

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO” –
UNESP
Instituto de Biociências, Letras e Ciências Exatas - Campus de São José do
Rio Preto**

WELLINGTON REGUERA GOUVEIA

**VISUALIZAÇÃO DE DADOS NÃO ESTRUTURADOS DA SAÚDE APOIADA POR
PROCESSAMENTO DE LINGUAGEM NATURAL**

São José do Rio Preto

2026



WELLINGTON REGUERA GOUVEIA

**VISUALIZAÇÃO DE DADOS NÃO ESTRUTURADOS DA SAÚDE APOIADA POR
PROCESSAMENTO DE LINGUAGEM NATURAL**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Computação Aplicada

Orientador(a): Prof. Dr. Carlos Roberto Valêncio

São José do Rio Preto

2026

G719 Gouveia, Wellington Reguera
Visualização de dados não estruturados da saúde apoiada por
processamento de linguagem natural / Wellington Reguera
Gouveia. --São José do Rio Preto, 2026
127 p. : il., tabs., fotos

Dissertação (mestrado) - Universidade Estadual Paulista
(UNESP), Instituto de Biociências Letras e Ciências Exatas, São
José do Rio Preto
Orientador: Carlos Roberto Valêncio

1. Ciência da computação. 2. Processamento de linguagem
natural.
3. Inteligência Artificial. I. Título.

IMPACTO POTENCIAL DESTA PESQUISA

Este trabalho propõe uma abordagem de estruturação de dados de textos médicos a partir de um modelo de linguagem capaz de reconhecer as informações mais relevantes e da geração de gráficos temporais interativos. A abordagem é viável na redução da carga cognitiva de profissionais da saúde, principalmente na área psiquiátrica.

POTENTIAL IMPACT OF THIS RESEARCH

This work proposes a data structuring approach for medical texts based on a language model capable of recognizing the most relevant information and generating interactive temporal graphs. This approach is viable for reducing the cognitive load of healthcare professionals, particularly in the psychiatric field.

WELLINGTON REGUERA GOUVEIA

**VISUALIZAÇÃO DE DADOS NÃO ESTRUTURADOS DA SAÚDE APOIADA POR
PROCESSAMENTO DE LINGUAGEM NATURAL**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Instituto de Biociências, Letras e Exatas, São José do Rio Preto, para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Computação Aplicada

Data da defesa: 23/06/2026

Banca Examinadora:

Prof. Dr. Carlos Roberto Valêncio
UNESP – Instituto de Biociências, Letras e Ciências Exatas - Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Geraldo Francisco Donegá Zafalon
UNESP – Instituto de Biociências, Letras e Ciências Exatas - Câmpus de São José do Rio Preto

Prof. Dr. Pedro Luiz Pizzigatti Corrêa
Universidade de São Paulo – Escola Politécnica

AGRADECIMENTOS

Agradeço a todos que de alguma forma me ajudaram e estiveram ao meu lado durante essa jornada. Em especial, agradeço aos meus pais, Evaldo e Sandra, por todo amor, pela educação que proporcionaram e por tornarem esse momento possível, à minha namorada e companheira de vida, Sarah, por estar sempre ao meu lado, nos momentos bons e ruins, aos meus amigos de infância Gustavo e Leandro que me acompanharam durante toda a minha jornada até os dias atuais, ao João Quadrado, pela amizade e pela trajetória do mestrado que passamos. Agradeço ao Prof. Dr. Carlos Roberto Valêncio, pela confiança, pelas oportunidades que me foram proporcionadas e pela orientação durante a graduação e nesta dissertação de mestrado. Por fim, agradeço aos queridos amigos do Grupo de Banco de Dados (GBD), dos mais antigos aos mais novos, por todos os momentos, ensinamentos e experiências que compartilhamos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.969117/2024-00.

“Os sonhos das pessoas nunca acabam!”
(One Piece).

RESUMO

A informatização na área da saúde permite que os profissionais analisem o histórico do paciente de maneira mais rápida e eficiente, o que possibilita um atendimento mais preciso e assertivo. Nesse sentido, a Visualização de Dados permite a criação de representações visuais que auxiliam a interpretação e a compreensão dos dados alvos da análise. Contudo, grande parte das informações médicas encontra-se no formato de texto e, devido a característica não estruturada desse tipo de dado, dificulta a aplicação direta de algoritmos de extração de informação. A área de Processamento de Linguagem Natural tem como objetivo a análise automática de textos escritos na linguagem humana, mais especificamente em pesquisas recentes, têm direcionado na visualização de dados e na subárea de Reconhecimento de Entidades Nomeadas. No entanto, a maioria desses estudos são voltados para os idiomas com mais recursos, como o inglês. O presente trabalho propõe uma abordagem capaz de extrair informações relevantes de textos clínicos psiquiátricos em língua portuguesa por meio de um modelo de linguagem ajustado para o reconhecimento de entidades nomeadas, com posterior estruturação dos dados extraídos em um banco de dados relacional. Os resultados experimentais demonstram que a abordagem proposta obteve desempenho superior em relação a abordagens comparáveis, com 0,80 de precisão, 0,84 de revocação e 0,82 de F1-Score. Esses valores evidenciam a qualidade do modelo na identificação de entidades clínicas relevantes comparados com outros modelos da literatura. A avaliação do recurso de visualização mostrou uma redução média de 29,8% no tempo de entendimento do caso, de modo que se mantiveram os valores das conclusões obtidas na análise textual. A principal contribuição científica consiste no desenvolvimento de uma abordagem que agrega o processo de extração automática de informações médicas e uma ferramenta de visualização baseada em gráfico de Gantt, que representa os eventos clínicos de forma cronológica e interativa.

Palavras-chave: processamento de linguagem natural; reconhecimento de entidades nomeadas; reconhecimento de entidades médicas; visualização de dados.

ABSTRACT

The computerization of healthcare allows professionals to analyze patient histories more quickly and efficiently, enabling more precise and assertive care. In this sense, Data Visualization allows the creation of visual representations that aid in the interpretation and understanding of the data targeted for analysis. However, much of the medical information is in text format, and due to the unstructured nature of this type of data, it hinders the direct application of information extraction algorithms. The field of Natural Language Processing aims at the automatic analysis of texts written in human language; more specifically, recent research has focused on data visualization and the sub-area of Named Entity Recognition. However, most of these studies are geared towards languages with more resources, such as English. This work proposes an approach capable of extracting relevant information from psychiatric clinical texts in Portuguese using a language model adjusted for named entity recognition, with subsequent structuring of the extracted data in a relational database. Experimental results demonstrate that the proposed approach outperformed comparable approaches, with a precision of 0.80, a recall of 0.84, and an F1-Score of 0.82. These values highlight the model's quality in identifying relevant clinical entities compared to other models in the literature. Evaluation of the visualization feature showed an average reduction of 29.8% in case comprehension time, maintaining the values of the conclusions obtained in the textual analysis. The main scientific contribution lies in the development of an approach that combines the automatic extraction of medical information and a Gantt chart-based visualization tool, which represents clinical events chronologically and interactively.

Keywords: natural language processing; named entity recognition; medical entity recognition; data visualization.

LISTA DE FIGURAS

Figura 1 – As atuações da Inteligência Artificial na saúde	20
Figura 2 – Tipos de Aprendizado de Máquina.....	22
Figura 3 – Modelo de Aprendizado Supervisionado	23
Figura 4 – Algoritmo K-means Clustering.....	24
Figura 5 – Aprendizado por reforço.....	25
Figura 6 – Relação entre as subáreas da Inteligência Artificial	26
Figura 7 – Arquitetura Transformer	34
Figura 8 – Arquitetura do <i>Encoder</i>	35
Figura 9 – Arquitetura do <i>Decoder</i>	37
Figura 10 – Exemplo de tarefas dispostas em um gráfico de Gantt	46
Figura 11 – Diagrama do trabalho desenvolvido	55
Figura 12 – Exemplo da aplicação do modelo em um texto médico	59
Figura 13 – Modelo Entidade-Relacionamento	63
Figura 14 – Visualização em gráfico da avaliação médica	65
Figura 15 – Visualização textual da avaliação médica	66
Figura 16 – Visualização em gráfico de Gantt da anotação de enfermagem	67
Figura 17 – Utilização do filtro no gráfico de Gantt da anotação de enfermagem	68
Figura 18 – Representação textual da anotação de enfermagem.....	69
Figura 19 – Visualização em gráfico de Gantt da doença atual e evolução	70
Figura 20 – Gráficos de linha	71
Figura 21 – Gráfico de caixa dos valores de tempo	91
Figura 22 – Comparação dos tempos obtidos por categoria de tarefa	92
Figura 23 – Comparação dos valores de <i>insight</i> obtidos por categoria de tarefa.....	93
Figura 24 – Médias de contribuição de usabilidade por pergunta	97
Apêndice A	
Figura A 1 – Código da função responsável pelo ajuste fino do modelo	111
Figura A 2 – Trecho do código responsável pela validação cruzada	112
Figura A 3 – Exemplo de uma nota clínica do tipo "Doença Atual e Evolução".....	113
Figura A 4 – Exemplo de uma nota clínica do tipo "Anotação de Enfermagem"	114
Figura A 5 – Exemplo de uma nota clínica do tipo "Avaliação Médica"	115
Figura A 6 – Modelo Relacional do banco de dados desenvolvido	116

LISTA DE ILUSTRAÇÕES

Quadro 1 – Comparação entre os trabalhos correlatos.....	53
Quadro 2 – Representação IOB2 dos Tokens das Sentenças	57
Quadro 3 – Parâmetros selecionados	74
Quadro 4 – Critérios para a definição de valores do <i>insight</i>	82
Quadro 5 – Questões e níveis de <i>insight</i> para o Caso 1	83
Quadro 6 – Questões e níveis de <i>insight</i> para o Caso 2	85
Quadro 7 – Questões e níveis de <i>insight</i> para o Caso 3	86
Quadro 8 – Questões e níveis de <i>insight</i> para o Caso 4	87
Quadro 9 – Comparação final entre os trabalhos correlatos	103

LISTA DE TABELAS

Tabela 1 – Distribuição das categorias	75
Tabela 2 – Desempenho dos modelos avaliados.....	77
Tabela 3 – Desempenho do BERTugues por categoria	77
Tabela 4 – Comparação de desempenho entre os correlatos.....	78
Tabela 5 – Comparação do desempenho por método de representação	90
Tabela 6 – Comparação do desempenho por categoria de pergunta	92
Tabela 7 – Testes estatísticos.....	94

LISTA DE ABREVIATURAS E SIGLAS

API	Interface de Programação de Aplicações
BERT	Bidirectional Encoder Representations from Transformers
BioBERT	Bidirectional Encoder Representations from Transformers for Biomedical Text Mining
CAAE	Certificado de Apresentação de Apreciação Ética
CEP	Comitê de Ética em Pesquisa
CNN	Redes Neurais Convolucionais
DL	Aprendizado Profundo
Google Colab	Google Colaboratory
GEP	Sistema de Gestão de Evolução de Prontuários
GPU	Unidade de Processamento Gráfico
HIV	Vírus da Imunodeficiência Humana
IA	Inteligência Artificial
IOB2	Inside-Outside-Beginning 2
JSON	JavaScript Object Notation
LLM	Grandes Modelos de Linguagem
LSTM	Memória de Longo e Curto Prazo
MER	Modelo Entidade-Relacionamento
ML	Aprendizado de Máquina
MLM	Modelo de Linguagem com Máscara
NER	Reconhecimento de Entidade Nomeada
NLM	Modelo de Linguagem Neural
NSP	Previsão da Próxima Sentença
PLM	Modelos de Linguagem Pré-treinados
PLN	Processamento de Linguagem Natural
RAG	Retrieval-Augmented Generation

RNN	Redes Neurais Recorrentes
SLM	Modelos de Linguagem Estatísticos
SUS	System Usability Scale
SVM	Máquina de Vetores de Suporte

SUMÁRIO

1.	INTRODUÇÃO	15
1.1	MOTIVAÇÃO E ESCOPO.....	16
1.2	OBJETIVOS.....	17
1.3	METODOLOGIA.....	17
1.4	CONTRIBUIÇÕES.....	17
1.5	ORGANIZAÇÃO DA DISSERTAÇÃO.....	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	PRONTUÁRIOS MÉDICOS ELETRÔNICOS	19
2.2	INTELIGÊNCIA ARTIFICIAL.....	19
2.3	APRENDIZADO DE MÁQUINA	21
2.3.1	Aprendizado supervisionado.....	22
2.3.2	Aprendizado não supervisionado.....	23
2.3.3	Aprendizado por reforço	24
2.4	APRENDIZADO PROFUNDO	25
2.5	PROCESSAMENTO DE LINGUAGEM NATURAL.....	27
2.5.1	Processamento de linguagem natural em prontuários médicos	28
2.6	RECONHECIMENTO DE ENTIDADES NOMEADAS	30
2.6.1	Reconhecimento de entidades nomeadas em prontuários médicos.....	31
2.7	MODELOS DE LINGUAGEM	32
2.7.1	Transformador	33
2.8	REPRESENTAÇÃO DE CODIFICADORES BIDIRECIONAIS DE TRANSFORMERS	38
2.8.1	Pré-treinamento.....	38
2.8.2	Ajuste fino.....	40
2.8.3	Métricas de avaliação	40
2.8.4	Modelo de linguagem de domínio biológico.....	41
2.8.5	Modelo de linguagem de domínio biológico em português.....	42
2.9	CIÊNCIA DE DADOS E VISUALIZAÇÃO	43
2.9.1	Visualização analítica	44
2.9.2	Gráficos baseados em linha do tempo	45
2.10	TRABALHOS CORRELATOS	47

2.10.1	Abordagem visual para reconhecimento de entidades biomédicas nomeadas	47
2.10.2	Abordagem baseada em linha do tempo para agregação de prontuários médicos eletrônicos	48
2.10.3	Visualização do histórico clínico do paciente em linha do tempo a partir de textos médicos	49
2.10.4	Linha do tempo das medicações prescritas ao paciente	50
2.10.5	Extração e padronização de dados clínicos não estruturados com uso de processamento de linguagem natural	50
2.10.6	Modelo de reconhecimento de entidades clínicas para narrativas psiquiátricas	51
2.10.7	Análise dos trabalhos correlatos	52
2.11	CONSIDERAÇÕES FINAIS	53
3	ABORDAGEM DE VISUALIZAÇÃO DE DADOS APOIADA POR PROCESSAMENTO DE LINGUAGEM NATURAL DESENVOLVIDA	54
3.1	VISÃO GERAL	54
3.2	TECNOLOGIAS	56
3.3	BASE DE DADOS ANOTADOS PARA O AJUSTE FINO	56
3.4	AJUSTE FINO DO MODELO	58
3.5	AQUISIÇÃO DAS NOTAS CLÍNICAS	59
3.6	ESTRUTURAÇÃO DOS DADOS	62
3.7	FERRAMENTA DE VISUALIZAÇÃO	64
3.8	AMBIENTE DE DESENVOLVIMENTO E DIMENSIONAMENTO DO SOFTWARE	71
3.9	CONSIDERAÇÕES FINAIS	72
4	AVALIAÇÃO EXPERIMENTAL	73
4.1	AVALIAÇÃO DO MODELO DE EXTRAÇÃO	73
4.1.1	Ambiente e configuração experimental	73
4.1.2	Corpus clínico anotado	75
4.1.3	Resultados e análise comparativa	76
4.2	AVALIAÇÃO DO SUPORTE À DECISÃO VISUAL	79
4.2.1	Amostragem	80
4.2.2	Organização das sessões de teste	80
4.2.3	Cenários clínicos e especificações das tarefas	81

4.3	RESULTADOS E DISCUSSÃO	89
4.4	ANÁLISE ESTATÍSTICA	93
4.5	ANÁLISE QUALITATIVA E PERCEPÇÕES DOS USUÁRIOS.....	95
4.6	AValiação DA USABILIDADE.....	96
4.7	ANÁLISE COMPARATIVA.....	98
4.8	CONSIDERAÇÕES FINAIS.....	99
5	CONCLUSÃO.....	101
5.1	CONTRIBUIÇÃO CIENTÍFICA	102
5.2	TRABALHOS FUTUROS.....	103
	REFERÊNCIAS	105
	APÊNDICE A – ASPECTOS TÉCNICOS DA IMPLEMENTAÇÃO DA ABORDAGEM	110
	ANEXO A – PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA EM PESQUISA	118
	DADOS CURRICULARES.....	124

1. INTRODUÇÃO

A visualização dos dados por meio de interfaces gráficas consegue combinar efetivamente múltiplas fontes de dados clínicos e reduzir a carga cognitiva que relatórios geram nos profissionais de saúde. Deste modo, a representação visual dos dados é capaz de auxiliar na tomada de decisões e gerar percepções importantes para o tratamento do paciente, com mais subsídios na busca por tratamentos mais precisos (Dabek; Jimenez; Caban, 2017; Ledesma et al., 2019; Keshta; Odeh, 2021; Dai et al., 2019).

Os sistemas de visualizações analíticas têm se mostrado eficazes no enfrentamento dos desafios relacionados à sobrecarga de informações presentes nos prontuários médicos eletrônicos, ao integrar métodos analíticos com representações visuais interativas. Essa abordagem possui o potencial de aprimorar o processo de tomada de decisões clínicas por parte dos especialistas da área da saúde, por meio de técnicas oriundas de diversos campos, como estatística, aprendizado de máquina e mineração de dados. Adicionalmente, as visualizações interativas permitem que os profissionais investiguem os dados subjacentes, modifiquem as formas de exibição e direcionem os procedimentos analíticos para realizar suas atividades (Rostamzadeh; Abdullah; Sedig, 2021).

Os dados do prontuário médico eletrônico, por sua vez, são frequentemente classificados com base na forma em que foi coletado e sua representação em duas classes: estruturados ou não estruturados. Os dados estruturados consistem em diferentes fontes como diagnósticos, medicamentos e valores laboratoriais em campos numéricos ou categóricos fixos. Enquanto os dados não estruturados, referem-se a textos de formato livre escritos pelos profissionais da área da saúde, como notas clínicas. Os dados não estruturados representam cerca de 80% do total de dados dos prontuários médicos eletrônicos e denotam um desafio para seu processamento e extração de informações (Li et al., 2022).

Atualmente, as aplicações na área da saúde são uma grande força no desenvolvimento de Inteligência Artificial (IA). A introdução da área de Processamento de Linguagem Natural (PLN) no campo da Inteligência Artificial resultou em uma mudança devido à sua capacidade de converter textos não estruturados em formato estruturado (Jiang et al., 2022; Pn; M, 2021). O Reconhecimento de Entidade Nomeada, do inglês *Named Entity Recognition* (NER), é uma subárea da PLN muito

utilizada para a extração de informações. No contexto médico, é utilizada para identificar as entidades médicas nomeadas e classificá-las em categorias predefinidas como doenças, medicamentos e operações (Dai et al., 2019).

O modelo BERT (*Bidirectional Encoder Representations from Transformers*) faz uso da abordagem de duas etapas para realizar o NER, assim como outras tarefas de PLN. A primeira etapa, na fase de pré-treinamento, ele é treinado para aprender representações de linguagens a partir de um conjunto de dados não rotulados, como o Wikipédia e o *BooksCorpus* que juntos somam 3,3 bilhões de palavras. Em seguida, na etapa de ajuste fino, os parâmetros pré-treinados do modelo são adaptados para uma tarefa específica, como o NER, através do treinamento com dados rotulados e a adição de uma única camada de saída (Devlin et al., 2018).

As características arquiteturais do BERT permitem sua adaptação a domínios específicos, conforme o conjunto de dados utilizados para o pré-treinamento e ajuste fino (Lee et al., 2020). Tal adaptabilidade se encontra na existência de modelos especializados tanto para o contexto médico, como o BioBERT (Lee et al., 2020), quanto para a língua portuguesa, como o BERTIMBAU (Souza, Nogueira, Lotufo, 2023) e o BERTugues (Zago, Pedotti, 2024). Nesse contexto, o presente trabalho atua entre esses dois domínios, a proposta do ajuste fino de um modelo BERT para a tarefa de NER em notas clínicas. O objetivo é extrair e estruturar os dados clínicos relevantes no contexto psiquiátrico para viabilizar a criação de painéis de visualização de dados.

1.1 MOTIVAÇÃO E ESCOPO

A dificuldade do rápido entendimento do histórico do paciente é um desafio crítico na prática clínica, amplificado pelo fato dos dados clínicos serem majoritariamente compostos por textos não estruturados, como notas de evolução e laudos. A falta de estrutura afeta tanto a recuperação de informações quanto também pode levar a incidentes clínicos devido à falta de padronização e a variação nos estilos de escrita, além da dificuldade do entendimento da cronologia dos eventos (Yada; Aramaki, 2023). Técnicas como o NER são fundamentais para extrair o valor das informações textuais.

Segundo o levantamento bibliográfico realizado, um método de reconhecimento de entidades nomeadas a partir de dados médicos, juntamente com

a representação visual baseado em linha do tempo, se apresenta como uma boa solução para lidar com os desafios deste contexto (Yada; Aramaki, 2023). Porém, destaca-se uma lacuna significativa na literatura pela falta de sistemas que realizam essa tarefa tanto no contexto psiquiátrico quanto no contexto da língua portuguesa. Portanto, a motivação desse trabalho está no preenchimento dessa lacuna ao desenvolver uma solução que torna o histórico clínico do paciente, a partir dos dados escritos em textos livres.

1.2 OBJETIVOS

O objetivo deste trabalho é a apresentação de uma abordagem capaz de extrair as informações mais relevantes de textos médicos e estruturá-las em um banco de dados relacional, para que seja possível a geração de gráficos interativos. Com isso, busca-se reduzir a carga cognitiva dos profissionais da saúde durante a interpretação de prontuários em cenários de emergência.

1.3 METODOLOGIA

Inicialmente foi realizado o levantamento bibliográfico do estado da arte sobre visualização de dados de prontuários médicos eletrônicos, modelos de linguagem pré-treinados e ajuste fino para a tarefa de reconhecimento de entidades nomeadas. Após essa análise, foram estudados mais profundamente os principais trabalhos correlatos da área.

Em seguida, foi proposta a abordagem para visualização de dados não estruturados de notas clínicas. O presente trabalho foi validado por meio de um estudo comparativo entre modelos baseados em BERT, após uma etapa de ajuste fino. Em seguida, o modelo selecionado foi aplicado em notas clínicas anonimizadas para serem estruturadas e dispostas em gráficos visuais. Por fim, os gráficos desenvolvidos foram avaliados por profissionais da saúde da área de enfermagem.

1.4 CONTRIBUIÇÕES

A principal contribuição científica deste trabalho é a redução da carga cognitiva de profissionais da saúde durante a interpretação de prontuários médicos

complexos. A abordagem proposta preenche uma lacuna na literatura ao integrar as técnicas de Processamento de Linguagem Natural para a estruturação de textos livres com a visualização cronológica das informações.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

Neste capítulo foram introduzidos os desafios relacionados à visualização de dados em prontuários médicos eletrônicos, assim como foram apresentadas as motivações e objetivos do trabalho. Além disso, foram apresentadas as principais contribuições científicas do trabalho em relação a literatura.

No Capítulo 2 é apresentada a fundamentação teórica do trabalho, com conceitos de processamento de linguagem natural, reconhecimento de entidades nomeadas visualização de dados e os principais trabalhos correlatos pertencentes à área de extração de conhecimento de prontuários médicos.

No Capítulo 3 é detalhado o desenvolvimento da proposta, assim como a descrição da abordagem e o tipo de informações que será extraído das notas clínicas. No Capítulo 4 é descrita a metodologia dos experimentos realizados para a validação da proposta e a discussão sobre os resultados alcançados. No Capítulo 5 são apresentadas as considerações finais referentes ao trabalho e a descrição das contribuições científicas.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados os conceitos essenciais para o entendimento deste trabalho, além de técnicas amplamente utilizadas na literatura para a processamento de linguagem natural em prontuários médicos e visualização de dados. Os conceitos são descritos em torno dos tópicos de dados da saúde, com detalhes a respeito dos prontuários médicos eletrônicos, e ciência de dados, com o foco nos conceitos relacionados a inteligência artificial e a visualização analítica. Por fim, são apresentados os trabalhos correlatos na área de reconhecimento de entidades nomeadas e visualização de dados da saúde.

2.1 PRONTUÁRIOS MÉDICOS ELETRÔNICOS

O prontuário médico é definido como uma versão eletrônica do histórico médico do paciente constituído por um conjunto de documentos padronizados que contém informações geradas a partir de acontecimentos e situações sobre a saúde, além da assistência prestada ao paciente, como registros de progresso, medicações, problemas, exames médicos, entre outros (Keshta; Odeh, 2021; Thofehrn; Lima, 2006).

Associado ao conceito de prontuário, um registro médico eletrônico é feito de registros legais preenchidos em ambientes hospitalares. Os conteúdos presentes nesses registros são usados como a principal fonte de dados para o prontuário eletrônico (Keshta; Odeh, 2021). As informações desses registros médicos são consideradas relevantes diante do contexto estudado, pois auxiliam diretamente no processo de atendimento do paciente (Silva, 2021).

A informatização na área da saúde contempla aplicações desde sistemas típicos de gestão de informações a sistemas de automatização e apoio de tarefas de diagnóstico (Silva, 2021). O uso dessas ferramentas contribui para coleta/armazenamento de informações e a troca entre profissionais, o que proporciona qualidade no atendimento (Bezerra, 2009).

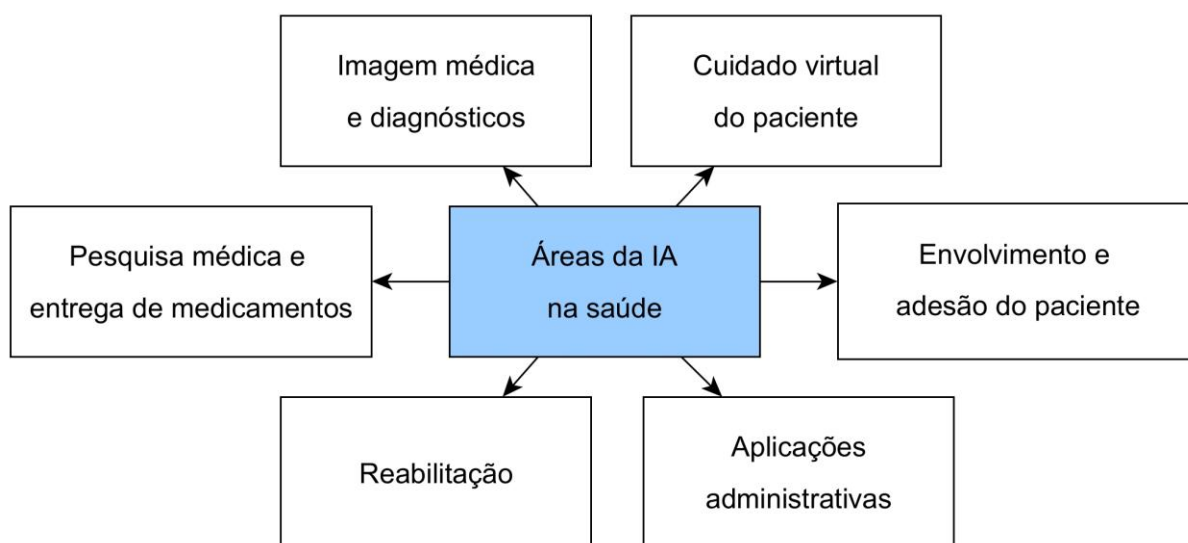
2.2 INTELIGÊNCIA ARTIFICIAL

Em sua definição, a Inteligência artificial (IA) é uma técnica para simular o intelecto humano a partir de uma coleção de algoritmos que visam criar um computador que consiga concluir tarefas humanas enquanto é executado de maneira paralela (Praful Bharadiya, 2023). As descrições de IA são variadas, com seu núcleo considerado como as teorias de pesquisa, métodos, tecnologias e aplicações com os objetivos de simular a inteligência humana. Atualmente, o conceito de IA tem um impacto cada vez mais profundo na vida humana (Jiang et al., 2022).

Os campos de aplicação de IA são extensos e abrangem diversas áreas de conhecimento, como reconhecimento de fala, processamento de imagens, processamento de linguagem natural, saúde, entre outras. Hoje em dia, aplicações na área da saúde atuam como uma força ativa para o desenvolvimento de IA devido às grandes quantidades de dados médicos que são coletadas diariamente. Contudo, essas informações possuem complexidades como a alta dimensionalidade e a não linearidade, o que dificulta no processo de desenvolvimento das tecnologias de IA (Jiang et al., 2022).

Os prontuários eletrônicos abrangem diversos dados de pacientes que podem ser associados com tecnologias de IA para obter *insights* precisos para o auxílio nos cuidados médicos (Al Kuwaiti et al., 2023). Várias áreas da saúde podem ser impactadas com o uso da IA, na Figura 1 são apresentados exemplos de áreas que podem ser exploradas.

Figura 1 – As atuações da Inteligência Artificial na saúde



Fonte: Adaptado de Al Kuawiti et al. (2023).

Comparado com outros campos técnicos, as pesquisas em IA na área da saúde apresentam desafios únicos. Diagnóstico médico é uma área na qual modelos quantitativos são escassos e no que diz respeito aos casos clínicos, as respostas dos médicos podem ser bastante diferentes umas das outras. Consequentemente, se torna um desafio o treinamento de ferramentas baseadas em IA com respostas subjetivas e que carregam o viés individual dos médicos (Kumar; Chauhan; Awasthi, 2023).

2.3 APRENDIZADO DE MÁQUINA

Uma área importante da Inteligência Artificial é o Aprendizado de máquina ou *Machine Learning* (ML). O ML é um subtipo de tecnologia de IA que tem sido utilizado de maneira multidisciplinar para automatizar processos de tomada de decisão e resolução de problemas complexos em diferentes contextos (Praful Bharadiya, 2023). No Aprendizado de máquina, os modelos são treinados a partir de dados existentes, de modo que, ao fornecer novos dados, o modelo irá identificá-los com base no que aprendeu anteriormente na fase de treinamento. O ML é considerado uma das técnicas mais comuns utilizadas na Inteligência Artificial (Manne; Kantheti, 2021).

Na área de análise semântica, é o processo de relacionar estruturas sintáticas de parágrafos, sentenças e palavras para o nível do texto como um todo (Shinde; Shah, 2018). Estes subdomínios serão explorados detalhadamente ao decorrer da pesquisa. No contexto da saúde, o ML é uma ferramenta utilizada para auxiliar os profissionais no cuidado dos pacientes e no gerenciamento de dados clínicos, da mesma forma que visa melhorar a velocidade e a precisão do trabalho dos médicos (Al Kuwaiti et al., 2023).

A classificação dos algoritmos de Aprendizado de Máquina é feita em três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço (Sharma; Sharma; Jindal, 2021). Na Figura 2 são apresentadas as três categorias citadas, e nos subtópicos a seguir, o detalhamento destas categorias.

Figura 2 – Tipos de Aprendizado de Máquina



Fonte: Adaptado de Sharma et al. (2021).

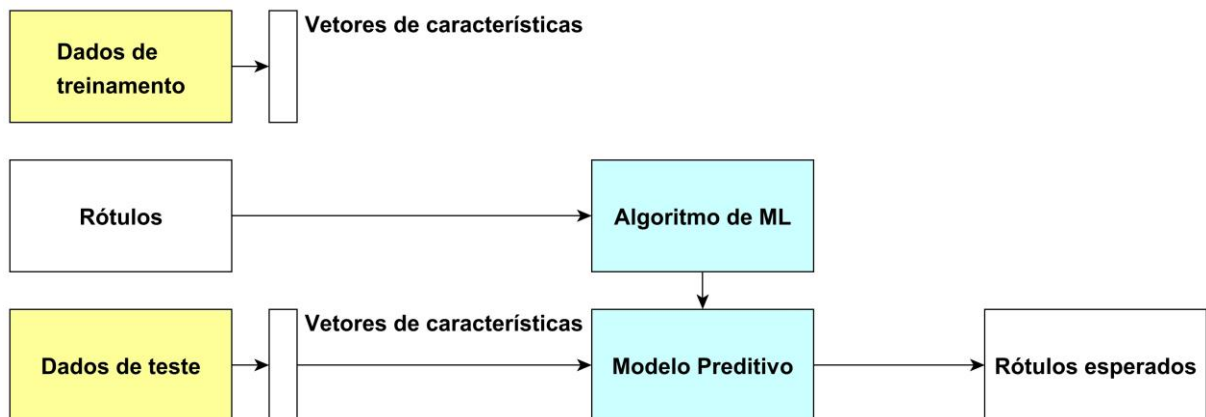
2.3.1 Aprendizado supervisionado

O aprendizado supervisionado é um método de ML que utiliza o conteúdo aprendido no passado a partir de dados rotulados para aplicar em um novo conjunto de dados (Manne; Kantheti, 2021). Para isto, os dados de entrada são separados em dados de treinamento e dados de teste, enquanto os dados de saída são previstos ou classificados a partir do conjunto de dados de treinamento. Desta maneira, o algoritmo tenta aprender formatos durante o treinamento e implantar esses padrões aprendidos para os dados de teste (Sharma; Sharma; Jindal, 2021).

Essa forma de aprendizado exige que as saídas do algoritmo sejam conhecidas e que os dados utilizados para o treinamento estejam rotulados. A partir destas respostas, o algoritmo compara a resposta correta com a resposta gerada e aprende com o resultado incorreto (Manne; Kantheti, 2021).

Para a rotulação dos dados é necessário um supervisor ou um especialista para fornecer as instruções para o programa de computador de maneira que o algoritmo aprenda as associações a partir da amostra de dados definida. Após serem aprendidas, a testagem é realizada para prever futuros exemplos e serem calculadas as métricas de avaliação do modelo (Kumar; Chauhan; Awasthi, 2023). Na Figura 3 é possível observar o modelo de aprendizado supervisionado.

Figura 3 – Modelo de Aprendizado Supervisionado



Fonte: Adaptado de Muhammad e Yan. (2015).

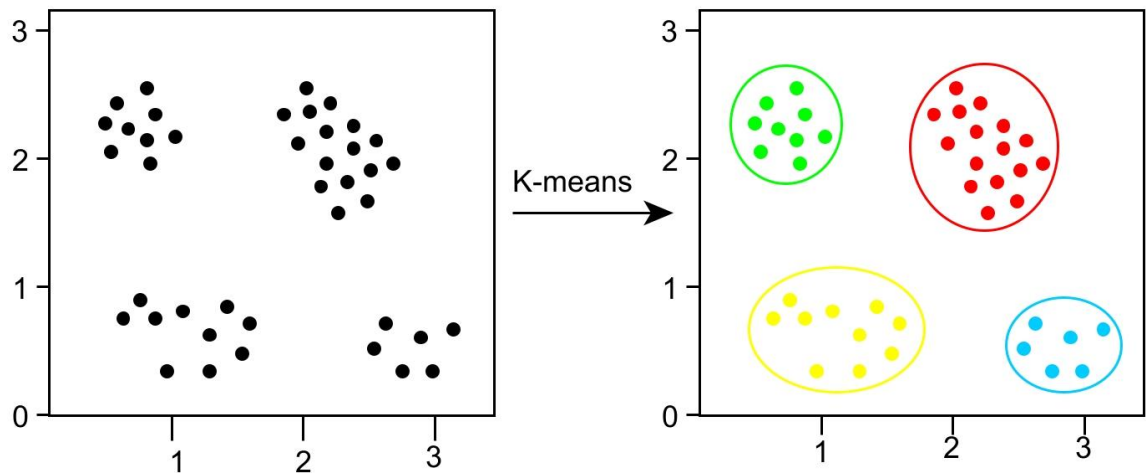
2.3.2 Aprendizado não supervisionado

No aprendizado não supervisionado, são utilizados os dados sem nenhum rótulo. O modelo não receberá nenhuma correlação entre entrada e saída ou resposta correta, o objetivo é que o algoritmo aprenda por si mesmo. Este tipo de aprendizado é mais complexo que o aprendizado supervisionado pelas suas características de rotulação (Manne; Kantheti, 2021).

Ao utilizar aprendizado não supervisionado, os computadores são capazes de descobrir padrões nos dados sem nenhum ensinamento explícito. Deste modo, geralmente é utilizado para algoritmos de agrupamento, processo de encontrar correlações desconhecidas entre os conjuntos de dados, com o objetivo de criar conjuntos com características compartilhadas (Kumar; Chauhan; Awasthi, 2023).

Um exemplo de algoritmo de agrupamento é o *K-means clustering*, nele são realizadas iterações para particionar pontos de dados em *clusters* distintos e que não estejam sobrepostos. Esse algoritmo visa minimizar a variação dentro do *cluster* ou a diferença entre as observações em um *cluster*. Para isso, é necessário especificar o número de agrupamentos necessários antes da sua execução (Eckhard, et al., 2023). Na Figura 4 é ilustrado um exemplo do funcionamento de um algoritmo *K-means clustering*.

Figura 4 – Algoritmo K-means Clustering



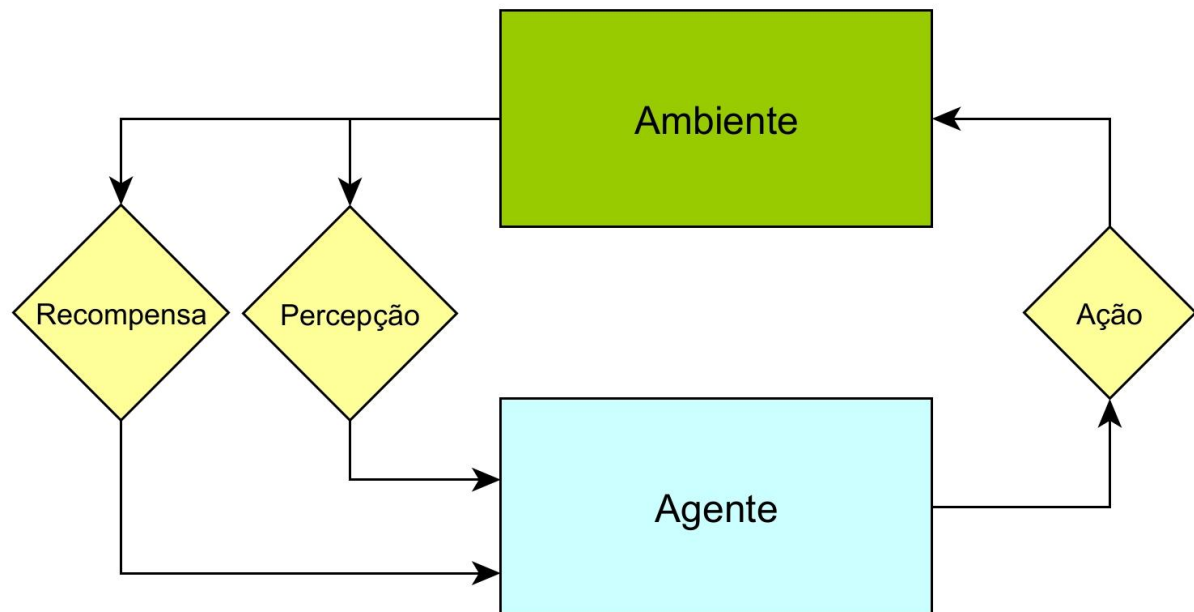
Fonte: Adaptado de Eckhard et al. (2023).

2.3.3 Aprendizado por reforço

O aprendizado por reforço ou *Reinforcement Learning* é baseado no conceito de recompensa e punição para treinar os algoritmos. Neste método de aprendizado, o modelo irá receber recompensas por executar corretamente e penalidades ao fazer incorretamente, o que faz com que o algoritmo aprenda as melhores recompensas e minimizar as penalidades (Manne; Kantheti, 2021).

No entanto, o algoritmo não possui as informações prévias dos dados, assim que uma situação é recebida, o agente aprende a decidir qual decisão tomar, o que gera consequências tanto para a situação presente quanto para as situações futuras (Sharma; Sharma; Jindal, 2021). Na Figura 5 é possível observar o funcionamento do aprendizado por reforço.

Figura 5 – Aprendizado por reforço

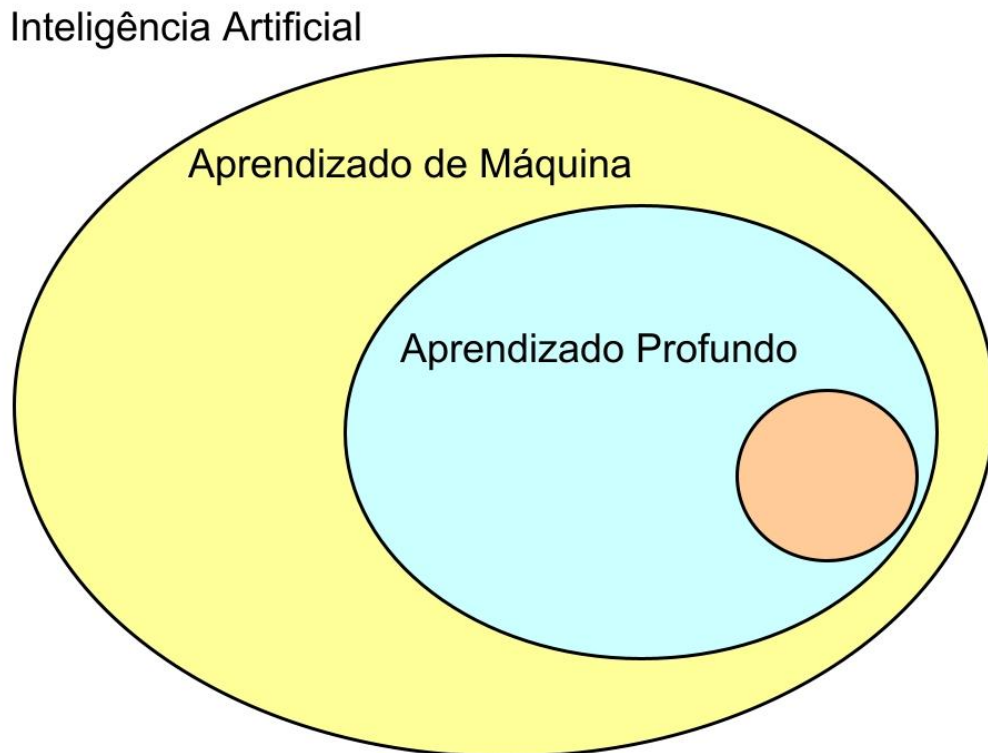


Fonte: Adaptado de Matsuo et al. (2022).

2.4 APRENDIZADO PROFUNDO

Aprendizado profundo ou *Deep Learning* (DL) é uma subdivisão do aprendizado de máquina (Jain; Prajapati, 2021). Sua característica fundamental é o uso de modelos computacionais compostos por múltiplas camadas de processamento para aprender representações de dados com vários níveis de abstração (LeCun, Bengio, Hinton, 2015). Na Figura 6 é ilustrada a relação entre o aprendizado profundo, inteligência artificial e o aprendizado de máquina.

Figura 6 – Relação entre as subáreas da Inteligência Artificial



Fonte: Adaptado de Shinde e Shah. (2018).

Diferente das técnicas de ML tradicionais, que frequentemente exigiam que especialistas humanos projetassem extratores de características para converter dados brutos em uma forma adequada para análise, os métodos de DL descobrem automaticamente as representações necessárias para detecção ou classificação diretamente dos dados (LeCun, Bengio, Hinton, 2015). O DL utiliza arquiteturas de modelos estabelecidas para representar a abstração de dados, de maneira a simular as funções de processamento de dados do cérebro humano, por meio de criação de padrões (Praful Bharadiya, 2023).

Em resumo, o aprendizado profundo utiliza múltiplas camadas de unidades de processamento não linear para a extração e transformação de características. As camadas inferiores, próximas à entrada de dados, aprendem as características mais simples, enquanto as camadas superiores aprendem características mais complexas (Shinde; Shah, 2018). Este processo torna o DL robusto em comparação com os métodos tradicionais de aprendizado de máquina (Sharma; Sharma; Jindal, 2021).

Deste modo, essa técnica é adequada para analisar e extrair conhecimento de grandes quantidades de dados de diferentes fontes. O termo “profundo” se refere

ao número de camadas em uma rede neural (Shinde; Shah, 2018). O processo de treinamento desses modelos é viabilizado principalmente pelo algoritmo de retropropagação.

Na aprendizagem supervisionada, o modelo recebe os dados de entrada, produz uma saída e compara essa saída com o resultado correto. O algoritmo de retropropagação calcula como cada parâmetro interno da rede contribui para o erro dessa previsão. Em seguida, os parâmetros são ajustados na direção que reduz esse erro. O processo é repetido milhões de vezes com um grande volume de dados rotulados (Wang et al., 2024; LeCun, Bengio, Hinton, 2015).

Os trabalhos em processamento de textos têm mostrado alta capacidade de ajuste, assim como uma capacidade poderosa de representação ao processar dados de alta dimensionalidade (Wang et al., 2024). Para dados sequenciais, como falas e texto, as Redes Neurais Recorrentes (RNNs) mantêm um “vetor de estado” que funciona como uma memória, o que permite processar sequências de comprimento variável. Uma evolução importante das RNNs são as redes *Long Short-Term Memory* (LSTM), que incorporam unidades de memória especiais capazes de aprender dependências de longo prazo, o que supera uma das principais limitações das RNNs tradicionais (LeCun, Bengio, Hinton, 2015).

Atualmente, o DL é aplicado em quase todos os campos, por isso é frequentemente denominado como um método de aprendizado universal. As áreas de Processamento de Linguagem Natural, Recuperação de informações e análise semântica são técnicas de aprendizado de máquina que já foram exploradas no passado, porém, em novas pesquisas podem ser aplicadas técnicas de DL (Sharma; Sharma; Jindal, 2021).

2.5 PROCESSAMENTO DE LINGUAGEM NATURAL

A complexidade da linguagem humana torna difícil a compreensão, seja pela sintaxe, semântica ou expressões, que são as tarefas mais difíceis para o aprendizado de um ser humano (Sharma; Sharma; Jindal, 2021). Com a ajuda do aprendizado profundo, o Processamento de Linguagem Natural (PLN) tem se tornado uma área de pesquisa em alta e atraído ampla atenção de muitas comunidades de pesquisa nos últimos anos (Zhou et al., 2022).

Em sua definição, o PLN é uma subárea da ciência da computação e inteligência artificial que atua na análise automática e compreensão da linguagem humana (Zhou et al., 2022). É utilizada para descrever o processo da utilização de algoritmos para identificar elementos chave na linguagem cotidiana e extrair significado a partir das entradas escritas e faladas (Jain; Prajapati, 2021).

Durante os anos 1950 e 1980, a pesquisa na área de PLN era focada na abordagem baseada em regras, na qual o computador segue instruções pré-definidas no programa. Porém, mesmo regras bem projetadas são limitadas para cobrir a linguagem humana em um todo, devido sua flexibilidade e complexidade. Para isso, desde a década de 1980, as pesquisas em PLN são focadas em abordagens baseadas em aprendizado de máquina, na qual é possível aplicar métodos de aprendizado supervisionado e não supervisionado (Jain; Prajapati, 2021; Zhou et al., 2022).

A análise de textos em IA utiliza PLN para transformar textos livres em dados estruturados após análise semântica e sintática apropriada. Esta técnica e a mineração de textos são capazes de converter dados não estruturados em formas estruturadas por meio de PLN. A introdução desta técnica no campo da IA resultou em grandes mudanças (Pn; M, 2021).

Para realizar a classificação de textos, o PLN ajuda na classificação de textos complexos em formatos estruturados a partir de etiquetas em dados não estruturados. As técnicas mais populares são análise de sentimentos, detecção de idioma, detecção de tópico e detecção de intenção. Por outro lado, para a extração de textos, o PLN atua em extrair entidades particulares do texto, como nome, localização, endereço, entre outras. Os métodos mais conhecidos são Reconhecimento de Entidade Nomeada, extração de atributos, e extração de palavras-chave (Pn; M, 2021).

2.5.1 Processamento de linguagem natural em prontuários médicos

Como uma grande quantidade de prontuários médicos eletrônicos está contida em narrativas clínicas, técnicas de PLN e DL têm sido utilizadas para a extração de informações (Jain; Prajapati, 2021). Os textos clínicos são originários de todos os cenários clínicos e são compostos principalmente por registros de textos não

estruturados, o que inclui notas médicas, diagnósticos, prescrições, entre outros (Zhou et al., 2022).

No contexto de dados médicos, os algoritmos de PLN extraem informações ou conceitos dos prontuários eletrônicos, depois processam as informações extraídas e, por fim, classificam os pacientes em subgrupos de acordo com as regras e aprendizados. Os procedimentos são complexos pois consistem em múltiplas técnicas combinadas (Jain; Prajapati, 2021). Apesar destes dados possuírem grandes quantidades de informações úteis em cuidados médicos, existem seus desafios e dificuldades que necessitam ser enfrentadas (Li et al., 2022), as quais incluem:

- a) Privacidade: devido a sensibilidade extrema da natureza das informações presentes em um prontuário e a existência de leis reguladoras, a manutenção da privacidade é imperativa. Portanto, antes que qualquer tarefa possa ser realizada ou que os dados possam ser compartilhados com outros, é preciso adotar etapas adicionais para a preservação da privacidade (Li et al., 2022).
- b) Falta de anotações: muitos modelos de aprendizado de máquina e aprendizado profundo são supervisionados e necessitam de dados rotulados para o treinamento. A anotação de dados de prontuários médicos é desafiadora devido à complexidade cognitiva da tarefa e à variabilidade na qualidade dos dados. Estes modelos necessitam de grandes quantidades de dados para o treinamento, porém, para algumas tarefas, apenas anotadores qualificados podem realizar as anotações, o que torna escassa a quantidade de dados úteis (Li et al., 2022).
- c) Interpretabilidade: redes neurais profundas são capazes de alcançar resultados superiores em relação a outros métodos, porém, em muitos campos são frequentemente tratadas como “caixas-pretas”. Normalmente, um modelo de rede neural tem uma grande quantidade de parâmetros treináveis, o que causa dificuldades para a interpretabilidade do modelo (Li et al., 2022).

Em geral, a extração de informações médicas temporais para transformar textos clínicos em uma linha do tempo é um desafio. Essa dificuldade se dá pela informação temporal estar escrita de maneira implícita, como “nas últimas 3 semanas”, “1 semestre após a cirurgia”, etc. Para construir uma linha do tempo é necessário extrair estas frases e normalizá-las em uma data exata. Avanços recentes em PLN

conseguiram ótimo desempenho para esses problemas por meio de modelos de linguagem (Yada; Aramaki, 2023).

2.6 RECONHECIMENTO DE ENTIDADES NOMEADAS

A extração de informação é a tarefa de identificar conteúdos importantes em textos não estruturados de forma automática. Neste sentido, o Reconhecimento de entidades nomeadas, *Named Entity Recognition* (NER) é importante para este tipo de tarefa em processamento de linguagem natural (Li et al., 2022; Sun et al., 2018).

O reconhecimento de entidades nomeadas é uma subtarefa do PLN que atua na identificação e categorização de entidades referenciadas em um texto em categorias predeterminadas, como pessoa, localização, organização e até expressões numéricas de data, moedas e percentuais (Yenkikar et al., 2024; Sun et al., 2018). O NER tem o objetivo de identificar semânticas de interesse em textos não estruturados, o que serve como base para muitas outras áreas de gerenciamento de informação, como população de ontologias, anotação semântica e mineração de opiniões (Marrero et al., 2013).

Para obter melhores desempenhos, é necessário que os sistemas NER compreendam as características das linguagens alvo. Estas linguagens são decisivas para as estratégias abordadas. O reconhecimento de entidades nomeadas monolíngues refere-se aos sistemas NER que utilizam apenas uma linguagem alvo, por outro lado, o reconhecimento de entidades nomeadas bilíngues se refere aos sistemas com duas linguagens alvo. Por fim, o reconhecimento de entidades multilíngues se refere aos sistemas que lidam com mais de duas linguagens alvo (Sun et al., 2018).

As abordagens de NER dependem muito da disponibilidade de conteúdos rotulados. Apesar dos métodos baseados em regras serem adotados, a maioria dos pesquisadores focam em aprendizado de máquina (Sun et al., 2018). Existem uma série de abordagens que podem ser utilizadas para a utilização do NER, entre elas, destacam-se as:

- a) Baseada em regras : as abordagens baseadas em regras são intuitivas para reconhecer entidades com a ajuda de recursos específicos da língua. Existem três maneiras de utilizar regras para lidar com essas tarefas: as regras (regras lexicais, contextuais e léxico morfológicas), a

combinação de regras e dicionários e a combinação de regras e algoritmos de seleção de atributos (Sun et al., 2018).

- b) Aprendizado de máquina: as abordagens de aprendizado de máquina são baseadas em conjuntos de dados rotulados de grande escala. O método de aprendizado supervisionado adquire conhecimento por meio de dados de treinamento rotulados e é uma das técnicas predominantes para abordar problemas de NER. O método de aprendizado semi-supervisionado aproveita dados não rotulados para moderar o impacto dos dados rotulados na precisão do sistema. Por fim, o método de aprendizado não supervisionado é baseado em agrupamentos, nos quais a informação de correferência desempenha um papel importante na tarefa do NER (Sun et al., 2018).
- c) Aprendizado profundo: o NER necessita de grandes quantidades de conhecimento linguístico na forma de léxicos para obter um bom desempenho. Para isto, o aprendizado profundo demonstrou eficácia para as tarefas de NER, ao detectar características em nível de palavras e caracteres de forma automática (Sun et al., 2018).

Pela razão da limitação da disponibilidade de recursos linguísticos, é necessário utilizar diferentes abordagens para as diferentes condições dos recursos. Para línguas ricas em recursos, os sistemas se expandem para domínios mais amplos, como notas médicas, por exemplo. Enquanto para as línguas com baixo recursos, diminuir a necessidade de dados rotulados e transferir o conhecimento de linguagens ricas em recurso são as melhores opções (Sun et al., 2018).

2.6.1 Reconhecimento de entidades nomeadas em prontuários médicos

Devido as características não estruturadas dos textos provenientes de prontuários médicos eletrônicos, o reconhecimento de entidades nomeadas se apresenta como uma importante solução no âmbito da extração de informações. Os registros médicos possuem uma grande quantidade de informações úteis, como sintomas de doenças e diagnóstico de pacientes, o que auxilia na criação de modelos que fazem previsões de doenças ou grafos de conhecimento médicos (Dai et al., 2019).

No contexto médico, o NER tem como objetivo identificar as entidades médicas do prontuário do paciente e classificá-las em categorias predefinidas, como doenças, medicamentos e operações (Dai et al., 2019). O NER atua em muitas atividades de PLN, como interação entre medicamentos e completude de bases de conhecimento (Yenkikar et al., 2024).

Em contextos genéricos, o NER possui a dificuldade da disponibilidade de dados rotulados para o treinamento de seus modelos. No contexto médico, este problema se torna mais dificultoso ainda devido à falta de grandes volumes destes dados rotulados e à complexidade do entendimento do domínio (Yenkikar et al., 2024).

A dificuldade da aquisição de grandes quantidades de dados rotulados tem influência da dificuldade da anotação de dados médicos, devido à complexidade da tarefa à variabilidade na qualidade dos dados. Determinadas tarefas podem ser realizadas somente por profissionais capacitados e outras precisam ser realizadas por múltiplos profissionais, o que pode gerar divergências nas anotações e comprometer a confiabilidade dos dados (Li et al., 2022).

2.7 MODELOS DE LINGUAGEM

A modelagem de linguagem é uma das maiores abordagens para o avanço da inteligência linguística em máquinas. Em geral, essa abordagem visa modelar a probabilidade de sequências de palavras, o que permite a previsão de tokens futuros ou ausentes. As pesquisas voltadas para Modelos de Linguagem, têm recebido atenção na literatura e podem ser divididos em quatro estágios de desenvolvimento, Modelos de linguagem estatísticos (SLM), Modelos de linguagem neural (NLM), Modelos de linguagem pré-treinados (PLM) e os Modelos de linguagem de grande escala (LLM) (Zhao et al., 2023).

Os Modelos de linguagem pré-treinados são baseados em aprendizado profundo e utilizam geralmente a arquitetura de *Transformers*, que são treinados em vastos corpora de texto não rotulados por meio de métodos supervisionados. O processo de treinamento ocorre em duas etapas principais: primeiro os modelos são pré-treinados em uma tarefa de modelagem de linguagem em grande escala, em seguida, são ajustados com dados rotulados em menor escala para tarefas específicas de PLN (Hu et al., 2023).

Por último, os Modelos de linguagem de grande escala são modelos com uma quantidade massiva de parâmetros que passa por tarefas de pré-treinamento, assim como as PLMs, para compreender e processar a linguagem humana. Um modelo eficaz deve possuir quatro características: compreensão profunda do contexto da linguagem natural, capacidade de gerar textos semelhantes aos seres humanos, sensibilidade ao contexto e uma forte habilidade de seguir instruções (Yao et al., 2024).

2.7.1 Transformador

A Rede Neural Recorrente foi uma das abordagens que se estabeleceram no estado da arte em PLN. A arquitetura das RNNs possui um laço de *feedback* nas conexões da rede, o que permite que a informação se propague de uma etapa para outra e as tornam ideais para problemas de modelagem de sequência e tradução automática (Tunstall et al., 2022; Vaswani et al., 2017).

Estes modelos recorrentes geralmente computam os cálculos ao longo das posições dos símbolos nas sequências de entradas e saídas. Para calcular os valores de uma sequência, cada estado depende do estado anterior, ou seja, o significado de cada palavra depende da palavra anterior. Essa característica, além de impedir a paralelização, torna o modelo incapaz de obter bons resultados a partir de sequências longas, pois ao percorrer por ela, as informações das primeiras palavras passam a ser esquecidas (Vaswani et al., 2017).

Para contornar esses problemas, mecanismos de atenção se tornaram uma abordagem essencial para modelos de sequências e transduções pela sua natureza de não considerar a distância entre as palavras nas sequências de entradas. Nesse contexto, foi criada a arquitetura transformador, do inglês *Transformer*, que se baseia inteiramente em um mecanismo de atenção (Vaswani et al., 2017).

O *Transformer* é baseado na arquitetura *encoder-decoder*, muito utilizada para tarefas de tradução automática, na qual uma sequência de palavras é traduzida de um idioma para outro (Tunstall et al., 2022). Essa arquitetura é composta por dois componentes:

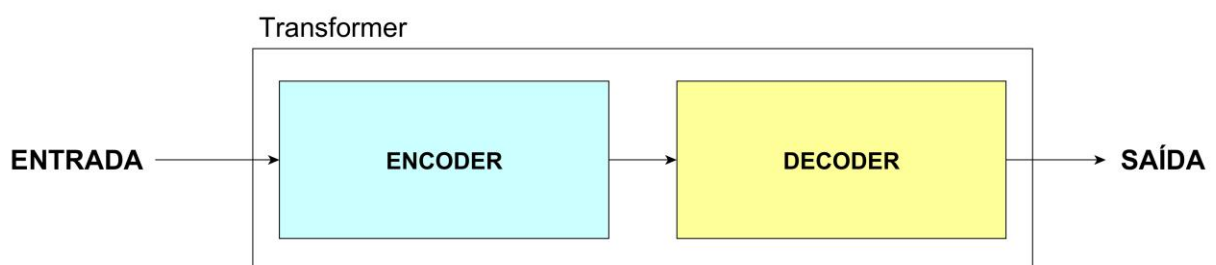
- a) *Encoder*: também conhecido como codificador, converte a sequência de palavras de entrada em uma sequência de vetores de valores para essas palavras. Esses valores presentes nos vetores são chamados de

embedding e os vetores de *embeddings* são chamados de estados ocultos ou contexto (Tunstall et al., 2022).

- b) *Decoder*: o *Decoder* ou decodificador, utiliza os estados ocultos do codificador para gerar uma sequência de saída de maneira iterativa (Tunstall et al., 2022).

Na Figura 7 é possível visualizar os dois componentes da arquitetura, que serão detalhados a seguir.

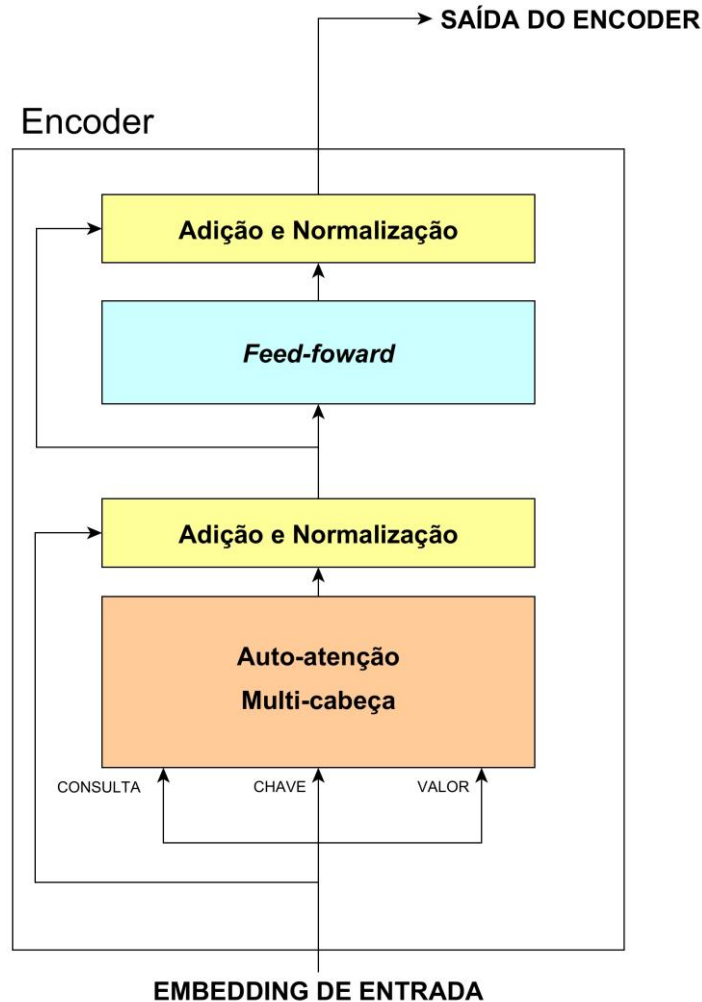
Figura 7 – Arquitetura Transformer



Fonte: Adaptado de Vaswani et al. (2017).

A estrutura do *Encoder* proposto por Vaswani et al., (2017) é composta por 6 camadas empilhadas as quais possuem duas subcamadas cada uma, como apresentado na Figura 8. A primeira subcamada é o mecanismo de atenção chamado de Auto-Atenção Multi-cabeças, do inglês *multi-head self-attention* e a segunda é uma subcamada *feed-forward* totalmente conectada.

Além disso, em torno de cada uma dessas subcamadas é aplicada uma técnica para somar a informação da camada anterior com a nova, chamada de conexão residual, para que a informação contida antes dos processos seja preservada. Além da conexão residual, é aplicada a normalização para que os valores mantenham no intervalo adequado. A saída de cada subcamada é $\text{LayerNorm}(x + \text{Sublayer}(x))$, onde $\text{Sublayer}(x)$ é a função implementada pela própria subcamada.

Figura 8 – Arquitetura do *Encoder*

Fonte: Adaptado de Vaswani et al. (2017).

A atenção é um mecanismo que permite que as redes neurais definam um peso para cada elemento em uma sequência. Para cada sequência de texto, os elementos são *embeddings* de tokens, nos quais cada token é mapeado para um vetor em uma dimensão fixa. A ideia principal do conceito de Auto-Atenção é utilizar a sequência como um todo para computar a média dos pesos de cada *embedding* ao invés de utilizar um vetor fixo para cada elemento da sequência (Tunstall et al., 2022).

A camada de atenção aplica três transformações lineares para obter os vetores de consulta, chave e valor. Cada projeção é associada a um conjunto de parâmetros ajustáveis, o que permite que a camada de atenção foque em diferentes

aspectos semânticos da sequência (Tunstall et al., 2022). Neste sentido, ao invés de aplicar uma única função de atenção, é vantajoso projetar esses vetores de consulta, chave e valor em várias projeções diferentes. Nessas projeções é possível executar a função de atenção de maneira paralela, o que é uma grande vantagem em comparação a outras arquiteturas como as RNNs. No final os valores são concatenados para gerar os valores finais. Essa técnica de aplicar o mecanismo de atenção em diversas projeções permite que o modelo possa atender as informações de diversos subespaços da representação em diferentes posições (Vaswani et al., 2017).

Existem diversas maneiras de se implementar uma camada de atenção, mas a mais comum é a “Atenção por Produto Escalar Escalonado”, do inglês *Scaled Dot-Product Attention*, mostrada na Fórmula (1). Esse processo passa por quatro etapas principais: na primeira, cada *embedding* é projetado em três vetores chamados de consulta(Q), chave(K) e valor(V). Para um determinado token, seu vetor de consulta é comparado com os vetores de chave de todos os outros tokens na sequência. A similaridade (calculada pelo produto escalar) entre a consulta de um token e a chave de outro determina o peso da atenção. Este peso é usado para ponderar o vetor de valor do segundo token. A representação final do token é a soma ponderada de todos os vetores de valor na sequência (Vaswani et al., 2017).

$$\text{Atenção}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

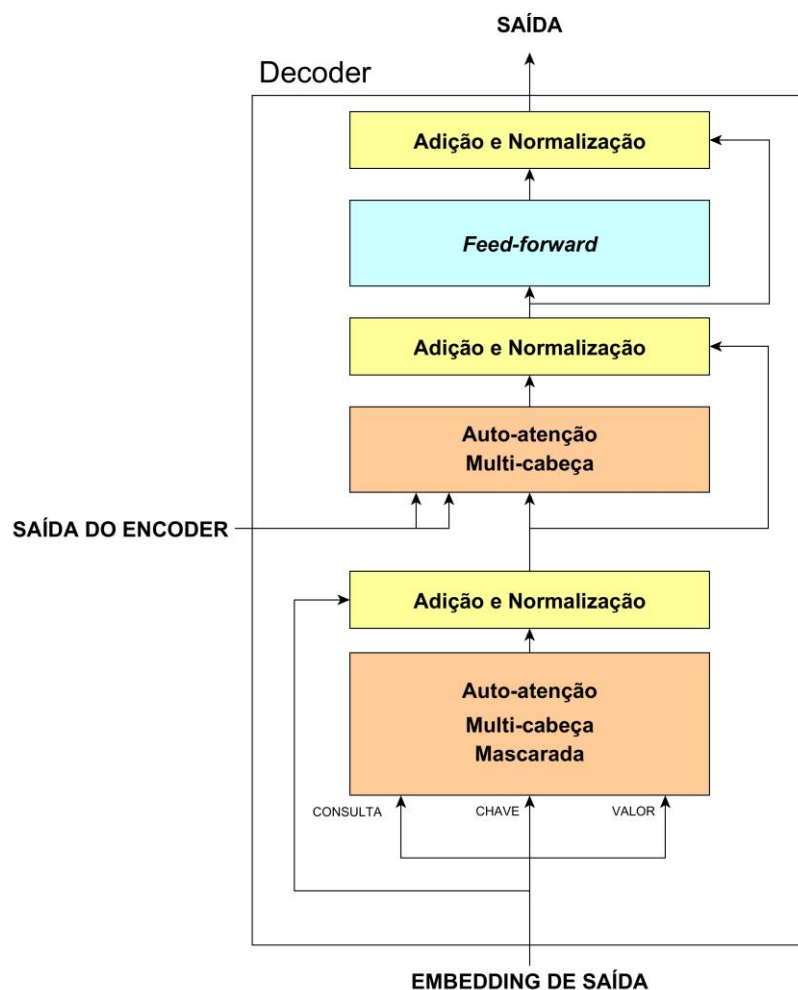
A escala por $\sqrt{d_k}$ (raiz quadrada da dimensão dos vetores de chave) é um detalhe crucial que estabiliza o treinamento, que evita que os produtos escalares se tornem muito grandes e saturam a função *softmax* (Vaswani et al., 2017).

Após isso, esses valores são ajustados por um fator de escala e normalizados a partir da função de *softmax*, o que resulta nos pesos de atenção. Por fim, os pesos de atenção são multiplicados pelo vetor dos valores para gerar um novo *embedding* com as representações atualizadas de cada token (Tunstall et al., 2022).

A subcamada de *feed-forward* é uma rede neural simples totalmente conectada de duas camadas com uma particularidade: ao invés de processar toda a sequência de *embeddings* como um único vetor, ela processa cada vetor individualmente (Tunstall et al., 2022).

A principal diferença entre o *encoder* e o *decoder* é que o *decoder* possui duas subcamadas de atenção. A primeira é chamada de “Auto-Atenção Multi-cabeças Mascarada” (*Masked Multi-Head Self-Attention*) e serve para garantir que o modelo gere os tokens baseados nas saídas anteriores e atuais, pois o decodificador poderia apenas copiar a sequência que está sendo tratada. Mascarar a entrada garante que o modelo aprenda de forma correta. A segunda é chamada de “*Encoder-decoder attention*” que realiza a atenção multi-cabeças sobre os vetores de chave e valor gerados pelo *encoder* em conjunto com as representações geradas pelo *decoder* como vetor de consultas. Dessa maneira, essa camada de atenção aprende a relacionar os tokens de duas sequências diferentes, como dois idiomas diferentes (Tunstall et al., 2022). O processo do *Decoder* pode ser visualizado pela Figura 9, no qual suas subcamadas são detalhadas.

Figura 9 – Arquitetura do *Decoder*



Fonte: Adaptado de Vaswani et al. (2017).

A arquitetura *Transformer* foi o primeiro modelo de transdução de sequência baseado inteiramente em atenção, o que obteve uma grande vantagem em relação aos modelos recorrentes mais comumente utilizados (Vaswani et al., 2017). A partir disso, diversos modelos de linguagem puderam usufruir desta arquitetura, além disso, esses modelos podem se basear em apenas um componente da arquitetura, *encoder* ou *decoder*, ou se basear nos dois, chamados de *encoder-decoder*.

2.8 REPRESENTAÇÃO DE CODIFICADORES BIDIRECIONAIS DE TRANSFORMERS

Um exemplo de modelo de linguagem que utiliza apenas o *encoder* de um *transformer* é o BERT. O BERT é um modelo de linguagem que significa *Bidirectional Encoder Representations from Transformers* (Representação de Codificadores Bidirecionais de Transformers), sua arquitetura é baseada no *encoder* da implementação original dos *transformers* (Devlin et al., 2018).

Ele é projetado para pré-treinar representações bidirecionais profundas a partir de textos não rotulados, ou seja, ao tratar de uma sequência, ele utiliza tanto a palavra à esquerda quanto a palavra à direita para considerar o significado dela no contexto. O modelo é conceitualmente simples e empiricamente poderoso (Devlin et al., 2018).

O BERT possui duas etapas em sua implementação: pré-treinamento e ajuste fino. Durante o pré-treinamento, o modelo é treinado em dados não rotulados sobre diferentes tarefas de pré-treinamento. Para o ajuste fino, o modelo é iniciado primeiramente com os parâmetros pré-treinados, e todos os são ajustados com os dados rotulados das tarefas seguintes. Cada tarefa tem um modelo ajustado separado, mesmo que sejam treinados com os mesmos parâmetros pré-treinados (Devlin et al., 2018).

2.8.1 Pré-treinamento

Ao invés, de utilizar modelos tradicionais esquerda-para-direita ou direita-para-esquerda, o BERT é pré-treinado a partir de duas tarefas não supervisionadas (Devlin et al., 2018).

A tarefa número 1 é chamada de “Modelo de Linguagem com Máscara”. É intuitivo que um modelo bidirecional profundo é mais poderoso que um modelo unidirecional da esquerda para a direita ou da direita para a esquerda, porém, os modelos de linguagem condicional podem ser treinados apenas de forma unidirecional pois o condicionamento bidirecional poderia fazer com que cada palavra “olhasse a si mesma”, isto é, cada palavra poderia ter acesso a si mesma através das camadas do modelo. O que compromete o treinamento, pois, ao invés do modelo aprender padrões linguísticos novos, ele prevê a própria palavra (Devlin et al., 2018).

Para treinar uma representação bidirecional, é preciso mascarar uma porcentagem dos tokens de entrada de maneira aleatória e, após isso, prever esses tokens mascarados. Esse processo é chamado de “Modelo de Linguagem com Máscara” (MLM). Nos experimentos do BERT, foram mascarados aleatoriamente 15% dos tokens em cada sequência (Devlin et al., 2018).

Apesar desta técnica permitir a criação de um modelo bidirecional, uma desvantagem é a discrepância entre o pré-treinamento e o ajuste fino, uma vez que o token [MASK] não aparece durante a fase do ajuste fino. Para contornar este problema, as palavras “mascaradas” nem sempre são substituídas pelo token real (Devlin et al., 2018).

A tarefa número 2 é chamada de “Previsão da Próxima Sentença”, do inglês *Next Sentence Prediction* (NSP). Muitas tarefas importantes como Perguntas e Respostas e Inferência de Linguagem Natural são baseadas no entendimento da relação entre duas sentenças, o que não é capturado diretamente pelo modelo de linguagem (Devlin et al., 2018).

Para que o modelo seja treinado de modo que compreenda essa relação entre sentenças, o pré-treinamento é feito para uma tarefa binária de previsão da próxima sentença, que pode ser gerada de forma simples a partir de qualquer corpus monolíngue. Por exemplo, ao escolher duas sentenças A e B para cada exemplo do pré-treinamento, 50% do tempo B é a próxima sentença que segue A (rotulada como *IsNext*), enquanto 50% do tempo é uma sentença aleatória do corpus (rotulada como *NotNext*) (Devlin et al., 2018).

Por fim, o procedimento de pré-treinamento do BERT segue em grande parte a literatura existente sobre modelos de linguagem. Os dados utilizados foram o *BookCorpus* com uma quantidade de 800 milhões de palavras e o site da *Wikipedia*

no idioma em inglês, com cerca de 2,5 bilhões de palavras, sem considerar listas, tabelas e cabeçalhos (Devlin et al., 2018).

2.8.2 Ajuste fino

No BERT o ajuste fino é feito de maneira direta, uma vez que o mecanismo de atenção do *transformer* permite que o BERT modele muitas tarefas, tanto as que envolvem um único texto ou pares de texto, por meio da troca entre as entradas e saídas apropriadas. Para aplicações que envolvem pares de textos, um padrão comum é codificar os pares de forma independente antes de aplicar a atenção cruzada (Devlin et al., 2018).

No entanto, o BERT usa o mecanismo de atenção para unificar essas duas etapas, pois a codificação de um par de texto concatenado com atenção inclui a atenção cruzada bidirecional entre duas sentenças. Para cada tarefa, é necessário conectar as entradas e saídas específicas no BERT e ajustar todos os parâmetros de ponta a ponta (Devlin et al., 2018).

2.8.3 Métricas de avaliação

A avaliação de modelos ajustados para a atividade de NER é um aspecto crítico para garantir sua confiabilidade e eficácia na extração e categorização de entidades em dados textuais não estruturados (Peddavenkatagar, 2024). Embora a acurácia seja uma métrica amplamente utilizada em modelos de classificação baseados em arquiteturas como o BERT, sua aplicação no contexto de reconhecimento de entidades é inadequada devido à natureza esparsa das classes de interesse (Manning, 2008).

Em tarefas que a classe alvo é rara frente à classe majoritária, como o rótulo *Outside* no NER, um modelo poderia atingir uma acurácia superior a 99% simplesmente ao classificar todos os termos como não-entidades, o que resulta em uma ferramenta funcionalmente inútil para a extração de informações. Esta limitação é implicitamente reconhecida na literatura ao focar exclusivamente em métricas de erros baseadas em acertos reais de entidades e não na contagem total de palavras corretas (Manning, 2008).

Deste modo, uma das principais métricas usadas na avaliação de um sistema NER é a Precisão. A precisão tem como característica medir a correção das entidades nomeadas extraídas. Para isso, é realizado o cálculo de proporção de entidades corretamente identificadas em relação ao número total de entidades previstas pelo sistema (Peddavenkatagar, 2024; Manning, 2008). Matematicamente, ela é expressa pela razão entre Verdadeiros Positivos (VP) e a soma de Verdadeiros Positivos com Falsos Positivos (FP), conforme mostrado em (2).

$$P = \frac{VP}{VP + FP} \quad (2)$$

A métrica de Revocação, por sua vez, mede a completude ou capacidade de cobertura do modelo. Ela calcula a proporção de entidades corretamente identificadas em relação ao volume total de entidades efetivamente presentes no conjunto de dados (Peddavenkatagar, 2024). Sua fórmula é dada pela razão entre VP e a soma de Verdadeiros Positivos com Falsos Negativos (FN), apresentada em (3).

$$R = \frac{VP}{VP + FN} \quad (3)$$

Por fim, o F-Score (ou F-measure) é definido como a média harmônica entre a precisão e a revocação, a qual oferece uma avaliação equilibrada e abrangente do desempenho do sistema (Peddavenkatagar, 2024). Geralmente, utiliza-se o F1-Score, em que o parâmetro B é igual a 1, para atribuir pesos iguais a ambas as métricas (Niero et al., 2023). A fórmula geral do F_β é representada por (4).

$$F_\beta = (1 + \beta^2) \cdot \frac{P \cdot R}{(\beta^2 \cdot P) + R} \quad (4)$$

2.8.4 Modelo de linguagem de domínio biológico

Devido as características arquiteturais do BERT, é possível adaptá-lo para diferentes contextos conforme o conjunto de dados utilizado no pré-treinamento e no

ajuste fino do modelo, assim como as tarefas de PLN. Entre esses diferentes domínios, encontra-se o biomédico, no qual são utilizados dados clínicos e biomédicos em seu desenvolvimento. Um modelo atuante neste domínio é o *BioBERT*, seu desenvolvimento é baseado na arquitetura do BERT com o foco em extrair informações de textos médicos por meio das tarefas de PLN (Lee et al., 2020).

Durante o desenvolvimento do modelo, o *BioBERT* foi inicializado com os pesos do BERT padrão e, após isso, foi realizada a etapa de pré-treinamento a partir de 4 corpus distintos, a Wikipédia no idioma inglês com 2.5 bilhões de palavras e o *BooksCorpus*, um conjunto de dados de livros, com cerca de 800 milhões de palavras, ambos com conteúdo de domínio geral, e os dados da *PubMed Abstracts* com 4.5 bilhões de palavras e a *PMC Full-text articles* com cerca de 13.5 bilhões de palavras, ambos os corpus no domínio biomédico (Lee et al., 2020).

Em seguida, para mostrar a eficiência do modelo em tarefas como mineração de textos, foi realizado um ajuste fino no *BioBERT* para que atendesse à 3 tarefas populares na área: o Reconhecimento de Entidades Nomeadas, Extração de relações e Perguntas e respostas (Lee et al., 2020).

O Reconhecimento de Entidades Nomeadas foi adaptado para o reconhecimento de entidades de categorias biomédicas, como doenças e medicamentos. A Extração de Relações é a tarefa de classificar as relações entre as entidades nomeadas no corpus biomédico. Por fim, Perguntas e Respostas é a tarefa de responder as questões escritas na linguagem natural (Lee et al., 2020).

2.8.5 Modelo de linguagem de domínio biológico em português

Embora o desempenho das tarefas de PLN, como o NER, em corpus em inglês tenha melhorado com a criação de modelos de linguagem contextualizados, as pesquisas realizadas em textos clínicos em idiomas com poucos recursos, como o português, são escassas. Nesse sentido, parte dessas pesquisas voltadas para a criação de modelos de linguagem se baseiam em arquiteturas conceituadas na literatura. O *BioBERTpt* é um modelo baseado na arquitetura do BERT que foi criado para atuar no contexto de dados clínicos e biomédicos em português (Schneider et al., 2020).

No total, os dados clínicos foram coletados em hospitais brasileiros no ano de 2002 até 2018 e contém 3.8 milhões de sentenças com 27.7 milhões de palavras.

Os conjuntos de dados biomédicos utilizam títulos e resumos de artigos científicos em português publicados pela *PubMed* e a *Scielo* (*Scientific Electronic Library Online*), nos quais contém cerca de 812 mil e 15 milhões de palavras, respectivamente (Schneider et al., 2020).

Em seu desenvolvimento, foi realizada a etapa do ajuste fino em três modelos baseados no BERT a partir dos dados clínicos e biomédicos extraídos. Para avaliar como o domínio pode influenciar no desempenho do NER, os modelos baseados foram divididos em: um modelo com os dados clínicos de hospitais brasileiros, um modelo com dados biomédicos retirados de artigos científicos e, por fim, um modelo com os dois tipos de dados. Esses modelos foram nomeados, respectivamente por *BioBERTpt(clin)*, *BioBERTpt(bio)* e *BioBERTpt(all)* (Schneider et al., 2020).

Em geral, o desenvolvimento de pesquisas na área de PLN a partir de textos clínicos em português é inferior se comparado ao inglês, contudo, a elaboração de trabalhos na área se mostra necessário uma vez que os modelos em inglês não conseguem atender os desafios e necessidades presentes em hospitais no Brasil.

2.9 CIÊNCIA DE DADOS E VISUALIZAÇÃO

Em termos gerais, a ciência de dados pode ser definida como um conjunto de princípios essenciais que sustentam e direcionam a obtenção de informações e conhecimentos a partir dos dados (Prevost; Fawcett, 2013). Trata-se de uma área que envolve a aplicação e o aprimoramento de algoritmos, rotinas, metodologias e estratégias para a compreender fenômenos com o objetivo de aprimorar o processo decisório (Salazar-Reyna et al., 2020). Estruturada sobre as etapas de coleta, análise e interpretação de dados, a área desempenha um papel crucial no setor da saúde, o grande volume de dados clínicos gerados representa uma enorme oportunidade para a análise por meio de algoritmos especializados (Subrahmanya et al., 2021; Chinthamu; Karukuri, 2023).

Como parte integrante do processo analítico, a visualização de dados pode ser definida como um procedimento para descrever uma grande quantidade de dados em um formato acessível, e compreensível (Muskan et al., 2022). Além de converter dados de várias fontes em visuais (comprimento, localização, forma, cor, entre outros) é uma ferramenta importante que oferece o caminho certo para uma tomada de

decisões com menos erros quando são utilizados dados precisos (Muskan et al., 2022; Muniswamaiah; Agerwala; Tappert, 2023).

As ferramentas de visualização devem considerar as habilidades de percepção das pessoas pois os usuários são acostumados a reconhecerem padrões ao observar. Nesse caso, deve auxiliar na clareza e na aparência da informação mostrada pelo sistema (Muniswamaiah; Agerwala; Tappert, 2023). Diferentes tipos de dados utilizam diferentes tecnologias para a representação visual, o que inclui tecnologias baseadas em gráficos, projeção geométrica, ícones e pixels. Exceto pelos dados tradicionais, surgiram tecnologias baseadas em interação humano computador que permitem a melhor exploração dos dados (Qu; Wang, 2021).

Visualização de dados é a combinação perfeita para oferecer uma visão geral abrangente de grandes quantidades de dados e tornar a compreensão de descobertas muito mais fácil, por meio de gráficos, mapas e grafos (Muskan et al., 2022). Em alguns casos, é um desafio detectar anomalias nos dados, o que pode implicar em aspectos negativos nas decisões e compreensões. O sistema de visualização deve utilizar dados precisos para enriquecer a clareza e a qualidade das decisões feitas (Muniswamaiah; Agerwala; Tappert, 2023).

2.9.1 Visualização analítica

A visualização analítica é um campo de pesquisa multidisciplinar fundamentado principalmente em visualização, análise algorítmica de dados e raciocínio analítico, que utiliza tanto de representações gráficas quanto interações como instrumentos para adequar o julgamento humano ao processo de extração de conhecimento (Cui, 2019). As ferramentas e abordagens de visualização analítica são utilizadas para sintetizar informações e gerar percepções a partir de grandes volumes de dados dinâmicos, ambíguos e, muitas vezes, conflitantes. Além disso, são utilizadas para fornecer avaliações rápidas e compreensíveis, comunicar os resultados de maneira eficaz para apoiar a tomada de decisões, identificar padrões esperados e revelar o inesperado (Thomas; Cook, 2006).

A visualização analítica explora a visualização de dados como um meio para integrar a cognição, as habilidades de raciocínio e a inteligência humanas ao processo de análise de dados para obter resultados explicáveis. Em comparação à simples visualização, a visualização analítica prioriza a interpretação e a descoberta

de conhecimento a partir dos dados, em vez de limitar-se à apenas apresentá-los e compreendê-los (Cui, 2019).

Esse tipo de recurso deve viabilizar técnicas analíticas que possibilitem a formulação de hipóteses e cenários, além de apoiar o usuário na avaliação com base nas evidências disponíveis. A visualização analítica deve contribuir para que o analista identifique relações inesperadas ou ausentes, capazes de gerar *insights* importantes. Ademais, é importante permitir que os usuários registrem sua progressão durante o processo de análise e compartilhem essas informações como referência para futuras análises (Thomas; Cook, 2006).

Na área da saúde, existem diversas aplicações potenciais da visualização analítica, como na detecção de doenças, desenvolvimento e pesquisa de medicamentos, e em múltiplos aspectos do atendimento ao paciente em hospitais, clínicas, laboratório, hospitais psiquiátricos ou em ambientes domiciliares (Thomas; Kielman, 2009)

Nesse cenário, há uma necessidade de apoiar pesquisas e projetos que explorem maneiras eficazes de utilizar a visualização analítica para auxiliar na análise de grandes conjuntos de dados médicos. Para isso, essa abordagem pode fornecer aos profissionais da saúde meios mais eficientes de integrar dados clínicos longitudinais com informações geradas pelos próprios pacientes, o que fornece uma compreensão mais aprofundada da evolução clínica. Sistemas de visualização analítica apresentam o potencial de apoiar análises intuitivas para todos esses perfis de usuários, de maneira a mascarar a complexidade subjacente dos dados (Caban; Gotz, 2015).

2.9.2 Gráficos baseados em linha do tempo

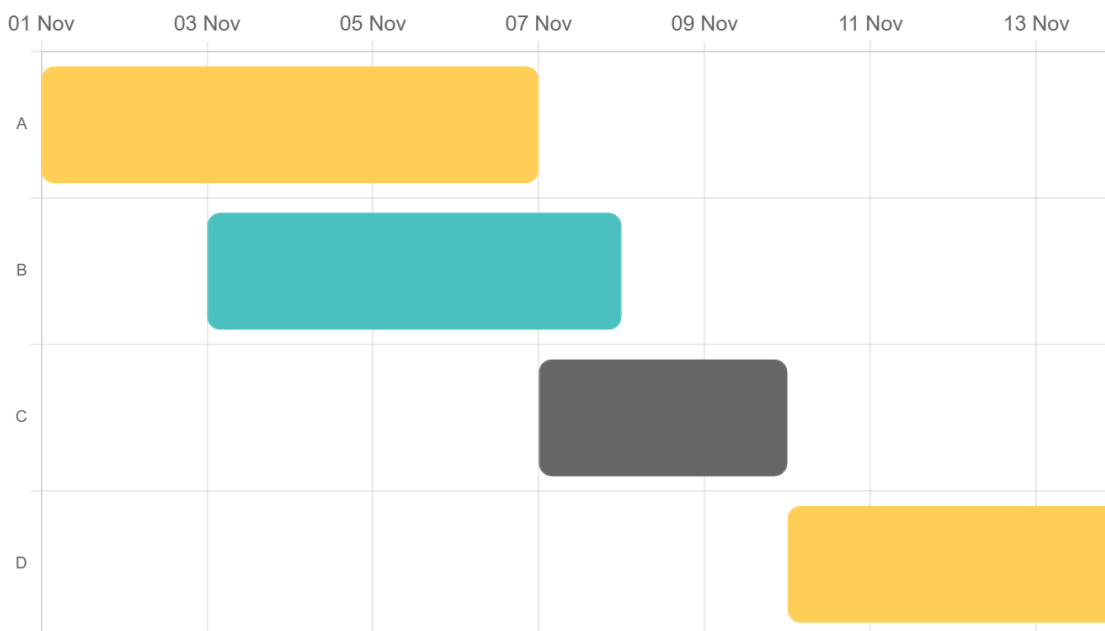
A área da visualização de dados possui diversos objetivos, um exemplo é a criação de um ambiente que seja representativo do conjunto de dados estudado, motivado a dar contexto às representações brutas, o que permite que o usuário tenha uma compreensão mais completa da significância dos dados. Apesar da visualização de um único conjunto de dados ser bastante direta, múltiplos conjuntos simultaneamente levantam diversos desafios para a criação de representações com clareza e compreensibilidade. Neste contexto, os gráficos de linha do tempo se

mostraram eficazes na representação visual de múltiplos conjuntos de dados, nos quais possuem dados temporais (Aoyama et al., 2007).

O gerenciamento de dados temporais é um tema antigo na história, desde seus primórdios, as pessoas utilizam meios para visualizar uma lista de eventos em ordem cronológica, com o objetivo de entender os acontecimentos e tendências a partir dessas informações. Eventos são basicamente acontecimentos orientados pelo tempo que incluem outras informações contextuais. Deste modo, a visualização dos dados por meio de linhas do tempo é mais natural e intuitiva para a compreensão do ser humano. Do ponto de vista psicológico, a ação mental não é instantânea e requer uma quantidade de tempo, além de seu comportamento permanecer sintonizado com as ações proporcionados pelo ambiente (Hasan et al., 2013).

Um exemplo de gráfico baseado em linha do tempo é o gráfico de Gantt, no qual foi desenvolvido por volta dos anos 1900 e foi projetado para visualizar o progresso de tarefas realizadas em fábricas (Jo et al., 2014). Na Figura 10 é possível observar um exemplo tradicional do gráfico com uma simulação de um projeto fictício.

Figura 10 – Exemplo de tarefas dispostas em um gráfico de Gantt



Fonte: Elaborado pelo autor.

No gráfico mostrado na Figura 10, o eixo X representa o tempo analisado e o eixo Y as tarefas a serem realizadas nesse projeto. As barras presentes no gráfico representam as tarefas A, B, C e D neste exemplo. Neste tipo de visualização, as

tarefas podem ser executadas de maneira paralela, como as tarefas A e B. Além disso, é possível representar a dependência de tarefas, na qual a tarefa D depende da tarefa C ser concluída para que se inicie (Jo et al., 2014).

No contexto da saúde, esses gráficos podem representar eventos clínicos de um paciente, como testes e tratamentos a partir de dados pré-formatados, ou seja, os eventos associados com as suas informações de data e hora (Yada; Aramaki, 2023). Além disso, a utilização do gráfico de Gantt permite que o profissional da saúde tenha uma visão mais clara a respeito do histórico do paciente, o que facilita na sua tomada de decisões.

2.10 TRABALHOS CORRELATOS

Neste subcapítulo são apresentados os principais trabalhos correlatos da literatura que realizam o reconhecimento de entidades a partir de textos para a visualização da informação ou então utilizam para outros contextos. Também é apresentado um quadro comparativo com as principais características dos trabalhos selecionados.

2.10.1 Abordagem visual para reconhecimento de entidades biomédicas nomeadas

O trabalho publicado por Yenikar et al., (2024) apresenta uma análise sobre o uso da ferramenta *spaCy* como alternativa para a visualização e identificação de entidades biomédicas. A ferramenta *spaCy* foi aplicada com o objetivo de alavancar as informações sintáticas auto processadas em diversas tarefas. Para os experimentos, utilizou-se a base de dados *BioNLP Cancer Genetics*, voltada para estudos em processamento de linguagem natural, focada em cânceres genéticos.

Além do *spaCy*, outros quatro modelos de Aprendizado de Máquina foram aplicados: máquina de detores de suporte (*Support Vector Machine* – SVM), correspondência de padrões conjuntos, abordagem baseada em regras e baseada em *parsing* (análise sintática). Esses modelos foram treinados e avaliados para identificar entidades biomédicas e comparados segundo métricas como acurácia, F1-Score, precisão e revocação.

A análise exploratória envolveu visualizações quantitativas, com o objetivo de comparar os valores numéricos das informações extraídas pelos modelos. Desse

modo, foram elaborados gráficos para a distribuição de frequência das categorias extraídas, total de palavras por cada categoria, tamanho total dos textos extraídos, entre outros.

Durante a análise do desempenho dos modelos, foi constatado que o modelo baseado em máquina de vetores de suporte (SVM) obteve um resultado melhor em comparação aos outros, com uma acurácia de 67,98% enquanto o *spaCy* obteve 45,31%. Essa avaliação demonstra limitações do *spaCy* quando aplicado especificamente ao contexto biomédico, que demanda de uma maior especialização dos modelos para capturar detalhes e complexidades do vocabulário científico e clínico.

O estudo concluiu que o *spaCy* se destaca como uma excelente ferramenta para visualização e análise exploratória de dados biomédicos. Contudo, como ferramenta de reconhecimento de entidades biomédicas, o *spaCy* é considerado menos eficiente quando comparado a outros modelos dedicados, como o SVM. O estudo sugere que em trabalhos futuros, deve-se explorar integrações entre *spaCy* e modelos especializados, como *transformers* e variantes do BERT voltadas para o domínio biomédico.

2.10.2 Abordagem baseada em linha do tempo para agregação de prontuários médicos eletrônicos

O trabalho publicado por Dabek, Jimenez e Caban (2017) propõe uma abordagem para o gerenciamento e análise de dados clínicos presentes em prontuários médicos eletrônicos, por meio da combinação de modelos analíticos com interfaces gráficas interativas e técnicas avançadas de visualização de dados. O principal objetivo é integrar modelos analíticos com interfaces gráficas em conjunto com a visualização de dados para criar uma abordagem efetiva para combinar múltiplas fontes de dados clínicos e reduzir a dificuldade cognitiva dos usuários finais ao acessar esses relatórios.

No framework desenvolvido, as informações clínicas registradas em um único dia são organizadas em nós de sumarização, que encapsulam atributos referentes a eventos e intervenções diárias. Essa estrutura é expandida por meio de uma abordagem baseada em árvores, que conecta os nós individuais ao longo de semanas, meses e anos, o que permite uma análise hierárquica e longitudinal dos

registros clínicos. Dessa forma, a visualização do histórico do paciente deixa de ser restrita ao nível diário e passa a respeitar uma organização temporal flexível, o que facilita comparações entre diferentes períodos e a identificação de tendências ou alterações clínicas relevantes.

Além da modelagem hierárquica dos dados, o sistema incorpora técnicas modernas de visualização para resultados laboratoriais e diagramas intuitivos para prescrições e intervenções clínicas. O uso de filtros, seleção de atributos e navegação dinâmica ao longo da linha do tempo possibilitam ao usuário limitar e refinar as informações visualizadas, o que garante foco nos aspectos mais importantes do histórico médico. A combinação dessas técnicas com métodos de sumarização baseados em aprendizado de máquina otimiza o processo de interpretação de dados e torna a análise mais precisa e eficiente.

2.10.3 Visualização do histórico clínico do paciente em linha do tempo a partir de textos médicos

O trabalho publicado por Yada e Aramaki (2023) propõe um sistema de registro por linha do tempo que permite a visualização do histórico clínico do paciente diretamente de textos em japonês na linguagem natural. A proposta tem como objetivo a superação dos desafios causados pela ausência de padronização e estruturação nos textos médicos. Para isso, o trabalho busca integrar modelos de linguagem que são capazes de reconhecer entidades médicas e suas relações temporais a partir de textos de prontuários eletrônicos e registros de radiologia.

Para que essa visualização seja realizada, os dados não estruturados do prontuário médico são aplicados em um modelo chamado *JaMIE* baseado no BERT, neste modelo estão presentes técnicas de NER e de extração de relações (RE). Para o treinamento, o conjunto de dados foi separado em dois tipos: registros de saúde e relatórios de radiologia. Os tipos de entidades extraídas foram: Doenças (e sintomas), Anatomia (partes), Característica (para modificar doenças e partes anatômicas), Mudança, Tempo, Teste, Medicamentos, e Contexto Clínico.

Para alinhar cronologicamente as entidades clínicas extraídas pelos registros médicos e facilitar sua interpretação, o sistema implementa um módulo de inferência de linha do tempo, capaz de agrupar as entidades extraídas em conjuntos de tempo, isto é, foram agrupadas as entidades que pertencem ao mesmo ponto de

tempo ou duração, baseada na relação temporal. A abordagem utilizada para a visualização foi baseada no formato do gráfico de Gantt, no qual os eventos estão dispostos na linha horizontal agrupadas em múltiplas linhas de categorias.

2.10.4 Linha do tempo das medicações prescritas ao paciente

O trabalho proposto por Ledesma et al., (2019) tem como objetivo a criação de uma ferramenta chamada *Health Timeline*, que busca facilitar a visualização e compreensão do histórico clínico dos pacientes por meio de uma linha do tempo interativa. A pesquisa utiliza uma metodologia baseada em *insights*, para mensurar a qualidade dos raciocínios clínicos que podem ser desenvolvidos pelos profissionais da saúde ao utilizar o sistema.

O *Health Timeline* representa eventos clínicos como prescrições, consultas e exames de forma cronológica e agrupa em linhas adjacentes. Cada evento está presente em uma caixa que indica sua categoria. No caso da linha do tempo dos medicamentos, essas caixas são consideradas os tipos dos medicamentos, como antidepressivos, estabilizadores de humor, antipsicóticos, entre outros.

Para a avaliação da ferramenta, as métricas utilizadas foram definidas pelo número total de *insights*, o tempo necessário para conseguir o primeiro *insight* e o valor atribuído para cada *insight*, conforme os critérios específicos de complexidade e utilidade diagnóstica. Ao comparar os insights obtidos com e sem o modelo proposto, foi analisado que ao utilizar o modelo, 7 *insights* alcançaram o maior valor possível. O trabalho proposto melhorou efetivamente a compreensão dos dados clínicos e ajudou os participantes a reconhecerem padrões complexos dos dados.

2.10.5 Extração e padronização de dados clínicos não estruturados com uso de processamento de linguagem natural

O estudo publicado por Wulff et al., (2020) tem como objetivo promover a extração automática de informações cruciais de textos médicos livres e transformar os dados clínicos não estruturados em uma representação padronizada e estruturada, por meio da implementação de uma *pipeline* para o processamento de histórico médico pediátrico. Essa *pipeline* possibilita a reutilização de textos médicos de forma estruturada e padronizada, através da seleção e modelagem de arquétipos

apropriados como modelos de informações clínicas padrões, definição de um dicionário com marcadores de texto cruciais para servir como base de conhecimento especializado para o PLN e criação regras de mapeamento entre a saída do processamento de linguagem natural e os arquétipos.

A abordagem desenvolvida utilizou um conjunto de 3055 entradas de marcadores de texto, 132 eventos de texto, 66 expressões regulares e um corpus de texto que consiste em 776 entradas para a correção automática de erros de ortografia. Essa *pipeline* realiza tarefas como segmentação de informações, classificação clínica, correção de erros, e integração dos dados extraídos ao repositório, o que permite consultas e análises avançadas conforme os padrões internacionais de interoperabilidade.

Na avaliação inicial, o sistema atingiu resultados como 97% de precisão e 94% de revocação na extração automática de informações dos históricos médicos pediátricos. O trabalho demonstra que a combinação de regras especializadas, modelagem de arquétipos e técnicas de PLN resulta em ganho de qualidade e interoperabilidade na utilização dos dados clínicos.

2.10.6 Modelo de reconhecimento de entidades clínicas para narrativas psiquiátricas

O trabalho publicado por Niero et al., (2023) tem como objetivo principal contribuir para a redução do déficit de corpora para o português na área psiquiátrica, por meio da criação, publicação e disponibilização de um corpus anotado de prontuários médicos e do treinamento de um modelo de linguagem especializado para reconhecimento de entidades nomeadas, chamado *PsyBERTpt*. A pesquisa propõe duas contribuições principais: a criação de um corpus anotado de 300 notas clínicas psiquiátricas e um modelo baseado em BERT para a realização do NER a partir do corpus desenvolvido.

O estudo abrange desde um levantamento do estado da arte sobre corpora clínico em português e outros idiomas, ferramentas para o NER clínico, até os processos detalhados utilizados para a geração do corpus, definição de *guidelines* de anotação, treinamento dos anotadores e validação das anotações. A metodologia engloba práticas rigorosas de anotação, etapa de treinamento, fases de validação e revisão das anotações pelos especialistas.

O corpus anotado desenvolvido pelos autores consiste em 300 notas clínicas de admissão e consulta de pacientes em um serviço de emergência psiquiátrica de um hospital especializado no Brasil. Neste corpus, foram anotadas categorias de entidades relacionadas a: comportamentos autodestrutivos, diagnósticos, medicamentos, uso de drogas, histórico do paciente, histórico familiar, entre outros.

No desenvolvimento do modelo de linguagem responsável pela tarefa de NER, foi realizado o ajuste fino em um modelo pré-treinado em português baseado no BERT com os dados anotados presentes no corpus clínico desenvolvido. A abordagem proposta obteve 65% de acurácia, 69% de revocação e 67% de F1-score, métricas compatíveis com o estado da arte sobre reconhecimento de entidades clínicas em português.

2.10.7 Análise dos trabalhos correlatos

A partir da análise dos trabalhos, foi possível verificar a existência de diferentes técnicas de extração e estruturação das informações médicas, assim como diferentes disposições visuais dos dados, além do gráfico de Gantt, foram apresentadas visualizações em linha do tempo e árvores. Os aspectos definidos para a análise representam requisitos técnicos e clínicos para a redução de raciocínio.

A realização da tarefa de NER está presente na etapa de extração de informações dos dados textuais, o tratamento de linguagem informal auxilia na recuperação de detalhes e nuances presentes nos textos escritos por humanos, a visualização dos dados por meio de linhas do tempo com o objetivo de reduzir a carga cognitiva do usuário representam a usabilidade da interface, o uso de dados psiquiátricos explora o domínio linguístico com subjetividade narrativa e a estruturação dos dados garante o armazenamento das informações extraídas.

Desta forma, com base na amostragem e análise dos correlatos, valida-se a necessidade de abordagens que consigam integrar essas duas áreas distintas. É exibida no Quadro 1 a comparação entre os trabalhos correlatos a partir das suas características descritas.

Quadro 1 – Comparação entre os trabalhos correlatos

	Yenkikar et al., (2024)	Dabek, Jimenez e Caban (2017)	Yada e Aramaki (2023)	Ledesma et al., (2019)	Wulff et al., (2020)	Niero et al., (2023)
Utiliza modelos baseados em BERT	Não	Não	Sim	Não	Não	Sim
Utiliza Reconhecimento de Entidades	Sim	Não	Sim	Não	Não	Sim
Tratamento de linguagem informal/abreviada	Não	Não	Sim	Não	Sim	Sim
Redução da carga cognitiva do usuário	Não	Sim	Sim	Sim	Não	Não
Realiza a estruturação dos dados	Não	Sim	Não	Não	Sim	Não
Utiliza dados psiquiátricos	Não	Não	Não	Sim	Não	Sim
Visualização em linha do tempo	Não	Sim	Sim	Sim	Não	Não

Fonte: Elaborado pelo autor.

2.11 CONSIDERAÇÕES FINAIS

Neste capítulo são apresentados os conceitos teóricos utilizados neste trabalho, assim como as principais tecnologias utilizadas na área e a suas aplicações no contexto da saúde, de modo a destacar os conceitos mais indicados para a pesquisa. Adicionalmente, foram apresentados os principais trabalhos correlatos da literatura voltados a extração de informações em textos médicos e visualização de dados.

3 ABORDAGEM DE VISUALIZAÇÃO DE DADOS APOIADA POR PROCESSAMENTO DE LINGUAGEM NATURAL DESENVOLVIDA

Neste capítulo é detalhado o desenvolvimento da abordagem proposta, a qual tem como objetivo a estruturação de dados provenientes de textos médicos escritos de maneira livre e a sua representação visual em gráficos interativos. A abordagem se baseia em três etapas, em que, na primeira é realizado o ajuste fino de um modelo de linguagem baseado na arquitetura BERT para a realização da tarefa de NER especializado em categorias psiquiátricas. Na segunda etapa é detalhado o processo do desenvolvimento do algoritmo responsável pela estruturação dos dados presentes nas notas.

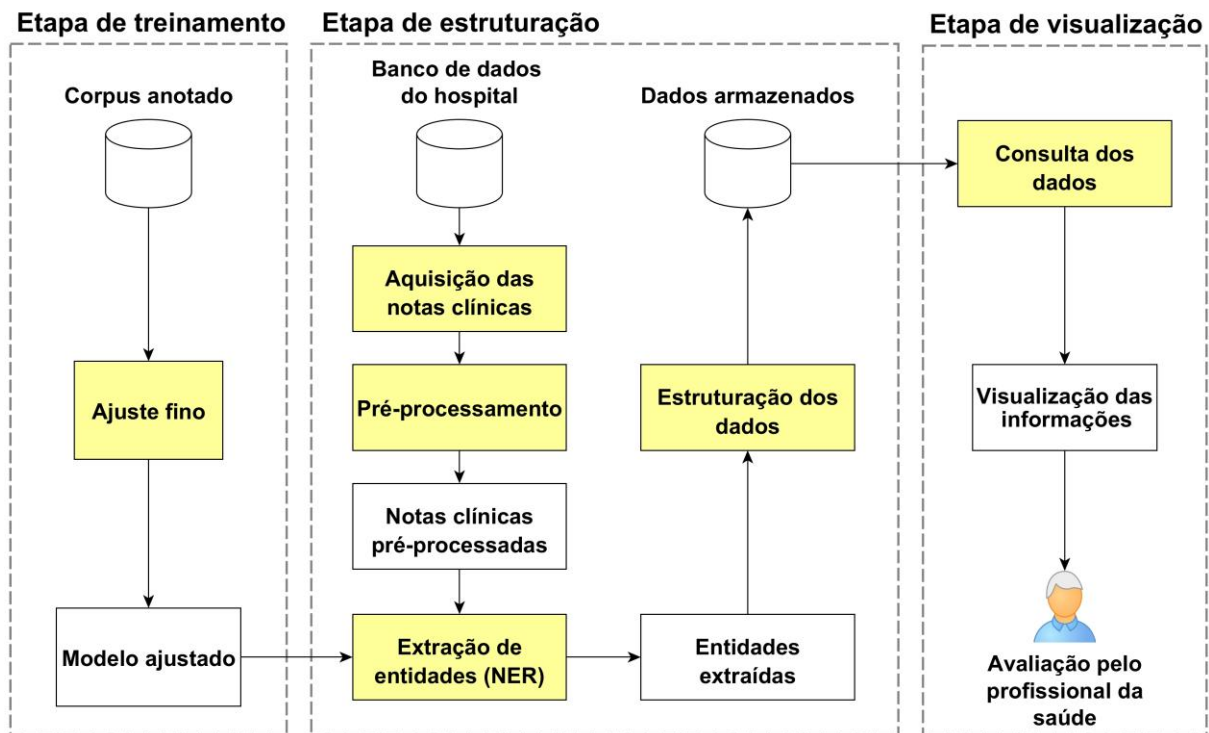
Após a estruturação dessas informações em um banco de dados relacional, os dados obtidos são representados em gráficos de acordo com a característica dos formulários estruturados. Nos quais são divididos em gráficos de Gantt para mapear eventos temporais, gráficos de barras para sinalizar a quantidade de ocorrências de determinadas categorias e gráficos de séries temporais para representar os sinais vitais registrados do paciente. Deste modo, a abordagem integra um modelo de linguagem e técnicas de visualização analítica para representar os dados de prontuários médicos.

3.1 VISÃO GERAL

Este trabalho apresenta uma abordagem computacional para a extração e visualização de informações contidas em notas clínicas de prontuários psiquiátricos que foram armazenadas de forma não estruturada. Tais documentos representam uma fonte detalhada do histórico do paciente, mas seu formato textual impede a análise sistemática e a extração de informações em larga escala. Logo, é proposta uma abordagem que utiliza o NER como técnica central para identificar e classificar informações clinicamente relevantes, seguida da estruturação desses dados em um banco de dados relacional para posterior análise e visualização.

A abordagem desenvolvida pode ser dividida em três etapas principais, detalhadas no diagrama apresentado na Figura 11. Destaca-se que os retângulos em amarelos ilustrados na figura representam as ações realizadas nos processos e os retângulos brancos os resultados dessas ações realizadas.

Figura 11 – Diagrama do trabalho desenvolvido



Fonte: Elaborado pelo autor.

A primeira etapa consiste no ajuste fino de um modelo de linguagem baseado na arquitetura do BERT. Para obter um melhor desempenho em textos psiquiátricos, o modelo foi especializado por meio do treinamento com um corpus customizado e anotados por especialistas. Isso garante que o modelo possa identificar com alta precisão as entidades selecionadas dentro do contexto clínico específico.

Na segunda etapa o modelo treinado é aplicado em um conjunto de notas clínicas reais devidamente anonimizadas para garantir a privacidade dos pacientes. As entidades identificadas pelo modelo são extraídas e organizadas em um banco de dados relacional. Essa estruturação transforma o texto livre em um formato quantificável, de modo que sejam associadas as datas dessas notas com as informações médicas presentes no texto, o que permite consultas complexas e a correlação de variáveis que antes estavam isoladas no texto.

Por fim, a terceira etapa foca em traduzir os dados estruturados em gráficos visuais intuitivos. Para isso, foram desenvolvidos gráficos de barras, séries temporais e diagramas de Gantt. O gráfico de barras permite visualizar a quantidade de ocorrências de cada categoria reconhecida nos textos das avaliações médicas. Já os

diagramas de Gantt são utilizados para visualizar a duração e a sobreposição de intervenções, como os períodos de medição, estado psíquico ou sintomas e queixas, o que oferece uma visão clara do histórico. As séries temporais permitem a visualização dos sinais vitais dos pacientes, o que torna possível verificar as variações do estado físico do paciente.

3.2 TECNOLOGIAS

A etapa do ajuste fino do modelo baseado em BERT foi integralmente implementada em *Python* na versão 3.11.13, por possuir um amplo conjunto de bibliotecas especializadas para o desenvolvimento de aplicações de Inteligência Artificial. Para a implantação, foram utilizadas bibliotecas como a *Hugging Face Transformers*, que simplifica o acesso e o ajuste fino de modelos pré-treinados, e o *TensorFlow* (ou *PyTorch*), que fornece a infraestrutura computacional para o treinamento e a inferência em larga escala. Essa combinação permitiu o desenvolvimento de uma solução especializada para tarefas de Processamento de Linguagem Natural (PLN).

Para o armazenamento dos dados extraídos pelo modelo, foi selecionado o sistema de gerenciamento de banco de dados relacional PostgreSQL (versão 17). A comunicação com o banco de dados foi implementada em *Python* por meio de bibliotecas como *psycopg2*, para a conexão direta, e a *SQLAlchemy* para abstrair as consultas SQL e facilitar a manutenção e escalabilidade do código de acesso aos dados.

A interface de visualização foi desenvolvida com o *framework* React.js, escolhido por sua integração com a biblioteca D3.js (*Data-Driven Documents*), utilizada para a criação de gráficos interativos. Por fim, a comunicação entre a aplicação web e banco de dados foi orquestrada por uma Interface de Programação de Aplicações (API) construída com o *framework* FastAPI. Este foi selecionado pelas suas funcionalidades que aceleram o desenvolvimento e garante uma comunicação eficiente e bem documentada entre o cliente e o servidor.

3.3 BASE DE DADOS ANOTADOS PARA O AJUSTE FINO

A eficácia de um modelo de reconhecimento de entidades é dependente da qualidade e da especificidade do corpus utilizado em seu treinamento. Para a tarefa do ajuste fino do modelo BERT, a utilização de um conjunto de dados genéricos seria inadequada, pois não distinguiria as nuances e o contexto semântico do domínio psiquiátrico. Nesse sentido, a utilização de um corpus em um outro idioma poderia afetar a capacidade do modelo. Portanto, a seleção de um corpus especializado foi um requisito fundamental para alcançar o desempenho necessário.

Diante disso, o trabalho utilizou o corpus *PSYclinBR*, uma base de dados publicamente disponível e especificamente desenvolvida para o PLN em psiquiatria clínica no contexto brasileiro. Este recurso é composto por 300 prontuários de notas de admissão de um serviço de emergência psiquiátrico que, por consequência, oferece um retrato fiel da linguagem empregada em cenários reais. O corpus é anotado em 9 categorias semânticas de alta relevância clínica: comportamento autodestrutivo, diagnóstico, droga, fármaco, função psíquica, histórico familiar, histórico do paciente, observação e sintoma e queixa psíquica. Ao todo, são 2.480 sentenças e 5.822 entidades anotadas que forneceram uma base sólida para o ajuste fino do modelo (Niero et al., 2023).

O formato de anotação do corpus é o IOB2 (*Inside-Outside-Begin 2*), e tem como característica uma forma simples de representação. A versão mais simples dessa representação possui apenas 2 informações: a letra “I” para identificar se a palavra está contida em uma entidade anotada e a letra “O” para o oposto. Na versão IOB2, a letra “B” é utilizada para sinalizar o token que está na posição inicial de qualquer entidade (Niero et al., 2023). No Quadro 2 é apresentado um exemplo de uma sentença anotada do corpus com a representação IOB2.

Quadro 2 – Representação IOB2 dos Tokens das Sentenças

Token	Representação
SENTE	O
MUITA	O
TRISTEZA,	B-Funcao_psiquica
CHORO	B-Funcao_psiquica
FÁCIL,	I-Funcao_psiquica
ISONE	B-Funcao_psiquica
Tratamento:	O
SERTRALINA	B-Farmacos

50MG	O
(100)	O
Cids:	O
F43,2	O
Transtornos	B-Diagnosticos
De	I-Diagnosticos
Adaptação	I-Diagnosticos

Fonte: Adaptado de Niero et al., (2023).

3.4 AJUSTE FINO DO MODELO

Foi utilizado um modelo de linguagem baseado na estrutura BERT pré-treinado especificamente para a língua portuguesa. Esse modelo foi submetido a uma etapa de ajuste fino com uma base de dados rotulados com o objetivo de possibilitar a realização do NER para 6 tipos de entidades encontradas em textos psiquiátricos: comportamento autodestrutivo, diagnóstico, droga, fármaco, função psíquica e sintoma e queixa psíquica.

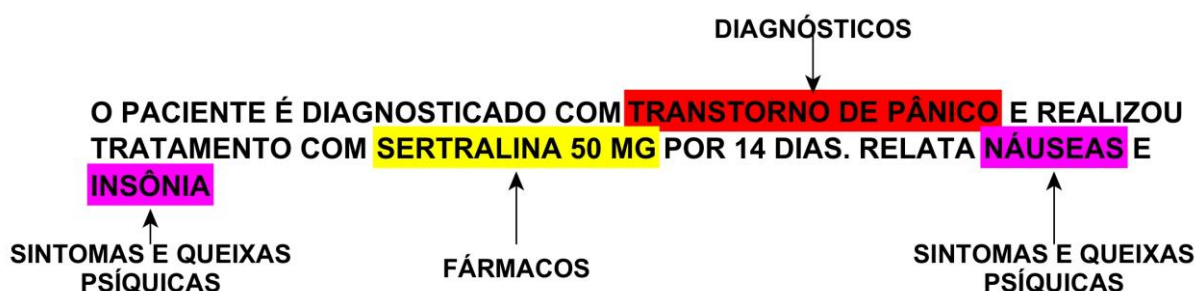
A etapa inicial do desenvolvimento do modelo é a preparação do ambiente e a limpeza dos dados. Primeiramente, é estabelecida uma semente fixa para garantir a reprodutibilidade dos experimentos. O pré-processamento remove as etiquetas irrelevantes para o escopo do trabalho e realiza o mapeamento dos módulos para identificadores numéricos, o que viabiliza o processamento pela rede neural.

A arquitetura BERT utiliza a técnica de tokenização por subpalavras (*WordPiece*), o que faz com que seja necessário um tratamento especial para os rótulos de cada entidade. Como o tokenizador segmenta as palavras complexas ou fora do vocabulário em fragmentos menores, ocorre um desalinhamento entre o texto original e as etiquetas fornecidas no conjunto de dados. A palavra “Psiquiatria”, por exemplo, seria decomposta em “Psi”, “##quia” e “##tria”. Para isso, é feita uma preservação de rótulos para garantir que as etiquetas de todos os fragmentos de uma palavra decomposta pelo tokenizador estejam vinculadas à sua categoria correta.

Após o ajuste fino e a validação, o modelo identifica trechos relevantes em documentos clínicos, mesmo sob condições de ambiguidade ou uso de linguagem informal característica de anotações psiquiátricas. Os aspectos técnicos da implementação das funções de treinamento e validação podem ser vistos nas Figuras A 1 e A 2 presentes no Apêndice A. Na Figura 12 é possível visualizar um exemplo de como as entidades reconhecidas são associadas às expressões originais e

armazenadas em um banco de dados relacional. Neste exemplo, o modelo reconhece que o “TRANSTORNO DE PÂNICO” é considerado como diagnóstico, a “SERTRALINA 50 MG” como um fármaco, “NÁUSEAS” e “INSÔNIA” como sintomas e queixas psíquicas. Este repositório centraliza informações sobre o tipo de entidade, o contexto do texto e o documento de origem, o que permite consultas rápidas e precisas por meio de filtros específicos por paciente.

Figura 12 – Exemplo da aplicação do modelo em um texto médico



Fonte: Elaborado pelo autor.

3.5 AQUISIÇÃO DAS NOTAS CLÍNICAS

Para a segunda fase do projeto, focada no desenvolvimento e validação da ferramenta de visualização, foi necessário um conjunto de dados distinto que simulasse o ambiente de produção. Deste modo, foi utilizada uma cópia da base de dados do Sistema de Gestão e Evolução de Prontuários (GEP), cedida pelo hospital parceiro, Dr. Adolfo Bezerra de Menezes, em São José do Rio Preto. A utilização destes dados foi precedida por um rigoroso protocolo de anonimização. Este processo envolveu a supressão de todas as Informações Pessoais Identificáveis (IPI), como nomes, CPFs, datas de nascimento e endereços, para impossibilitar a reidentificação dos pacientes.

Na anonimização dos formulários foi realizada a retirada dos identificadores dos pacientes, de modo que só se manteve um número como identificador no sistema e o número de prontuário, ambos gerados de forma aleatória para não ser possível a associação com pessoas reais. Apesar da remoção desses identificadores, alguns textos escritos em formato livre possuíam nomes de familiares e responsáveis pelos pacientes, além de médicos e outros profissionais do hospital. Para isso, foi realizada uma verificação individual em cada nota utilizada no trabalho.

Tal procedimento foi executado em estrita conformidade com as diretrizes da Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018), de modo a garantir a total confidencialidade e a preservação da privacidade dos indivíduos. Após a anonimização dos dados, a base foi submetida ao Comitê de Ética em Pesquisa (CEP) do Instituto de Biociências, Letras e Ciências Exatas (IBILCE), Câmpus de São José do Rio Preto. Os dados em questão foram submetidos via plataforma Brasil e foram aprovados para a pesquisa sob o Certificado de Apresentação de Apreciação Ética (CAAE) número 89963525.0.0000.5466, conforme apresentado no Anexo A.

A base de dados selecionada armazena as informações sobre 73637 pacientes com doenças mentais e dependências químicas entre os anos de 1971 e 2025, em que ocorreram 51243 registros de internações, 72631 registros de atendimento emergencial e 10073 consultas médicas. Dentre esses valores, 12900 pacientes tiveram mais de uma passagem em atendimento emergencial, enquanto 8297 foram internados mais de uma vez e 1551 pacientes tiveram mais de uma consulta médica. Além disso, houve pacientes que foram atendidos em mais de um módulo, como nos casos em que o paciente foi para a emergência do hospital e em seguida teve que ser internado devido ao seu estado de saúde.

Apesar da grande quantidade de documentos relacionados às internações, o trabalho tem como foco nos registros de emergência, pois os formulários presentes nesse módulo têm como característica a duração de poucos dias. Isso faz com que os gráficos gerados na etapa da visualização não fiquem desproporcionais, ao contrário dos formulários de uma internação que pode ultrapassar meses de duração. Adicionalmente, o trabalho tem como foco a agilidade no entendimento das informações dos textos, de modo que processos de uma emergência médica se tornem mais rápidos.

O processo de aquisição das notas clínicas consistiu em duas etapas principais: primeiramente, a seleção dos documentos mais relevantes para os objetivos do trabalho e, em segundo lugar, a aplicação dos métodos de pré-processamento para garantir a padronização dos dados. A padronização envolveu a limpeza do texto e a remoção de identificações pessoais para assegurar o anonimato dos pacientes, de modo que seja preparado o corpus textual para a análise e visualização subsequente. Para facilitar a manipulação desses dados, as informações dos formulários processados foram organizadas no formato JSON. Nesse novo

formato, é possível associar o texto do documento com o paciente, sem precisar da identificação.

O critério de seleção dos documentos médicos foi baseado na frequência temporal de registro e o conteúdo descritivo em texto livre. Observações preenchidas com alta frequência, como as diárias, mostraram-se ideais para representação em gráficos de Gantt, pois permitem mapear a continuidade do cuidado e a duração de certas condições ou intervenções. Em contrapartida, registros preenchidos de forma periódica ou esporádica os gráficos de barras foram considerados mais adequados, pois destacam as ocorrências de entidades das categorias específicas. Nesse sentido, os documentos "Doença Atual e Evolução" e "Anotações de Enfermagem" foram escolhidos para uma representação contínua, enquanto o documento "Avaliação Médica" foi selecionado para uma visualização pontual.

O documento "Doença Atual e Evolução" detalha a jornada clínica do paciente. Nele, a equipe de enfermagem registra a progressão da condição, respostas aos tratamentos prescritos e surgimento de intercorrências, assim como as queixas relatadas pelo paciente e as respectivas condutas adotadas. Esse registro constrói a narrativa que permite compreender o raciocínio clínico e as variações do estado de saúde do paciente ao longo do tempo.

As "Anotações de Enfermagem", por sua vez, combina dados objetivos e subjetivos. Nesse tipo de formulário são documentadas informações qualitativas sobre o estado físico e mental, como níveis de dor, humor, padrões de comportamento, nível de consciência e a qualidade da interação social. Somado a isso, o formulário também armazena informações quantitativas como os sinais vitais, que incluem aferições de pressão arterial, saturação de oxigênio, frequência respiratória, frequência cardíaca e temperatura.

Por fim, a "Avaliação Médica" representa um registro de alta importância preenchido em momentos cruciais, como a admissão do paciente na emergência do hospital. Portanto, existe a ausência de um intervalo regular entre os preenchimentos desse tipo de formulário. Esse documento possui os dados da anamnese, hipóteses diagnósticas, definição do diagnóstico principal e o tratamento inicial. Nele, são formalizadas as principais queixas dos pacientes e a análise inicial feita pelo profissional. Exemplos das notas adquiridas no formato JSON podem ser vistos nas Figuras A 3, A 4 e A5, presentes no Apêndice A.

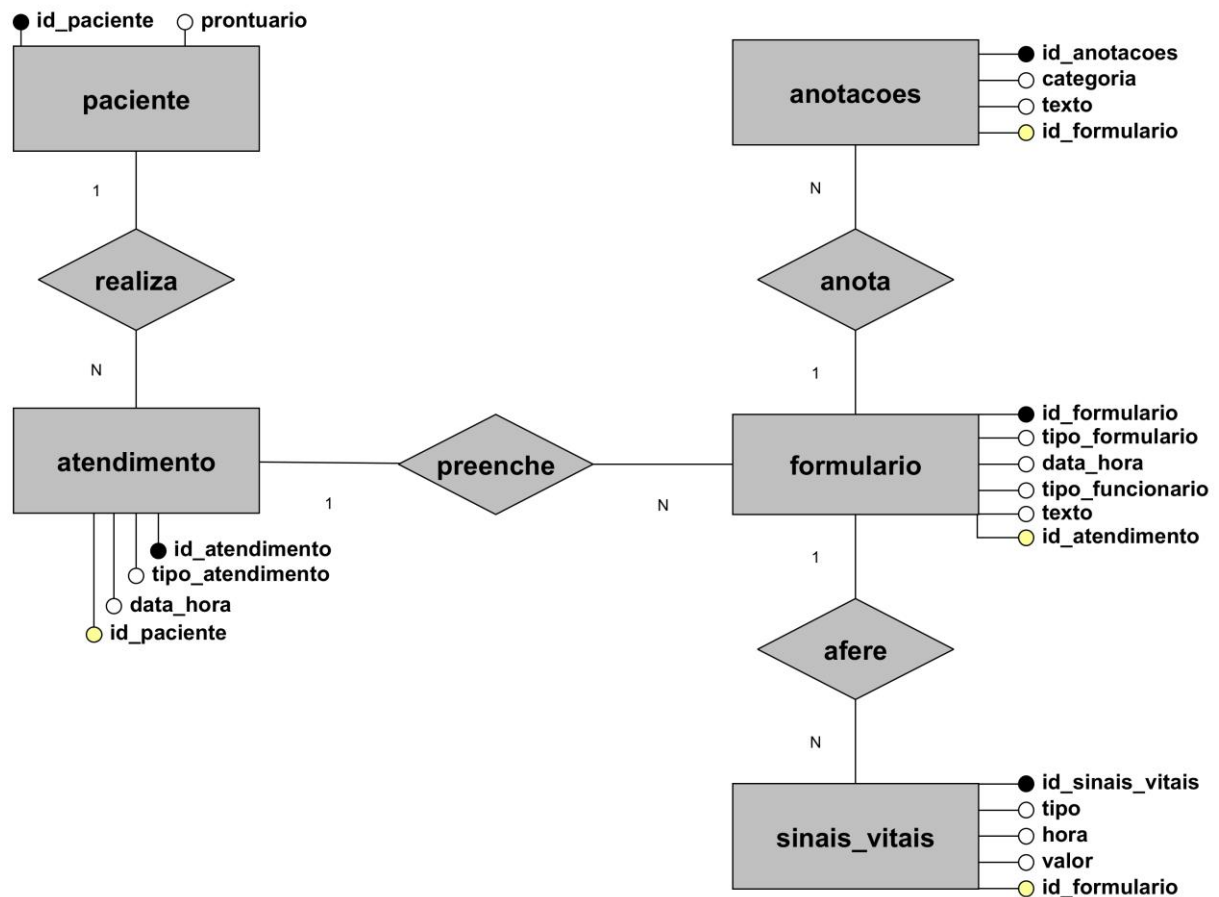
3.6 ESTRUTURAÇÃO DOS DADOS

Após a aquisição e o devido pré-processamento das notas clínicas psiquiátricas, o modelo treinado foi operacionalizado dentro de um pipeline de extração de dados. Nessa etapa, o modelo é aplicado em formulários como avaliação médica (elaborados por médicos), doença atual/evolução e anotações de enfermagem (preenchido pela equipe de enfermagem) e transforma os textos livres em informações estruturadas e quantificáveis.

Antes da implementação, foi realizada uma fase de estudo para definir o escopo dos dados. Essa análise buscou identificar os atributos essenciais não apenas para armazenar o resultado do NER, mas também para permitir a reconstrução contextual do histórico do paciente. Deste modo, para a eficácia da visualização, era necessário extrair: a entidade textual exata e sua categoria semântica, a origem precisa do registro (tipo de formulário e data/hora), o contexto do episódio clínico (atendimento de emergência ou internação) e um identificador único e anônimo que servisse como a ligação para todos os dados de um mesmo paciente. Adicionalmente, foi decidido incluir os dados de sinais vitais, pois a correlação entre estes marcadores fisiológicos objetivos e as observações textuais subjetivas extraídas pelo NER oferece um potencial analítico maior, o qual permite uma visão abrangente do estado do paciente.

Para ilustrar essa estrutura desenvolvida, foi projetado um Modelo Entidade-Relacionamento (MER), representado na Figura 13, e implementado em um banco de dados relacional PostgreSQL. A arquitetura foi definida para garantir a integridade dos dados e otimizar a performance de consultas. Diante disso, foram criadas as entidades: Anotações, Formulário, Atendimento, Paciente e Sinais Vitais. Cada uma com seus devidos relacionamentos com outras entidades.

Figura 13 – Modelo Entidade-Relacionamento



Fonte: Elaborado pelo autor.

As entidades de pacientes e atendimento permitem que o sistema organize o histórico clínico de forma individualizada e cronológica. Para isso, pacientes possuem apenas um número de identificação gerado aleatoriamente, de modo que não seja possível associá-lo com alguma pessoa real e ser possível associar com os atendimentos, que possuem as informações temporais.

A relação entre as anotações e o formulário constitui a base que sustenta a funcionalidade da ferramenta de visualização. O formulário armazena o tipo e o texto original da nota clínica analisada, além de suas informações temporais como data e hora. A partir disso, a entidade pode ser associada a diversas anotações. Para cada formulário processado, são geradas múltiplas entradas nesta entidade. Cada registro vincula um trecho específico do texto à sua respectiva categoria semântica.

Complementarmente, a entidade de sinais vitais armazena medições quantitativas vinculadas exclusivamente aos formulários do tipo “Anotações de

Enfermagem”, o que permite a correlação entre os indicadores fisiológicos e as observações subjetivas extraídas pelo modelo.

A arquitetura desenvolvida para banco de dados constitui um framework lógico para a extração de conhecimento. A estrutura relacional foi planejada para suportar as consultas temporais que são pré-requisito para a geração de linhas do tempo dinâmicas e gráficos de Gantt, de modo a transformar efetivamente um amontoado de dados textuais em uma narrativa clínica coerente e visualmente interpretável. O mapeamento detalhado do projeto conceitual para o respectivo projeto lógico, com a especificação técnica das chaves e tipos de dados, é apresentado na Figura A 6 localizada no Apêndice A.

3.7 FERRAMENTA DE VISUALIZAÇÃO

Para o desenvolvimento da ferramenta de visualização de dados clínicos, a seleção dos métodos gráficos foi pautada na natureza e na frequência das informações contidas nos diferentes tipos de formulário. Essa abordagem garante que cada tipo de dado seja representado da maneira mais clara e funcional possível para o profissional da saúde.

Para a representação das avaliações médicas, que consistem em registros pontuais e de alta densidade informacional, optou-se pelo gráfico de barras horizontais em conjunto com a disposição estruturada das entidades reconhecidas. Esta abordagem permite a sumarização quantitativa das ocorrências por categoria clínica, o que facilita a percepção imediata dos domínios predominantes em um atendimento específico.

Nas Figuras 14 e 15 são exemplificadas as aplicações do gráfico de barras em comparação com o texto da avaliação médica do paciente 42206. Na figura 14, é apresentada a avaliação preenchida no dia 28 de novembro de 2020. Nessa emergência, a ferramenta identificou e extraiu automaticamente categorias como "diagnósticos", "fármacos", "função psíquica", "sintomas e queixas psíquicas" e "comportamento autodestrutivo" do texto livre. A visualização não apenas lista estas categorias, mas também permite a análise contextual das palavras que levaram a essa classificação, de modo que proporcione agilidade para a percepção dos elementos mais críticos do relato médico.

Figura 14 – Visualização em gráfico da avaliação médica



Fonte: Elaborado pelo autor.

Na Figura 15, por sua vez, é apresentado o texto escrito referente a essa avaliação médica. A partir do texto é possível verificar os detalhes da emergência médica que a representação visual não foi capaz de representar. Além disso, a representação textual permite a verificação de cada termo reconhecido e apresentado no gráfico, de modo que seja possível entender o contexto referente a cada queixa e relato do paciente.

Figura 15 – Visualização textual da avaliação médica

DATA	HORA	PROFISSIONAL
28 De Novembro De 2020	01:00	Médico Psiquiatra

CLIENTE COM RETARDO MENTAL, VEM TRAZIDA PELO SAMU, DESACOMPANHADA. PACIENTE RECEBEU ALTA HÁ UM DIA DA OBSERVAÇÃO DESTA INSTITUIÇÃO, RELATA TER IDO VISITAR SUA MÃE, PORÉM NÃO A ENCONTROU EM CASA E FOI EM FRENTE AO HB PARA TENTAR UMA 'CARONA' PARA COSMORAMA PARA TENTAR ACHA-LA. ALEGA TER FICADO NERVOSA E TENTOU TOMAR IMOSEC PARA SE MATAR, QUANDO FOI IMPEDIDA POR POPULARES QUE ACIONARAM O SAMU. PACIENTE TEM INÚMERAS PASSAGENS NA EMERGÊNCIA DO HABM PELO MESMO MOTIVO E TEM POUCO SUPORTE FAMILIAR - MÃE TEM MEDIDA PROTETIVA CONTRA PACIENTE DEVIDO AGRESSÃO.

EM USO DE :

- 1-) CLORPROMAZINA 25 MG (1 0 1)
- 2-) PROMETAZINA 25 MG (1 0 1)
- 3-) ÁCIDO VALPRÓICO 250 MG (1 0 1)

AP: INTOLERANCIA A HALOPERIDOL. NEGA PATOLOGIAS CLÍNICAS.

TEM 9 INTERNAÇÕES PRÉVIAS NESTE SERVIÇO SENDO A ÚLTIMA ALTA DIA 23/11/2020

UM INTERNAÇÃO NA HOSPITAL DE JACI, NO PERÍODO DE 21/07/2020 A 10/08/2020.

EXAME FÍSICO: LÚCIDA, ORIENTADA, PORÉM INFANTILIZADA.

TRATAMENTO: OBSERVAÇÃO.

HIPÓTESE DIAGNÓSTICA: F06.2/F71.1

Fonte: Elaborado pelo autor.

Para os formulários que registram informações de natureza mais contínua, como “Doença Atual e Evolução” e “Anotações de Enfermagem”, foi realizada a escolha do gráfico de Gantt. Sua estrutura demonstra superioridade à linha do tempo no quesito de visualização da duração e sobreposição de estados ou eventos. Isto é útil em cenários de observação constante, como as emergências, em que o estado do paciente é monitorado diversas vezes.

Na Figura 16, que ilustra uma anotação de enfermagem na emergência, observa-se a proeminência da categoria “Sintomas e Queixas Psíquicas”. Isso reflete a importância de monitorar o estado mental do paciente ao longo do seu período de observação. O gráfico de Gantt representa de forma eficaz a persistência de certas condições, por meio de barras contínuas que se estendem por um ou mais dias, o que enfatiza a cronicidade ou continuidade de um sintoma ou estado.

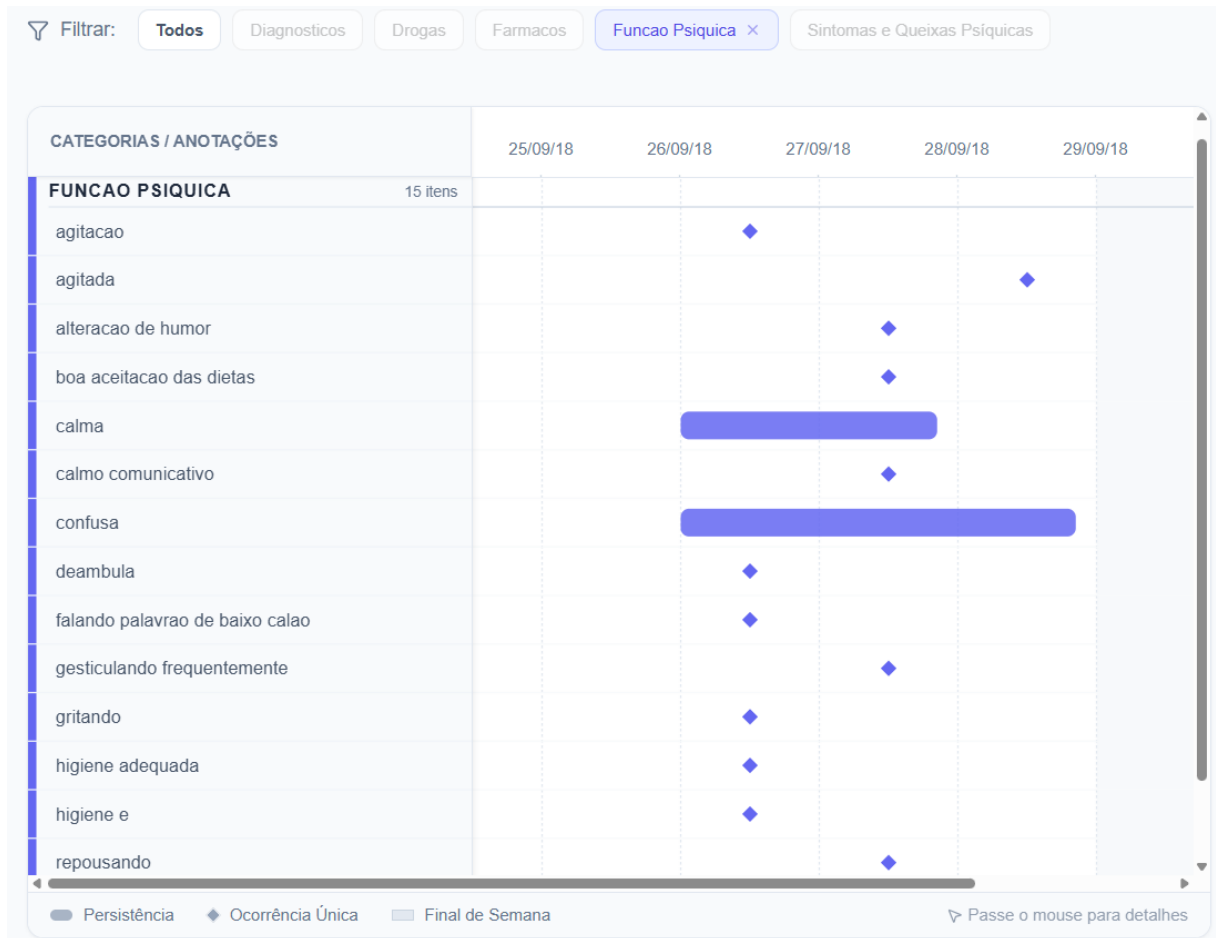
Figura 16 – Visualização em gráfico de Gantt da anotação de enfermagem



Fonte: Elaborado pelo autor.

Além da representação da continuidade dos casos pelas barras verticais, o gráfico possui filtros para a seleção das categorias/anotações desejadas para a observação. Na Figura 17 é apresentado um exemplo do mesmo gráfico, porém, com apenas a categoria “Função Psíquica” visível. Os filtros auxiliam aos profissionais selecionarem as informações mais relevantes no momento da consulta, o que diminui a necessidade de “rolar” o gráfico para baixo com o objetivo de procurar os dados de determinadas categorias.

Figura 17 – Utilização do filtro no gráfico de Gantt da anotação de enfermagem



Fonte: Elaborado pelo autor.

Na Figura 18 é possível observar a representação textual referente às informações dos formulários de “Anotação de Enfermagem”. A disposição das informações dessa forma torna necessária a navegação extensa do histórico do paciente, o que gera, por consequência, uma demora para o entendimento de determinados casos que o gráfico representa em uma única tela.

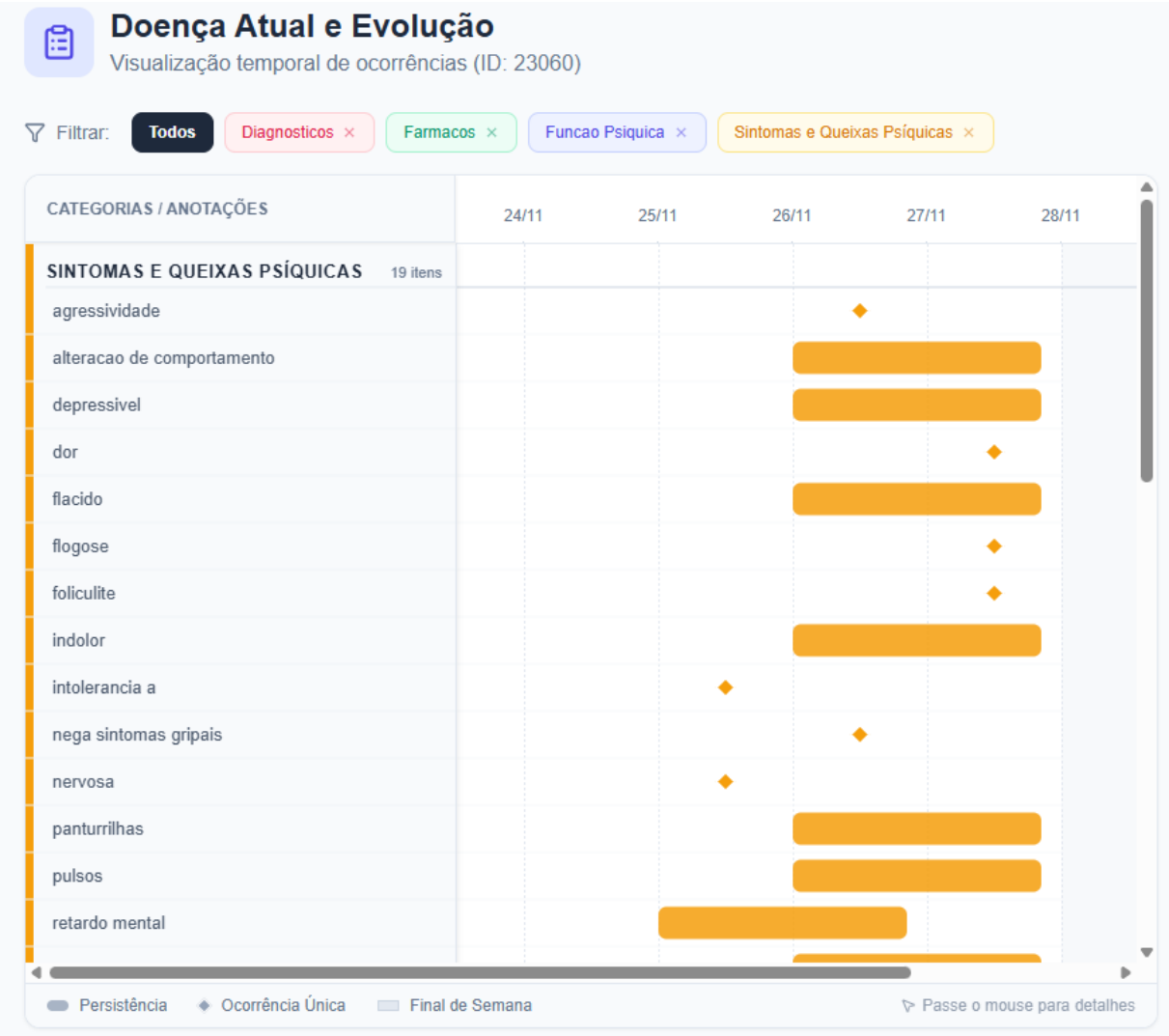
Figura 18 – Representação textual da anotação de enfermagem



Fonte: Elaborado pelo autor.

Deste mesmo modo, o formulário “Doença Atual e Evolução” também se beneficia da visualização em Gantt. No exemplo de emergência apresentado na Figura 19, a alta frequência de textos associados a “Função Psíquica” e “Sintomas e Queixas Psíquicas” permite ao clínico correlacionar o estado mental do paciente com as hipóteses diagnósticas em tempo real, uma análise crucial para o tratamento. Assim como as anotações, este gráfico possui filtros para a seleção das informações necessárias e a visualização textual em um formato de linha do tempo.

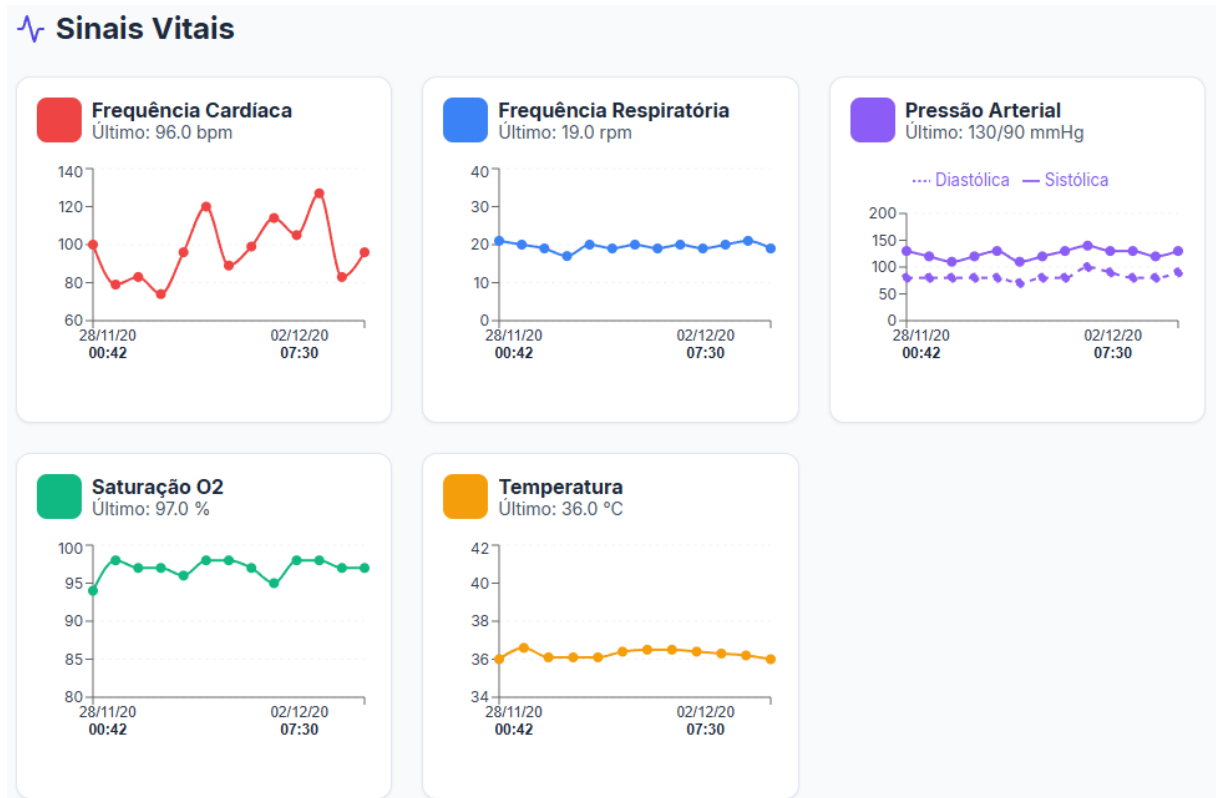
Figura 19 – Visualização em gráfico de Gantt da doença atual e evolução



Fonte: Elaborado pelo autor.

Por fim, para os dados quantitativos e seriais, como os sinais vitais, o gráfico de linha foi o método selecionado. A escolha desse método como padrão para a visualização de séries temporais se deve à sua clareza para revelar tendências, picos e flutuações. Na Figura 20 são apresentadas, respectivamente, a evolução da frequência cardíaca, frequência respiratória, pressão arterial, saturação e temperatura.

Figura 20 – Gráficos de linha



Fonte: Elaborado pelo autor.

Ao integrar os dados quantitativos dos gráficos de linha com informações contextuais e qualitativas dos gráficos de Gantt, o trabalho proposto oferece um panorama clínico completo. Isso permite os profissionais a realizarem associações complexas, como a correlação de um sinal vital como pressão e o relato de uma queixa psíquica, o que fundamenta uma tomada de decisão mais ágil e segura.

3.8 AMBIENTE DE DESENVOLVIMENTO E DIMENSIONAMENTO DO SOFTWARE

A implementação da infraestrutura computacional totalizou 5987 linhas de código desenvolvido. Esse volume total refere a integração de três módulos principais. O primeiro deles é a interface gráfica, que contém a maior parte das linhas de código do trabalho, com 3936 linhas desenvolvidas em *TypeScript/React* e *D3.js*.

O segundo engloba os serviços e o banco de dados, composto por 1089 linhas em *python* com o *framework* *FastAPI* para a API, além de 815 linhas para o script de leitura de arquivos JSON e armazenamento no banco de dados relacional.

Por fim, o desenvolvimento do modelo de inteligência artificial contém 147 de linhas de código.

Toda infraestrutura local, banco de dados, API e interface gráfica foi executada sob o sistema operacional Windows 11 em uma estação de trabalho com processador *Intel Core i5* de 10^a geração, 16 gigas de memória RAM e placa de vídeo *Nvidea GTX 1050ti*, enquanto o ajuste fino do modelo foi processado em nuvem na plataforma *Google Collaboratory*.

3.9 CONSIDERAÇÕES FINAIS

Este capítulo detalhou a arquitetura desenvolvida e implementada para extrair e visualizar informações de registros médicos em formato de texto livre. O foco principal foi a conversão de dados qualitativos não estruturados em representações gráficas interativas e intuitivas.

Inicialmente foi realizada a aquisição dos dados, seguida pela fase de pré-processamento, que incluiu a anonimização para garantir a privacidade do paciente e a formatação das notas clínicas. Com os dados preparados, a etapa seguinte envolveu a aplicação do modelo para analisar o conteúdo textual e extrair as informações clínicas relevantes. Por fim, a etapa de visualização foi realizada para representar esses dados estruturados previamente em ferramentas gráficas interativas.

4 AVALIAÇÃO EXPERIMENTAL

Neste capítulo são apresentadas, de forma detalhada, a metodologia de experimentação, assim como as métricas do treinamento do modelo selecionado, a discussão dos resultados obtidos e as devidas comparações com os modelos da literatura. Além disso, são descritos os testes de visualização realizados com os profissionais da saúde sobre uma base de dados real.

4.1 AVALIAÇÃO DO MODELO DE EXTRAÇÃO

A validação da abordagem proposta neste trabalho foi fundamentada em um processo de experimentação cujo objetivo é demonstrar a eficácia do modelo de PLN implementado. A eficácia do modelo em extrair e classificar informações impacta diretamente a qualidade da estruturação dos dados clínicos. Consequentemente, a qualidade e utilidade da ferramenta de visualização são afetadas diretamente pela qualidade do modelo.

Nesse contexto, foi realizado um estudo comparativo entre 8 arquiteturas do modelo BERT, todas pré-treinadas para o idioma português. Para assegurar a validade da comparação, todos os modelos foram ajustados com a mesma base de dados e o mesmo conjunto de parâmetros. A seleção do melhor modelo foi baseada em métricas de avaliação quantitativas, o qual foi escolhido aquele com o desempenho estatisticamente superior nas métricas de precisão, F1-Score e revocação. A metodologia de experimentação foi baseada nos trabalhos correlatos, como o *PsyBERTpt* (Niero, et al., 2023).

A seguir, são descritos o ambiente de execução e os materiais e métodos utilizados nessa etapa de experimentação, assim como a especificação dos parâmetros utilizados no treinamento dos modelos e como impactam no resultado final.

4.1.1 Ambiente e configuração experimental

Como informado anteriormente, a etapa do treinamento do modelo foi desenvolvida na linguagem de programação *Python*, dessa forma, o ambiente de desenvolvimento escolhido foi o *Google Colaboratory (Colab)* na versão *Pro*. Este

ambiente oferece um conjunto de ferramentas que auxiliam no desenvolvimento de projetos com inteligência artificial. Em particular, o acesso a Unidades de Processamento Gráfico (GPUs) de alto desempenho que foram cruciais para viabilizar os tempos de execução e acelerar os ciclos de experimentação.

A escolha correta dos parâmetros para o ajuste fino do modelo exerce um impacto direto e significativo na qualidade final. Uma configuração inadequada pode comprometer todo o conhecimento pré-adquirido, o que resulta em um modelo com desempenho insatisfatório ou incapaz de executar a tarefa para a qual foi designado (Devlin et al., 2018; Schneider et al., 2020).

Dentre os hiperparâmetros, o parâmetro mais influente é a taxa de aprendizado. Se configurado com um valor muito alto, pode causar um fenômeno que o modelo perde o conhecimento adquirido no pré-treinamento ao tentar se adaptar aos novos dados chamado “esquecimento catastrófico”. Por outro lado, uma taxa muito baixa torna o treinamento ineficaz e extremamente lento (Koroteev, 2019).

Outros hiperparâmetros fundamentais como o tamanho do lote e o número de épocas, também são importantes e devem ser balanceados com a taxa de aprendizado. O tamanho do lote influencia a estabilidade da atualização dos gradientes, enquanto o número de épocas determina o grau de exposição do modelo ao conjunto de dados, o que impacta diretamente os riscos de subajuste ou sobreajuste (Devlin et al., 2018; Lee et al., 2019; Koroteev, 2019).

Não existe uma configuração universal destes parâmetros, a combinação ideal varia conforme a tarefa e conjunto de dados e exige experimentação para cada caso específico (Devlin et al., 2018; Lee et al., 2019). Dessa forma, após a verificação dos valores ideais, por meio da comparação entre os valores obtidos pelos modelos selecionados para a etapa de testes, o conjunto de parâmetros que demonstrou melhor desempenho é apresentado no Quadro 3.

Quadro 3 – Parâmetros selecionados

Parâmetro	Valor
Comprimento da sequência	128
Lote de treinamento	4

Lote de validação	2
Número de épocas	5
Taxa de aprendizado	1e-05
Recorte de gradiente	10

Fonte: Elaborado pelo autor.

4.1.2 Corpus clínico anotado

Os dados utilizados para o treinamento do modelo foram obtidos através do corpus clínico anotado *PSYclinBR* proposto por Niero et al., 2023. Tal base de dados é constituída por 300 prontuários de notas de admissão de emergência psiquiátrica e anotado em 9 categorias semânticas: comportamento autodestrutivo, diagnóstico, droga, fármaco, função psíquica, histórico familiar, histórico do paciente, observação, e sintoma e queixa psíquica. No total, é formado por 2480 sentenças e 5822 entidades.

Para este trabalho, foi necessário realizar um refinamento no conjunto de dados original. Com o objetivo de focar nas informações mais clinicamente acionáveis, foram selecionadas seis das nove categorias semânticas disponíveis: diagnóstico, função psíquica, droga, fármaco, comportamento autodestrutivo, e sintoma e queixa psíquica. A distribuição final das entidades utilizadas neste trabalho, após a seleção, é detalhada na Tabela 1.

Tabela 1 – Distribuição das categorias

Categoria	Quantidade
Diagnóstico	819
Função psíquica	1687
Droga	367
Fármaco	853
Comportamento autodestrutivo	379
Sintoma e queixa psíquica	711

Fonte: Elaborado pelo autor.

A escolha do modelo para ser utilizado na etapa de estruturação dos dados baseou-se em uma avaliação comparativa entre as métricas. Para isso, foram selecionados e submetidos ao processo de ajuste fino os modelos baseados na arquitetura BERT e pré-treinados com dados na língua portuguesa.

Dentre os modelos monolíngues pré-treinados, isto é, desenvolvidos especificamente para português, o *BERTIMBAU* foi um dos principais modelos avaliados, projetado para superar as limitações dos modelos multilíngues em tarefas específicas do idioma. Suas duas versões, *base cased* e a *large cased* foram testadas. A versão *base* é mais leve computacionalmente e menos custosa, enquanto a *large* possui uma arquitetura mais complexa, com mais parâmetros, o que potencialmente obtém um desempenho superior (Souza, Nogueira, Lotufo, 2023). Adicionalmente, foi utilizado o *BERTugues*, outro modelo monolíngue desenvolvido para o português, se destaca por ter 7000 novos tokens que o *BERTIMBAU* (Zago, Pedotti, 2024).

Além dos modelos em português, modelos especializados em textos biomédicos também foram ajustados. O *BioBERTpt* representa um conjunto de modelos pré-treinados em português com foco no contexto médico. As três versões do *BioBERT* foram avaliadas, como o *BioBERT(clin)*, focado para textos clínicos, o *BioBERTpt(bio)*, focado em textos biomédicos, e, por fim, o *BioBERTpt(all)* que é especializado em ambos os contextos (Schneider et al., 2020).

Por fim, como base de comparação, foi utilizado o BERT multilíngue. Esse modelo foi pré-treinado em um conjunto de dados de 104 idiomas diferentes, incluindo o português, com o objetivo de criar uma base de conhecimento linguístico generalista. Suas duas versões foram avaliadas, a *cased*, que diferencia letras maiúsculas e minúsculas e a versão *uncased* que não realiza essa diferenciação (Devlin et al., 2018).

4.1.3 Resultados e análise comparativa

Após a realização do ajuste fino com a mesma base de dados e o mesmo conjunto de parâmetros, o resultado final se baseia nas métricas obtidas por cada um dos modelos nas etapas de validação do treinamento. Os resultados desta análise comparativa estão detalhados na Tabela 2, com os valores de maior destaque para cada métrica evidenciados em negrito.

Tabela 2 – Desempenho dos modelos avaliados

Modelo	Precisão	Revocação	F1-Score
BERTIMBAU large cased	0,79	0,81	0,80
BERTIMBAU base cased	0,78	0,79	0,79
BERTugues	0,80	0,84	0,82
BioBERTpt(clin)	0,80	0,82	0,81
BioBERTpt(bio)	0,75	0,77	0,76
BioBERTpt(all)	0,78	0,83	0,80
BERT multilíngue cased	0,76	0,77	0,76
BERT multilíngue uncased	0,81	0,83	0,82

Fonte: Elaborado pelo autor.

A análise dos resultados revela um desempenho superior dos modelos *BERTugues* e BERT multilíngue *uncased*, que alcançaram o maior valor de F1-Score com 0,82. Embora ambos tenham apresentado uma performance geral equivalente, o *BERTugues* foi o modelo selecionado para a etapa de estruturação dos dados devido ao seu desempenho superior nas métricas de revocação (0,84) e sua característica focada na língua portuguesa.

Uma vez selecionado o *BERTugues*, foi realizada uma análise mais particular para avaliar a eficácia do modelo para o reconhecimento de cada estrutura semântica. Para isso, foi calculada a média das métricas de precisão, revocação e F1-Score em todos os *folds* do treinamento. Os resultados desta análise por categoria são apresentados na Tabela 3.

Tabela 3 – Desempenho do BERTugues por categoria

Categoria	Precisão	Revocação	F1-Score
Comportamento autodestrutivo	0,56	0,71	0,60
Diagnósticos	0,87	0,90	0,88
Drogas	0,59	0,68	0,63
Fármacos	0,96	0,97	0,97
Função psíquica	0,79	0,82	0,80
Sintomas e queixas psíquicas	0,56	0,67	0,61

Fonte: Elaborado pelo autor.

Após a análise de desempenho por categoria, é possível observar que o modelo obteve as melhores métricas para o reconhecimento de Fármacos, Diagnósticos e Função Psíquica. Isso sugere que as entidades nessas categorias possuem padrões linguísticos mais consistentes e menos ambíguos no conjunto de dados utilizado para o treinamento, o que facilita sua identificação. Este desempenho nas categorias centrais é um indicador positivo, pois a alta fidelidade no reconhecimento dessas entidades é fundamental para garantir que a etapa de estruturação dos dados seja eficiente e que a ferramenta de visualização apresente informações precisas e completas.

Para validar a eficácia do modelo proposto e posicioná-lo no contexto da literatura atual, foi conduzida uma análise de desempenho comparativa. Foram selecionados modelos presentes em trabalhos correlatos em diferentes níveis de especialização para a tarefa de NER no contexto médico na língua portuguesa.

Foram escolhidas as três versões do *BioBERTpt* (Schneider et al., 2020) que foram ajustadas para a realização da tarefa e o *PsyBERTpt* (Niero et al., 2023) que foi selecionado como comparativo mais direto, por se tratar de um modelo já especializado e ajustado especificamente para a extração de entidades em prontuários psiquiátricos.

Para essa comparação, foram analisadas as métricas de precisão, revocação e F1-Score dos modelos. Na Tabela 4 são apresentados os desempenhos de cada modelo ajustado especificamente para a extração de entidades em dados médicos.

Tabela 4 – Comparação de desempenho entre os correlatos

Modelo	Precisão	Revocação	F1-Score
<i>BioBERTpt(clin)</i>	0,61	0,60	0,60
<i>BioBERTpt(bio)</i>	0,62	0,59	0,60
<i>BioBERTpt(all)</i>	0,61	0,61	0,60
<i>PsyBERTpt</i>	0,61	0,62	0,66
Trabalho proposto	0,80	0,84	0,82

Fonte: Elaborado pelo autor.

A análise comparativa dos resultados, apresentados na Tabela 4, demonstram a superioridade do modelo proposto nas métricas avaliadas. A eficácia do modelo

confirma o valor da abordagem adotada e valida o modelo proposto para extração de entidades clínicas em textos psiquiátricos em português.

4.2 AVALIAÇÃO DO SUPORTE À DECISÃO VISUAL

Para avaliar o recurso de visualização desenvolvido neste trabalho, foi adotada uma abordagem experimental comparativa de natureza qualitativa e quantitativa. O objetivo constituiu em mensurar a eficácia da visualização analítica, composta por gráficos interativos, em comparação ao método tradicional de interpretação de prontuários e notas médicas baseados em textos convencionais. Para isso, a fundamentação metodológica seguiu a Metodologia Baseada em *Insights* proposta originalmente por North, (2006) e aplicada em domínios clínicos por trabalhos correlatos como o de Ledesma et al., (2019).

Um *insight* pode ser definido como a capacidade de obter uma compreensão precisa e profunda. Enquanto abordagens comuns de avaliação de interfaces clínicas costumam focar em entrevistas subjetivas ou no desempenho do usuário na execução de um conjunto de tarefas predefinidas, a metodologia baseada em *insights* concentra-se no reconhecimento e na quantificação das descobertas obtidas a partir do uso exploratório da visualização de dados. Para realizar o registro dessas percepções, foram utilizados os mecanismos de “processo de verbalização do pensamento” e o uso de um diário escrito para registrar as etapas realizadas durante a análise de dados (Ledesma et al., 2019).

Deste modo, foram conduzidas sessões experimentais com profissionais da saúde em que foram acessados casos clínicos anonimizados para responder a questionamentos sobre condições físicas e psiquiátricas dos pacientes. As sessões foram gravadas para permitir a posterior transcrição e análise dos dados brutos. Esse procedimento possibilitou a quantificação de duas métricas essenciais: o Tempo Ativo de *Insight*, intervalo necessário para que o profissional processe a informação e formule uma conclusão, e o Valor do *Insight*, classificado de acordo com a sua complexidade e relevância para o diagnóstico e cuidado do paciente.

Em conjunto com a análise de eficácia analítica, foi aplicado o questionário *System Usability Scale* (SUS), desenvolvido por Brooke, (1986). Essa ferramenta consiste em uma escala de usabilidade amplamente validada na literatura capaz de fornecer uma métrica quantitativa sobre a percepção de satisfação e facilidade de uso,

mesmo quando aplicada a amostras reduzidas. A utilização do SUS permite verificar se a complexidade das visualizações propostas não impõe barreiras de aprendizado, de modo a assegurar que a ferramenta seja não apenas tecnicamente eficiente, mas também satisfatória para a experiência do usuário.

4.2.1 Amostragem

Um total de 6 profissionais da saúde da área da psiquiatria foram convidados a participar do experimento de maneira voluntária. Esses profissionais são enfermeiros no Hospital Psiquiátrico Dr. Adolfo Bezerra de Menezes presente na cidade de São José do Rio Preto, todos com experiência prévia na área da psiquiatria e que já trabalharam com casos parecidos com os apresentados na etapa de teste. Além disso, os enfermeiros convidados expressaram de maneira escrita o consentimento para a participação desta pesquisa.

4.2.2 Organização das sessões de teste

As sessões experimentais foram conduzidas individualmente, com duração média entre 45 e 60 minutos por participante. O protocolo de testes foi estruturado de forma que cada enfermeiro fosse exposto a dois cenários clínicos distintos, com a alternância entre os métodos de representação. Na primeira etapa da sessão, o participante interagiu com o cenário experimental e respondeu a cinco questões direcionadas para evidenciar diferentes níveis de profundidade de *insight* por meio da interface gráfica desenvolvida. Na etapa seguinte, o profissional realizou a análise de um segundo cenário clínico a partir da consulta dos registros clínicos escritos em textos.

Para garantir o rigor estatístico e mitigar o viés de aprendizado, no qual o conhecimento prévio de um caso poderia influenciar o desempenho no método seguinte, foi adotado um design experimental intragrupos com contrabalanceamento. Para isso, foram estruturados 4 casos clínicos de complexidade equivalente, distribuídos de modo que nenhum enfermeiro analisasse o mesmo caso sob as duas formas de representação, gráfica e textual. Essa organização permitiu uma comparação direta entre os resultados obtidos pelo uso do recurso de visualização e

pela leitura tradicional, o que isolou a variável do formato de apresentação dos dados como o fator principal na eficiência e eficácia do raciocínio clínico.

Cada sessão de teste obedeceu à seguinte estrutura:

- a) Apresentação: leitura e exposição da metodologia com a explicação detalhada dos objetivos da pesquisa.
- b) Treinamento: demonstração das funcionalidades do sistema com o uso de um caso-exemplo neutro.
- c) Execução de Tarefas: etapa na qual o participante analisou os dois casos propostos, experimental e controle.
- d) Avaliação Final: aplicação de instrumentos para coleta de percepção de uso e satisfação.

Durante a execução das tarefas, utilizou-se o protocolo *Think-Aloud*, no qual os profissionais de saúde verbalizaram de forma contínua as ações executadas, os pensamentos por trás de cada decisão e as conclusões provenientes da utilização do sistema. Tais informações foram registradas em áudio e posteriormente transcritas para a análise dos resultados. Ao término, aplicou-se o questionário de usabilidade *System Usability Scale* (SUS), composto por 10 perguntas objetivas e 2 perguntas de resposta livre para a coleta do *feedback* qualitativo.

4.2.3 Cenários clínicos e especificações das tarefas

Foram estruturados 4 casos clínicos de complexidade semelhante para serem avaliados, baseados em pacientes que passaram por várias emergências. Para a obtenção das informações de cada caso clínico, foram extraídos dados do prontuário médico de pacientes psiquiátricos do Hospital Dr. Adolfo Bezerra de Menezes. A avaliação da eficácia analítica do método de visualização e a quantidade de casos clínicos selecionados, seguiu a metodologia baseada em *insights*, com a escala de codificação proposta por North, (2006) e adaptada ao contexto psiquiátrico por Ledesma et al., (2019).

Para cada cenário clínico, foram definidas cinco questões direcionadas com o objetivo de evidenciar respostas em diferentes tipos de profundidade e complexidade de raciocínio. Dessa forma, o valor de um *insight* deriva de uma escala de um a cinco pontos, com base na complexidade e profundidade. A complexidade refere-se à capacidade de associação entre múltiplos elementos de dados, como a

relação entre a dose de um medicamento e o desfecho comportamental do paciente. A profundidade, por sua vez, mensura o grau de explicação das razões subjacentes aos dados, ou seja, se existe uma justificativa clínica plausível para o estado observado (Ledesma et al., 2019).

Neste estudo, as tarefas obedeceram aos critérios de codificação estabelecidos na literatura. Os valores 1 e 2 referem-se a percepções descritivas ou padrões simples, a exemplo da localização de uma comorbidade física ou da contagem de intervenções. Os valores de 3 a 5 contemplam a formulação de hipóteses e teorias explicativas. Dessa maneira, o participante deve combinar informações fragmentadas para produzir uma síntese clínica válida, como a correlação entre um surto psicótico e a interrupção do tratamento medicamentoso.

Diferente da exploração livre comum nos estudos citados, o uso de perguntas com níveis-alvo permitiu avaliar a capacidade do sistema em dar suporte a conclusões complexas sob demanda. Essa escolha metodológica reduz a subjetividade da análise e viabiliza a comparação estatística direta entre as representações gráficas e visuais. No Quadro 4 são mostrados exemplos extraídos das transcrições e os critérios dos valores atribuídos.

Quadro 4 – Critérios para a definição de valores do *insight*

Valor	Critério	Exemplo
1	Observação Simples – descrição de um dado isolado. Sem informação temporal ou conexões.	"A paciente apresenta um quadro de pediculose (piolhos)."
2	Padrão e Frequência – identificação de repetições ou contagem de eventos.	"O medicamento Midazolam foi administrado sete vezes durante o mês de setembro."
3	Tendência ou Mudança – análise da evolução dos dados ao longo do tempo ou identificação de quebra de padrão.	"O comportamento da paciente oscilou entre crises de agitação e períodos de calma durante o mês de julho."
4	Correlação Causal – relação entre	"A agitação psicomotora diminuiu

	duas ou mais variáveis distintas.	logo após a administração do esquema de Haloperidol com Prometazina."
5	Síntese Complexa – formulação de hipóteses clínicas profundas ou correção de contexto a partir de dados fragmentados.	"A paciente é portadora de HIV. Embora o diagnóstico não esteja explícito, identifiquei o uso contínuo de Tenofovir."

Fonte: Adaptado de Ledesma et al., (2019).

O primeiro cenário clínico apresenta um paciente com deficiência intelectual moderada e psicose crônica, com histórico marcado por episódios de agressividade e hiperatividade motora. A seleção deste caso teve como objetivo testar a capacidade do sistema na representação de eventos de curta duração e alta intensidade, ou seja, a necessidade de análise de intervenções medicamentosas múltiplas e sucessivas em um único período. Desta forma, no Quadro 5 são apresentadas as questões formuladas para o caso e seus respectivos valores de *insight* alvo para cada resposta, cada enunciado foi validado para garantir que a resposta correta estivesse presente tanto na representação textual quanto na gráfica.

Quadro 5 – Questões e níveis de *insight* para o Caso 1

Pergunta (Categoria)	Enunciado	Valor
Detecção de Comorbidade Física	"Em Fevereiro de 2019, além do surto psicótico, foi identificada e tratada uma condição de parasitose (problema físico/higiene). Qual foi a condição e o tratamento?"	1
Frequência de Medicação	"Analisar exclusivamente o mês de Setembro de 2018. Devido à instabilidade do quadro, a equipe precisou intervir com medicação injetável (Midazolam) várias vezes. Quantas vezes essa medicação (Midazolam) foi registrada nesse mês?"	2

Identificação de Risco de Autoagressão	"Este paciente tem histórico de ingestão de objetos ou substâncias? Se sim, cite um exemplo do que foi ingerido."	3
Correlação Clínica em Episódio Agudo	"Analise a manhã do dia 25/09/2018. A paciente estava tumultuando o plantão. A primeira intervenção medicamentosa (Haloperidol + Prometazina) foi suficiente para acalmá-la imediatamente ou foi necessário administrar uma nova medicação sedativa logo em seguida?"	4
Perfil Clínico Predominante	"Observando a frequência e os tipos de anotações nas categorias 'Função Psíquica' e 'Sintomas' ao longo de todos esses anos, como você classificaria o perfil psiquiátrico predominante desta paciente durante as crises: a) Perfil Depressivo/Hipoativo (Isolamento, choro, recusa alimentar, silêncio) b) Perfil Expansivo/Hiperativo (Deambulação excessiva, verborragia, agitação, agressividade)? Justifique visualmente."	5

Fonte: Elaborado pelo autor.

O segundo cenário descreve um paciente com esquizofrenia e transtorno bipolar, em acompanhamento para o HIV, uma comorbidade infecciosa crônica. A inclusão deste caso teve como propósito avaliar a eficácia do sistema na detecção de diagnósticos implícitos, identificáveis apenas por meio do cruzamento de classes medicamentosas com sintomas. Além disso, a extensão temporal dos dados de 2018 a 2024 forneceu a base para verificar a capacidade da ferramenta para a síntese de riscos graves, como a associação entre alucinações de comando e tentativas de autoextermínio ao longo de vários anos. No Quadro 6 são apresentadas as questões definidas para o caso e seus respectivos valores alvo de *insight*.

Quadro 6 – Questões e níveis de *insight* para o Caso 2

Pergunta (Categoria)	Enunciado	Valor
Detecção de Comorbidade Física	"Além do quadro psiquiátrico, essa paciente possui uma doença infecciosa crônica grave. Identifique qual é essa condição ou cite a classe de medicamentos de uso contínuo (não psiquiátricos) que ela deveria estar tomando."	1
Frequência de Medicação	"Durante as várias passagens pelo serviço no mês de Março de 2019, quantas vezes foi registrada antipsicótico Haloperidol?"	2
Identificação de Padrão Temporal	"Analise o comportamento da paciente durante o mês de Julho de 2019 (do dia 04/07 ao dia 26/07). Ela apresentou uma melhora progressiva linear ou manteve um quadro de instabilidade com flutuações constantes de humor e consciência?"	3
Correlação Clínica	"Na madrugada do dia 26 de Março de 2019 (01:34h), a paciente estava inquieta e deambulando. Qual medicação injetável foi administrada e qual foi o estado da paciente na manhã seguinte (06:49h)?"	4
Padrão de Risco e Conteúdo da Psicose	"Olhando o histórico geral da paciente ao longo dos anos, nota-se a presença frequente de 'Alucinações' (auditivas e visuais). Observando o gráfico nesses exatos momentos de crise alucinatória, que tipo de ideação ou comportamento de alto risco costuma aparecer simultaneamente (sobreposto), indicando o provável conteúdo dessas alucinações?"	5

Fonte: Elaborado pelo autor.

O terceiro caso retrata o histórico de um paciente com psicose crônica e insuficiência renal terminal, o que exige sessões frequentes de hemodiálise e

cuidados com acesso vascular. A seleção deste cenário teve como objetivo a apresentação de um quadro no qual os dados de contenção mecânica tivessem uma justificativa de autoproteção como risco de queda por sedação ou fraqueza física e não de controle comportamental agressivo. O intuito foi observar se a justaposição de sinais físicos e psiquiátricos auxilia o profissional na formulação de hipóteses mais precisas e menos enviesadas. No Quadro 7 são apresentadas as questões definidas para o terceiro caso e seus respectivos valores alvo de *insight*.

Quadro 7 – Questões e níveis de *insight* para o Caso 3

Pergunta (Categoria)	Enunciado	Valor
Identificação de Comorbidade Física	"Observando o histórico geral da paciente, nota-se que ela possui uma condição física crônica severa (não psiquiátrica) que exige uma estrutura física específica no corpo dela para tratamento. Qual é a doença clínica e qual é o nome dessa estrutura/acesso corporal?"	1
Análise Quantitativa (Vulnerabilidade Social)	"A equipe de enfermagem frequentemente registra o estado de higiene dos pacientes. Quantas vezes a paciente deu entrada no serviço em 'péssimas condições de higiene' ou foi explicitamente citada em 'situação de rua'?"	2
Padrão Temporal	"Concentre-se no período entre 19/06/2019 e 24/06/2019. O padrão comportamental da paciente nesses 6 dias foi de estabilidade ou houve uma crise aguda prolongada? Que sintoma específico (ação vocal) predominou nesses dias?"	3
Correlação Interdisciplinar	"No final de semana dos dias 13 e 14 de Abril de 2019, além da confusão mental, a paciente apresentou uma emergência clínica (física) que exigiu medicações não-psiquiátricas. Qual foi o	4

	problema clínico e cite um dos remédios dados para isso?"	
Justificativa Clínica de Contenção	"Geralmente, na psiquiatria, a 'contenção mecânica' está associada à extrema agressividade. No entanto, ao analisar os dias em que esta paciente ficou 'Restrita' ou em 'Contenção' (como em 19/06/2019 e 15/07/2019), o gráfico mostra a sobreposição de outros sintomas físicos e motores muito específicos no mesmo momento. Baseado nesse agrupamento visual, formule uma hipótese: por qual risco específico de segurança a equipe de enfermagem precisou restringir essa paciente ao leito nesses dias?"	5

Fonte: Elaborado pelo autor.

Por fim, no caso 4, é retratado o histórico de um paciente em situação de rua com um padrão de múltiplas internações breves e recorrentes. Este perfil foi escolhido para avaliar a eficiência da visualização macroscópica da linha do tempo, tarefa que exige alta carga cognitiva em prontuários textuais devido à fragmentação das datas. O cenário também serviu para medir a capacidade de varredura visual da ferramenta na busca por eventos físicos isolados e atípicos, como infestações parasitárias e protocolos de isolamento, em meio a um histórico socialmente ruidoso. No Quadro 8 são apresentadas as questões definidas para o quarto caso e seus respectivos valores alvo de *insight*.

Quadro 8 – Questões e níveis de *insight* para o Caso 4

Pergunta (Categoria)	Enunciado	Valor
Análise Quantitativa (Vulnerabilidade)	"Acesse a emergência do dia 23 de junho de 2020. Neste dia, além dos sintomas psiquiátricos e do uso de drogas (álcool/crack), a paciente apresentou uma	1

Social)	queixa física que motivou seu encaminhamento para o isolamento. Qual foi esse sintoma não-psiquiátrico?"	
Frequência de Medicação	"Abra a emergência do dia 06 de Abril de 2019. Durante os dias dessa internação, quantas vezes a enfermagem precisou intervir com medicação específica para o trato gastrointestinal (Bromoprida) devido a queixas de vômito/náusea?"	2
Identificação de Padrão	"Sem entrar no detalhe de nenhuma anotação, observe apenas a distribuição de datas na lista de emergências do ano de 2018. Qual é o padrão de internação dessa paciente? Ela teve poucas internações que duraram meses, ou apresentou um padrão de múltiplas idas e vindas ao longo do ano?"	3
Correlação Interdisciplinar	"Acesse a emergência do dia 12 de Maio de 2021. A paciente chegou trazida pelo SAMU com uma alteração motora descrita como 'aparentemente robotizada' (sintoma extrapiramidal comum). Qual foi o esquema de antipsicóticos administrados nesse dia para tentar conter o quadro?"	4
Contexto Subjetivo	"Abra a emergência do dia 27 de Outubro de 2023. Geralmente, ela chega com o SAMU, nua ou em surto. Neste dia, o registro mostra que ela veio andando por meios próprios. O gráfico mostrará 'Solilóquios'. Mas para entender a mente da paciente, é preciso ler as palavras dela. Segundo o texto, qual foi o motivo exato (a frase que ela disse) para ter entrado no hospital nesse dia?"	5

Fonte: Elaborado pelo autor.

Devido ao tamanho da amostra de 6 enfermeiros e ao uso de 4 casos clínicos para evitar o efeito de memória, a distribuição dos casos entre as representações apresentou uma leve variação, alguns casos analisados 3 vezes e outros 2 vezes. Contudo, essa assimetria foi neutralizada pela equivalência de complexidade entre os cenários e pelo uso de testes estatísticos pareados, que priorizam a variação de desempenho individual em detrimento da frequência absoluta de cada cenário.

4.3 RESULTADOS E DISCUSSÃO

A análise dos resultados fundamenta-se na transcrição de um total de 237 minutos de áudio, coletados por meio das gravações das sessões individuais de cada participante dos testes. Os registros verbais e as observações foram estruturados em uma planilha que associa, de forma unificada, a identificação anônima de cada enfermeiro, o cenário clínico analisado e o respectivo método de representação utilizado. Para cada tarefa executada, foi registrado o tempo exato de processamento cognitivo e o nível de complexidade da resposta, o que viabiliza o cruzamento de informações e a posterior aplicação de testes estatísticos.

Para avaliar a eficácia e eficiência do sistema, adotou-se um conjunto de métricas quantitativas e inferenciais. A eficiência foi avaliada por meio da média do tempo ativo de *insight*, definido como o intervalo entre a carga dos dados na interface e a verbalização da conclusão pelo participante. A eficácia analítica foi verificada pela média do valor da conclusão, baseada na escala de cinco níveis. Dada a natureza não-paramétrica da amostra e o desenho experimental pareado, aplicou-se o teste de postos sinalizados de Wilcoxon para verificar a significância estatística das diferenças entre o método gráfico e o textual. Complementarmente, calculou-se o tamanho do efeito, chamado de r de Rosenthal, com o intuito de determinar a magnitude do impacto da ferramenta no raciocínio clínico (Ledesma et al., 2019).

A partir da planilha registrada com os dados resultantes dos testes, foi feito o cálculo das médias de desempenho para cada uma das variáveis de interesse. Desse modo, é possível realizar a comparação do impacto da interface na velocidade de recuperação da informação e na qualidade da conclusão obtida. A média dos resultados agregados de todos os participantes e cenários clínicos são apresentados na Tabela 5.

Tabela 5 – Comparação do desempenho por método de representação

Métrica de Avaliação	Método Gráfico	Método Textual	Diferença
Tempo Ativo de <i>Insight</i>	82,1s	116,9s	-29,8%
Valor do <i>Insight</i>	3,23	3,16	+2,16%
Taxa de sucesso	88,3%	91,7%	-3,4%

Fonte: Elaborado pelo autor.

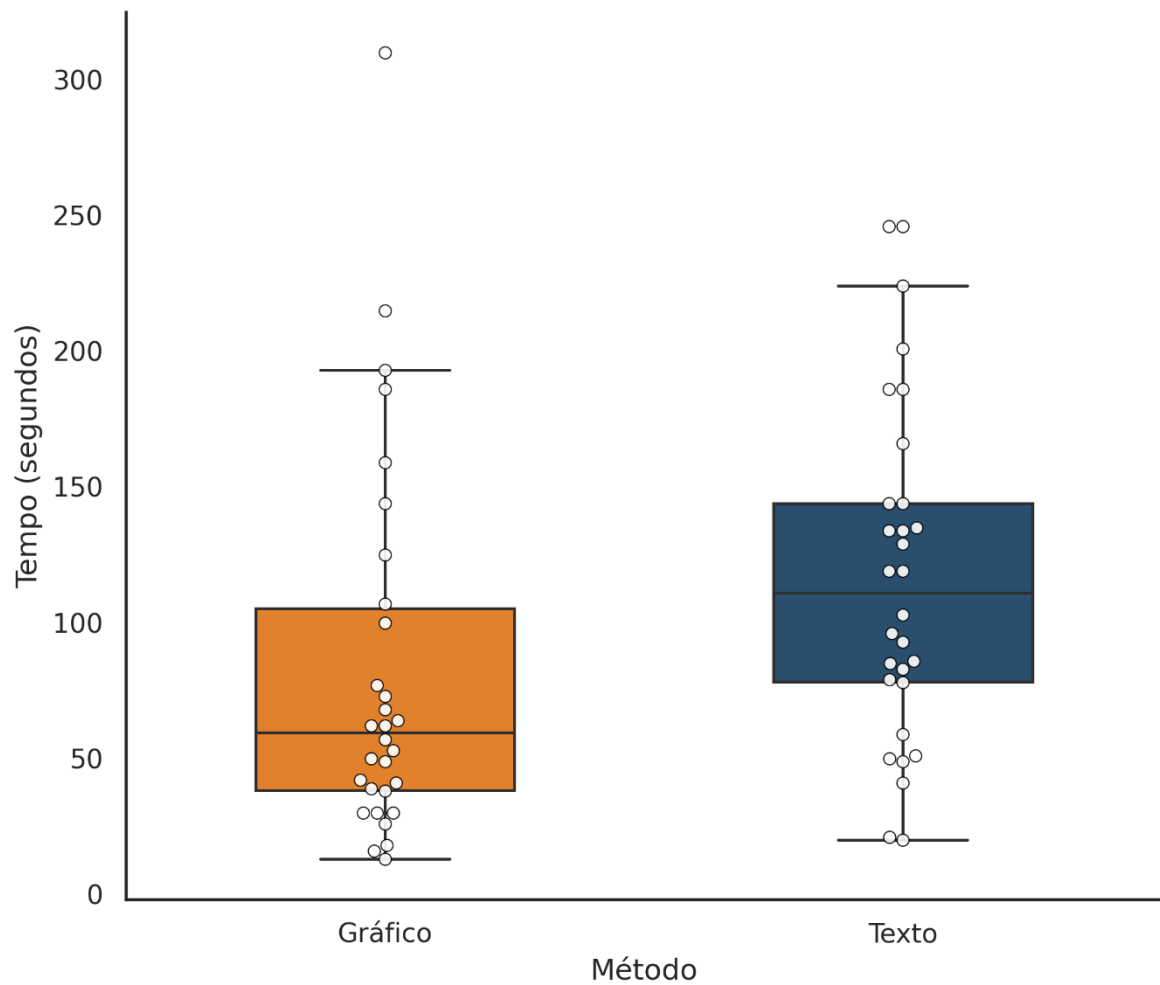
Com base nos resultados da Tabela 5, o método gráfico obteve uma redução de 29,8% no tempo médio necessário para a conclusão das tarefas em relação ao método textual. Além do ganho de tempo, a qualidade dessas respostas se manteve, visto que o valor dos *insights* permaneceu estável, com uma leve vantagem a favor da representação visual.

A representação textual apresentou uma taxa de sucesso geral de 91,7%, enquanto o método gráfico obteve 88,3%. A diferença de 3,4% que favorece o método textual é consequência da curva de aprendizado inicial, da densidade de informações apresentadas pelo gráfico, além da existência de informações que podem não ter sido estruturadas corretamente pelo modelo NER.

Em contrapartida, as falhas e sucessos parciais no método textual relacionaram-se à fadiga de leitura e à fragmentação de documentos, o que resultou em tempos de execução significativamente maiores para o alcance de resultados semelhantes.

Na Figura 21, a distribuição detalhada dos tempos de resposta é apresentada por meio de um gráfico de caixa. A partir da observação da figura, é possível confirmar que a representação gráfica registrou uma mediana de tempo inferior à do método textual. Essa maior concentração dos dados em torno de valores baixos indica que o recurso de visualização proporciona um desempenho mais homogêneo entre os profissionais. Enquanto o método textual exibe uma dispersão elevada, reflexo da variabilidade nas habilidades individuais de leitura e busca, o método gráfico atua como um facilitador cognitivo, mesmo diante de uma curva de aprendizado inicial evidenciada por alguns registros de maior duração.

Figura 21 – Gráfico de caixa dos valores de tempo



Fonte: Elaborado pelo autor.

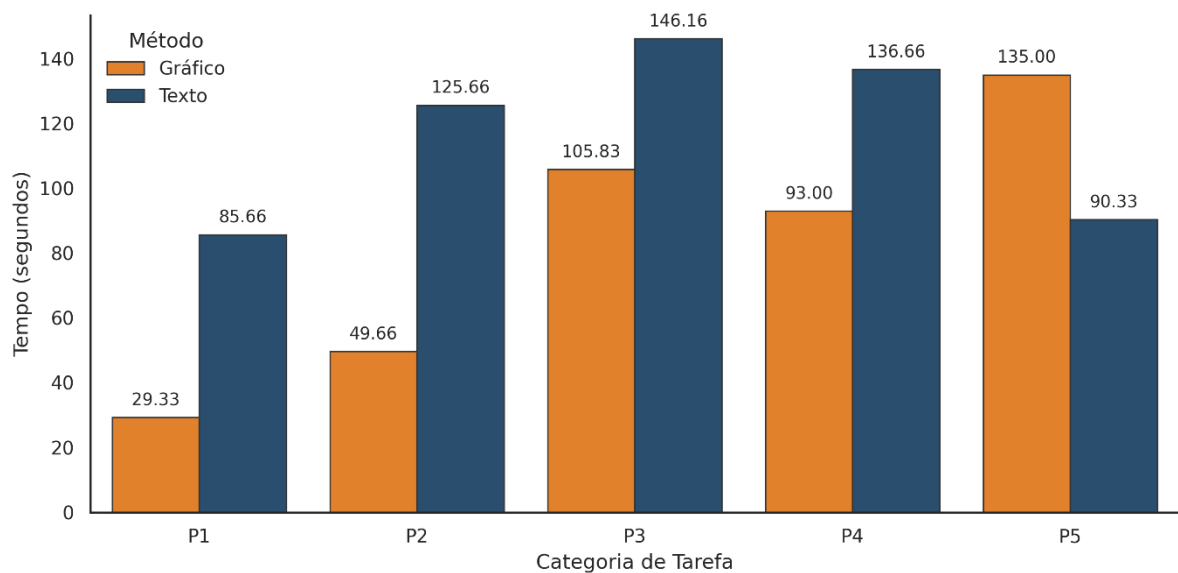
Após a análise dos valores de média geral, foi realizado o cálculo individual para cada categoria de tarefa para verificar as forças e as fraquezas de cada representação conforme a natureza do caso clínico. Na Tabela 6, são apresentadas as médias de tempo e valor de *insight* para cada uma das cinco classes de perguntas formuladas, enquanto nas Figuras 22 e 23 são ilustrados os comparativos de tempo ativo e a profundidade dos *insights* alcançados em cada categoria.

Tabela 6 – Comparação do desempenho por categoria de pergunta

Categoria da tarefa	Tempo do gráfico (s)	Tempo do texto (s)	Valor do gráfico	Valor do texto
P1 – Detecção de Dado	29,33	85,66	3,33	2,66
P2 – Padrão Temporal	49,66	125,66	2,5	1,83
P3 – Tendência	103,5	146,16	2,5	2,16
P4 – Correlação Causal	93	136,66	2,83	4,33
P5 – Síntese Complexa	135	90,33	5	4,16

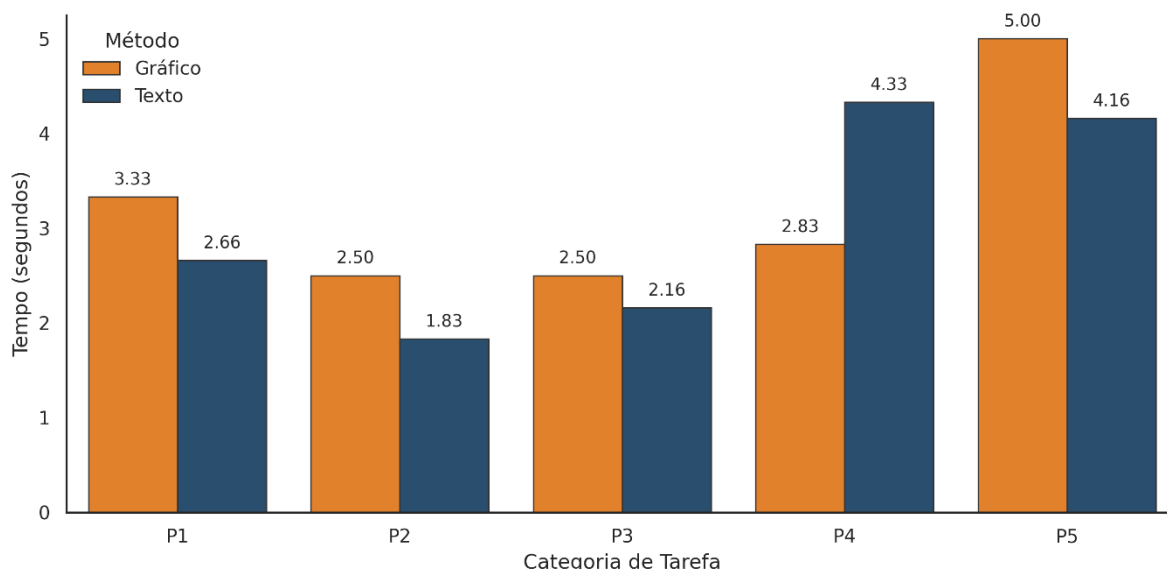
Fonte: Elaborado pelo autor.

Figura 22 – Comparação dos tempos obtidos por categoria de tarefa



Fonte: Elaborado pelo autor.

Figura 23 – Comparação dos valores de *insight* obtidos por categoria de tarefa



Fonte: Elaborado pelo autor.

A partir da análise da Tabela 6, é possível observar que o método visual superou o método textual nas tarefas de triagem e detecção. Nas perguntas P1 de padrão temporal e P2 de detecção de dado, o uso de gráficos reduziu o tempo de resposta em mais de 65,7% e 60,4%, respectivamente. Em contrapartida, a tarefa P4 de correlação causal apresentou um comportamento distinto. Embora o gráfico tenha mantido a rapidez, o método textual alcançou uma média de valor de *insight* superior. Conforme registrado nos áudios das sessões, a estrutura narrativa do texto facilitou a percepção de conexões de causa e efeito, do mesmo modo que a representação visual gerou uma fragmentação que dificultou a síntese lógica por parte de alguns participantes. Por fim, na tarefa P5 de Síntese Complexa o gráfico exigiu um tempo maior para a curadoria dos dados, porém garantiu o alcance da nota máxima de *insight* em todos os casos. Portanto, apesar do gráfico diminuir o tempo para a busca de padrões, o texto oferece a densidade necessária para a validação final e o entendimento do contexto do paciente.

4.4 ANÁLISE ESTATÍSTICA

Com o objetivo de validar o sistema tanto na eficiência temporal quanto na qualidade de interpretação dos profissionais da saúde, além de verificar se a superioridade da interface visual ocorreu por fatores aleatórios, foi aplicado um

conjunto de testes estatísticos inferenciais. O primeiro teste estatístico utilizado foi o Teste de Postos Sinalizados de Wilcoxon pela sua característica não-paramétrica adequada para amostras reduzidas e dados que não apresentam uma distribuição normal (Siegel; Castellan Jr, 2006). Na Tabela 7, são mostrados os resultados obtidos pelo teste que foi aplicado tanto com os resultados da análise temporal quanto com os dados referentes aos valores de *insight* obtidos pelos enfermeiros.

Tabela 7 – Testes estatísticos

Variável analisada	Estatística W	Valor de p	r de Rosenthal	Significância
Tempo Ativo	0,0	0,0312	0,6218	Sim
Valor do <i>Insight</i>	8,5	0,8316	0,1161	Não

Fonte: Elaborado pelo autor.

Com base na análise da Tabela 7, é possível observar que o método gráfico obteve uma redução estatisticamente significativa no tempo de resposta do usuário em relação ao método textual, com $W = 0$ e $p = 0,0312$. A estatística W nula indica que, em todos os testes realizados, o tempo de processamento no cenário gráfico foi inferior ao registrado no método textual. O valor de p obtido representa o limite mínimo de significância alcançável para uma amostra de 6 participantes em testes bicaudais, o que comprova a redução de tempo de processamento cognitivo de forma consistente (Siegel; Castellan Jr, 2006). Adicionalmente, foi calculado o tamanho de efeito de Rosenthal para comprovar a relevância prática desta redução. A representação gráfica atingiu o índice de 0,6218, o que representa um efeito de grande magnitude.

O teste de Wilcoxon aplicado aos valores de *insight* não revelou diferença estatisticamente significativa entre os métodos com o $p = 0,83$. A ausência da significância estatística nesse quesito demonstra a equivalência de eficácia entre as representações. Deste modo, a simplificação visual e a extração de dados por meio do modelo NER não causaram a perda de nuances críticas nem a redução da qualidade do julgamento clínico dos especialistas, os quais atingiram valores de compreensão equivalentes em ambas as modalidades.

Tais índices demonstram que, embora o texto tenda a equalizar os profissionais pela lentidão da leitura linear, o gráfico permite que se manifestem as diferenças individuais de proatividade e percepção. Durante a realização dos

experimentos, cada enfermeiro explorou as funcionalidades disponíveis de forma diferente, por meio da utilização de filtros ou apenas pela observação. Enquanto a leitura textual propõe uma narrativa linear que induz os participantes a conclusões mais homogêneas, o que gera como consequência um tempo de processamento maior, o recurso gráfico estimula a autonomia analítica. Este fenômeno indica que a visualização proposta suporta a descoberta de conhecimento de forma mais plural e profunda, mesmo sob uma curva de aprendizado inicial.

4.5 ANÁLISE QUALITATIVA E PERCEPÇÕES DOS USUÁRIOS

A análise dos resultados quantitativos fundamenta-se nos registros verbais obtidos por meio das gravações de áudio durante o processo de verbalização dos pensamentos. As transcrições revelam os processos mentais implícitos à descoberta de informações e permitem identificar como a representação visual influenciou o raciocínio clínico dos participantes.

A visualização de entidades extraídas pelo modelo NER funcionou como um impulsionador para o conhecimento prévio dos enfermeiros. Em diversos momentos, o gráfico permitiu que o profissional realizasse conexões clínicas complexas a partir de rótulos isolados, mesmo quando a informação não estava explícita na tela. O Participante 1, ao visualizar o medicamento Tenofovir no gráfico, verbalizou imediatamente: “Pelo que eu já posso identificar aqui, ela é portadora de HIV (Vírus da Imunodeficiência Humana), o diagnóstico aqui está faltando”. Esse comportamento mostra que a disposição categórica dos fármacos atua como um gatilho cognitivo. No método textual, informações de mesma natureza exigiram uma varredura de parágrafos densos, o que resultou em um tempo maior de resposta.

Além disso, as sessões confirmaram a eficácia de um modelo híbrido, no qual o gráfico serve como um guia temporal para o texto original. O Participante 2 demonstrou ao identificar um evento adverso no gráfico: “Aqui fala que foi engolido, mas não fala o que. Agora vou no gráfico de doença atual e evolução para ver. Aqui já foi identificado: ela ingeriu duas baterias”. Este relato evidencia que a visualização analítica contribui para sinalizar anomalias de forma rápida, enquanto a consulta do texto preserva a riqueza de detalhes necessária para o entendimento completo do caso. A superioridade do texto para a compreensão de nuances subjetivas e sociais ficou evidente na tarefa sobre o motivo de internação, na qual o Participante 4 relatou:

“A paciente relata que veio em consulta porque ‘o hospital é lindo e elegante’”. Tais detalhes, são complexos e difíceis de codificar em barras ou pontos em representações visuais, o que justifica a coexistência das duas representações.

Por fim, o uso de eixos temporais integrados permitiu que os profissionais justificassem condutas de enfermagem de forma mais precisa. Na análise de episódios de restrição física, o Participante 3 atingiu um *insight* de alta profundidade ao cruzar os sinais motores com a medicação: “Debilidade, hiper medicada por conta da parte psiquiátrica, sonolenta e apresenta risco de queda, por isso contida no leito”. Diferente do prontuário tradicional, no qual a justificativa de uma contenção pode estar separada da anotação da medicação por várias páginas, a visualização gráfica colaborou nesse sentido. Isso permitiu ao profissional deduzir que a intervenção possuía caráter protetivo como prevenção de queda e não punitivo como agressões.

4.6 AVALIAÇÃO DA USABILIDADE

A avaliação da usabilidade percebida ocorreu por meio da aplicação do *System Usability Scale* (SUS). A escolha desta escala foi realizada pela sua capacidade de extrair métricas sobre a satisfação do usuário, mesmo em amostras reduzidas (Brooke, 1996). O questionário é composto por 10 afirmações, as respostas seguiram uma escala *Likert* de 1 a 5, na qual o valor 1 indica discordância total e o valor 5 representa concordância total. Adicionalmente, o instrumento contou com duas perguntas, representadas pelas questões 11 e 12, de resposta livre para coleta de *feedback* qualitativo.

As questões foram apresentadas conforme a lista abaixo:

- 1) Eu gostaria de usar esse sistema frequentemente no meu trabalho.
- 2) Eu achei o sistema desnecessariamente complexo.
- 3) Eu achei o sistema fácil de usar.
- 4) Eu acho que precisaria de ajuda de um técnico para conseguir usar o sistema.
- 5) Eu achei que as várias funções do sistema estavam bem integradas.
- 6) Eu achei que havia muita inconsistência no sistema.
- 7) Eu imagino que a maioria dos enfermeiros aprenderia a usar esse sistema muito rapidamente.
- 8) Eu achei o sistema muito confuso/trabalhoso de usar.

9) Eu me senti muito confiante usando o sistema.

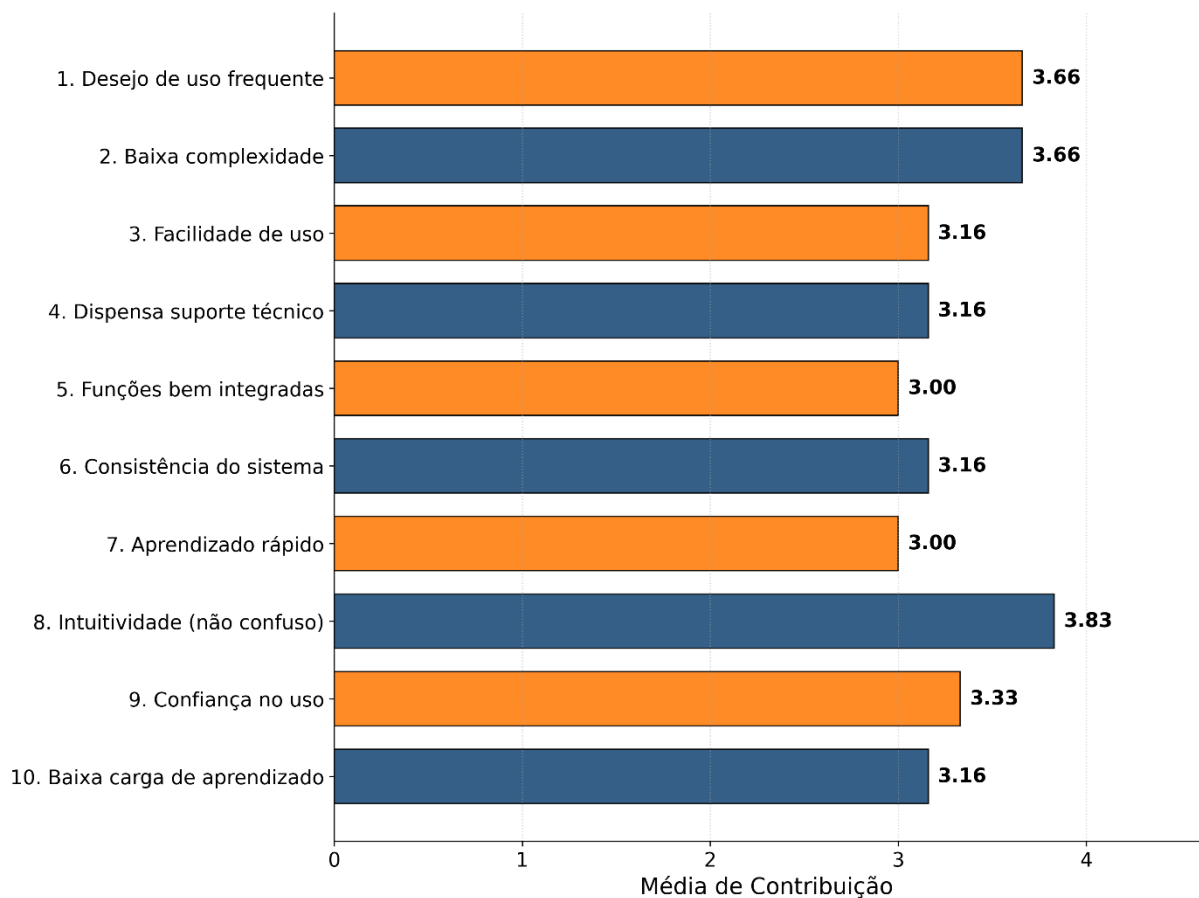
10) Eu precisei aprender muitas coisas antes de conseguir lidar com esse sistema.

11) Qual gráfico ou funcionalidade você achou mais útil?

12) O que você mudaria para tornar a ferramenta melhor?

Para o cálculo do score final, foi realizada a normalização dos valores de cada resposta, em uma escala de 0 a 4 pontos. Para as perguntas ímpares, denominadas como positivas, o cálculo é feito a partir da subtração do valor obtido pelo usuário por 1. Para as perguntas pares, consideradas negativas, o valor foi obtido pela subtração de 5 pelo valor registrado pelo usuário. Depois disso, os valores são somados e multiplicados por 2.5 para resultar no score individual da usabilidade do sistema, que possui um intervalo de 0 a 100 pontos (Brooke, 1996). Na Figura 24 são apresentadas as médias de contribuição para cada uma das 10 questões.

Figura 24 – Médias de contribuição de usabilidade por pergunta



Fonte: Elaborado pelo autor.

Com base na observação da Figura 25, é possível visualizar que os itens relacionados à intuitividade como a questão 8, ao desejo de uso frequente como a questão 1 e à baixa complexidade, como a questão 2, obtiveram os índices de satisfação mais elevados. O sistema alcançou um score médio global de 82,92 pontos, valor considerado excelente de acordo com os critérios de classificação da literatura. Este índice supera o limiar de 68 pontos, considerado a média de mercado (Bangor; Kortum; Miller, 2009).

Nas respostas da questão 11, houve consenso entre os profissionais sobre a utilidade do gráfico de Gantt para a percepção rápida da evolução do paciente, o qual se destacou pela disposição dos dados de medicamentos e diagnósticos. Em relação às melhorias propostas nas respostas da questão 12, os participantes relataram dificuldade na navegação entre as emergências dos pacientes. Os casos com alto volume de registros geraram uma lentidão no acesso dessas informações, o que motivou a sugestão de novos filtros para a seleção de datas e ocorrências.

Além disso, os enfermeiros recomendaram a inclusão de horários específicos no gráfico das anotações de enfermagem para solucionar ambiguidades presentes em registros nos quais o paciente estava descrito como “Calmo” e “Agitado” no mesmo dia. Essa medida permitiria a distinção cronológica em situações em que o paciente estava agitado em primeiro momento e, após alguma intervenção medicamentosa, houve a melhora do quadro.

4.7 ANÁLISE COMPARATIVA

Com o objetivo de situar a visualização proposta perante a literatura, foi realizada uma análise comparativa com estudos de referência. A aplicação da Metodologia Baseada em *Insights* permitiu uma comparação direta com trabalhos correlatos. O trabalho proposto seguiu o uso de tarefas direcionadas, enquanto outros trabalhos como o de Ledesma et al., (2019), adotaram uma abordagem de exploração livre.

Enquanto o correlato em questão obteve uma média de *insight* de 1,70 para o método gráfico, o presente estudo obteve uma média de 3,23. Essa diferença é consequência da natureza estrutural das questões formuladas, que guiaram o raciocínio analítico dos profissionais para níveis de complexidade maiores.

Em relação ao trabalho de Yada e Aramaki (2023), que propõe a visualização de históricos radiológicos a partir de textos via NER, este trabalho introduz a disposição integrada de dados de múltiplas categorias. A maior diferença está na disposição dos sinais vitais e anotações de enfermagem em um único painel, funcionalidade que amplia o suporte à decisão clínica em comparação aos visualizadores focados em relatórios médicos isolados.

4.8 CONSIDERAÇÕES FINAIS

Neste capítulo, foram apresentados os testes e validações efetuados para validar a metodologia proposta. O objetivo principal foi verificar tanto a qualidade geral do modelo ajustado para a tarefa de reconhecimento de entidades quanto o desempenho do recurso de visualização na geração de percepções valiosas. Para isso, os testes foram separados em duas etapas: a avaliação do modelo proposto para o NER e a avaliação dos gráficos gerados.

A primeira etapa consistiu em uma análise comparativa para selecionar qual modelo obteve melhor desempenho. Para isto, um total de 8 modelos pré-treinados com texto em português foram submetidos a um processo de ajuste fino com a mesma base de dados e com o mesmo conjunto de parâmetros. A partir disso, o *BERTugues* obteve um desempenho superior nas métricas de avaliação com as médias de 0,80 de precisão, 0,84 de revocação e 0,82 de F1-Score.

Em sequência foi realizada a comparação com outros modelos presentes na literatura que realizam a mesma tarefa de reconhecimento de entidades em textos médicos na língua portuguesa, no qual o modelo proposto obteve um melhor desempenho. Por fim, por meio dos resultados obtidos pelo ajuste fino do *BERTugues* valida o método proposto como uma solução robusta e de alta performance para a estruturação de dados médicos contidos em texto livre. Essa eficiência na estruturação contribui para a qualidade da ferramenta de visualização desenvolvida.

A segunda etapa consistiu na comparação da visualização das informações dispostas por meio de gráficos e a leitura das informações pelo método tradicional textual. Para isto, foi utilizada a metodologia baseada em *insights*, na qual 6 enfermeiros foram convidados para participar de sessões experimentais dos dois métodos. Nessas sessões, os participantes responderam perguntas pré-definidas a partir de casos clínicos reais. Durante as respostas, seus pensamentos foram

verbalizados com o objetivo de calcular o tempo em que os profissionais entendiam as informações dispostas e a qualidade dessas conclusões obtidas.

Após os cálculos de média dos valores obtidos, o trabalho proposto registrou uma redução média de 29,8% no tempo ativo do *insight* em comparação com o método tradicional de leitura de informações por meio dos textos. Foi verificado que o método gráfico registrou uma média de 82,10 segundos e 3,23 de valor de *insight*, enquanto o método textual registrou uma média de 116,90 segundos e 3,17 pontos, respectivamente.

Para comprovar essa melhora, foram realizados cálculos estatísticos por meio do Teste de Wilcoxon e r de Rosenthal. Como resultado, foi observado que o método gráfico obteve um valor estatisticamente significativo na redução de tempo de obtenção de *insight* com o $p = 0,0312$ e um impacto de grande magnitude com o $r = 0,6218$, porém, não obteve um valor estatisticamente significativo na diferença entre os valores obtidos por ambas as representações. Isto mostra que a diminuição no tempo de resposta não afetou a qualidade das conclusões.

Por último, os participantes preencheram um questionário com 12 perguntas para verificar a escala de usabilidade do sistema, que por sua vez, teve índices de satisfação elevados em itens relacionados à intuitividade, ao desejo de uso frequente e à baixa complexidade. Além disso, foi registrado um índice médio global de 82,92 pontos, valor considerado excelente de acordo com os critérios de classificação da literatura. Esses dados indicam que a interface gráfica acelera a percepção de padrões e riscos, enquanto o acesso ao texto original garante a segurança da informação, o que permite a validação de detalhes subjetivos e a correção de eventuais ambiguidades provenientes da extração automática pelo NER.

5 CONCLUSÃO

Devido ao avanço das pesquisas nas áreas de cuidados médicos e inteligência artificial, a extração de informações provenientes do prontuário médico eletrônico se apresenta como uma técnica importante para o entendimento do histórico do paciente. Contudo, devido às características não estruturadas dos prontuários, a utilização de PLN se tornou uma tarefa necessária. Além disso, a visualização de dados por meio de interfaces gráficas permite a integração múltiplas fontes de dados e auxilia os profissionais da saúde na tomada de decisões sobre o tratamento da paciente (Dabek, Jimenez, Caban, 2017; Ledesma et al., 2019).

Com isso, diversas estratégias têm sido propostas para enfrentar os desafios relacionados à estruturação dos dados provenientes de notas clínicas. Contudo, os estudos presentes na literatura geralmente não possuem o foco em trabalhos que lidem diretamente com os dados clínicos no contexto psiquiátrico na língua portuguesa e que ainda proponham uma abordagem visual para auxiliar especialistas nos processos decisivos.

Neste contexto, o objetivo deste trabalho foi buscar contornar estes problemas de modo a se desenvolver um modelo capaz de reconhecer categorias semânticas de um texto psiquiátrico em português. A partir disso, foi criado um algoritmo para estruturar os dados em um banco de dados relacional, para possibilitar o desenvolvimento de uma ferramenta de visualização em linha do tempo e em gráfico de Gantt para representar eventos clínicos de cada paciente individualmente.

Na fundamentação teórica foram apresentados os principais conceitos relacionados à inteligência artificial, em particular, foram apresentados os métodos de PLN e arquiteturas baseadas em *transformers*. Também foi discutido o papel da visualização de dados na identificação de padrões e na facilitação da análise clínica. O trabalho realizou uma comparação entre arquiteturas BERT pré-treinadas em português para selecionar o modelo com melhor desempenho durante o ajuste fino para a tarefa de reconhecimento de entidades psiquiátricas. O modelo selecionado foi o *BERTugues*, que também demonstrou desempenho superior em comparação com trabalhos correlatos, com métricas de 0,93 de acurácia, 0,80 de precisão, 0,84 de revocação e 0,82 de F1-Score.

Em seguida, a ferramenta de visualização desenvolvida foi comparada com a representação textual escrita em forma livre. Para isso, 6 enfermeiros do Hospital

Psiquiatra Dr. Adolfo Bezerra de Menezes, da cidade de São José do Rio Preto, foram convidados a participar de sessões individuais, nas quais foram expostos a dois cenários clínicos distintos com a alternância entre os métodos de representação. Foram elaborados 4 casos clínicos no total, em que cada caso possuía 5 questões direcionadas para diferentes níveis de valor de percepção. Após a análise dos casos, os profissionais responderam questionários com o intuito de avaliar a usabilidade do recurso visual proposto.

Os resultados indicaram que os gráficos desenvolvidos foram benéficos no tempo de conclusão e se manteve equivalente no valor dos pensamentos, com uma média de 82,1 segundos e 3,23 de valor, uma diminuição de 29,8% no tempo obtido pela análise dos prontuários em texto. O valor de $p = 0,0312$ obtido nos Testes de Wilcoxon representa alta significância no contexto da comparação de tempo dos dois métodos, comprovados, também, pelo valor $r = 0,6218$ de Rosenthal. Em contrapartida, os testes realizados com os valores médios obtidos nos *insights* não demonstraram ser estatisticamente relevantes, com o $p = 0,8316$ e $r = 0,1161$.

Embora o método textual ainda possua os detalhes e nuances relevantes sobre o histórico do paciente, a visualização das informações médicas por meio dos gráficos diminui o tempo do entendimento do estado clínico do paciente, como observado nos resultados obtidos. Esses dados indicam que a integração de modelos de NER com a visualização analítica de dados médicos psiquiátricos é capaz de diminuir o tempo de raciocínio para obtenção de *insights*.

5.1 CONTRIBUIÇÃO CIENTÍFICA

Ao realizar a comparação deste trabalho com os correlatos da literatura, é possível destacar como principal diferencial a apresentação de recursos visuais para reduzir a carga cognitiva de profissionais da saúde durante a interpretação de notas médicas, apoiada por um modelo de linguagem especializado para extrair entidades psiquiátricas na língua portuguesa.

No Quadro 9 é possível observar as principais características dos trabalhos correlatos em relação à abordagem desenvolvida. As características selecionadas para essa contribuição representam requisitos técnicos e clínicos para a redução do tempo de percepção. A utilização de modelos BERT e o NER atua no processamento dos dados textuais, o tratamento de linguagem informal auxilia na recuperação de

informações dos prontuários escritos por humanos, a redução da carga cognitiva e visualização da linha do tempo analisam a usabilidade da interface, o uso de dados psiquiátricos explora o domínio linguístico com subjetividade narrativa e a estruturação dos dados garante a persistência das informações a longo prazo.

A partir da análise do quadro, é possível concluir que o trabalho atual preenche uma lacuna na literatura ao propor uma abordagem que reduz o tempo de percepção dos profissionais da saúde por meio da visualização de dados não estruturadas em conjunto com a estruturação das informações por meio de um modelo de reconhecimento de entidades. Enquanto parte dos trabalhos focam apenas no processamento de texto médico e outra parte apenas na visualização dos dados já estruturados, o presente trabalho apresenta uma solução que une as duas áreas.

Quadro 9 – Comparação final entre os trabalhos correlatos

	Yenkikar et al., (2024)	Dabek, Jimenez e Caban (2017)	Yada e Aramaki (2023)	Ledesma et al., (2019)	Wulff et al., (2020)	Niero et al., (2023)	Este trabalho
Utiliza modelos baseados em BERT	Não	Não	Sim	Não	Não	Sim	Sim
Utiliza Reconhecimento de Entidades	Sim	Não	Sim	Não	Não	Sim	Sim
Tratamento de linguagem informal/abreviada	Não	Não	Sim	Não	Sim	Sim	Sim
Redução da carga cognitiva do usuário	Não	Sim	Sim	Sim	Não	Não	Sim
Realiza a estruturação dos dados	Não	Sim	Não	Não	Sim	Não	Sim
Utiliza dados psiquiátricos	Não	Não	Não	Sim	Não	Sim	Sim
Visualização em linha do tempo	Não	Sim	Sim	Sim	Não	Não	Sim

Fonte: Elaborado pelo autor.

5.2 TRABALHOS FUTUROS

Por fim, alguns possíveis trabalhos futuros incluem a implementação do recurso visual em conjunto com a consulta textual, na qual os usuários utilizam as informações escritas como complemento para as percepções obtidas na visualização do gráfico. Além disso, existe a possibilidade de agregar outros modelos de linguagem

capazes de realizar outros tipos de tarefa que auxiliam o processo dos profissionais da saúde. A criação de um *Retrieval-Augmented Generation* (RAG) especializado nos dados médicos pode permitir a análise e síntese de informações de fontes externas, o que otimiza ainda mais o entendimento dos casos clínicos.

Adicionalmente, incluem considerar o reconhecimento das datas das notas clínicas no próprio texto livre do documento, para não houver a necessidade de considerar a data do preenchimento do formulário em si, assim como aplicar diferentes técnicas no ajuste fino para conseguir detectar negações. Outros aspectos que podem ser explorados envolvem a extração de relações, extrair relações semânticas entre as entidades pode identificar conexões e gerar um grafo de conhecimento muito mais rico.

Em relação à ferramenta de visualização, foram levadas em consideração apenas representações temporais das informações clínicas de um único paciente. Logo, ao observar outros fatores clínicos em comum dos pacientes, é possível definir uma representação visual capaz de auxiliar os profissionais da saúde em um contexto mais amplo. Além disso, o *feedback* interativo dos profissionais pode ser útil para a melhoria do modelo de reconhecimento de entidades, uma vez que o próprio conhecimento do usuário pode contribuir para a adição de informações relevantes para o entendimento do caso do paciente, o que gera uma abordagem cíclica de análise de informações.

REFERÊNCIAS

- AL KUWAITI, Ahmed; NAZER, Khalid; AL-REEDY, Abdullah; AL-SHEHRI, Shaher; AL-MUHANNA, Afnan; SUBBARAYALU, Arun Vijay; AL MUHANNA, Dhoha; AL-MUHANNA, Fahad A. A review of the role of artificial intelligence in healthcare. **Journal of personalized medicine**, v. 13, n. 6, p. 951, 2023.
- AOYAMA, David A.; HSIAO, Jen-Ting T.; CÁRDENAS, Alfonso F.; PON, Raymond K. TimeLine and visualization of multiple-data sets and the visualization querying challenge. **Journal of Visual Languages & Computing**, v. 18, n. 1, p. 1–21, 2007.
- BANGOR, Aaron; KORTUM, Philip; MILLER, James. Determining what individual SUS scores mean: Adding an adjective rating scale. **Journal of usability studies**, v. 4, n. 3, p. 114-123, 2009.
- BROOKE, John. SUS-A quick and Dirty usability scale. **Usability evaluation in industry**, v. 189, n. 194, p. 4-7, 1996.
- CABAN, Jesus J.; GOTZ, David. Visual analytics in healthcare - opportunities and research challenges. **Journal of the American Medical Informatics Association**, v. 22, n. 2, p. 260–262, 1 mar. 2015.
- CHINTHAMU, Narender; KARUKURI, Manideep. Data Science and Applications. **Journal of Data Science and Intelligent Systems**, 11 jul. 2023.
- CUI, Wenqiang. Visual Analytics: A Comprehensive Overview. **IEEE Access**, v. 7, p. 81555–81573, 2019.
- BEZERRA, Selene Maria. Prontuário Eletrônico do Paciente: uma ferramenta para aprimorar a qualidade dos Serviços de Saúde. **Revista Meta: Avaliação**, v. 1, n. 1, p. 73, 2009.
- DABEK, Filip; JIMENEZ, Elizabeth; CABAN, Jesus J. A timeline-based framework for aggregating and summarizing electronic health records. *In*: **2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)**. Phoenix, AZ: IEEE, 2017, p. 55–61.
- DAI, Zhenjin; WANG, Xutao; NI, Pin; LI, Yuming; LI, Gangmin; BAI, Xuming. Named Entity Recognition Using BERT BiLSTM CRF for Chinese Electronic Health Records. *In*: **2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)**. Suzhou, China: IEEE, 2019, p. 1–5.
- DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018.
- ECKHARDT, Cristina; MADJAROVA, Sophia; WILLIAMS, Rilley; OLLIVER, Mattheu; KARLSSON, Jón; PAREEK, Ayoosh; NWACHUKWU, Benedict. Unsupervised machine learning methods and emerging applications in healthcare. **Knee Surgery, Sports Traumatology, Arthroscopy**, v. 31, p. 376-381, 2022

HASAN, Khandaker, Tabin; SHAUGAT ABDULLAH, Sheikh; AHMED, Rezwan; GIUNCHIGLIA, Fausto. The History of Temporal Data Visualization and a Proposed Event Centric Timeline Visualization Model. **International Journal of Computer Applications**, v. 70, n. 27, p. 27–33, 2013.

HU, Linmei; LIU, Zeyi; ZHAO, Ziwang; HOU, Lei; NIE, Liqiang; LI, Juanzi. A Survey of Knowledge Enhanced Pre-Trained Language Models. **IEEE Transactions on Knowledge and Data Engineering**, v. 36, n. 4, p. 1413–1430, 2024.

JAIN, Kamal; PRAJAPATI, Vishal. NLP/deep learning techniques in healthcare for decision making. **Primary Health Care: Open Access**, v. 11, n. 3, p. 373-380, 2021.

JIANG, Yuchen; LI, Xiang; LUO, Hao; YIN, Shen; KAYNAK, Okyay. Quo vadis artificial intelligence? **Discover Artificial Intelligence**, v. 2, n. 1, p. 4, 2022.

JO, Jaemin; HUH, Jaeseok; PARK, Jonghun; KIM, Bohyoung; SEO, Jinwook. LiveGantt: Interactively Visualizing a Large Manufacturing Schedule. **IEEE Transactions on Visualization and Computer Graphics**, v. 20, n. 12, p. 2329–2338, 2014.

KESHTA, Ismail; ODEH, Ammar. Security and privacy of electronic health records: Concerns and challenges. **Egyptian Informatics Journal**, v. 22, n. 2, p. 177–183, 2021.

KOROTEEV, M. V. BERT: A Review of Applications in Natural Language Processing and Understanding. **arXiv (Cornell University)**, 2021.

KUMAR, Pranjal; CHAUHAN, Siddhartha; AWASTHI, Lalit Kumar. Artificial Intelligence in Healthcare: Review, Ethics, Trust Challenges & Future Research Directions. **Engineering Applications of Artificial Intelligence**, v. 120, p. 105894, 2023.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep Learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.

LEDESMA, Andres; BIDARGADDI, Niranjan; STROBEL, Jörg; SCHRADER, Geoffrew; NIEMINEN, Hannu; KORHONEN, Ilkka; ERMES, Miikka. Health timeline: an insight-based study of a timeline visualization of clinical data. **BMC Medical Informatics and Decision Making**, v. 19, n. 1, p. 170, 2019.

LEE, Jinhyuk; YOON, Wonjin; KIM, Sungdong; KIM, Donghyeon; KIM, Sunkyu; HO SO, CHAN; KANG, Jaewoo. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. **Bioinformatics**, v. 36, n. 4, p. 1234–1240, 2020.

LI, Irene; PAN, Jessica; GOLDWASSER, Jeremy; VERMA, Neha; PAN WONG, Wai; NUZUMLALI, Muhammed Yavuz; ROSAND, Benjamin; LI, Yixin; ZHANG, Matthew; CHANG, David; TAYLOR, R. Andrew; KRUMHOLZ, Harlan M.; RADEV, Dragomir. Neural Natural Language Processing for unstructured data in electronic health records: A review. **Computer Science Review**, v. 46, p. 100511, 2022.

MANNE, Ravi; KANTHETI, Sneha C. Application of Artificial Intelligence in Healthcare: Chances and Challenges. **Current Journal of Applied Science and Technology**, p. 78–89, 2021.

MANNING, Christopher D. Introduction to information retrieval. **Syngress Publishing**, 2008.

MARRERO, Mónica; URBANO, Julián; SÁNCHEZ-CUADRADO, Sonia; MORATO, Jorge; GÓMEZ-BERBÍS, Juan Miguel. Named Entity Recognition: Fallacies, challenges and opportunities. **Computer Standards & Interfaces**, v. 35, n. 5, p. 482–489, 2013.

MATSUO, Yutaka; LECUN, Yann; SAHANI, Maneesh; PRECUP, Doina; SILVER, David; SUGIYAMA, Masashi; UCHIBE, Eiji; MORIMOTO, Jun. Deep learning, reinforcement learning, and world models. **Neural Networks**, v. 152, p. 267–275, 2022.

MUHAMMAD, Iqbal; YAN, Zhu. SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY. **ICTACT Journal on Soft Computing**, v. 05, n. 03, p. 946–952, 2015.

MUNISWAMAIAH, Manoj; AGERWALA, Tilak; TAPPERT, Charles C. Big Data and Data Visualization Challenges. *In*: **2023 IEEE International Conference on Big Data (BigData)**. Sorrento, Italy: IEEE, 2023, p. 6227–6229.

MUSKAN; SINGH, Gurpreet; SINGH, Jaspreet; PRABHA, Chander. Data Visualization and its Key Fundamentals: A Comprehensive Survey. *In*: **2022 7th International Conference on Communication and Electronics Systems (ICCES)**. Coimbatore, India: IEEE, 2022, p. 1710–1714.

NIERO, Luiz Henrique Pereira; GUILHERME, Ivan Rizzo; OLIVEIRA, Lucas Emanuel Silva E; FILHO, Gerardo Maria de Araújo. PsyBERTpt: A Clinical Entity Recognition Model for Psychiatric Narratives. *In*: **2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)**. L'Aquila, Italy: IEEE, 2023, p. 672–677.

NORTH, Chris. Toward measuring visualization insight. **IEEE computer graphics and applications**, v. 26, n. 3, p. 6-9, 2006.

PEDDAVENKATAGARI, Chakradhar. Evaluation Methodologies and Performance Metrics in Supervised NER for Journal Articles: A Critical Analysis. **International Research Journal of Modernization in Engineering Technology and Science**, v. 6, 2024.

PN, Anjana; M, Narayanamoorthi. Secured Natural Language Processing for Conversion of Unstructured Text Into Structured Intelligence. *In*: **2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)**. Coimbatore, India: IEEE, 2021, p. 957–962.

PRAFUL BHARADIYA, Jasmin. A Comparative Study of Business Intelligence and Artificial Intelligence with Big Data Analytics. **American Journal of Artificial Intelligence**, 2023.

PROVOST, Foster; FAWCETT, Tom. Data Science and its Relationship to Big Data and Data-Driven Decision Making. **Big Data**, v. 1, n. 1, p. 51–59, 13 fev. 2013.

QU, Kecheng; WANG, Lina. Research on Visual Data Mining Technology. **Journal of Physics: Conference Series**, v. 1748, n. 3, p. 032056, 2021.

ROSTAMZADEH, N.; ABDULLAH, S. S.; SEDIG, K. Visual Analytics for Electronic Health Records: A Review. **Informatics**, v. 8, n. 1, p. 12, 23 fev. 2021.

SALAZAR-REYNA, Roberto; GONZALEZ-ALEU, Fernando; GRANDA-GUTIERREZ, Edgar M.A.; DIAZ-RAMIREZ, Jenny; GARZA-REYES, Jose Arturo; KUMAR, Anil. A systematic literature review of data science, data analytics and machine learning applied to healthcare engineering systems. **Management Decision**, 7 dez. 2020.

SCHNEIDER, Elisa Terumi Rubel; DE SOUZA, João Vitor Andrioli; KNAFOU, Julien; COPARA, Jenny; OLIVEIRA, Lucas E.S. e; GUMIEL, Yohan B.; OLIVEIRA, Lucas D.A. de; TEODORO, Douglas; PARAISO, Emerson Cabrera; MORO, Claudia. BioBERTpt - A Portuguese Neural Language Model for Clinical Named Entity Recognition. *In: Proceedings of the 3rd Clinical Natural Language Processing Workshop*. Online: Association for Computational Linguistics, 2020, p. 65–72.

SHARMA, Neha; SHARMA, Reecha; JINDAL, Neeru. Machine Learning and Deep Learning Applications-A Vision. **Global Transitions Proceedings**, v. 2, n. 1, p. 24–28, 2021.

SHINDE, Pramila P.; SHAH, Seema. A Review of Machine Learning and Deep Learning Applications. *In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. Pune, India: IEEE, 2018, p. 1–6.

SIEGEL, Sidney; CASTELLAN JR, N. John. Estatística não paramétrica para ciências do comportamento. **Artmed Editora**, 2006.

SILVA, Cristiane Rodrigues Da. História do Prontuário Médico: Evolução do Prontuário Médico Tradicional ao Prontuário Eletrônico do Paciente – PEP. **Research, Society and Development**, v. 10, n. 9, p. e28510918031, 2021.

SOUZA, Fábio; NOGUEIRA, Rodrigo Fassetto; LOTUFO, Roberto. BERT models for Brazilian Portuguese: Pretraining, evaluation and tokenization analysis. **Applied soft computing**, v. 149, p. 110901–110901, 2023.

SUBRAHMANYA, Sri Venkat Gunturi; SHETTY, Dasharathraj K.; PATIL, Vathsala; HAMEED, B.M. Zeeshan; PAUL, Rahul; SMRITI, Komal; NAIK, Nithesh; SOMANI, Bhaskar K. The role of data science in healthcare advancements: applications, benefits, and future prospects. **Irish Journal of Medical Science**, v. 191, n. 4, 16 ago. 2021.

SUN, Peng; YANG, Xuezhen; ZHAO, Xiaobing; WANG, Zhijuan. An Overview of Named Entity Recognition. *In: 2018 International Conference on Asian Language Processing (IALP)*. Bandung, Indonesia: IEEE, 2018, p. 273–278.

THOFEHRN, Claudia; LIMA, Walter Celso De. Prontuário Eletrônico do Paciente – A Importância da Clareza da Informação. **Revista Eletrônica de Sistemas de Informação**, v. 5, n. 1, 2006.

THOMAS, James J.; COOK, Kristin A. A visual analytics agenda. **IEEE Computer Graphics and Applications**, v. 26, n. 1, p. 10–13, jan. 2006.

THOMAS, Jim.; KIELMAN, Joe. Challenges for Visual Analytics. **Information Visualization**, v. 8, n. 4, p. 309–314, jan. 2009.

TUNSTALL, Lewis; VON WERRA, Leandro; WOLF, Thomas. **Natural language processing with transformers**. " O'Reilly Media, Inc.", 2022.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, LUKASZ; POLOSUKHIN, Illia. **Attention Is All You Need**. arXiv.org.

WANG, Xu; WANG, Sen; LIANG, Xingxing; ZHAO, Dawei; HUANG, Jincan; XU, Xin. Deep Reinforcement Learning: A Survey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 35, n. 4, p. 5064–5078, 2024.

WULFF, Antje; MAST, Marcel; HASSLER, Marcus; MONTAG, Sara; MARSCHOLLEK, Michael; JACK, Thomas. Designing an openEHR-Based Pipeline for Extracting and Standardizing Unstructured Clinical Data Using Natural Language Processing. **Methods of Information in Medicine**, v. 59, n. S 02, p. e64–e78, 2020.

YADA, Shuntaro; ARAMAKI, Eiji. HearT: Health Record Timeliner to visualise patients' medical history from health record text. **arXiv preprint arXiv:2306.14379**, 2023.

YAO, Yifan; DUAN, Jinhao; XU, Kaidi; CAI, Yuanfang; SUN, Zhibo; ZHANG, Yue. A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly. **High-Confidence Computing**, v. 4, n. 2, p. 100211, 2024.

YENKIKAR, Anuradha; BALI, Manish; PATIL, Rutuja Rajendra; MIRAJKAR, Riddhi; ARA, Tabassum. Biomedical Named Entity Recognition through spaCy: A Visual Exploration. In: **2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)**. IEEE, 2024. p. 17-22.

ZAGO, Ricardo Mazza; PEDOTTI, Luciane Agnoletti dos Santos. BERTugues: A Novel BERT Transformer Model Pre-trained for Brazilian Portuguese. **Semina: Ciências Exatas e Tecnológicas**, v. 45, p. e50630, 2024.

ZHAO, Wayne Xin; ZHOU, Kun; LI, Junyi; TANG, Tianyi; WANG, Xiaolei; HOU, Yupeng; MIN, Yingqian; ZHANG, Beichen; ZHANG, Junjie; DONG, Zican; DU, Yifan; YANG, Chen; CHEN, Yushuo; CHEN, Zhipeng; JIANG, Jinhao; REN, Ruiyang; LI, Yifan; TANG, Xinyu; LIU, Zikang; LIU, Peiyu; NIE, Jian-Yun; WEN, Ji-Rong. A survey of large language models. **arXiv preprint arXiv:2303.18223**, 2023.

ZHOU, Binggui; YANG, Guanghua; SHI, Zheng; MA, Shaodan. Natural Language Processing for Smart Healthcare. **IEEE Reviews in Biomedical Engineering**, v. 17, p. 4–18, 2024.

APÊNDICE A – ASPECTOS TÉCNICOS DA IMPLEMENTAÇÃO DA ABORDAGEM

O apêndice A foi criado com o objetivo de apresentar alguns conceitos técnicos da implementação da abordagem desenvolvida. Nesse processo, são apresentados os códigos das funções responsáveis pelo ajuste fino do modelo de linguagem, exemplos das notas clínicas no formato JSON e, por fim, o modelo relacional do banco de dados desenvolvido com suas entidades e relações detalhadas.

Após a utilização da técnica de tokenização e a preservação das etiquetas de todos os fragmentos das palavras decompostas pelo tokenizador, a classe *dataset* gerencia a preparação final dos dados para o modelo. Nessa classe, o sistema adiciona os tokens especiais obrigatórios [CLS], para sinalizar o início da sequência, e [SEP] para indicar o término. Para viabilizar o processamento em lotes, a classe aplica a técnica de preenchimento, que estende todas as sentenças até o limite de 128 tokens por meio do preenchimento com o token [PAD].

Simultaneamente, o sistema gera a máscara de atenção que instrui o modelo a ignorar os tokens de preenchimento e focar o processamento apenas no conteúdo textual real. Por fim, as palavras e rótulos são convertidos em tensores numéricos para serem processados em GPU. O ciclo de treinamento utiliza funções modulares para otimizar o aprendizado dos pesos da rede neural.

Na Figura A1 é apresentada a parte do código responsável pela função do treinamento do modelo. A função *train* coordena a passagem para frente (*forward pass*), etapa na qual os dados de entrada atravessam as camadas do BERT para a geração das predições brutas. A partir desses resultados, o sistema efetua o cálculo da perda (*loss*), que calcula a diferença entre as predições do modelo e os rótulos reais por meio da função de entropia cruzada.

Após o cálculo do erro, é iniciada a retropropagação (*backward pass*). Nesta fase, o otimizador calcula os gradientes de erro em relação a cada parâmetro do modelo com o uso da regra da cadeia. Para garantir a estabilidade do processo e evitar que as atualizações excessivamente grandes desestabilizem a rede, aplica-se a técnica de *gradient clipping*. Este recurso limita a norma dos gradientes a um valor máximo, o que mitiga o problema da “explosão do gradiente”.

Figura A 1 – Código da função responsável pelo ajuste fino do modelo

```
def train(model, training_loader, optimizer):
    tr_loss, tr_accuracy = 0, 0
    nb_tr_examples, nb_tr_steps = 0, 0
    # Coloca o modelo em modo de treinamento
    model.train()

    for idx, batch in enumerate(training_loader):
        # Move os dados (IDs, Máscara e Alvos) para o dispositivo de processamento
        ids = batch['ids'].to(device, dtype=torch.long)
        mask = batch['mask'].to(device, dtype=torch.long)
        targets = batch['targets'].to(device, dtype=torch.long)

        # Passagem para frente (Forward Pass): o modelo processa os dados e calcula a perda
        outputs = model(input_ids=ids, attention_mask=mask, labels=targets)
        loss, tr_logits = outputs.loss, outputs.logits
        tr_loss += loss.item() # Acumula o valor da perda do lote atual

        nb_tr_steps += 1
        nb_tr_examples += targets.size(0)

        # Cálculo da acurácia de treinamento (Nível de Token)
        # Achata as matrizes para listas longas para facilitar a comparação
        flattened_targets = targets.view(-1)
        active_logits = tr_logits.view(-1, model.num_labels)
        # Obtém o índice do maior valor nos logits (a classe predita pelo modelo)
        flattened_predictions = torch.argmax(active_logits, axis=1)

        # Filtro de Máscara para identificar onde existem palavras reais
        active_accuracy = mask.view(-1) == 1
        targets_active = torch.masked_select(flattened_targets, active_accuracy)
        predictions_active = torch.masked_select(flattened_predictions, active_accuracy)

        # Calcula a acurácia apenas para os tokens relevantes
        tmp_tr_accuracy = accuracy_score(targets_active.cpu().numpy(), predictions_active.cpu().numpy())
        tr_accuracy += tmp_tr_accuracy

        # Processo de otimização
        # Limita o tamanho dos gradientes para evitar o problema de "Exploding Gradients"
        torch.nn.utils.clip_grad_norm_(parameters=model.parameters(), max_norm=MAX_GRAD_NORM)

        # Zera os gradientes acumulados do passo anterior
        optimizer.zero_grad()
        # Calcula os gradientes do erro em relação aos pesos (Retropropagação)
        loss.backward()
        # Atualiza os pesos do modelo com base nos gradientes calculados e na taxa de aprendizado
        optimizer.step()

    # Calcula as médias de perda e acurácia ao final da época de treinamento
    epoch_loss = tr_loss / nb_tr_steps
    tr_accuracy = tr_accuracy / nb_tr_steps
```

Fonte: Elaborado pelo autor.

Complementarmente, a função *valid*, realiza a inferência em dados não vistos para avaliar a capacidade de generalização do sistema. Durante este processo, o modelo opera em modo de avaliação e o cálculo de gradientes é desativado para economizar recursos computacionais. Diferente de tarefas de classificação simples, a avaliação utiliza a biblioteca *seqval*, que adota o esquema de marcação IOB2. Para

garantir a qualidade dos resultados, foi aplicada a técnica de validação cruzada com cinco partições (*5-fold Cross-Validation*). Como é possível observar na Figura A2, o conjunto de dados é dividido em subconjuntos distintos, de modo que cada partição atue como base de teste em um ciclo diferente. Ao término das iterações, o sistema calcula a média e o desvio padrão da precisão, revocação e F1-Score.

Figura A 2 – Trecho do código responsável pela validação cruzada

```
kfold = KFold(n_splits=5, shuffle=True, random_state=42)

for fold, (train_ids, val_ids) in enumerate(kfold.split(data)):
    print(f"\n===== FOLD {fold + 1}/{n_splits} =====\n")

    # Selecionam os índices de sentenças que serão usados especificamente para TREINO e para VALIDAÇÃO nesta partição atual.
    train_subsampler = torch.utils.data.SubsetRandomSampler(train_ids)
    val_subsampler = torch.utils.data.SubsetRandomSampler(val_ids)

    # Gerenciadores da entrega dos dados em lotes (batches) para a memória
    training_loader = DataLoader(full_dataset, batch_size=TRAIN_BATCH_SIZE, sampler=train_subsampler)
    testing_loader = DataLoader(full_dataset, batch_size=VALID_BATCH_SIZE, sampler=val_subsampler)

    # A cada novo fold, o modelo é carregado do zero para evitar o vazamento de dados entre as partições
    model = BertForTokenClassification.from_pretrained(
        MODEL_NAME, num_labels=len(id2label), id2label=id2label, label2id=label2id
    )
    model.to(device)
    optimizer = torch.optim.Adam(params=model.parameters(), lr=LEARNING_RATE)

    for epoch in range(EPOCHS):
        print(f"--- Epoch {epoch + 1}/{EPOCHS} ---")
        train(model, training_loader, optimizer)

    print("\nValidando o fold...")
    report, accuracy, final_labels, final_preds = valid(model, testing_loader)

    # Extrai métricas do relatório da sequeval
    precision = report['weighted avg']['precision']
    recall = report['weighted avg']['recall']
    f1 = report['weighted avg']['f1-score']

    print(f"\nResultados do Fold {fold + 1}:")
    print(f"  Acurácia (token-level): {accuracy:.4f}")
    print(f"  Precisão (ponderada): {precision:.4f}")
    print(f"  Revocação (ponderada): {recall:.4f}")
    print(f"  F1-Score (ponderado): {f1:.4f}")

    print("\nRelatório de Classificação Detalhado (sequeval):")
    print(classification_report(final_labels, final_preds))

    fold_history.append({
        'accuracy': accuracy,
        'precision': precision,
        'recall': recall,
        'f1': f1,
    })
})
```

Fonte: Elaborado pelo autor.

Nas Figuras A3, A4 e A5 são apresentados os exemplos de como as notas clínicas adquiridas foram organizadas no formato JSON. A partir desse formato, é possível associar o texto do documento com os atendimentos do paciente, sem a necessidade de um identificador pessoal.

Figura A 3 – Exemplo de uma nota clínica do tipo "Doença Atual e Evolução"

Doença Atual e Evolução

```
{
  "cabecalho": {
    "paciente": {
      "prontuario": "4529"
    },
    "atendimento": {
      "tipo": "emergencia",
      "data": "2019-09-19 11:24:30.842"
    },
    "profissional": {
      "cargo": "Supervisor de Enfermagem"
    },
    "data_preenchimento": "22/09/2019",
    "hora_preenchimento": "06:31:08"
  },
  "evolucao": {
    "observacoes": "CLIENTE REATIVO CONTACTUANTE DISCURSO ACELERADO, MANTENDO PERÍODOS DE ANSIEDADE, ACEITOU DIETA OFERTADA, HIGIENE E VESTE ADEQUADA, PROSSEGUE COM RETAGUARDA FAMILIAR ACOMPANHADO PELA GENITORA."
  }
}
```

Fonte: Elaborado pelo autor.

Figura A 4 – Exemplo de uma nota clínica do tipo "Anotação de Enfermagem"

Anotação de Enfermagem

```
{
  "cabecalho": {
    "paciente": {
      "prontuario": "4529"
    },
    "atendimento": {
      "tipo": "emergencia",
      "data": "2019-09-19 11:24:30.842"
    },
    "profissional": {
      "cargo": "Técnico em Enfermagem"
    },
    "data_preenchimento": "19/09/2019",
    "hora_preenchimento": "12:47:09"
  },
  "evolucao": {
    "frequencia_cardiaca": { "valor": "85.0", "horario": "12:40" },
    "frequencia_respiratoria": { "valor": "21.0", "horario": "12:40" },
    "pressao": { "valor": "110/70", "horario": "12:40" },
    "saturacao": { "valor": "97.0", "horario": "12:40" },
    "temperatura": { "valor": "36.2", "horario": "12:40" },
    "observacoes": {
      "observacao": "PACIENTE ADMITIDO NA EMERGÊNCIA POR VOLTA DAS 11:20 VINDO DE
      REMOÇÃO DE OUTRO SERVIÇO,CHEGOU CALMO,RESPONSIVO,SONOLENTO,COM
      ACOMPANHANTE,SIC DA MAE PACIENTE VEM FAZENDO USO DE ÁLCOOL E SPA,NA
      DATA DE ONTEM FICOU AGITADO AGRESSIVO FISICAMENTE COM O PAI ONDE
      TENTOU MATA-LO COM VENENO DE RATO NA COMIDA,FICA EM OBSERVAÇÃO EM
      NOSSO SERVIÇO."
    }
  }
}
```

Fonte: Elaborado pelo autor.

Figura A 5 – Exemplo de uma nota clínica do tipo "Avaliação Médica"

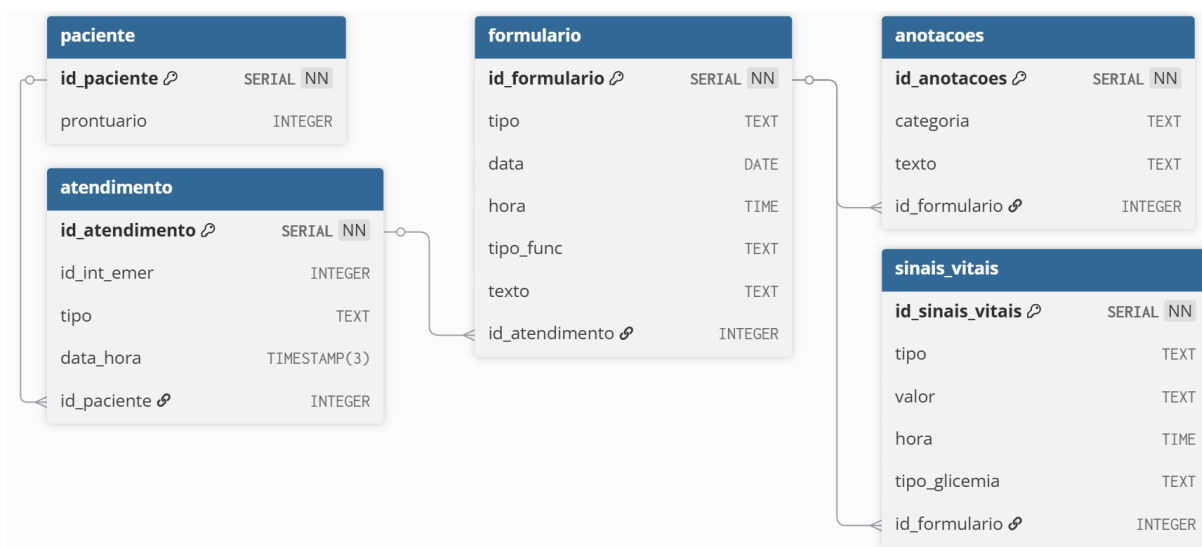
Avaliação Médica

```
{
  "cabecalho": {
    "paciente": {
      "prontuario": "33928"
    },
    "atendimento": {
      "tipo": "emergencia",
      "data": "2023-03-09 15:14:31.954"
    },
    "profissional": {
      "cargo": "Médico Psiquiatra"
    },
    "data_preenchimento": "09/03/2023",
    "hora_preenchimento": "18:33:31"
  },
  "evolucao": {
    "historico_doenca": "PACIENTE DESACOMPANHADA. REFERINDO IDEAÇÃO SUICIDA COM PLANEJAMENTO. RELATA QUE HÁ CERCA DE 05 MESES COMEÇOU APRESENTAR PENSAMENTOS RECORRENTES DE MORTE E SUICÍDIO, E HÁ 04 MESES PLANEJOU SUICÍDIO POR ENFORCAMENTO, CHEGOU A POSICIONAR CORDA, MAS FICOU COM MEDO. DESDE ENTÃO VEM APRESENTANDO BAIXA AUTO-ESTIMA, TRISTEZA, PENSAMENTOS DE CULPA RECORRENTES, CONSEGUINDO CONCILIAR SUAS ATIVIDADES COTIDIANAS, PORÉM REFERE DESÂNIMO, ANERGIA, FADIGA. PERDA O APETITE. CONCILIANDO O SONO, COM HIPERSONIA DIURNA. PAROU DE REALIZAR ATIVIDADES PRAZEIROAS (COMO FREQUENTAR A IGREJA, OUVIR MÚSICA, COZINHAR E CANTAR). EPISÓDIO DEPRESSIVO COM TTO PSICOTERÁPICO EM 2016 - INTERROMPEU POR CONTA
  }
}
```

Fonte: Elaborado pelo autor.

Após a aquisição e a aplicação do modelo nas informações textuais extraída das notas clínicas no formato JSON, as entidades extraídas são estruturadas em um banco de dados de acordo com o modelo relacional representado na Figura A6.

Figura A 6 – Modelo Relacional do banco de dados desenvolvido



Fonte: Elaborado pelo autor.

A entidade das anotações possui os atributos de categoria e texto, para registrar a categoria semântica reconhecida no texto e o próprio texto em que essa categoria foi reconhecida, respectivamente. Além disso, a entidade possui o relacionamento de cardinalidade N para 1 com a entidade de formulários, de modo que uma anotação está presente em apenas um formulário e um formulário pode ter várias anotações associadas a ele. Desta forma, é possível associar de quais formulários estão presentes determinadas informações reconhecidas no texto.

A entidade de formulários possui atributos como: tipo, texto, data, hora e o tipo de funcionário que preencheu, como médico, enfermeiro, auxiliar, entre outros. O atributo de tipo serve para identificar qual formulário é e quais informações são mais relevantes para ele. Os tipos selecionados para o trabalho foram: Avaliação médica, Anotações de enfermagem e Doença atual e evolução. Além do tipo, o formulário possui o atributo “texto” que serve para armazenar o texto original presente no formulário. Essa informação armazenada permite a comparação posterior do texto com as entidades reconhecidas apresentadas no gráfico. O formulário possui uma relação de cardinalidade de N para 1 com a entidade dos atendimentos, pois um formulário só pode ser preenchido em um atendimento, enquanto um atendimento pode ter vários formulários associados.

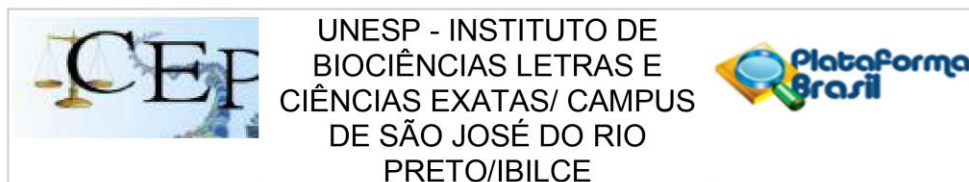
A entidade de atendimentos possui atributos de tipo, data e hora. Do mesmo modo que os formulários, o tipo de atendimento auxilia na identificação de informações mais importantes para cada atendimento, como a emergência ou a

internação do paciente. O atendimento possui uma relação N para 1 com a entidade de pacientes, pois um paciente pode passar por vários atendimentos e um atendimento só pode estar associado a um único paciente.

Os pacientes possuem apenas um número de identificação gerado aleatoriamente, de modo que não seja possível associá-lo com alguma pessoa real e ser possível associar com os atendimentos e outras entidades relacionadas para que, posteriormente, esses dados sejam visualizados em gráficos.

Por fim, a entidade de sinais vitais foi projetada para armazenar medições quantitativas e possuem uma relação de 1 para 1 com os formulários. Isso se deve à condição de que ela só existe caso o formulário for do tipo “Anotações de enfermagem”, de maneira a representar o fluxo real de como essas informações são representadas e armazenadas. Essa entidade possui atributos de tipo, valor e hora. O tipo de sinal vital é importante para determinar o valor dele, os tipos selecionados são: pressão, frequência cardíaca, frequência respiratória, temperatura e saturação.

ANEXO A – PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA EM PESQUISA



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Extração de conhecimento em Bases de Dados para Hospitais com foco no atendimento em patologias mentais e dependência química

Pesquisador: CARLOS ROBERTO VALENCIO

Área Temática:

Versão: 2

CAAE: 89963525.0.0000.5466

Instituição Proponente: Instituto de Biociências Letras e Ciências Exatas/ Campus de

Patrocinador Principal: HOSPITAL DR ADOLFO BEZERRA DE MENEZES

DADOS DO PARECER

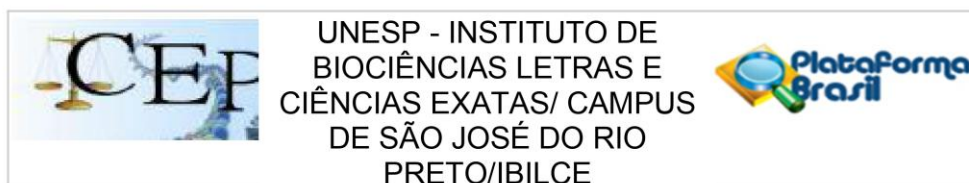
Número do Parecer: 7.887.245

Apresentação do Projeto:

O projeto intitulado "Extração de conhecimento em Bases de Dados para Hospitais com foco no atendimento em patologias mentais e dependência química" considera que o processo de extração de conhecimento em banco de dados se tornou um vasto campo de estudo em meio a grandes repositórios de dados e a busca por informações implícitas nessas estruturas [1, 2]. Segundo o pesquisador, tal atividade faz parte de um processo mais amplo

conhecido como Knowledge Discovery in Database e KDD, cujas demais atividades envolvidas constituem-se do pré-processamento de dados, etapa de preparação dos dados, e pós-processamento e etapa de análise e compreensão dos resultados obtidos [3]. Diante do grande volume de dados armazenados, a extração de conhecimento deve ser realizada com apoio de técnicas e tecnologias de Data Clean, Data Mining, Spatial Data Mining, Visual Data Mining e outras, uma vez que o ser humano é incapaz de realizar esse

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro: JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município:** SAO JOSE DO RIO PRETO
Telefone: (17)3221-2567 **Fax:** (17)3221-2409 **E-:** cep.ibilce@unesp.br

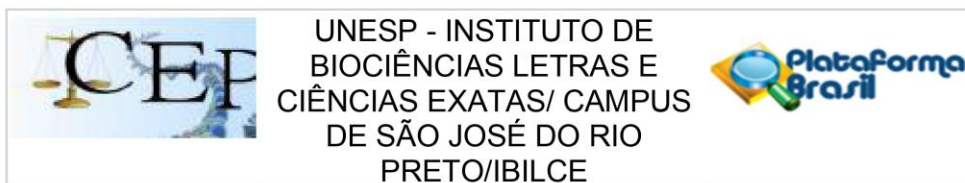


Continuação do Parecer: 7.887.245

processo manualmente [1], sendo considerado útil para aplicação em diversas áreas do conhecimento.

Além das coleções de técnicas, algoritmos e estratégias tradicionais que suportam o KDD, o pesquisador destaca que recentemente incrementou-se aos recursos da área de Inteligência Artificial (IA) que vem se apresentado como ingrediente relevante no processo de extração de conhecimento. Assim, segundo o pesquisador, algoritmos de aprendizado de máquina e redes neurais são utilizados para identificar padrões complexos a partir de dados [4]. A combinação de KDD e IA possibilita uma análise mais profunda e eficiente dos dados, na qual permite avanços significativos na área da saúde. Uma subárea importante da IA aplicada ao KDD, segundo o pesquisador, é o processamento de linguagem natural (PLN). O PLN permite que máquinas compreendam e interpretem a linguagem humana, o que possibilita a extração de conhecimento a partir de grandes volumes de texto não estruturado [5]. A área da saúde, por meio dos sistemas computacionais conectados a sensores e demais coletores de dados, o advento das mídias sociais e outros novos recursos, configuram um novo cenário para os dados, denominado Big Data [6][7][8]. O Grupo de Banco de Dados - GBD, em parceria com o Hospital Filantrópico Dr. Adolfo Bezerra de Menezes - HABM, de São José do Rio Preto, têm desenvolvido o sistema GEP para informatizar os prontuários médicos dos pacientes, e demais setores do Hospital que estão ligados aos prontuários médicos, com o objetivo de tornar-se possível a elaboração de estratégias de análise dos dados armazenados. Assim, segundo o pesquisador, o desenvolvimento de pesquisas com foco na extração de conhecimento em bases de dados relacionadas a saúde permite a identificação de padrões e correlações entre as informações, o que oferece suporte tanto na tomada de decisão sobre ações de prevenção quanto medidas de tratamento. A pesquisa a ser realizada

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município** SAO JOSE DO RIO PRETO
Telefone (17)3221-2567 **Fax:** (17)3221-2409 **E-** cep.ibilce@unesp.br



Continuação do Parecer: 7.887.245

disponibilizará ao Hospital Dr. Adolfo Bezerra de Menezes a adição de recursos computacionais avançados e de apoio ao atendimento, tratamento e acompanhamento de pacientes com patologias mentais ou dependência química.

Objetivo da Pesquisa:

A hipótese apresentada é de que a aplicação de técnicas de extração de conhecimento na base de dados do Hospital Dr. Adolfo Bezerra de Menezes permite a identificação de correlações implícitas nos dados, bem como padrões e comportamentos de pacientes com dependências químicas ou doenças mentais. O entendimento das relações possibilita o desenvolvimento de políticas preventivas e melhores abordagens para tratamentos de pacientes específicos no hospital.

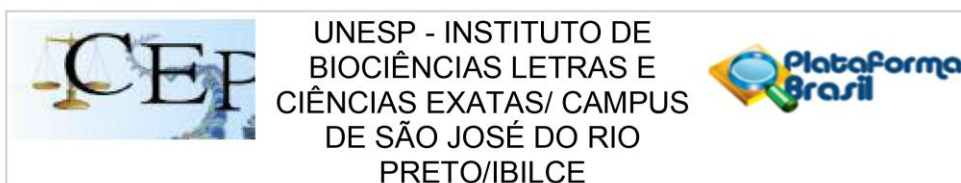
O objetivo primário:

As atividades principais consistem em confeccionar infraestrutura computacional que possibilite o desenvolvimento de um sistema de gestão e evolução de prontuários médicos e, em paralelo, os trabalhos de pesquisas que permitam a extração de conhecimento nas bases de dados de suporte à infraestrutura computacional citada. Para tanto, serão desenvolvidas ferramentas para sistematização dos dados sobre pacientes e prontuários, além da elaboração de técnicas para execução das etapas necessárias para o processo de extração de conhecimento em banco de dados.

O objetivo secundário:

Infraestrutura computacional para o armazenamento de dados e gerenciamento dos processos médicos e hospitalar; Desenvolvimento de tecnologias computacionais para o processo de extração de conhecimento em banco de dados; e eliminação de informações impressos em papel conforme as restrições pertinentes legais e de segurança.

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município** SAO JOSE DO RIO PRETO
Telefone (17)3221-2567 **Fax:** (17)3221-2409 **E-** cep.ibilce@unesp.br



Continuação do Parecer: 7.887.245

Avaliação dos Riscos e Benefícios:

ASPECTOS ÉTICOS

Riscos, providências e cautelas

A pesquisa não oferece riscos aos pacientes, uma vez que as análises realizadas sobre os dados preenchidos pelos médicos e enfermeiros não possuem a identificação dos pacientes, pois se tratam de dados anonimizados de maneira que não podem ser associados à pessoas reais. Deste modo, será solicitada a dispensa do Termo de Consentimento Livre e Esclarecido.

Benefícios

A área da saúde gera um grande volume de dados que necessitam de estratégias de armazenamento, gerenciamento e análise, entre outras, para disponibilizarem informações úteis aos usuários especialistas. Assim, o desenvolvimento de pesquisas com foco no processo de extração de conhecimento em banco de dados relacionadas a saúde permite a identificação de padrões e correlações entre as informações, o que oferece suporte tanto na tomada de decisão sobre ações de prevenção quando medidas de tratamento. Dessa forma, a pesquisa a ser realizada disponibilizará ao hospital uma importante ferramenta de apoio ao atendimento de pacientes com patologias mentais ou dependência química, segundo consta no projeto.

Comentários e Considerações sobre a Pesquisa:

A pesquisa vem somar para a melhoria do atendimento do paciente.

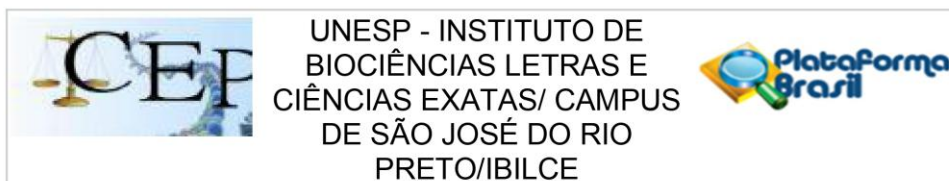
Considerações sobre os Termos de apresentação obrigatória:

Os termos de apresentação obrigatória estão em ordem.

Conclusões ou Pendências e Lista de Inadequações:

O pesquisador esclareceu todas as dúvidas

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município** SAO JOSE DO RIO PRETO
Telefone (17)3221-2567 **Fax:** (17)3221-2409 **E-** cep.ibilce@unesp.br



Continuação do Parecer: 7.887.245

Considerações Finais a critério do CEP:

Deliberação: O Comitê de Ética em Pesquisa, em reunião ordinária de 29 de setembro de 2025, deliberou, por unanimidade, pela aprovação da emenda ao presente Projeto de Pesquisa.

Os relatórios deverão ser encaminhados anualmente, contando a partir desta data, bem como o relatório final ao término da pesquisa, conforme modelo em nossa página: <https://www.ibilce.unesp.br/#!/comite/etica-em-pesquisa/relatorio-projeto/>

Situação: aprovado

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2578661.pdf	13/08/2025 12:52:54		Aceito
Outros	cartaanuencia.pdf	13/08/2025 12:51:58	CARLOS ROBERTO VALENCIO	Aceito
Parecer Anterior	carta_respostaassinado.pdf	13/08/2025 12:50:45	CARLOS ROBERTO VALENCIO	Aceito
Projeto Detalhado / Brochura Investigador	GEP_Brochura_do_Investigador_Projeto_Detalhado_Atualizadoassinado.pdf	13/08/2025 12:50:02	CARLOS ROBERTO VALENCIO	Aceito
Folha de Rosto	folha_de_rosto.pdf	18/06/2025 17:07:28	CARLOS ROBERTO VALENCIO	Aceito

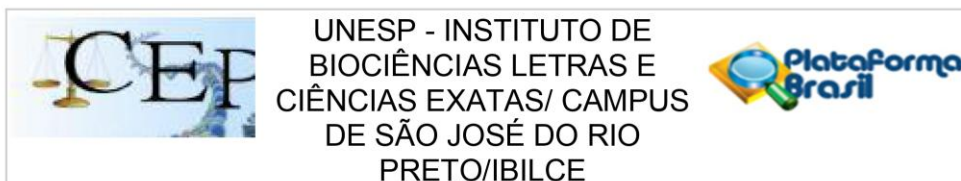
Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município** SAO JOSE DO RIO PRETO
Telefone (17)3221-2567 **Fax:** (17)3221-2409 **E-** cep.ibilce@unesp.br



Continuação do Parecer: 7.887.245

SAO JOSE DO RIO PRETO, 07 de Outubro de 2025

Assinado por:
Lília Santos AbreuTardelli
(Coordenador(a))

Endereço: Rua Cristóvão Colombo, 2265 - Prédio principal, Ala 1, 1º andar, sala 9.
Bairro JARDIM NAZARETH **CEP:** 15.054-000
UF: SP **Município** SAO JOSE DO RIO PRETO
Telefone (17)3221-2567 **Fax:** (17)3221-2409 **E-** cep.ibilce@unesp.br

DADOS CURRICULARES

IDENTIFICAÇÃO	
	Wellington Reguera Gouveia Data de nascimento: 28/05/2001
Nacionalidade	Brasileiro
Nome em citações bibliográficas	Gouveia, W. R. GOUVEIA, WELLINGTON REGUERA
Currículo Lattes	https://lattes.cnpq.br/4421326383467375
ORCID	https://orcid.org/0009-0005-9419-177X
FORMAÇÃO ACADÊMICA	
2019/2023	Bacharel em Ciência da Computação Universidade Estadual Paulista “Júlio de Mesquita Filho”
2024/2026	Mestre em Ciência da Computação Universidade Estadual Paulista “Júlio de Mesquita Filho”
PRODUÇÃO BIBLIOGRÁFICA	
Valêncio, C.R.; Gouveia, W.R.; Zafalon, G.F.D.; Colombini, A.C.; Tronco, M.L.; de Andrade, T.L. An Architectural Optimization Framework for Scalable Spatial Clustering in High-Redundancy Environments. Technologies 2026, 14, 171. https://doi.org/10.3390/technologies14030171	
PARTICIPAÇÃO EM BANCAS E ORIENTAÇÕES	
Bancas de trabalhos de conclusão	
Orientações	
PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS	