no NER existem várias classes em uma mesma sentença o que impede a distribuição proporcional, sem a quebra da mesma.

A BioBERTptRT foi treinada para NER usando a implementação BERT do PyTorch proposta por Hugging Face rodando em uma GPU GeForce, RTX 3070 SG, 8GB, GDDR6 com os seguintes hiperparâmetros de treinamento:

- *Maximum token size: 256;*
- *Batch Size: 32;*
- *Number of train epochs: 5;*
- *Learning rate: 2e-5.*

Experimentos foram conduzidos usando mais épocas, mas não foram observados ganhos ou perdas substanciais.

6.2.3 Resultados

A Tabela 29 resume todos os resultados obtidos do treinamento, exibindo para cada classe os seguintes indicadores: positivo verdadeiro, negativo verdadeiro, falso positivo, falso negativo, acurácia, precisão, revocação e medida-F. Também são apresentadas as médias simples e balanceadas.

Como resultado, a rede apresentou uma precisão de 0,9223, acurácia de 0,9207, revocação de 0,9207 e Medida-F de 0,9210. As classes mais comuns na base de dados, além de O, que representa a ausência de classe, são aquelas relacionadas aos resultados da classificação SOAP Objetivo (B-exame, I-exame, B-resultado, I-resultado). A classe com o melhor desempenho é a que representa as medições (B-medicao e I-medicao), que, apesar de não apresentar uma amostra muito grande, é composta por um subconjunto pequeno de palavras (peso, altura, T.A., etc.), o que facilita o reconhecimento pela rede.

Figura 13 – Versão final da GUI desenvolvida para criação do padrão-ouro.

id	SOAP	Texto
1	Subjetivo	<p>PACIENTE RECLAMA DE DORES NO PEITO E DOR DE CABEÇA .</p> <p>Histórico Familiar: Doença Familiar Negativa Observação</p> <p>Doença pré-existente: Doença Temporalidade Parte do corpo Negativa Observação</p> <p>Medição em uso: Medicamento Posologia Parte do corpo Temporalidade Observação Negativa</p> <p>DORES PEITO Descrição Parte do corpo Temporalidade Observação Negativa</p> <p>DOR CABEÇA Descrição Parte do corpo Temporalidade Observação Negativa</p> <p>Sintoma: Descrição Parte do corpo Temporalidade Observação Negativa</p> <p><input type="checkbox"/> Sem Itens</p> <p>Salvar Atualizar Resetar</p>
6	Objetivo	<p>AO EXAME : PESO : 95 KG (ROUPA + CALÇADO) ; ALTURA : 163 CM ; TA : 180 / 120 .</p> <p>Resultado de Exame: Nome do exame Resultado Parte do corpo Observação Data Negativa</p> <p>Exame Físico: Nome do exame físico Resultado Parte do corpo Observação Negativa</p> <p>PESO 95 KG (CALÇADO + ROUPA) Nome da medição Resultado Observação</p> <p>ALTURA 163 CM Nome da medição Resultado Observação</p> <p>TA 180 / 120 Nome da medição Resultado Observação</p> <p>Medição: Nome da medição Resultado Observação</p> <p><input type="checkbox"/> Sem Itens</p> <p>Salvar Atualizar Resetar</p>
7	Avaliação	<p>HD : HAS SEM CONTROLE ADEQUADO</p> <p>HAS SEM CONTROLE ADEQUADO Diagnóstico Grau de certeza Parte do corpo Observação Temporalidade Negativa</p> <p>Diagnóstico: Diagnóstico Grau de certeza Parte do corpo Observação Temporalidade Negativa</p> <p><input type="checkbox"/> Sem Itens</p> <p>Salvar Atualizar Resetar</p>
8	Plano	<p>SOL . ECG + RX DE TORAX .</p> <p>ECG Nome do Procedimento Parte do corpo Observação Data</p> <p>RX TORAX Nome do Procedimento Parte do corpo Observação Data</p> <p>Solicitação de Procedimento: Nome do Procedimento Parte do corpo Observação Data</p> <p>Encaminhamento: Especialidade Observação</p> <p>Prescrição: Medicamento Posologia Parte do corpo Observação Negativa</p> <p>Atestado: tempo Observação</p> <p><input type="checkbox"/> Sem Itens</p> <p>Salvar Atualizar Resetar</p>

Tabela 29 – Resultados da avaliação do componente de reconhecimento de entidade.

Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação	Medida-F
O	134546	220091	7666	16432	0,8912	0,9461	0,8912	0,9178
B-certeza	388	378197	101	49	0,8879	0,7935	0,8879	0,8380
I-certeza	145	378529	29	32	0,8192	0,8333	0,8192	0,8262
B-data	3184	374970	382	199	0,9412	0,8929	0,9412	0,9164
I-data	10229	367341	834	331	0,9687	0,9246	0,9687	0,9461
B-diagnostico	2750	375073	563	349	0,8874	0,8301	0,8874	0,8578
I-diagnostico	1200	376623	459	453	0,7260	0,7233	0,7260	0,7246
B-especialidade	1443	376873	270	149	0,9064	0,8424	0,9064	0,8732
I-especialidade	101	378478	70	86	0,5401	0,5906	0,5401	0,5642
B-exame	28159	348175	1735	666	0,9769	0,9420	0,9769	0,9591
I-exame	3824	374309	388	214	0,9470	0,9079	0,9470	0,9270
B-exameFisico	4954	373538	154	89	0,9824	0,9699	0,9824	0,9761
I-exameFisico	602	378091	26	16	0,9741	0,9586	0,9741	0,9663
B-medicamento	5477	372495	495	268	0,9534	0,9171	0,9534	0,9349
I-medicamento	559	377745	219	212	0,7250	0,7185	0,7250	0,7218
B-medicao	2251	376444	21	19	0,9916	0,9908	0,9916	0,9912
I-medicao	3436	375279	5	15	0,9957	0,9985	0,9957	0,9971
B-negativa	187	378470	32	46	0,8026	0,8539	0,8026	0,8274
I-negativa	15	378688	7	25	0,3750	0,6818	0,3750	0,4839
B-observacao	3007	373201	1315	1212	0,7127	0,6957	0,7127	0,7041
I-observacao	6638	365732	3768	2597	0,7188	0,6379	0,7188	0,6759
B-parteCorpo	5722	371996	697	320	0,9470	0,8914	0,9470	0,9184
I-parteCorpo	3334	374809	338	254	0,9292	0,9080	0,9292	0,9185
B-posologia	3568	374002	610	555	0,8654	0,8540	0,8654	0,8597
I-posologia	13642	362302	1637	1154	0,9220	0,8929	0,9220	0,9072
B-resultado	36136	338817	2372	1410	0,9624	0,9384	0,9624	0,9503
I-resultado	68706	302914	4836	2279	0,9679	0,9342	0,9679	0,9508
B-solicitacaoProcedimento	3692	374027	676	340	0,9157	0,8452	0,9157	0,8790
I-solicitacaoProcedimento	634	377640	270	191	0,7685	0,7013	0,7685	0,7334
B-temporalidade	101	378585	25	24	0,8080	0,8016	0,8080	0,8048
I-temporalidade	90	378601	15	29	0,7563	0,8571	0,7563	0,8036
Média (Balanceada)					0,9207	0,9223	0,9207	0,9210
Média (Simples)					0,8570	0,8475	0,8570	0,8502

Em geral a rede apresentou um bom desempenho na tarefa de NER, principalmente nas classes com o maior número de amostras ou menor número de entidades únicas. A Figura 14 apresenta a matriz de confusão com detalhes do desempenho. A diagonal indica o número de predições corretas, evidenciando o bom desempenho da rede, especialmente nas classes B-exame e I-exame. Isso sugere que o modelo é particularmente eficaz em reconhecer esses tipos de dados e suas sequências quando apresentados. Além disso, apesar dos bons resultados das classes B-resultado (início do resultado) e I-resultado (continuação do resultado), houve confusão entre si, indicando uma dificuldade do modelo em distinguir o início de um resultado de sua continuação.

Além das confusões notáveis entre as classes "B-" e "I-", o que é esperado, pois esses prefixos indicam se a entidade é o início de uma menção ou parte interna de uma menção, as confusões mais comuns mostram que certas categorias foram mais desafiadoras para o modelo identificar corretamente. Isso é evidenciado pelo número de predições incorretas para a categoria O em comparação com predições corretas para categorias específicas. Por exemplo, I-observação (continuação de observação) foi frequentemente confundido com O (fora de entidade), com 2.929 instâncias erroneamente classificadas como O, e B-observação (início de observação) também teve 816 instâncias classificadas incorretamente como O. Isso pode ser explicado pelo fato de que, como observação é uma categoria ampla e semi-estruturada, ela pode ser confundida com outras entidades que não foram estruturadas e, por isso, foram classificadas como O.

6.2.3.1 Comparação dos resultados obtidos com o estado da arte

Embora uma comparação direta com outros trabalhos não seja possível, devido às diferenças nas bases de dados e nos critérios de rotulagem, realizou-se um comparativo com estudos recentes que representam o estado da arte do NER na medicina, em língua portuguesa. Selecionaram-se três estudos: ROCHA et al.(ROCHA et al., 2022), que desenvolveu uma rede neural para extrair informações de registros médicos não estruturados, abrangendo sintomas, diagnósticos, medicações, condições, exames e tratamentos, a partir de um corpus de 1200 textos clínicos do Hospital Clínico da Faculdade de Medicina de Botucatu, São Paulo, Brasil; SILVA(SILVA, 2023), que propôs um modelo NER integrando ontologia, avaliado em 1.622 documentos com 146.769 entidades, provenientes de uma empresa de Oncologia; e SCHNEIDER et al.(SCHNEIDER et al., 2020), que introduziu a BioBERTpt e avaliou seu modelo NER usando a base de dados pública SemClinBr (OLIVEIRA et al., 2020). A Tabela 30 resume este comparativo, destacando as entidades comuns ou similares e os valores de medida-F, precisão e revocação, conforme disponibilizados nos estudos. Para a comparação, as entidades deste trabalho foram consolidadas, omitindo os prefixos "B-" e "I-". A análise sugere que a rede alcançou resultados promissores, superando os demais estudos em todos os aspectos comparados, especialmente nas entidades relacionadas a diagnósticos, procedimentos e partes do corpo, que foram mais desafiadoras para as outras redes.

Tabela 30 – Comparação de Desempenho na tarefa NER com outros trabalhos (valores em porcentagem).

Entidade	(ROCHA et al., 2022)			(SCHNEIDER et al., 2020)			(SILVA, 2023)			Componente		
	Medida-F	Precisão	Revocação	Medida-F	Medida-F	Precisão	Revocação	Medida-F	Precisão	Revocação		
Diagnóstico	49,27	54,42	45,01					82,53	80,71	84,44		
Exame	54,66	72,60	43,83		78,77	78,92	77,27	95,60	93,86	97,41		
Medicamento/Droga	80,96	87,47	75,36	91,00				91,58	90,00	93,23		
Procedimento				70,00				86,85	83,44	90,57		
Membro/Parte Corpo					78,25	78,26	77,74	92,9	90,78	95,12		
Resultado					70,59	70,01	70,38	95,87	94,37	97,43		
Tempo/Temporalidade					73,56	73,33	72,17	85,05	87,44	82,78		
Data					76,33	77,76	75,14	94,77	92,54	97,10		

6.2.4 Discussão dos Resultados

A criação de um padrão-ouro, por meio da extensão GUI já desenvolvida, permitiu não apenas estabelecer o padrão-ouro para o NER do componente, mas também estruturar as informações para o treinamentos e avaliações dos demais componentes.

Após o treino e avaliação, a rede treinada para o componente de reconhecimento de entidade apresentou um ótimo desempenho, particularmente na identificação de classes com o maior número de amostras, evidenciando a eficácia da abordagem adotada.

A comparação dos resultados obtidos com trabalhos similares na área médica, embora limitada pelas diferenças em bases de dados e critérios de rotulagem, mostra que a rede alcançou resultados promissores, superando os estabelecidos por estudos anteriores, especialmente em entidades desafiadoras como diagnósticos, procedimentos e partes do corpo, validando a eficácia da BioBERTptRT e das técnicas de treinamento empregadas para implementar esse componente.

6.3 Treinamento e Avaliação do Componente de Relacionamento de Entidades

Para implementar o Componente de Relacionamento de Entidades, adotou-se a mesma estratégia de reaprendizagem da BioBERTptRT, empregada anteriormente no módulo de reconhecimento de entidades. Assim, o modelo foi retreinado para tarefa de classificação de sentenças, onde a entrada compreende a sentença e duas entidades, resultando em uma classificação binária que indica a existência de uma relação entre elas. Nas subseções a seguir, detalham-se os procedimentos de treinamento e avaliação desse componente, abrangendo o pré-processamento, a rotulação da base de dados, o treinamento em si, as métricas de avaliação empregadas, os resultados obtidos e, finalmente, as conclusões preliminares do processo de treinamento e avaliação do componente.

6.3.1 Pré-processamento e rotulagem

Figura 15 – Exemplo do layout para exportação do Componente de Relacionamento de Entidades.



Para adaptar a classificação de sentenças à tarefa de relação entre entidades, foram utilizados dois recursos do BERT: a divisão do token em contexto e pergunta e o separador especial [SEP],

que demarca segmentos de texto dentro de uma entrada fornecida ao modelo. Desta forma, o contexto é formado pela sentença completa, a pergunta é formada pelas duas entidades que estão sendo avaliadas e suas posições na sentença (para evitar eventuais ambiguidades), todos separados pelo [SEP]. Sendo a resposta binária se as duas entidades estão relacionadas entre si. A Figura 15 mostra um exemplo desse layout.

As entidades são formadas pelo mesmo grupo de palavras do componente de reconhecimento de entidade, isto é, uma entidade formada por B-T com um ou mais I-T se tornam uma entidade no relacionamento. A Tabela 31 mostra todas as relações de uma sentença, formada por 6 entidades (“Hb”, “13,2”, “Ht”, “7,9%”, “leuco” e “20”) que geram 15 entradas para o treinamento, onde três são relacionadas e as demais não.

Tabela 31 – Exemplo de sentença rotulada para o Componente de Relacionamento de Entidades.

Contexto	Pergunta	Resposta
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Hb [SEP] 1 [SEP] 13 , 2 [SEP] 2	1
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Hb [SEP] 1 [SEP] Ht [SEP] 6	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Hb [SEP] 1 [SEP] 37 , 9 % [SEP] 7	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Hb [SEP] 1 [SEP] leuco [SEP] 12	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Hb [SEP] 1 [SEP] 20 [SEP] 13	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	13 , 2 [SEP] 2 [SEP] Ht [SEP] 6	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	13 , 2 [SEP] 2 [SEP] 37 , 9 % [SEP] 7	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	13 , 2 [SEP] 2 [SEP] leuco [SEP] 12	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	13 , 2 [SEP] 2 [SEP] 20 [SEP] 13	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Ht [SEP] 6 [SEP] 37 , 9 % [SEP] 7	1
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Ht [SEP] 6 [SEP] leuco [SEP] 12	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	Ht [SEP] 6 [SEP] 20 [SEP] 13	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	37 , 9 % [SEP] 7 [SEP] leuco [SEP] 12	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	37 , 9 % [SEP] 7 [SEP] 20 [SEP] 13	0
Hb 13 , 2 ; Ht 37 , 9 % ; leuco 20	leuco [SEP] 12 [SEP] 20 [SEP] 13	1

6.3.2 Treinamento e métricas de avaliação

Também foi utilizada a validação cruzada k-fold (k=5) não estratificada para o treinamento e a coleta de métricas de desempenho da rede. A BioBERTptRT foi treinada para classificação de sentenças utilizando o mesmo hardware e hiperparâmetros dos demais modelos, com exceção do *Maximum token size*, que foi aumentado de 256 para 512, a fim de acomodar o aumento do tamanho do token com a pergunta.

6.3.3 Resultados obtidos

A Tabela 32 resume todos os resultados obtidos do treinamento, exibindo para cada classe os valores de verdadeiro positivo, verdadeiro negativo, falso positivo, falso negativo, acurácia, precisão, revocação e medida-F para cada classe, além das médias simples e balanceadas. Como

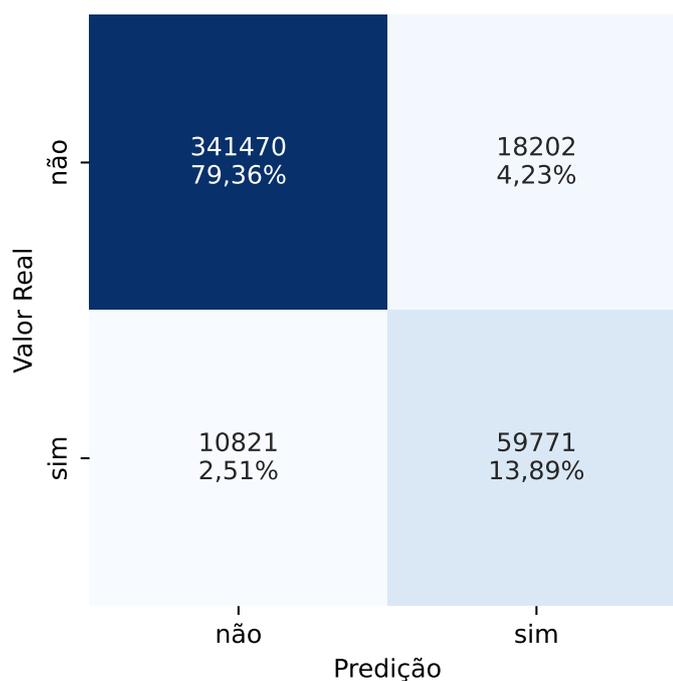
Tabela 32 – Resultados de desempenho para o componente de Relacionamento de Entidade.

Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação	Medida-F
Não	341470	59771	10821	18202	0,9494	0,9693	0,9494	0,9592
Sim	59771	341470	18202	10821	0,8467	0,7666	0,8467	0,8046
Média (Balanceada)					0,9325	0,9360	0,9325	0,9339
Média (Simples)					0,8981	0,8679	0,8981	0,8819

resultado, a rede apresentou uma acurácia de 0,9325, precisão de 0,936, revocação de 0,9325 e medida-F de 0,9339, considerando as medias simples. Levando em conta as métricas balanceadas, a acurácia foi de 0,8981, precisão de 0,8679, revocação de 0,8981 e medida-F de 0,8819.

Especificamente, o modelo foi mais eficaz em identificar a ausência de relações, com uma precisão de 96,93% e uma medida-F1 de 95,92%, refletindo um excelente equilíbrio entre precisão e revocação. Por outro lado, a capacidade do modelo para detectar a presença de relações é menos precisa, com uma precisão de 76,66% e uma pontuação F1 de 80,46%. Isso sugere que o modelo é melhor em prever a ausência de relações do que em identificar quando elas existem, possivelmente devido ao desequilíbrio entre as classes. Isso pode ser visto pela análise da matriz de confusão apresentada na Figura 16, onde a classe que representa a ausência constitui 83,59% das entidades, e a presença de relações representa somente 16,41%, o que é esperado devido à natureza da tarefa, onde é mais comum a ausência de relação que sua existência.

Figura 16 – Matriz de Confusão do componente de relacionamento de entidades com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.



6.3.3.1 Comparação dos resultados com o estado da arte

Somente foi encontrado um trabalho que realiza a tarefa de relacionamento de entidades na área médica, em língua portuguesa, SILVA(SILVA, 2023). Esta proposta inclui a integração de ontologia, além da tarefa de NER comparada ao componente anteriormente mencionado. Este trabalho também treinou e avaliou uma tarefa de relacionamento de entidades. A Tabela 33 apresenta um comparativo entre os trabalhos. É possível analisar apenas os resultados totais, visto que o trabalho SILVA(SILVA, 2023) aborda múltiplas relações e o componente analisado realiza somente a relação binária. O modelo proposto apresentou um resultado superior, desconsiderando as diferenças das bases de dados. Esse ganho pode ser atribuído a dois fatores: a simplificação na identificação das relações e o maior número de amostras utilizadas no treinamento.

Tabela 33 – Comparação de Desempenho na tarefa de relacionamento de entidade com outros trabalhos (valores em porcentagem).

	SILVA(SILVA, 2023)			Componente		
	Acurácia	Revocação	Medida-F	Acurácia	Revocação	Medida-F
Balanceada	76,33%	78,17%	78,22%	93,25%	93,25%	93,39%
Média (Simples)	76,78%	76,00%	75,10%	89,81%	89,81%	88,19%
Amostras	25.584			430.264		

6.3.4 Discussão dos Resultados

A implementação e avaliação do Componente de Relacionamento de Entidades, utilizando a abordagem de retreino com a BioBERTptRT, revelou-se eficaz na tarefa de relacionamento de entidade, através da classificação binária de sentenças para identificar relações entre as entidades.

Comparativamente, o desempenho do componente superou o trabalho similar comparado, mesmo levando em consideração as variações nas bases de dados e na definição das tarefas de relacionamento. Isso sugere que a abordagem adotada neste trabalho, particularmente a simplificação na identificação de relações e a utilização de uma base de dados de treinamento mais amplo, foi fundamental para o bom desempenho do componente.

6.4 Implementação e Avaliação do Pipeline de Estruturação

Essa seção apresenta a implementação e avaliação do pipeline de estruturação para avaliar o modelo proposto, sendo implementado a partir da integração dos demais componentes desenvolvidos e avaliados.

A primeira seção detalha o Estruturador de Sentença, que incorpora o Divisor de Sentença e o Classificador de Sentença, para divisão e classificação das sentenças conforme o padrão SOAP. Na sequência, descreve-se o Estruturador de Entidade, que apresenta a integração dos

componentes já desenvolvidos de Reconhecimento de Entidades, Relacionamento de Entidades e apresenta a implementação do componente de Agrupamento de Entidades.

Por fim, é apresentada a nova base de dados de teste, composta por registros de evoluções dos pacientes não previamente usados, e os resultados desta avaliação são apresentados detalhadamente. A discussão dos resultados realça tanto o potencial quanto as limitações identificadas na implementação.

6.4.1 Implementação do Estruturador de Sentença

Como descrito no modelo do Estruturador de Sentença, ele é composto de dois componentes principais: o Divisor de Sentença e o Classificador de Sentença. Para a implementação do componente Divisor de Sentença, após a avaliação das funcionalidades, recursos e a realização de alguns testes, optou-se pelo Stanza. A ferramenta tem uma acurácia de 93,10% na tarefa de divisão sentença em língua portuguesa e possui um excelente *toolkit* em Python para a tarefa. Para a implementação do Classificador de Sentença, foi selecionada a rede BioBERTptRT, treinada conforme descrito na subseção 6.1, para a tarefa. Visto que o processo de treino e avaliação do *k-fold* (com $k=5$) resultou em 5 modelos de BioBERTptRT para a tarefa, foi selecionado o modelo com o melhor resultado entre os cinco. Desta forma, o componente de Estruturador de Sentença recebe como entrada uma evolução completa de um paciente, e, através do Stanza, a evolução é dividida em uma lista de sentenças. Através do *BertTokenizer* e *AutoModelForSequenceClassification* da biblioteca *transformers*, as sentenças são transformadas em tokens e classificadas no padrão SOAP pela rede. Como resultado do componente, tem-se uma lista de tuplas, com a sentença e sua classificação SOAP.

6.4.2 Implementação do Estruturador de Entidade

O componente Estruturador de Entidades é composto por três subcomponentes: Reconhecimento de Entidades, Relacionamento de Entidades e Agrupamento de Entidades. Para os componentes de Reconhecimento e Relacionamento, foram utilizadas as redes treinadas nas subseções 6.2 e 6.3, também selecionando a melhor rede de cada *fold*. Para o componente de Agrupamento, foi implementado um algoritmo baseado em regras. Dessa forma, o componente recebe uma lista de tuplas, contendo uma sentença e sua classificação SOAP.

O Estruturador de Entidades recebe como parâmetro uma lista de tuplas, com a sentença e sua classificação SOAP. Através do *BertTokenizer* e *AutoModelForTokenClassification* da biblioteca *transformers*, as sentenças são transformadas em tokens e enviadas ao Reconhecimento de Entidades, que retorna uma lista contendo a classificação NER de cada token. Os tokens são agrupados a partir do início (B-) e continuidade da entidade (I-), retornando uma lista de tuplas, com a sentença, sua classificação SOAP e uma lista de tuplas de entidades com sua posição e a classe da entidade. Essa saída se torna a entrada do Relacionamento de Entidades, que para

cada tupla de entidades, monta e tokeniza pares de entidade como descrito na Figura 15, que através do *AutoModelForSequenceClassification* identifica se existe relação entre as duplas. A partir das relações, o componente monta lista de conjuntos contendo as entidades agrupadas entre si. Retornando a sentença, sua classificação SOAP e a lista de conjuntos da relação, com suas entidades, classes e posição.

Por fim, as informações são enviadas ao componente de Vinculação de Entidades, que finaliza a estruturação. A Figura 17 apresenta um pseudocódigo do comportamento do componente, ele percorre a lista de sentenças e para cada conjunto de relação, levando em conta a sua classificação SOAP, determina a qual grupo de entidades ele pertence. Após essa determinação, preenche os demais campos. Caso não seja encontrado nenhum grupo de entidade estruturável, é definido que a sentença atingiu o limite da estruturação.

Para realizar a vinculação das tabelas de contexto TUSS 20 (medicamentos), TUSS 22 (procedimentos) e CID-10, inicialmente foi previsto usar a distância de Levenshtein, mas devido às características da nomenclatura de procedimentos, diagnóstico e procedimentos, não se apresentou como a melhor solução. Por exemplo, o medicamento Denason é um remédio para infecções bacterianas e Dunason é um colírio, e entre si possuem uma distância de Levenshtein de 1. Além disto, outro problema comum é a abreviatura de nomes, desta forma, foi adotada uma abordagem de tabela de tradução para realizar a tarefa. Para isso, a partir da base de treino do componente Estruturador de Sentença, foram selecionadas e agrupadas visando unicidade, as entidades de diagnóstico, procedimento, exames e medicamento. Para cada entidade foi associado um ou mais códigos que representam a sua identificação no contexto. A necessidade de múltiplos códigos se deve ao fato de nem sempre ser possível determinar qual código o procedimento ou medicamento pertence ou mesmo possuir mais de uma codificação dependendo da quantidade ou da parte do corpo a que se refere. A Tabela 34 mostra o resultado da associação, indicando o total de entidades únicas, a quantidade de entidades que foi possível vincular e porcentagem de vinculação.

Tabela 34 – Tabelas de contexto para vinculação do componente Estruturar de Sentenças.

Classe/Tabela	Total	Vinculações	Porcentagem (%)
Diagnóstico - CID 10	1343	1012	75,35
Solicitação de Procedimento - TUSS 22	609	592	97,20
Resultado de Exame - TUSS 22	672	582	86,60
Prescrição - TUSS 20	1388	1310	94,38

6.4.3 Base de dados e métricas de avaliação

Para a avaliação final do modelo, foi construída uma nova base de testes, seguindo um processo similar ao das demais bases. Assim, foram selecionadas evoluções de 234.673 registros da base de dados original que ainda não tinham sido utilizadas em nenhuma outra avaliação.

Figura 17 – Pseudo-código do componente de Agrupamento de Entidades.

```
Para cada sentença:
  Para cada conjunto da relação:
    Se a sentença é da classe Objetivo:
      Se conjunto tem "exame" e "resultado":
        conjunto É "Resultado de Exame"
      Se conjunto tem "medicao" e "resultado":
        conjunto É "Medição"
      Caso contrário:
        Se conjunto tem "exameFisico" ou "parteCorpo":
          conjunto É "Exame Físico"
    Se a sentença é da classe Avaliação:
      Se conjunto tem "diagnostico":
        conjunto É "Diagnóstico"
    Se a sentença é da classe Plano:
      Se conjunto tem "solicitacaoProcedimento":
        conjunto É "Solicitação de exame"
      Se conjunto tem "especialidade":
        conjunto É "Encaminhamento"
      Se conjunto tem "medicamento":
        conjunto É "Prescrição"
      Se conjunto tem "tempo":
        conjunto É "Atestado"
    Com base no tipo do conjunto:
      Preecha os demais campos
    Se o tipo tem tabela de Contexto:
      Se a entidade tem equivalente na tabela de Contexto:
        Preecha os códigos de Contexto das entidades
    Se o conjunto não tem tipo:
      Descarta estruturação do conjunto
  Se nenhum conjunto foi estruturado:
    Sentença não pode ser mais estruturada
```

As evoluções foram selecionadas e divididas até atingir um total de 10.013 sentenças, em um conjunto final de 721 evoluções. Diferentemente das bases de dados construídas para os demais treinamentos, as sentenças duplicadas foram mantidas para que a avaliação refletisse a aplicação real do modelo. Utilizando a interface já desenvolvida, todas as sentenças foram classificadas manualmente em suas respectivas classes SOAP e, para as classes Objetivo, Avaliação e Plano, as entidades foram estruturadas quando possível, seguindo o mesmo padrão proposto para o treinamento dos componentes. A Tabela 35 mostra a distribuição da nova rotulação das 10.013 sentenças, das quais 3.226 tiveram as entidades estruturadas, totalizando 32,22%. Considerando apenas as classes que foram estruturadas no segundo nível para avaliação, somam-se 4.960 sentenças, das quais 3.226 tiveram as entidades estruturadas, totalizando 65,00%.

Após essa rotulação, as mesmas 721 evoluções foram processadas pela implementação do modelo, resultando em 10.046 sentenças. A diferença deve-se aos erros do divisor de sentenças, que, em algumas ocasiões, agrupou sentenças que não deveriam ser agrupadas ou dividiu outras incorretamente. Cada sentença das evoluções do padrão ouro foi associada à sua correspondente na saída do modelo, formando o par padrão verdadeiro versus predição para avaliação.

Tabela 35 – Padrão-ouro para avaliação do pipeline de estruturação.

Classe SOAP	Estruturação da sentença	Estruturação da entidade	Porcentagem
Assessment	487	268	55,03%
Objective	3109	2351	75,62%
Plan	1364	607	44,50%
Subjective	5053	0	0,00%
Total	10013	3226	32,22%

6.4.4 Resultados Obtidos

Nesta seção, são apresentados os resultados obtidos a partir das avaliações realizadas nos dois principais componentes da implementação do modelo. Inicialmente é feita a análise do Componente Estruturador de Sentenças, Seguida da avaliação do Componente Estruturador de Entidades e concluindo com a avaliação geral da implementação do modelo.

6.4.4.1 Avaliação do Componente Estruturador de Sentenças

A primeira avaliação realizada focou no componente Estruturador de Sentenças, composto pelo divisor de sentenças e pela classificação de sentenças. A Tabela 36 exhibe os resultados obtidos para cada classe SOAP. Este componente alcançou uma acurácia de 0,9036, uma revocação de 0,9036 e uma medida-F de 0,9264. Conforme esperado, os resultados foram inferiores aos da avaliação isolada do componente de classificação de sentenças, devido à inclusão da camada de divisão de sentenças. A Figura 18 ilustra a matriz de confusão do componente,

evidenciando os erros de divisão. No total, 501 sentenças apresentaram erro de divisão, o que representa 5% do total de sentenças. Embora as médias das classes Subjetivo, Objetivo e Plano tenham sido inferiores em comparação com o componente de classificação de sentenças, a classe Avaliação obteve um resultado superior. Ao analisar as bases de dados e os erros, essa diferença pode ser atribuída principalmente à remoção de duplicatas na fase de treinamento e validação do componente de classificação, resultando em uma base de avaliação mais heterogênea. Em contrapartida, na avaliação do componente estruturador, que manteve as duplicatas, 47% das sentenças apresentaram repetições, o que contribuiu para o melhor desempenho dessa classe.

Tabela 36 – Resultados de desempenho do componente de classificação sentença integrado ao divisor de sentenças.

Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação	Medida-F
Subjetivo	4555	4766	194	498	0,9014	0,9591	0,9014	0,9294
Objetivo	2906	6701	203	203	0,9347	0,9347	0,9347	0,9347
Avaliação	362	9514	12	125	0,7433	0,9679	0,7433	0,8409
Plano	1225	8594	55	139	0,8981	0,9570	0,8981	0,9266
Média (Balanceada)					0,9036	0,9517	0,9036	0,9264
Média (Simples)					0,8693	0,9546	0,8693	0,9079

Figura 18 – Matriz de confusão do componente Estruturador de Sentenças formado pelo divisor de sentença e o classificação sentença.

Valor Real	Subjetivo	4555 45,49%	179 1,79%	7 0,07%	44 0,44%	268 2,68%
	Objetivo	50 0,50%	2906 29,02%	5 0,05%	2 0,02%	146 1,46%
	Avaliação	77 0,77%	13 0,13%	362 3,62%	9 0,09%	26 0,26%
	Plano	67 0,67%	11 0,11%	0 0,00%	1225 12,23%	61 0,61%
		Subjetivo	Objetivo	Avaliação	Plano	Erro na Divisão
		Predição				

6.4.4.2 Avaliação do Componente Estruturador de Entidades

A segunda avaliação foi conduzida no componente Estruturador de Entidades, analisando os resultados da estruturação dos grupos de entidades em cada sentença. Os grupos de estrutura definidos como padrão ouro foram comparados com os resultados obtidos pelo pipeline de estruturação. Assim, identificou-se uma acumulação de possíveis erros no processo de estruturação, que não se limitam ao simples erro de classificação incorreta. Portanto, foram criadas quatro novas classes para representar esses erros, que são:

- **Erro de Sentença:** O erro ocorreu no nível da sentença ou pela falha em sua divisão pela classificação SOAP errada;
- **Não Estruturou:** O componente falhou em estruturar o grupo de entidades, considerando a sentença ou aquele grupo de entidades específicos não é passível de elevação estrutural;
- **Erro de Estruturação:** O componente classificou corretamente o grupo de entidades, mas falhou ao estruturar cada entidade no grupo. Por exemplo: cortou um pedaço da entidade, adicionou um pedaço de entidade de outra ou confundiu uma classe interna (considerou uma posologia, como observação);

A Tabela 38 resume os resultados obtidos na avaliação do componente. O componente alcançou uma acurácia de 0,7464, uma precisão de 0,9849 e uma medida-F de 0,8459. O melhor resultado foi obtido pelo grupo de entidades Medição, com uma medida-F de 0,9415, enquanto o pior resultado foi do grupo Diagnóstico, com uma medida-F de 0,7343. O grupo Atestado, apesar do excelente resultado, teve uma amostra de apenas cinco itens em toda a base de dados, o que pode afetar a precisão dos resultados. Outro aspecto que se destaca é a alta precisão do modelo, com baixa taxa de falsos positivos. Ao analisar individualmente os erros, foi identificado que quando um erro de confusão em alguns dos componentes do pipeline, normalmente resulta em uma não estruturação ou estruturação errada, em vez de uma estruturação correta mas na classe errada.

A matriz de confusão da Figura 19 destaca esse comportamento do componente, erros mais comuns Erro de Sentença (herdado do Estruturar de Sentenças), Não Estruturou ou Erro de Estruturação. Sendo muito incomum a confusão entre as classes, que só ocorreram em classes da mesma classificação SOAP da sentença, por exemplo a classe Exame Físico, só gerou confusão com as classes Medição e Resultado de Exame, que pertencem a mesma classificação SOAP Objetivo e nunca com a classe Diagnóstico que pertence a classe SOAP Avaliação.

Outra avaliação feita no componente foi em relação a vinculação das tabelas de contexto TUSS 20, TUSS 22 e CID-10 em suas respectivas entidades. A Tabela 37 sumariza os resultados encontradas, identificando para cada classe a quantidade de entidades, quantas foram vinculadas e porcentagem de vinculações.

Tabela 37 – Vinculação das tabelas de contexto realizadas pelo componente Estruturar de Sentenças

Grupo/Tabela	Total	Vinculações	Porcentagem (%)
Diagnóstico - CID 10	505	158	31,28
Solicitação de Procedimento - TUSS 22	295	204	69,15
Resultado de Exame - TUSS 22	2019	1037	51,36
Prescrição - TUSS 20	328	210	64,40
Total	3147	1609	51,12

Tabela 38 – Resultados de desempenho do componente Estruturador de Entidades.

Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação Ponderado/Macro	Medida-F
Exame Físico	712	3801	13	179	0,7991	0,9821	0,7991	0,8812
Medição	402	4253	4	46	0,8973	0,9901	0,8973	0,9415
Resultado de Exame	1445	2669	17	574	0,7157	0,9884	0,7157	0,8302
Diagnóstico	293	4200	0	212	0,5802	1,0000	0,5802	0,7343
Atestado	4	4700	0	1	0,8000	1,0000	0,8000	0,8889
Encaminhamento	131	4486	5	83	0,6121	0,9632	0,6121	0,7486
Prescrição	278	4368	9	50	0,8476	0,9686	0,8476	0,9041
Solicitação de Procedimento	247	4402	8	48	0,8373	0,9686	0,8373	0,8982
Média (Balanceada)					0,7464	0,9849	0,7464	0,8459
Média (Simples)					0,7612	0,9826	0,7612	0,8534

6.4.4.3 Avaliação Geral da Implementação do Modelo

Para a avaliação final do modelo, foi conduzida uma análise detalhada do desempenho de sentenças processadas integralmente pela implementação proposta. Essa análise envolveu a comparação entre cada sentença gerada pelo modelo e o padrão-ouro, examinando a correta divisão, classificação conforme o padrão SOAP, e estruturação apropriada das sentenças. Além dos erros relacionados à confusão entre as classes SOAP, foram definidas três categorias adicionais para identificar outros tipos de erros:

- **Erro na Divisão:** Ocorre quando a implementação do modelo falha na correta divisão das sentenças;
- **Não Estruturou:** Indica a falha em estruturar adequadamente um conjunto de entidades;
- **Erro na Estruturação:** Denota um equívoco na estruturação de algum grupo de entidades.

A Tabela 39 apresenta os resultados alcançados pelo modelo na avaliação de todas as classes SOAP. As métricas, calculadas como médias balanceadas, foram: acurácia de 0,8641, precisão de 0,9493, e medida-F de 0,9029. Dentre os resultados, a classe Subjetivo destacou-se com a maior acurácia (0,9014), enquanto a classe Avaliação apresentou o menor desempenho (0,5749). A Figura 20 exibe a matriz de confusão correspondente, incluindo as três categorias de erro adicionalmente definidas. A maioria dos erros foi observada na divisão de sentenças, com 501 ocorrências, seguida por erros de estruturação, com 253 ocorrências.

Figura 19 – Matriz de confusão do componente Estruturador de Entidades com a relação do rótulo verdadeiro e rótulo previsto das classes com o valor numérico e percentual em relação ao total.

Valor Real	Exame Físico	712 15,13%	1 0,02%	7 0,15%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	45 0,96%	100 2,13%	26 0,55%
	Medição	4 0,09%	402 8,54%	10 0,21%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	21 0,45%	3 0,06%	8 0,17%
	Resultado de Exame	9 0,19%	3 0,06%	1445 30,71%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	37 0,79%	274 5,82%	251 5,33%
	Diagnóstico	0 0,00%	0 0,00%	0 0,00%	293 6,23%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	113 2,40%	28 0,60%	71 1,51%
	Atestado	0 0,00%	0 0,00%	0 0,00%	0 0,00%	4 0,09%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	1 0,02%	0 0,00%
	Encaminhamento	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	131 2,78%	6 0,13%	3 0,06%	19 0,40%	2 0,04%	53 1,13%
	Prescrição	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	2 0,04%	278 5,91%	5 0,11%	7 0,15%	10 0,21%	26 0,55%
	Solicitação de Procedimento	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	3 0,06%	3 0,06%	247 5,25%	7 0,15%	18 0,38%	17 0,36%
		Exame Físico -	Medição -	Resultado de Exame -	Diagnóstico -	Atestado -	Encaminhamento -	Prescrição -	Solicitação de Procedimento -	Erro de Sentença -	Não Estruturou -	Erro de Estruturação -
		Predição										

Importante destacar que, como erros de estruturação na classe Subjetivo não foram contabilizados ao nível da entidade, eles não influenciam diretamente as métricas gerais. Esse fato pode potencialmente levar a uma interpretação distorcida do desempenho do modelo. Assim, para a determinação das métricas finais da avaliação da implementação, foram consideradas apenas as classes que englobam ambos os níveis de estruturação: Objetivo, Avaliação, e Plano. A Tabela 40 detalha as métricas finais, resultando em uma acurácia de 0,8354, precisão de 0,9382, e medida-F de 0,8815. Nessa configuração o erro mais comum passa a ser o erro na estruturação dos grupos de entidades com 253 que representam 5,79% das sentenças, seguido do erro na divisão 233 que representam 5,33% das sentenças.

6.4.5 Discussão

A seção detalhou a implementação e avaliação do pipeline de estruturação proposto para a avaliação do modelo, descrevendo a implementação dos dois componentes principais: o Estruturador

Tabela 39 – Resultados de desempenho após a execução completa do Pipeline.

Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação	Medida-F
Subjetivo	4555	4766	194	498	0,9014	0,9591	0,9014	0,9294
Objetivo	2695	6701	203	414	0,8668	0,9300	0,8668	0,8973
Avaliação	280	9514	12	207	0,5749	0,9589	0,5749	0,7189
Plano	1122	8594	55	242	0,8226	0,9533	0,8226	0,8831
Média (Balanceada)					0,8641	0,9493	0,8641	0,9029
Média (Simples)					0,7915	0,9503	0,7915	0,8572

Tabela 40 – Resultados de desempenho após a execução completa do Pipeline, contendo somente as classes com os dois níveis de estruturação.

Classe	Verdadeiro Positivo	Verdadeiro Negativo	Falso Positivo	Falso Negativo	Acurácia	Precisão	Revocação	Medida-F
Objetivo	2695	6701	203	414	0,8668	0,9300	0,8668	0,8973
Avaliação	280	9514	12	207	0,5749	0,9589	0,5749	0,7189
Plano	1122	8594	55	242	0,8226	0,9533	0,8226	0,8831
Média (Balanceada)					0,8354	0,9382	0,8354	0,8815
Média (Simples)					0,7548	0,9474	0,7548	0,8331

de Sentença e o Estruturador de Entidade, bem como a integração de seus subcomponentes.

Quanto à implementação do Estruturador de Sentença, a utilização do Stanza para divisão de sentenças e da rede BioBERTptRT para classificação de sentenças provou ser eficiente. A acurácia combinada dos dois subcomponentes (0,9464 e 0,931) indicava uma performance esperada próxima de 0,8810, mas o componente alcançou uma acurácia de 0,8981, demonstrando que não houve perda significativa na integração dos subcomponentes.

De modo similar, a integração dos subcomponentes do Estruturador de Entidades mostrou-se adequada. Este é composto pelos subcomponentes de Reconhecimento de Entidades, Relacionamento de Entidades e Agrupamento de Entidades. O cálculo do acúmulo de erro para este componente é mais complexo, uma vez que herda erros do componente anterior e nem todas as sentenças são estruturadas no nível da entidade. O erro esperado foi calculado considerando a proporção de sentenças com um segundo nível de estruturação, neste caso, 65%. Assim, o erro esperado do componente é uma proporção do acúmulo de erros do Estruturador de Sentença (0,8810), somado à proporção de erros do Estruturador de Entidade. O acúmulo de erro do Estruturador de Entidade resulta da soma dos erros do Estruturador de Sentença (0,8810) e dos subcomponentes de Reconhecimento de Entidades (0,9207) e Relacionamento de Entidades (0,9325), resultando em um erro acumulado esperado de 0,7564. Contudo, o resultado final do componente foi de 0,8354, evidenciando um bom desempenho na integração dos componentes. Outro fator que pode ter influenciado na diferença de desempenho foi a diferença no tratamento de sentenças repetidas, as quais foram excluídas nos treinamentos e avaliações individuais dos componentes, mas mantidas nas avaliações finais do pipeline para uma visão mais realista da tarefa.

Em conclusão, a implementação e avaliação do pipeline de estruturação validaram o modelo proposto, apresentando um bom desempenho após sua integração.

7 CONCLUSÃO

Essa tese apresentou a modelagem, desenvolvimento e avaliação de um modelo computacional multinível para a estruturação das informações contidas nas evoluções em texto livre do prontuário do paciente, sem perda de informações relevantes.

Levando em conta os requisitos de aplicação dessa proposta, que envolve a migração de um prontuário com evoluções não estruturadas para um prontuário com evoluções estruturadas e semi-estruturadas, foi realizada uma extensa revisão sistemática e análise comparativa. Assim, foi possível identificar as melhores práticas e tecnologias para a implementação do modelo.

Desta forma, o modelo foi proposto como um pipeline, formado por dois componentes principais: o Estruturador de Sentença e o Estruturador de Entidade. O Estruturador de Sentença é composto por dois subcomponentes: Divisor de Sentença e Classificador de Sentenças; e o Estruturador de Entidade, formado pelo Reconhecimento de Entidade, Relacionamento de Entidade e Agrupamento de Entidades.

Para a implementação dos componentes, foi proposta a aplicação de uma série de técnicas e tarefas de PLN, tais como detecção de limites de sentença, classificação de sentenças, reconhecimento de entidades nomeadas, extração de relações e correspondência de ontologias de informação estruturada. Como ferramentas para implementação dessas técnicas, foi proposta a utilização de ferramentas já desenvolvidas ou pré-treinadas, como o Stanza, e as redes BERT, ou a implementação de algoritmos baseados em regras.

Para avaliar a viabilidade do modelo, foi implementado um protótipo do pipeline do modelo, utilizando como base de dados para avaliação e treino das redes BERT uma base de dados cedida pela empresa que motivou o estudo. Esta base, formada por 10.000 prontuários e 234.673 evoluções clínicas, das quais 15.883 foram divididas em 100.021 sentenças, classificadas e estruturadas utilizando uma Interface Gráfica de Usuário (GUI) desenvolvida especificamente para essa tarefa, a fim de formar o padrão ouro usado no estudo. Por fim, após uma série de treinos e avaliações, as melhores redes resultantes para cada componente foram selecionadas e o pipeline foi implementado.

Para a avaliação final, foram utilizados 721 prontuários em total de 10.013 sentenças, que também foram classificadas usando a GUI e foram comparadas os seus resultados com os obtidos do pipeline. Como resultado, ao avaliar todas as classes SOAP, calculadas como médias balanceadas, obteve-se uma acurácia de 0,8641, precisão de 0,9493, e medida-F de 0,9029. Ao avaliar somente as classes que tiveram os dois níveis de estruturação avaliados (Objetivo, Avaliação e Plano), se obteve uma acurácia de 0,8354, precisão de 0,9382, e medida-F de 0,8815.

7.1 Questões de pesquisa

Durante o desenvolvimento do trabalho foi possível responder às questões de pesquisa propostas na motivação. Abaixo são retomadas essas questões e suas respostas:

- *Como modelar, desenvolver e avaliar um modelo computacional multinível para a estruturação das informações contidas nas evoluções em texto livre do prontuário do paciente, sem perda de informações relevantes?* A questão principal da pesquisa foi respondida durante todo o desenvolvimento do trabalho. Mais especificamente, o Capítulo 4 apresenta o modelo multinível proposto e os Capítulos 5 e 6, o desenvolvimento e avaliação de um protótipo do modelo para validar a viabilidade de sua implementação;
- *Quais são as principais técnicas e ferramentas utilizadas para a estruturação da informação?* Essa pergunta foi respondida através da revisão sistemática realizada no Capítulo 3, onde foi feita uma extensa revisão do estado da arte da estruturação da informação. Destacando como principais técnicas utilizadas as redes neurais artificiais e algoritmos baseados em regras, e como principais ferramentas: SciSpaCy, MetaMap, cTAKES, e redes neurais Bi-LSTM, CNN, CRF e BERT;
- *Como aplicar essas técnicas e ferramentas para a estruturação multinível do modelo proposto?* Em resposta a essa questão, o Capítulo 5 apresenta as propostas para aplicação destas técnicas e ferramentas no desenvolvimento do modelo, e o Capítulo 6 apresenta a aplicação de algumas delas no desenvolvimento do protótipo. Como exemplo, a utilização do Stanza para divisão de sentença e a utilização das redes BERT para implementação dos componentes: Classificador de Sentença, Reconhecimento de Entidade e Relacionamento de Entidade;
- *Qual o nível de estruturação que o modelo consegue atingir?* Essa pergunta é respondida pela avaliação geral da implementação do modelo, apresentada na subseção 6.4.4. O protótipo conseguiu estruturar no nível da sentença 100% da informação relevante com uma acurácia balanceada de 90,36% e conseguiu estruturar no nível dos grupos de entidade 65% das sentenças, com uma acurácia balanceada de 83,54%, considerando somente as classes com dois níveis de estruturação.

7.2 Contribuições

A principal contribuição desta tese é o modelo multinível de estruturação da evolução do paciente. Além do modelo em si, que pode ser aplicado para soluções de problemas de estruturação similares, a implementação já desenvolvida pode ser aplicada para a estruturação de outras bases de dados que tenham a necessidade de estruturação similar ou re-treinada para se adaptar a outros padrões de classificação de sentenças ou grupos de interesse.

Além destas contribuições esse trabalho obteve outras contribuições marginais, que são descritas abaixo:

- A revisão sistemática apresentada no capítulo 3 trouxe uma série de contribuições para o entendimento do estado da arte da estruturação de informações em prontuários eletrônicos,

assim como seus desafios para aplicação, principalmente em língua portuguesa;

- O modelo de rede neural BERT para classificação de sentenças no padrão SOAP proposto no capítulo 6 apresenta dois aspectos de ineditismos: é o primeiro estudo a utilizar um modelo BERT para classificação de sentenças no padrão SOAP; e é primeiro estudo a propor um modelo de classificação de sentenças no padrão SOAP em língua portuguesa;
- O processo de busca da melhor rede BERT para tarefa de classificação de sentenças, fez achados relevantes sobre o retreino dos modelos BERT e os ganhos obtidos com aumento de sua especialização, esses achados trazem contribuição para futuros trabalhos na área;
- A BioBERTptRT, que foi a rede retreinada a partir da base de dados utilizada, foi outra contribuição deste trabalho. Ela representa o estado da arte em redes BERTs para essa base de dados, sendo o modelo mais especializado, possuindo conhecimento específico da língua portuguesa, terminologia clínica e o contexto do grupo médico, e apresenta os melhores resultados nas tarefas avaliadas nessa base de dados;
- A GUI desenvolvida para a criação do padrão-ouro neste trabalho pode ser facilmente adaptada para estruturar outras informações, tanto na área de saúde quanto em outras áreas que necessitem desse tipo de estruturação. Ela foi desenvolvida de forma que somente um arquivo de configuração define quais são as classes para classificação de sentenças e quais são os grupos e entidades relevantes. Além disso, a ferramenta já exporta automaticamente as informações para o treinamento nas três redes;
- Da mesma forma os componentes Estruturador de Sentença e Estruturador de Entidade podem ser usados individualmente para tarefas similares em outros modelos.

Das contribuições apresentadas acima, duas foram publicadas e avaliadas por pares: uma é uma revisão sistemática e a outra, um trabalho inédito para classificação SOAP em notas clínicas de texto livre usando modelos de linguagem pré-treinados específicos do domínio. Abaixo, são descritas essas publicações:

- OLIVEIRA, J. M. de; COSTA, C. A. da; ANTUNES, R. S. Data structuring of electronic health records: a systematic review. *Health and Technology*, [S.l.], v. 11, n. 6, p. 1219–1235, 2021;
- OLIVEIRA, J. M. de; COSTA, C. A. da; ANTUNES, R. S. SOAP classifier for free-text clinical notes with domain-specific pre-trained language models. *Expert Systems with Applications*, 2024.

7.3 Limitações

Visto que a maioria dos componentes do modelo utiliza aprendizado de máquina a partir de treinamento, seu desempenho pode ser afetado pelas características dos dados usados. Desta

forma, a base de dados utilizada para treino e avaliação acaba por ser uma limitação no trabalho. Apesar da extensa base de dados adquirida, ela representa apenas um grupo médico, sendo preenchida por 83 médicos de dez especialidades diferentes. Seria necessário avaliar a implementação do modelo em outras bases de dados para aferir seus resultados nessas bases. Outro aspecto que traz limitações em relação à base de dados é que, por ser uma base privada, o acesso à mesma é limitado, o que reduziu o número de pessoas disponíveis para realizar a rotulagem dos dados. Desta forma, só foi possível rotular 100.000 sentenças no padrão SOAP e 51.237 sentenças no segundo nível de estruturação. Além disso, alguns grupos de entidades, como o grupo Atestado, apresentaram uma baixa amostra, o que pode afetar o treinamento e a avaliação da rede.

Outra limitação é o nível de estruturação de algumas entidades, apesar de o modelo ter estruturado com sucesso as entidades, algumas apresentam uma granularidade maior em suas estruturas. Um bom exemplo são as entidades da classe posologia, que contêm informações sobre o modo de usar o medicamento e possuem informações em suas entidades que poderiam ser estruturadas em mais um nível. Por exemplo, as posologias: "25 MG/DIA", "5mg 1cp se necessário", "30mg 2cps antes do café", "conforme prescrição"; poderiam ser estruturadas com uma granularidade maior, separando entidades como quantidade, espaço temporal e condicionais. Devido à complexidade e variabilidade destas estruturas, além da dificuldade em sua rotulação, optou-se por manter esses aspectos das posologias em uma mesma entidade.

A implementação do componente Agrupamento de Entidade foi desenvolvida através de um algoritmo baseado em regras, desta forma, qualquer alteração nos grupos de interesse requer uma alteração do algoritmo, diferentemente dos componentes que utilizam as redes neurais, os quais só precisam ser retreinados a partir da nova base de dados.

Além destas limitações, a implementação do modelo apresentou uma dificuldade maior ao classificar a classe SOAP Avaliação e os grupos de entidades Encaminhamento e Diagnóstico. Outro processo da implementação que apresentou um baixo desempenho foi a vinculação com as tabelas de contexto, apenas 51,12% das entidades passíveis de vinculação foram vinculadas às suas tabelas de domínio.

7.4 Trabalhos Futuros

A partir das conclusões, achados e limitações, foram identificadas diversas oportunidades para o aprimoramento e extensões que podem ser abordadas em trabalhos futuros. Abaixo são listadas essas oportunidades:

- **Ampliação da base de treinamento e avaliação:** Realizar a rotulação do padrão-ouro dos 48.784 itens do segundo nível de estruturação dos itens já classificados no primeiro nível e ampliar o número de sentenças para avaliar o ganho de desempenho em relação aos resultados obtidos;
- **Aplicar e avaliar o modelo no processo de migração da base de dados:** Utilizar o

protótipo para estruturar todos os registros da base de dados e importar para o novo sistema da empresa que forneceu os dados do trabalho. Após a importação, avaliar com os usuários do novo sistema os resultados obtidos;

- **Avaliar a aplicação do modelo em outras bases de dados:** Avaliar tanto a aplicação do modelo já treinado em bases de dados de outros grupos médicos, como o retreinamento desses modelos usando essas bases de dados. Buscando responder se o modelo necessita de retreinamento ao mudar de base ou se é robusto o suficiente para ser aplicado diretamente em outras bases;
- **Adicionar um novo nível de estruturação no modelo:** Modelar e avaliar um novo componente para lidar com a granularidade de algumas entidades, como por exemplo a posologia, permitindo estruturar suas demais informações, como quantidade, intervalo temporal e condicionais;
- **Automatizar o processo de treinamento e construção do modelo a partir da GUI de rotulação:** Propor e avaliar uma nova abordagem para implementação do componente de Agrupamento de Entidades, que possa ser treinado ou autoconfigurado para reconhecer e estruturar os grupos de entidades a partir da configuração da GUI de rotulação ou dados rotulados, sem a necessidade de reescrever o código;
- **Buscar novas abordagens e técnicas para a vinculação com as tabelas de contexto:** A utilização de tabelas de tradução mostrou-se ineficiente; além do baixo desempenho, torna-se necessário o cadastro manual. Desta forma, torna-se importante buscar novas abordagens e técnicas que melhorem o desempenho e automatizem o processo.

REFERÊNCIAS

- ADENIYI, A. O. et al. The impact of electronic health records on patient care and outcomes: a comprehensive review. **World Journal of Advanced Research and Reviews**, [S.l.], v. 21, n. 2, p. 1446–1455, 2024.
- AFSHAR, M. et al. Deployment of real-time natural language processing and deep learning clinical decision support in the electronic health record: pipeline implementation for an opioid misuse screener in hospitalized adults. **JMIR Med Inform**, [S.l.], v. 11, p. e44977, Apr 2023.
- ALDRAIMLI, M. et al. A data science approach for early-stage prediction of patient's susceptibility to acute side effects of advanced radiotherapy. **Computers in Biology and Medicine**, [S.l.], v. 135, p. 104624, 2021.
- ALMEIDA, J. R.; MATOS, S. Rule-based extraction of family history information from clinical notes. In: ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING, 35., 2020. **Proceedings...** ACM, 2020.
- ALODADI, M. S.; JANEJA, V. P. Clinical entities association rules (clear): untangling clinical notes in electronic health records. In: IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 2620–2627.
- ALSENTZER, E. et al. **Publicly available clinical bert embeddings**. 2019.
- ALVAREZ, E. E.; REINHART, J. M. Use of an interactive online teaching module improved students' ability to write a clinically appropriate soap note. **Journal of Veterinary Medical Education**, [S.l.], v. 47, n. 6, p. 700–708, 2020.
- AMIN, S. et al. Mlt-dfki at clef ehealth 2019: multi-label classification of icd-10 codes with bert. In: CLEF (WORKING NOTES), 2019. **Anais...** [S.l.: s.n.], 2019. p. 1–15.
- AMMAR, W. et al. Construction of the literature graph in semantic scholar. **arXiv preprint arXiv:1805.02262**, [S.l.], 2018.
- ARONSON, A. R.; LANG, F.-M. An overview of metamap: historical perspective and recent advances. **Journal of the American Medical Informatics Association**, [S.l.], v. 17, n. 3, p. 229–236, 2010.
- BALABAEVA, K.; KOVALCHUK, S. Experiencer detection and automated extraction of a family disease tree from medical texts in russian language. In: COMPUTATIONAL SCIENCE – ICCS 2020, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 603–612.
- BAMPA, M.; DALIANIS, H. Detecting adverse drug events from swedish electronic health records using text mining. In: LREC 2020 WORKSHOP ON MULTILINGUAL BIOMEDICAL TEXT PROCESSING (MULTILINGUALBIO 2020), 2020. **Proceedings...** [S.l.: s.n.], 2020. p. 1–8.
- BELTAGY, I.; LO, K.; COHAN, A. Scibert: a pretrained language model for scientific text. **arXiv preprint arXiv:1903.10676**, [S.l.], 2019.

BLANCO, A. et al. Multi-label clinical document classification: impact of label-density. **Expert Systems with Applications**, [S.l.], v. 138, p. 112835, dec 2019.

BLINOV, P. et al. Predicting clinical diagnosis from patients electronic health records using BERT-based neural networks. In: **Artificial intelligence in medicine**. [S.l.]: Springer International Publishing, 2020. p. 111–121.

BOSHNAKA, H.; ABDELGABERB, S.; AMANYABDOC, E. Y. Ontology-based knowledge modelling for clinical data representation in electronic health records. **International Journal of Computer Science and Information Security (IJCSIS)**, [S.l.], v. 16, n. 10, 2018.

BOYTCHEVA, S. Indirect association rules mining in clinical texts. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE: METHODOLOGY, SYSTEMS, AND APPLICATIONS, 2018. **Anais...** [S.l.: s.n.], 2018. p. 36–47.

BREISCHNEIDER, C. et al. Automatic extraction of breast cancer information from clinical reports. In: IEEE 30TH INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS (CBMS), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 213–218.

BROWNE, A. C. et al. Umls language and vocabulary tools: amia 2003 open source expo. In: AMIA ANNUAL SYMPOSIUM PROCEEDINGS, 2003. **Anais...** [S.l.: s.n.], 2003. v. 2003, p. 798.

BUDGEN, D.; BRERETON, P. Performing systematic literature reviews in software engineering. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, 28., 2006, New York, NY, USA. **Proceedings...** Association for Computing Machinery, 2006. p. 1051–1052. (ICSE '06).

CAI, T. et al. EXTraction of EMR numerical data: an efficient and generalizable tool to EXTEND clinical research. **BMC Medical Informatics and Decision Making**, [S.l.], v. 19, n. 1, Nov. 2019.

CAI, T.; ZHOU, Y.; ZHENG, H. Cost-quality adaptive active learning for chinese clinical named entity recognition. **arXiv preprint arXiv:2008.12548**, [S.l.], 2020.

CAMERON, S.; TURTLE-SONG, I. Learning to write case notes using the soap format. **Journal of Counseling & Development**, [S.l.], v. 80, n. 3, p. 286–292, 2002.

CASTILLO, V. H. et al. An interaction framework for supporting the adoption of ehrs by physicians. **Universal Access in the Information Society**, [S.l.], v. 18, n. 2, p. 399–412, 2019.

CESAR, C. L. G. et al. Uso da classificação internacional de doenças em inquéritos de saúde. **Revista Brasileira de Epidemiologia**, [S.l.], v. 4, p. 120–129, 2001.

CHAKRAVARTHI, B. R. et al. Corpus creation for sentiment analysis in code-mixed tamil-english text. **arXiv preprint arXiv:2006.00206**, [S.l.], 2020.

CHEN, L. et al. Using natural language processing to extract clinically useful information from chinese electronic medical records. **International Journal of Medical Informatics**, [S.l.], v. 124, p. 6 – 12, 2019.

CHEN, R.; HO, J. C.; LIN, J.-M. S. Extracting medication information from unstructured public health data: a demonstration on data from population-based and tertiary-based samples. **BMC Medical Research Methodology**, [S.l.], v. 20, n. 1, Oct. 2020.

- CHEN, Y. et al. Named entity recognition from chinese adverse drug event reports with lexical feature based bilstm-crf and tri-training. **Journal of Biomedical Informatics**, [S.l.], v. 96, p. 103252, 2019.
- CHENG, M. et al. A hybrid method to extract clinical information from chinese electronic medical records. **IEEE Access**, [S.l.], v. 7, p. 70624–70633, 2019.
- CHOWDHURY, S. et al. A multitask bi-directional RNN model for named entity recognition on chinese electronic medical records. **BMC Bioinformatics**, [S.l.], v. 19, n. S17, dec 2018.
- CHUANG, Y.-N. et al. Spec: a soft prompt-based calibration on performance variability of large language model in clinical notes summarization. **Journal of Biomedical Informatics**, [S.l.], v. 151, p. 104606, 2024.
- COLMENAREJO, G. Machine learning models to predict childhood and adolescent obesity: a review. **Nutrients**, [S.l.], v. 12, n. 8, p. 2466, 2020.
- DAI, H.-J.; SU, C.-H.; WU, C.-S. Adverse drug event and medication extraction in electronic health records via a cascading architecture with different sequence labeling models and word embeddings. **Journal of the American Medical Informatics Association**, [S.l.], v. 27, n. 1, p. 47–55, jul 2019.
- DALE, R.; MOISL, H.; SOMERS, H. **Handbook of natural language processing**. [S.l.]: CRC press, 2000.
- DANDALA, B. et al. Extraction of information related to drug safety surveillance from electronic health record notes: joint modeling of entities and relations using knowledge-aware neural attentive models. **JMIR Medical Informatics**, [S.l.], v. 8, n. 7, 2020.
- DASH, S. et al. **Handbook of research on modeling, analysis, and application of nature-inspired metaheuristic algorithms**. 1st. ed. USA: IGI Global, 2017.
- DE SILVA, K. et al. Clinical notes as prognostic markers of mortality associated with diabetes mellitus following critical care: a retrospective cohort analysis using machine learning and unstructured big data. **Computers in Biology and Medicine**, [S.l.], v. 132, p. 104305, 2021.
- DENG, L.; LIU, Y. **Deep learning in natural language processing**. [S.l.]: Springer, 2018.
- DESHPANDE, S.; PALSHIKAR, G. K.; ATHIAPPAN, G. An unsupervised approach to sentence classification. In: COMAD, 2010. **Anais...** [S.l.: s.n.], 2010. p. 88.
- DEVLIN, J. et al. Bert: pre-training of deep bidirectional transformers for language understanding. **CoRR**, [S.l.], v. abs/1810.04805, 2018.
- DEVLIN, J. et al. Bert: pre-training of deep bidirectional transformers for language understanding. In: CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: HUMAN LANGUAGE TECHNOLOGIES, VOLUME 1 (LONG AND SHORT PAPERS), 2019., 2019, Minneapolis, Minnesota. **Proceedings...** Association for Computational Linguistics, 2019. p. 4171–4186.
- DONG, X. et al. Deep learning for named entity recognition on chinese electronic medical records: combining deep transfer learning with multitask bi-directional LSTM RNN. **PLOS ONE**, [S.l.], v. 14, n. 5, p. e0216046, may 2019.

DONNELLY, K. Snomed-ct: the advanced terminology and coding system for ehealth. **Studies in health technology and informatics**, [S.l.], v. 121, p. 279, 2006.

DU, M.; PANG, M.; XU, B. Multi-task learning for attribute extraction from unstructured electronic medical records. In: SEMANTIC TECHNOLOGY, 2020, Singapore. **Anais...** Springer Singapore, 2020. p. 117–128.

EL RIFAI, O. et al. Blockchain-based personal health records for patients' empowerment. In: RESEARCH CHALLENGES IN INFORMATION SCIENCE, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 455–471.

FEATURE EXTRACTION METHOD FROM ELECTRONIC HEALTH RECORDS IN RUSSIA, 2020. **Anais...** [S.l.: s.n.], 2020. n. 26, p. 497–500.

FENG, J.; SHAIB, C.; RUDZICZ, F. Explainable clinical decision support from text. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP), 2020., 2020. **Proceedings...** [S.l.: s.n.], 2020. p. 1478–1489.

FENG, X. et al. Effective deep memory networks for distant supervised relation extraction. In: IJCAI, 2017. **Anais...** [S.l.: s.n.], 2017. v. 17.

FONFERKO-SHADRACH, B. et al. Using natural language processing to extract structured epilepsy data from unstructured clinic letters: development and validation of the exact (extraction of epilepsy clinical text) system. **BMJ Open**, [S.l.], v. 9, n. 4, 2019.

FORMAN, G. et al. An extensive empirical study of feature selection metrics for text classification. **J. Mach. Learn. Res.**, [S.l.], v. 3, n. Mar, p. 1289–1305, 2003.

FRIEDMAN, C. et al. Automated encoding of clinical documents based on natural language processing. **Journal of the American Medical Informatics Association**, [S.l.], v. 11, n. 5, p. 392–402, 2004.

FU, S. et al. Clinical concept extraction: a methodology review. **Journal of Biomedical Informatics**, [S.l.], p. 103526, 2020.

FU, T.-J.; LI, P.-H.; MA, W.-Y. Graphrel: modeling text as relational graphs for joint entity and relation extraction. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 57., 2019. **Proceedings...** [S.l.: s.n.], 2019. p. 1409–1418.

GALLANT, S. I.; GALLANT, S. I. **Neural network learning and expert systems**. [S.l.]: MIT press, 1993.

GALVÃO, M. C. B.; RICARTE, I. L. M. A classificação internacional de doenças e problemas relacionados à saúde (cid-11): características, inovações e desafios para implementação. **Asklepion: Informação em Saúde**, [S.l.], v. 1, n. 1, p. 104–118, 2021.

GOLDBERG, Y.; HIRST, G. **Neural network methods in natural language processing**. [S.l.]: Morgan & Claypool Publishers, 2017.

GORRELL, G.; SONG, X.; ROBERTS, A. Bio-yodie: a named entity linking system for biomedical text. **arXiv preprint arXiv:1811.04860**, [S.l.], 2018.

GUBERT, L.; COSTA, C. André da; RIGHI, R. Context awareness in healthcare: a systematic literature review. **Universal Access in the Information Society**, [S.l.], v. 19, 06 2020.

HAN, X.; RUONAN, R. The method of medical named entity recognition based on semantic model and improved svm-knn algorithm. In: SEVENTH INTERNATIONAL CONFERENCE ON SEMANTICS, KNOWLEDGE AND GRIDS, 2011., 2011. **Anais...** [S.l.: s.n.], 2011. p. 21–27.

HARRIS, Z. S. Distributional structure. **WORD**, [S.l.], v. 10, n. 2-3, p. 146–162, 1954.

HOUSHMAND, M.; KHORRAM, E. Similarity aggregation in ontology matching based on reliability maximization. In: IRANIAN CONFERENCE ON ELECTRICAL ENGINEERING, 2011., 2011. **Anais...** [S.l.: s.n.], 2011. p. 1–1.

HU, R. et al. Efficient hardware architecture of softmax layer in deep neural network. In: IEEE 23RD INTERNATIONAL CONFERENCE ON DIGITAL SIGNAL PROCESSING (DSP), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–5.

HUANG, C. et al. Transitions from one electronic health record to another: challenges, pitfalls, and recommendations. **Applied clinical informatics**, [S.l.], v. 11, n. 05, p. 742–754, 2020.

HUANG, H.-L.; HONG, S.-H.; TSAI, Y.-C. Approaches to text mining for analyzing treatment plan of quit smoking with free-text medical records: a prisma-compliant meta-analysis. **Medicine**, [S.l.], v. 99, n. 29, 2020.

ILAN, R. et al. Handover patterns: an observational study of critical care physicians. **BMC health services research**, [S.l.], v. 12, n. 1, p. 1–10, 2012.

IQBAL, E. et al. ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. **PLOS ONE**, [S.l.], v. 12, n. 11, p. e0187121, Nov. 2017.

JI, J.; CHEN, B.; JIANG, H. Fully-connected LSTM–CRF on medical concept extraction. **International Journal of Machine Learning and Cybernetics**, [S.l.], v. 11, n. 9, p. 1971–1979, Feb. 2020.

JIANG, J. Information extraction from text. **Mining text data**, [S.l.], p. 11–41, 2012.

JOUFFROY, J. et al. MedExt: combining expert knowledge and deep learning for medication extraction from french clinical texts (preprint). **JMIR Med Inform**, [S.l.], jan 2020.

KAIESKI, N. et al. Application of artificial intelligence methods in vital signs analysis of hospitalized patients: a systematic literature review. **Applied Soft Computing**, [S.l.], p. 106612, 2020.

KEELE, S. et al. **Guidelines for performing systematic literature reviews in software engineering**. [S.l.]: Technical report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.

KENEI, J. K. et al. Deep CNN with residual connections and range normalization for clinical text classification. **Computer Science and Information Technology**, [S.l.], v. 7, n. 4, p. 111–127, aug 2019.

KERSLOOT, M. G. et al. Automated SNOMED CT concept and attribute relationship detection through a web-based implementation of cTAKES. **Journal of Biomedical Semantics**, [S.l.], v. 10, n. 1, Sept. 2019.

KERSLOOT, M. G. et al. Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies. **Journal of biomedical semantics**, [S.l.], v. 11, n. 1, p. 1–21, 2020.

KIM, J.; CHOE, Y.; MUELLER, K. Extracting clinical relations in electronic health records using enriched parse trees. **Procedia Computer Science**, [S.l.], v. 53, p. 274–283, 2015. INNS Conference on Big Data 2015 Program San Francisco, CA, USA 8-10 August 2015.

KIM, M. C. et al. Mapping scientific landscapes in umls research: a scientometric review. **Journal of the American Medical Informatics Association**, [S.l.], v. 27, n. 10, p. 1612–1624, 2020.

KOLECK, T. A. et al. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. **Journal of the American Medical Informatics Association**, [S.l.], v. 26, n. 4, p. 364–379, 2019.

KORMILITZIN, A. et al. Med7: a transferable clinical natural language processing model for electronic health records. **CoRR**, [S.l.], 2020.

KRALJEVIC, Z. et al. Multi-domain clinical natural language processing with medcat: the medical concept annotation toolkit. **arXiv preprint arXiv:2010.01165**, [S.l.], 2020.

KUNDETI, S. R. et al. Clinical named entity recognition: challenges and opportunities. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 2016., 2016. **Anais...** [S.l.: s.n.], 2016. p. 1937–1945.

LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: EIGHTEENTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2001, San Francisco, CA, USA. **Proceedings...** Morgan Kaufmann Publishers Inc., 2001. p. 282–289. (ICML '01).

LAMMERS, E. J.; ZHENG, K. Characteristics associated with hospital health it vendor switching and dropping. In: AMIA ANNUAL SYMPOSIUM PROCEEDINGS, 2011. **Anais...** [S.l.: s.n.], 2011. v. 2011, p. 742.

LAMY, M. et al. Extracting clinical knowledge from electronic medical records. **Extracting clinical knowledge from electronic medical records**, [S.l.], n. 3, p. 488–493, 2018.

LAMY, M. et al. Extracting clinical information from electronic medical records. In: INTERNATIONAL SYMPOSIUM ON AMBIENT INTELLIGENCE, 2018. **Anais...** [S.l.: s.n.], 2018. p. 113–120.

LAURENTI, R. Análise da informação em saúde: 1893-1993, cem anos da classificação internacional de doenças. **Revista de saúde pública**, [S.l.], v. 25, p. 407–417, 1991.

LE GLAZ, A. et al. Machine learning and natural language processing in mental health: systematic review. **J Med Internet Res**, [S.l.], v. 23, n. 5, p. e15708, May 2021.

LEE, J. et al. Biobert: a pre-trained biomedical language representation model for biomedical text mining. **Bioinformatics**, [S.l.], v. 36, n. 4, p. 1234–1240, 09 2019.

- LEE, J. Y.; DERNONCOURT, F.; SZOLOVITS, P. MIT at semeval-2017 task 10: relation extraction with convolutional neural networks. **CoRR**, [S.l.], v. abs/1704.01523, 2017.
- LEE, W. et al. Conditional random fields for clinical named entity recognition: a comparative study using korean clinical texts. **Computers in Biology and Medicine**, [S.l.], v. 101, p. 7 – 14, 2018.
- LEITER, R. E. et al. Deep natural language processing to identify symptom documentation in clinical notes for patients with heart failure undergoing cardiac resynchronization therapy. **Journal of Pain and Symptom Management**, [S.l.], v. 60, n. 5, p. 948 – 958.e3, 2020.
- LERNER, I.; PARIS, N.; TANNIER, X. Terminologies augmented recurrent neural network model for clinical named entity recognition. **Journal of Biomedical Informatics**, [S.l.], v. 102, p. 103356, 2020.
- LI, J. et al. A comparative study of pre-trained language models for named entity recognition in clinical trial eligibility criteria from multiple corpora. **BMC medical informatics and decision making**, [S.l.], v. 22, n. 3, p. 1–10, 2022.
- LI, L. et al. An attention-based deep learning model for clinical named entity recognition of chinese electronic medical records. **BMC Medical Informatics and Decision Making**, [S.l.], v. 19, n. 5, p. 1–11, 2019.
- LI, P. et al. Improving relation extraction with knowledge-attention. **arXiv preprint arXiv:1910.02724**, [S.l.], 2019.
- LI, P. et al. Medical knowledge extraction and analysis from electronic medical records using deep learning. **Chinese Medical Sciences Journal**, [S.l.], v. 34, n. 2, p. 133 – 139, 2019.
- LI, P.; MAO, K. Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. **Expert Systems with Applications**, [S.l.], v. 115, p. 512–523, 2019.
- LI, Y. et al. Chinese clinical named entity recognition in electronic medical records: development of a lattice long short-term memory model with contextualized character representations. **JMIR Medical Informatics**, [S.l.], v. 8, n. 9, p. e19848, sep 2020.
- LI, Y. et al. Towards chinese clinical named entity recognition by dynamic embedding using domain-specific knowledge. **Journal of Biomedical Informatics**, [S.l.], v. 106, p. 103435, jun 2020.
- LI, Z. et al. A system for automatically extracting clinical events with temporal information. **BMC Medical Informatics and Decision Making**, [S.l.], v. 20, n. 1, Aug. 2020.
- LISENBY, K. M. et al. Ambulatory care preceptors' perceptions on soap note writing in advanced pharmacy practice experiences (appes). **Currents in Pharmacy Teaching and Learning**, [S.l.], v. 10, n. 12, p. 1574–1578, 2018.
- LIU, K. et al. Named entity recognition in chinese electronic medical records based on CRF. In: **WEB INFORMATION SYSTEMS AND APPLICATIONS CONFERENCE (WISA)**, 2017., 2017. **Anais...** IEEE, 2017.
- LIU, S. et al. Rxnorm: prescription for electronic drug information exchange. **IT professional**, [S.l.], v. 7, n. 5, p. 17–23, 2005.

- LIU, S. et al. An automated approach for clinical quantitative information extraction from chinese electronic medical records. In: **Health information science**. [S.l.]: Springer International Publishing, 2018. p. 98–109.
- LIU, X. et al. Ontology matching: state of the art, future challenges, and thinking based on utilized information. **IEEE Access**, [S.l.], v. 9, p. 91235–91243, 2021.
- LOPES, F.; TEIXEIRA, C.; OLIVEIRA, H. G. Named entity recognition in portuguese neurology text using crf. In: EPIA CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2019. **Anais...** [S.l.: s.n.], 2019. p. 336–348.
- LOPES, F.; TEIXEIRA, C.; OLIVEIRA, H. G. Comparing different methods for named entity recognition in portuguese neurology text. **Journal of Medical Systems**, [S.l.], v. 44, n. 4, feb 2020.
- LU, N. et al. Chinese clinical named entity recognition with word-level information incorporating dictionaries. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2019., 2019. **Anais...** IEEE, 2019.
- LUO, Y. Recurrent neural networks for classifying relations in clinical notes. **Journal of Biomedical Informatics**, [S.l.], v. 72, p. 85–95, aug 2017.
- LUO, Y. et al. Segment convolutional neural networks (seg-cnns) for classifying relations in clinical notes. **Journal of the American Medical Informatics Association**, [S.l.], v. 25, n. 1, p. 93–98, 2018.
- MACIEL, D. A.; FERREIRA, D. P.; FÁTIMA MARIN, H. de. Padrões de terminologias nacionais para procedimentos e intervenções na saúde. **Revista de Administração em Saúde**, [S.l.], v. 18, n. 71, 2018.
- MAGUIRE, F. B. et al. A text-mining approach to obtain detailed treatment information from free-text fields in population-based cancer registries: a study of non-small cell lung cancer in california. **PLOS ONE**, [S.l.], v. 14, n. 2, p. e0212454, feb 2019.
- MANNING, C.; SCHUTZE, H. **Foundations of statistical natural language processing**. [S.l.]: MIT press, 1999.
- MANSOURI, A.; AFFENDEY, L. S.; MAMAT, A. Named entity recognition approaches. **International Journal of Computer Science and Network Security**, [S.l.], v. 8, n. 2, p. 339–344, 2008.
- MAXIMILIAN, Z.; J., B. O.; MICHAEL, M. Using openehr archetypes for automated extraction of numerical information from clinical narratives. **Studies in Health Technology and Informatics**, [S.l.], v. 267, n. German Medical Data Sciences: Shaping Change – Creative Solutions for Innovative Medicine, p. 156–163, 2019.
- MCCAWLEY, J. D. Concerning the base component of a transformational grammar. **Foundations of language**, [S.l.], p. 243–269, 1968.
- MIAKE-LYE, I. M. et al. Transitioning from one electronic health record to another: a systematic review. **Journal of General Internal Medicine**, [S.l.], v. 38, n. Suppl 4, p. 956–964, 2023.

MOEN, H. et al. Evaluation of a prototype system that automatically assigns subject headings to nursing narratives using recurrent neural network. In: NINTH INTERNATIONAL WORKSHOP ON HEALTH TEXT MINING AND INFORMATION ANALYSIS, 2018. **Proceedings...** Association for Computational Linguistics, 2018.

MOEN, H. et al. Supporting the use of standardized nursing terminologies with automatic subject heading prediction: a comparison of sentence-level text classification methods. **Journal of the American Medical Informatics Association**, [S.l.], v. 27, n. 1, p. 81–88, oct 2019.

MOEN, H. et al. Assisting nurses in care documentation: from automated sentence classification to coherent document structures with subject headings. **Journal of Biomedical Semantics**, [S.l.], v. 11, n. 1, sep 2020.

MOWERY, D. et al. Building an automated soap classifier for emergency department reports. **Journal of Biomedical Informatics**, [S.l.], v. 45, n. 1, p. 71–81, 2012.

MUJTABA, G. et al. Clinical text classification research trends: systematic literature review and open issues. **Expert systems with applications**, [S.l.], v. 116, p. 494–520, 2019.

MUNKHDALAI, T.; LIU, F.; YU, H. Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning. **JMIR Public Health and Surveillance**, [S.l.], v. 4, n. 2, p. e29, apr 2018.

NATARAJAN, S. et al. Markov logic networks for adverse drug event extraction from text. **Knowledge and Information Systems**, [S.l.], v. 51, n. 2, p. 435–457, Aug. 2016.

NEUMANN, M. et al. Scispacy: fast and robust models for biomedical natural language processing. **arXiv preprint arXiv:1902.07669**, [S.l.], 2019.

NÉVÉOL, A. et al. Clinical natural language processing in languages other than english: opportunities and challenges. **Journal of biomedical semantics**, [S.l.], v. 9, n. 1, p. 12, 2018.

NEVES, K. et al. Benefícios e desvantagens da implementação do prontuário eletrônico do paciente para o serviço de saúde. **Research, Society and Development**, [S.l.], v. 9, p. 735974630, 06 2020.

NGUYEN, T. H.; GRISHMAN, R. Relation extraction: perspective from convolutional neural networks. In: OF THE 1ST WORKSHOP ON VECTOR SPACE MODELING FOR NATURAL LANGUAGE PROCESSING, 2015. **Proceedings...** [S.l.: s.n.], 2015. p. 39–48.

NORGEOT, B. et al. Protected health information filter (philter): accurately and securely de-identifying free-text clinical notes. **NPJ digital medicine**, [S.l.], v. 3, n. 1, p. 1–8, 2020.

NUTHAKKI, S. et al. Natural language processing of mimic-iii clinical notes for identifying diagnosis and procedures with neural networks. **arXiv preprint arXiv:1912.12397**, [S.l.], 2019.

OHNO-MACHADO, L.; SÉROUSSI, B. Automatic methods to extract prescription status quality measures from unstructured health records. In: MEDINFO 2019: HEALTH AND WELLBEING E-NETWORKS FOR ALL: PROCEEDINGS OF THE 17TH WORLD CONGRESS ON MEDICAL AND HEALTH INFORMATICS, 2019. **Anais...** [S.l.: s.n.], 2019. v. 264, p. 15.

- OLIVEIRA, J. M. de; ANTUNES, R. S.; COSTA, C. A. da. Soap classifier for free-text clinical notes with domain-specific pre-trained language models. **Expert Systems with Applications**, [S.l.], v. 245, p. 123046, 2024.
- OLIVEIRA, J. M. de; COSTA, C. A. da; ANTUNES, R. S. Data structuring of electronic health records: a systematic review. **Health and Technology**, [S.l.], v. 11, n. 6, p. 1219–1235, 2021.
- OLIVEIRA, L. E. S. e et al. Semclinbr - a multi institutional and multi specialty semantically annotated corpus for portuguese clinical NLP tasks. **CoRR**, [S.l.], v. abs/2001.10071, 2020.
- ORGANIZATION, W. H. et al. **The icd-10 classification of mental and behavioural disorders**: diagnostic criteria for research. [S.l.]: World Health Organization, 1993. v. 2.
- OSBORNE, J. D.; GYAWALI, B.; SOLORIO, T. Evaluation of ytex and metamap for clinical concept recognition. **arXiv preprint arXiv:1402.1668**, [S.l.], 2014.
- PÉREZ, A. et al. Semi-supervised medical entity recognition: a study on spanish and swedish clinical corpora. **Journal of Biomedical Informatics**, [S.l.], v. 71, p. 16–30, jul 2017.
- PETERSON, K. J.; LIU, H. Automating the transformation of free-text clinical problems into snomed ct expressions. **AMIA Summits on Translational Science Proceedings**, [S.l.], v. 2020, p. 497, 2020.
- QIN, X. et al. Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. **Journal of Clinical Epidemiology**, [S.l.], 2021.
- RAISKIN, Y.; EICKHOFF, C.; BEELER, P. E. Categorization of free-text drug orders using character-level recurrent neural networks. **International Journal of Medical Informatics**, [S.l.], v. 129, p. 20–28, sep 2019.
- REZA, F.; PRIETO, J. T.; JULIEN, S. P. Electronic health records: origination, adoption, and progression. In: PUBLIC HEALTH INFORMATICS AND INFORMATION SYSTEMS, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 183–201.
- RINK, B.; HARABAGIU, S.; ROBERTS, K. Automatic extraction of relations between medical concepts in clinical texts. **Journal of the American Medical Informatics Association**, [S.l.], v. 18, n. 5, p. 594–600, 2011.
- ROCHA, N. C. da et al. Natural language processing to extract information from portuguese-language medical records. **Data**, [S.l.], v. 8, n. 1, p. 11, 2022.
- SADVILKAR, N.; NEUMANN, M. Pysbd: pragmatic sentence boundary disambiguation. **CoRR**, [S.l.], v. abs/2010.09657, 2020.
- SAGHEB, E. et al. Use of natural language processing algorithms to identify common data elements in operative notes for knee arthroplasty. **The Journal of Arthroplasty**, [S.l.], 2020.
- SANH, V. et al. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. **arXiv preprint arXiv:1910.01108**, [S.l.], 2019.
- SANTOS, C. N. d.; XIANG, B.; ZHOU, B. Classifying relations by ranking with convolutional neural networks. **arXiv preprint arXiv:1504.06580**, [S.l.], 2015.

SANTOS., D. et al. **Assessing the effectiveness of multilingual transformer-based text embeddings for named entity recognition in portuguese.** [S.l.]: SciTePress, 2021. 473-483 p.

SARIPALLE, R.; SOOKHAK, M.; HAGHPARAST, M. An interoperable umls terminology service using fhir. **Future Internet**, [S.l.], v. 12, n. 11, p. 199, 2020.

SAVOVA, G. K. et al. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. **Journal of the American Medical Informatics Association**, [S.l.], v. 17, n. 5, p. 507–513, 2010.

SCHNEIDER, E. T. R. et al. Biobertpt-a portuguese neural language model for clinical named entity recognition. In: CLINICAL NATURAL LANGUAGE PROCESSING WORKSHOP, 3., 2020. **Proceedings...** [S.l.: s.n.], 2020. p. 65–72.

SEN, C. et al. Patient-level classification on clinical note sequences guided by attributed hierarchical attention. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), 2019., 2019. **Anais...** IEEE, 2019.

SEZGIN, E. et al. Extracting medical information from free-text and unstructured patient-generated health data using natural language processing methods: feasibility study with real-world data. **JMIR Formative Research**, [S.l.], v. 7, p. e43014, 2023.

SHAH, S. et al. Neural networks for mining the associations between diseases and symptoms in clinical notes. **Health Information Science and Systems**, [S.l.], v. 7, n. 1, nov 2018.

SHARMA, B. et al. Publicly available machine learning models for identifying opioid misuse from the clinical notes of hospitalized patients. **BMC Medical Informatics and Decision Making**, [S.l.], v. 20, n. 1, Apr. 2020.

SHEIKHALISHAHI, S. et al. Natural language processing of clinical notes on chronic diseases: systematic review. **JMIR medical informatics**, [S.l.], v. 7, n. 2, p. e12239, 2019.

SHI, X. et al. Family history information extraction via deep joint learning. **BMC Medical Informatics and Decision Making**, [S.l.], v. 19, n. S10, dec 2019.

SILVA, A. Cartão nacional de saúde e o prontuário eletrônico do paciente no brasil: uma visão do estado atual. **Resende: Associacao Educacional Dom Bosco**, [S.l.], 2007.

SILVA, D. P. d. **Reconhecimento de entidades nomeadas e extração de relações de registros de prontuários médicos para população de ontologia.** 2023. Tese (Doutorado em Ciência da Computação) — UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS, São Leopoldo, Rio Grande do Sul, Brasil, 2023.

SINGH, A. K. B. et al. **Multi-label natural language processing to identify diagnosis and procedure codes from mimic-iii inpatient notes.** 2020.

SINGH, G. et al. A neural candidate-selector architecture for automatic structured clinical text annotation. In: ACM ON CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 2017., 2017, New York, NY, USA. **Proceedings...** Association for Computing Machinery, 2017. p. 1519–1528. (CIKM '17).

- Song, B. et al. Un-apriori: a novel association rule mining algorithm for unstructured emrs. In: IEEE 19TH INTERNATIONAL CONFERENCE ON E-HEALTH NETWORKING, APPLICATIONS AND SERVICES (HEALTHCOM), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 1–6.
- SORIANO, I. M.; CASTRO, J. Dner clinical (named entity recognition) from free clinical text to snomed-ct concept. **WSEAS Transactions on Computers**, [S.l.], v. 16, p. 83–91, 2017.
- SORIANO, I. M.; PEÑA, J. L. C. Automatic medical concept extraction from free text clinical reports, a new named entity recognition approach. **International Journal of Computers**, [S.l.], v. 2, 2017.
- SOUZA, J. V. A. et al. A multilabel approach to portuguese clinical named entity recognition. **Journal of Health Informatics**, [S.l.], v. 12, 2021.
- SOYSAL, E. et al. Clamp—a toolkit for efficiently building customized clinical natural language processing pipelines. **Journal of the American Medical Informatics Association**, [S.l.], v. 25, n. 3, p. 331–336, 2018.
- SPANDORFER, A. et al. Deep learning to convert unstructured ct pulmonary angiography reports into structured reports. **European radiology experimental**, [S.l.], v. 3, n. 1, p. 37, 2019.
- SREEJITH, S.; Khanna Nehemiah, H.; KANNAN, A. Clinical data classification using an enhanced smote and chaotic evolutionary feature selection. **Computers in Biology and Medicine**, [S.l.], v. 126, p. 103991, 2020.
- STEINKAMP, J. M. et al. Toward complete structured information extraction from radiology reports using machine learning. **Journal of Digital Imaging**, [S.l.], v. 32, n. 4, p. 554–564, June 2019.
- STEINKAMP, J. M. et al. Task definition, annotated dataset, and supervised natural language processing models for symptom extraction from unstructured clinical notes. **Journal of Biomedical Informatics**, [S.l.], v. 102, p. 103354, feb 2020.
- STEVENSON, M.; GAIZAUSKAS, R. Experiments on sentence boundary detection. In: SIXTH APPLIED NATURAL LANGUAGE PROCESSING CONFERENCE, 2000. **Anais...** [S.l.: s.n.], 2000. p. 84–89.
- STUMPF, M. K.; FREITAS, H. M. de. A gestão da informação em um hospital universitário: o processo de definição do patient core record. **Revista de Administração Contemporânea**, [S.l.], v. 1, p. 71–99, 1997.
- SU, J. et al. Extraction of risk factors for cardiovascular diseases from chinese electronic medical records. **Computer Methods and Programs in Biomedicine**, [S.l.], v. 172, p. 1–10, apr 2019.
- SYED, L.; JABEEN, S.; MANIMALA, S. Telemammography: a novel approach for early detection of breast cancer through wavelets based image processing and machine learning techniques. In: ADVANCES IN SOFT COMPUTING AND MACHINE LEARNING IN IMAGE PROCESSING, 2018, Cham. **Anais...** Springer International Publishing, 2018. p. 149–183.
- SYMEONIDOU, A.; SAZONAU, V.; GROTH, P. Transfer learning for biomedical named entity recognition with biobert. In: SEMANTICS POSTERS&DEMOS, 2019. **Anais...** [S.l.: s.n.], 2019.

- TAFT, L. et al. Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery. **Journal of Biomedical Informatics**, [S.l.], v. 42, n. 2, p. 356–364, 2009.
- TANHA, J. et al. Boosting methods for multi-class imbalanced data classification: an experimental review. **Journal of Big Data**, [S.l.], v. 7, n. 1, p. 1–47, 2020.
- TARCAR, A. K. et al. Healthcare ner models using language model pretraining. In: INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, 13., 2020. **Proceedings...** [S.l.: s.n.], 2020. p. 12–18. (WSDM '20).
- THORAT, P.; TONGAONKAR, R.; JAGTAP, V. Towards designing the best model for classification of fish species using deep neural networks. In: PROCEEDING OF INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND APPLICATIONS, 2020, Singapore. **Anais...** Springer Singapore, 2020. p. 343–351.
- TIBBO, M. E. et al. Use of natural language processing tools to identify and classify periprosthetic femur fractures. **The Journal of arthroplasty**, [S.l.], v. 34, n. 10, p. 2216–2219, 2019.
- TOGNOLA, G.; MURRI, A.; CUDA, D. Cognitive computing for the automated extraction and meaningful use of health data in narrative medical notes: an application to the clinical management of hearing impaired aged patients. In: IEEE EMBS INTERNATIONAL CONFERENCE ON BIOMEDICAL & HEALTH INFORMATICS (BHI), 2018., 2018. **Anais...** IEEE, 2018.
- TOMANEK, K.; WERMTER, J.; HAHN, U. Sentence and token splitting based on conditional random fields. In: CONFERENCE OF THE PACIFIC ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 10., 2007. **Proceedings...** [S.l.: s.n.], 2007. v. 49, p. 57.
- UZUNER, O. et al. Community annotation experiment for ground truth generation for the i2b2 medication challenge. **Journal of the American Medical Informatics Association**, [S.l.], v. 17, n. 5, p. 519–523, 09 2010.
- VASWANI, A. et al. Attention is all you need. **arXiv preprint arXiv:1706.03762**, [S.l.], 2017.
- VIANI, N. et al. Information extraction from italian medical reports: an ontology-driven approach. **International Journal of Medical Informatics**, [S.l.], v. 111, p. 140 – 148, 2018.
- WANG, L. et al. An interdisciplinary literature classifier based on multi-task multi-label learning. In: INTERNATIONAL CONFERENCE ON ASIAN LANGUAGE PROCESSING (IALP), 2021., 2021. **Anais...** [S.l.: s.n.], 2021. p. 183–188.
- Wang, Q.; Zeng, L. Chinese symptom component recognition via bidirectional lstm-crf. In: TENTH INTERNATIONAL CONFERENCE ON ADVANCED COMPUTATIONAL INTELLIGENCE (ICACI), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 45–50.
- WANG, R. et al. Medical entity recognition of esophageal carcinoma based on word clustering. In: INTERNATIONAL CONFERENCE ON SECURITY, PATTERN ANALYSIS, AND CYBERNETICS (SPAC), 2018., 2018. **Anais...** IEEE, 2018.
- WANG, S. et al. A childhood disease database based on word segmentation technology: research and practice. In: IEEE 42ND ANNUAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE (COMPSAC), 2018., 2018. **Anais...** IEEE, 2018.

- WANG, S. et al. Information extraction for intestinal cancer electronic medical records. **IEEE Access**, [S.l.], v. 8, p. 125923–125934, 2020.
- WEEGAR, R. et al. Deep medical entity recognition for swedish and spanish. In: IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM), 2018., 2018. **Anais...** IEEE, 2018.
- WEEKS, H. L. et al. medExtractR: a targeted, customizable approach to medication extraction from electronic health records. **Journal of the American Medical Informatics Association**, [S.l.], v. 27, n. 3, p. 407–418, jan 2020.
- WILCOXON, F. et al. Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test. **Selected tables in mathematical statistics**, [S.l.], v. 1, p. 171–259, 1970.
- WILLIAMS, S.; PETROVICH, E. Natural language processing for unlocking insights from unstructured big data in the healthcare industry. **Emerging Trends in Machine Intelligence and Big Data**, [S.l.], v. 15, n. 10, p. 30–39, 2023.
- WONG, D. F.; CHAO, L. S.; ZENG, X. isentimizer-: multilingual sentence boundary detection model. **The Scientific World Journal**, [S.l.], v. 2014, 2014.
- WU, P.-H. et al. Keyword extraction and structuralization of medical reports. **Health Information Science and Systems**, [S.l.], v. 8, n. 1, Apr. 2020.
- WU, S.; DREDZE, M. **Are all languages created equal in multilingual bert?** [S.l.]: arXiv, 2020.
- WUNNAVA, S. et al. Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. **Drug Safety**, [S.l.], v. 42, n. 1, p. 113–122, Jan. 2019.
- XU, H. et al. Medex: a medication information extraction system for clinical narratives. **Journal of the American Medical Informatics Association**, [S.l.], v. 17, n. 1, p. 19–24, 2010.
- XUE, K. et al. Fine-tuning bert for joint entity and relation extraction in chinese medical text. In: IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 892–897.
- XUE, X.; WANG, Y. Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio. **Artificial Intelligence**, [S.l.], v. 223, p. 65–81, 2015.
- YADAV, S. et al. Relation extraction from biomedical and clinical text: unified multitask learning framework. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, [S.l.], 2020.
- YADAV, S. et al. Exploring disorder-aware attention for clinical event extraction. **ACM Transactions on Multimedia Computing, Communications, and Applications**, [S.l.], v. 16, n. 1s, p. 1–21, apr 2020.
- YANG, T. et al. Chinese data extraction and named entity recognition. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA ANALYTICS (ICBDA), 2020., 2020. **Anais...** IEEE, 2020.

- YANG, X. et al. MADEx: a system for detecting medications, adverse drug events, and their relations from clinical notes. **Drug Safety**, [S.l.], v. 42, n. 1, p. 123–133, Jan. 2019.
- YEHIA, E. et al. Ontology-based clinical information extraction from physician's free-text notes. **Journal of biomedical informatics**, [S.l.], v. 98, p. 103276, 2019.
- YIN, M. et al. Chinese clinical named entity recognition with radical-level feature and self-attention mechanism. **Journal of Biomedical Informatics**, [S.l.], v. 98, p. 103289, oct 2019.
- YUJIAN, L.; BO, L. A normalized levenshtein distance metric. **IEEE transactions on pattern analysis and machine intelligence**, [S.l.], v. 29, n. 6, p. 1091–1095, 2007.
- ZENG, D. et al. Relation classification via convolutional deep neural network. In: COLING 2014, THE 25TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS: TECHNICAL PAPERS, 2014. **Proceedings...** [S.l.: s.n.], 2014. p. 2335–2344.
- ZENG, Z. et al. Natural language processing for ehr-based computational phenotyping. **IEEE/ACM transactions on computational biology and bioinformatics**, [S.l.], v. 16, n. 1, p. 139–153, 2018.
- ZHANG, F. et al. Section classification in clinical notes with multi-task transformers. In: INTERNATIONAL WORKSHOP ON HEALTH TEXT MINING AND INFORMATION ANALYSIS (LOUHI), 13., 2022, Abu Dhabi, United Arab Emirates (Hybrid). **Proceedings...** Association for Computational Linguistics, 2022. p. 54–59.
- ZHANG, H. et al. Identifying transitional high cost users from unstructured patient profiles written by primary care physicians. In: BIOCOMPUTING 2020, 2019. **Anais...** WORLD SCIENTIFIC, 2019.
- ZHANG, R. et al. A text structuring method for chinese medical text based on temporal information. **International Journal of Environmental Research and Public Health**, [S.l.], v. 15, n. 3, 2018.
- ZHANG, T. et al. Constructing fine-grained entity recognition corpora based on clinical records of traditional chinese medicine. **BMC Medical Informatics and Decision Making**, [S.l.], v. 20, n. 1, Apr. 2020.
- ZHANG, X.; LIU, C.-A. Model averaging prediction by k-fold cross-validation. **Journal of Econometrics**, [S.l.], v. 235, n. 1, p. 280–301, 2023.
- ZHANG, Y. et al. Clinical named entity recognition from chinese electronic health records via machine learning methods. **JMIR medical informatics**, [S.l.], v. 6, n. 4, p. e50, 2018.
- ZHANG, Z. et al. Attention-based deep residual learning network for entity relation extraction in chinese EMRs. **BMC Medical Informatics and Decision Making**, [S.l.], v. 19, n. S2, apr 2019.
- ZHAO, B. Clinical data extraction and normalization of cyrillic electronic health records via deep-learning natural language processing. **JCO Clinical Cancer Informatics**, [S.l.], n. 3, p. 1–9, dec 2019.
- ZHU, L.; HUA, G.; GAO, W. Mapping ontology vertices to a line using hypergraph framework. **International Journal of Cognitive Computing in Engineering**, [S.l.], v. 1, p. 1–8, 2020.