



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DETECÇÃO DE REAÇÕES ADVERSAS A MEDICAMENTOS A PARTIR DA  
FUSÃO DE INFORMAÇÕES COM BASE NA CONFIABILIDADE DE SENSORES E  
FONTES

Luiz Alberto Pereira Afonso Ribeiro

**Orientador**

Profa. Dra. Ana Cristina Bicharra Garcia

**Co-Orientador**

Prof. Dr. Paulo Sérgio Medeiros dos Santos

RIO DE JANEIRO, RJ - BRASIL

FEVEREIRO DE 2022

DETECÇÃO DE REAÇÕES ADVERSAS A MEDICAMENTOS A PARTIR DA  
FUSÃO DE INFORMAÇÕES COM BASE NA CONFIABILIDADE DE SENSORES E  
FONTES

LUIZ ALBERTO PEREIRA AFONSO RIBEIRO

TESE DE DOUTORADO APRESENTADA COMO REQUISITO PARCIAL PARA OB-  
TENÇÃO DO TÍTULO DE DOUTOR PELO PROGRAMA DE PÓS-GRADUAÇÃO  
EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JA-  
NEIRO. APROVADA PELA COMISSÃO EXAMINADORA ABAIXO ASSINADA.

Aprovada por:



Ana Cristina Bicharra Garcia, Ph.D (Orientadora, UNIRIO)



Paulo Sérgio Medeiros dos Santos, D.Sc (Co-orientador, UNIRIO)

Carlos Eduardo Ribeiro de Mello, Ph.D - UNIRIO



Sean Wolfgang Siqueira, D.Sc. - UNIRIO



Jose Manuel Lopez Molina, Ph.D. - Universidad Carlos III Madrid

Luis Martí Orosa, Ph.D - INRIA-Chile

RIO DE JANEIRO, RJ - BRASIL  
FEVEREIRO DE 2022.

## Início da Ficha catalográfica

Ribeiro, Luiz Alberto Pereira Afonso

484 Detecção de reações adversas a medicamentos a partir da fusão de informações com base na confiabilidade de sensores e fontes

133.

Orientador: Ana Cristina Bicharra Garcia. Coorientador: Paulo Sérgio Medeiros dos Santos. Tese (Doutorado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Informática, 2022

1. Teoria da evidência. 2. Machine learning. 3. Reações adversas a medicamentos. 4. Inferência bayesiana. 5. Séries temporais. I. Garcia, Ana Cristina Bicharra, orient. II. dos Santos, Paulo Sérgio Medeiros, coorient. III. Título

*Dedico esse trabalho a minha amada mãe-  
zinha, Celeste, com muito amor.*

## **Agradecimentos**

Antes de mais nada quero agradecer a Deus por ter abençoado todos os dias da minha vida, por iluminar meu caminho e me dar forças para seguir sempre em frente. É um desafio agradecer em 2 páginas um esforço para atingir esse que, sem dúvida, foi o principal projeto da minha vida. Eu sabia que não seria uma corrida de 100 metros, seria mais que uma maratona, quase algo tipo “Iron Man”. Ninguém imaginava que iríamos passar por uma pandemia. Mas a resiliência está aí pra ser empregada.

Desejo exprimir os meus agradecimentos a todos aqueles que possibilitaram que esta tese se concretizasse. Independente do fato da autoria da tese ser individual, é uma construção fruto do esforço de várias pessoas indiretamente e algumas poucas de uma forma mais direta.

Em primeiro lugar quero agradecer a Profa. Dra. Ana Cristina Garcia, minha orientadora, por ter-me acolhido em seu grupo de trabalho e ter me dado o voto de confiança para o desenvolvimento desse projeto. Nunca tinha tido a oportunidade de estar próximo a uma pessoa tão inteligente e generosa. Poucas vezes na vida vi alguém que ame tanto o que faz como a Profa. Cristina. Certamente por isso o faz tão bem. Muito obrigado, serei eternamente grato.

Muito também agradeço ao Prof. Dr. Paulo Sérgio Medeiros dos Santos do PPGI UNIRIO. O professor Paulo entrou numa fase crucial do projeto e sua co-orientação foi decisiva para o desenrolar do trabalho, desde a etapa da qualificação. O pensar científico foi um dos grandes desafios de um projeto de doutorado e o Prof. Paulo me ajudou muito nesse sentido. O meu muito obrigado.

Agradeço ao Prof. e pesquisador da FIOCRUZ Dr. Fabrício Silva. Muito importante no levantamento inicial dessa pesquisa. Também agradeço ao Dr. Daniel Cinalli que atuou no ano de 2019 no grupo de pesquisa em uma época que ainda estudava RAM a partir de mídias sociais.

Agradeço aos professores do PPGI, Pimentel, Sean, Simone, Marcio, Adriana, Renata, Guta e Cadu. Todos vocês foram muito importantes nessa formação. Especialmente professores Sean e Pimentel, que lecionaram disciplinas que muito agregaram ao meu preparo, como comunicação científica e Design Science Research. Os feedbacks enriquecedores dos professores Sean, Carlos Eduardo, Molina e Luis Martí enriqueceram bastante esse trabalho.

Também fiz novos colegas, cito dois em especial, o Wagner Silva e a Patrícia Zamprogno, por dividirem horas de angústia, troca de ideias e conhecimento. Também agradeço a sempre produtiva troca com os colegas do grupo de pesquisa, José Orete (também colega do INCA), Fernando Durier, Fabio Daudt, Noemi Paixão, Luciano Santos e Daniel Costa.

Agradeço a dois colegas do INCA que foram muito importantes nesse trabalho: Sandra Gomes e André Cordeiro. Sandra foi muito generosa em me explicar vários conceitos importantes de farmacovigilância na fase inicial desse trabalho. O André me auxiliou sobremaneira no processo de análise e extração de dados dos registros eletrônicos de saúde.

Deixo os familiares por último. No meio dessa jornada minha mãezinha amada, a Celeste, voltou pra casa do Pai. A ela que foi, o meu melhor exemplo de vida, pois meu pai eu perdi ainda menino, eu dedico esse trabalho. E agradeço por toda a intercessão que certamente ela vem me proporcionando lá do céu. Como diz o poeta Drummond, “mães não deveriam morrer nunca”. Aliás, na verdade, elas não morrem.

Agradeço ainda a paciência e amor demonstrados, especialmente nos momentos mais difíceis dessa empreitada, pela minha companheira Lygia e meus filhos Fellipe e Fábio. Peço-vos desculpas pela atenção desviada e agradeço a paciência e acompanhamento nessa fase especial. Um abraço carinhoso e meu eterno obrigado por todo esse amor envolvido.

A todos os meus sinceros agradecimentos.

## RESUMO

A fusão de informações multissensor traz desafios como heterogeneidade de dados, precisão de fonte e a fusão de incertezas que impactam a qualidade dos classificadores. Uma abordagem amplamente utilizada para problemas de classificação em um contexto multissensor é a Teoria Dempster-Shafer. Essa abordagem considera as crenças vinculadas a cada fonte para consolidar as informações sobre as hipóteses para chegar a um classificador com maior precisão. No entanto, a premissa fundamental para usar a abordagem é que as fontes são independentes e que as hipóteses de classificação são mutuamente exclusivas. Algumas abordagens ignoram essa premissa, o que pode levar a resultados não confiáveis. Existem outras abordagens, baseadas em estatísticas e técnicas de aprendizado de máquina, que eliminam as dependências ou incluem um fator de desconto para mitigar o risco de dependências. Propomos uma nova abordagem baseada em rede Bayesiana, teste de Pearson e regressão linear para ajustar as crenças para uma fusão de dados mais precisa, mitigando possíveis correlações ou dependências. Testamos nossa abordagem aplicando-a no domínio da descoberta de reações adversas a medicamentos. O experimento utilizou nove bancos de dados contendo dados de 50.000 pacientes ativos de um hospital oncológico brasileiro, incluindo exames clínicos, exames laboratoriais, anamnese médica, prescrições médicas, notas clínicas, embalagens de bulas de medicamentos, classificação internacional de doenças e modelos de diagnóstico de doenças. Este estudo teve aprovação do comitê de ética do hospital. Obteve-se uma melhoria estatisticamente significativa na precisão e na recuperação dos resultados em comparação com as abordagens existentes. Os resultados obtidos mostram que o índice de credibilidade proposto pelo modelo aumenta significativamente a qualidade das evidências geradas com o algoritmo Random Forest. Foi realizado um benchmark entre três conjuntos de dados, incrementados gradativamente com atributos de um índice de credibilidade, obtendo uma precisão de 92%. Por fim, realizamos um benchmark com base pública de cardiopatias, obtendo bons resultados.

**Palavras-chave:** Teoria da evidência, machine learning, reações adversas a medicamentos, inferência bayesiana, séries temporais.

## ABSTRACT

Multisensor information fusion brings challenges such as data heterogeneity, source precision, and the merger of uncertainties that impact the quality of classifiers. A widely used approach for classification problems in a multisensor context is the Dempster–Shafer Theory. This approach considers the beliefs attached to each source to consolidate the information concerning the hypotheses to come up with a classifier with higher precision. Nevertheless, the fundamental premise for using the approach is that sources are independent and that the classification hypotheses are mutually exclusive. Some approaches ignore this premise, which can lead to unreliable results. There are other approaches, based on statistics and machine learning techniques, that expurgate the dependencies or include a discount factor to mitigate the risk of dependencies. We propose a novel approach based on Bayesian net, Pearson’s test, and linear regression to adjust the beliefs for more accurate data fusion, mitigating possible correlations or dependencies. We tested our approach by applying it in the domain of adverse drug reactions discovery. The experiment used nine databases containing data from 50,000 active patients of a Brazilian cancer hospital, including clinical exams, laboratory tests, physicians’ anamnesis, medical prescriptions, clinical notes, medicine leaflets packages, international classification of disease, and sickness diagnosis models. This study had the hospital’s ethical committee approval. A statistically significant improvement in the precision and recall of the results was obtained compared with existing approaches. The results obtained show that the credibility index proposed by the model significantly increases the quality of the evidence generated with the algorithm Random Forest. A benchmark was performed between three datasets, incremented gradually with attributes of a credibility index, obtaining a precision of 92%. Finally, we performed a benchmark with a public base of heart disease, achieving good results.

**Keywords:** Theory of evidence, machine learning, adverse drug reactions, Bayesian inference, time series.

## Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
1.1	Motivação . . . . .	3
1.2	Definição do Problema de Pesquisa . . . . .	5
1.3	Objetivos da Pesquisa . . . . .	7
1.4	Metodologia da Pesquisa . . . . .	8
1.5	Organização da Tese . . . . .	9
<b>2</b>	<b>Fundamentação Teórica</b>	<b>11</b>
2.1	Contextualização sobre RAM . . . . .	11
2.2	Introdução e Histórico . . . . .	13
2.3	Fusão de Informação . . . . .	14
2.4	Rede Bayesiana . . . . .	16
2.5	Séries Temporais . . . . .	17
2.6	Teoria da Evidência - Dempster and Shafer Theory (DST) . . . . .	17
2.6.1	Benefícios da DST . . . . .	22
2.6.2	Desafios da DST . . . . .	22
2.6.3	Considerações Finais . . . . .	23
<b>3</b>	<b>Revisão de Literatura</b>	<b>24</b>

3.1	Metodologia da revisão . . . . .	24
3.1.1	Estratégia de Busca . . . . .	24
3.2	Triagem da revisão . . . . .	25
3.2.1	Eixos da Revisão . . . . .	26
3.3	Resposta à 1ª questão Quais são as fontes alternativas e respectivos métodos para identificação de RAM? . . . . .	26
3.4	Resposta à 2ª questão - Qual a evolução das técnicas de classificação e análise de sentimentos para a extração de RAM? . . . . .	29
3.4.1	Modelos profundos e combinações - Uma evolução . . . . .	30
3.4.2	Tarefa PLN em destaque - reconhecimento de entidades nomeadas (REN) . . . . .	31
3.4.3	O uso de <i>embedding</i> no LSTM e associação com CRF . . . . .	32
3.5	Resposta à terceira questão - Quais são as principais bases de dados sobre RAM e como são criadas, mantidas e pesquisadas? . . . . .	35
3.5.1	Dicionários e abordagens baseadas em regras. . . . .	36
3.5.2	“Corpus” e ferramentas auxiliares para a formação de léxicos, integrando a bases preexistentes. . . . .	37
3.5.3	Conjunto de dados tradicionais . . . . .	38
3.5.4	Bancos de dados próprios utilizando processo de fusão de dados . . . . .	39
3.5.5	Fusão de informação multisensores FIMS . . . . .	40
3.6	Resposta à quarta questão - Que estruturas de raciocínio podem ser aplicadas para mitigar a incerteza pelas relações de dependência na formação das hipóteses em processos de fusão de dados? . . . . .	43
3.6.1	Iniciativas para reforçar a confiança e credibilidade . . . . .	44
3.6.2	Associação, combinação na análise de dados - trabalhando a condicionalidade e temporalidade . . . . .	46
3.6.3	Combinação de técnicas Bayesianas, Fuzzy e DST para melhora de precisão . . . . .	46

3.7	Críticas a BPA e origem das limitações de Dependência . . . . .	47
3.7.1	Dependências- causalidade - usando Bayesiano conjugado com DST . . . . .	48
3.8	Discutindo os resultados desta revisão de literatura . . . . .	49
3.9	Trabalhos relacionados . . . . .	53
3.10	Comentários finais sobre a revisão de literatura . . . . .	56
<b>4</b>	<b>DEP-DST: Método de geração de evidências com ajuste de dependências</b>	<b>57</b>
4.1	Modelo Arquitetônico . . . . .	57
4.2	Componentes do Índice de credibilidade para o tratamento das depen- dências . . . . .	59
4.2.1	DAG Bayesiano - Probabilidades Condicionais . . . . .	60
4.2.2	Análise de Correlação com Teste de Pearson . . . . .	61
4.2.3	Tamanho do Efeito- análise de interações . . . . .	63
4.2.4	Análise Temporal . . . . .	63
4.3	Método PCA - Integração dos Componentes do Índice de Credibilidade .	65
4.4	Algoritmos . . . . .	67
4.5	Ameaças à validade do modelo . . . . .	68
4.6	Considerações Finais sobre o modelo . . . . .	68
<b>5</b>	<b>Método DEP-DST aplicado no domínio de RAM</b>	<b>69</b>
5.1	Formação do Banco de dados - Conhecimento, Informações Contextuais e Semiestruturadas . . . . .	71
5.2	Materiais utilizados na implementação dessa tese . . . . .	72
5.3	FIMS - Processo de Fusão de Dados . . . . .	75
5.3.1	Nível 0 - Sinal e Pré-processamento . . . . .	75
5.3.2	Nível 1 - Refinamento de Objeto . . . . .	76

5.3.3	Nível 2 - Refinamento da situação . . . . .	77
5.3.4	Nível 3 - Avaliação de Impacto . . . . .	78
5.3.5	Nível 4- Refinamento do Processo - Calcula a Probabilidade de RAM e Aplica IC . . . . .	81
5.3.6	<b>Nível 5 Gerenciamento de Processos e Interface Homem-Computador (IHC)</b> . . . . .	84
5.4	Explicações sobre o a geração de BPA . . . . .	85
5.5	Ameaças à validade . . . . .	87
5.6	Considerações Finais sobre a implantação do modelo no domínio RAM .	88
<b>6</b>	<b>Resultados</b>	<b>90</b>
6.1	Resultados no domínio de RAM . . . . .	90
6.2	Benchmark com outros trabalhos no domínio de RAM . . . . .	92
6.3	Benchmark com base de domínio de deonças cardíacas . . . . .	92
6.4	Considerações Finais sobre os resultados . . . . .	95
<b>7</b>	<b>Conclusão</b>	<b>97</b>
7.1	Contribuições . . . . .	99
7.2	Trabalhos futuros . . . . .	99
7.3	Limitações . . . . .	100

## Lista de Figuras

2.1	Métodos comuns de fusão de dados - inspirado em . . . . .	14
2.2	Arquitetura focada na gestão e em usuários- inspirado em DFIG . . . . .	16
3.1	Critério de busca nas bases de artigos . . . . .	25
3.2	Análise de sentimento de polaridade X classificador SVM/NB adaptada de . . . . .	28
3.3	Dados espaço-temporais adaptado de . . . . .	41
3.4	Evolução de publicações FIMS com DST . . . . .	45
4.1	DEP-DST Modelo de geração de evidências DST com ajuste de depen- dências . . . . .	58
4.2	Esquema do Índice de Credibilidade . . . . .	59
4.3	Amostragem no DAG Bayesiano . . . . .	60
4.4	Simulação de teste de Pearson para análise de correlação . . . . .	62
4.5	Gerar variáveis componentes do IC . . . . .	67
4.6	Gerar a evidência baseada na técnica Random Forest . . . . .	68
5.1	Formação do banco de dados . . . . .	71
5.2	Mapa Mental - RAM a partir de RES . . . . .	73
5.3	Modelo DEP-DST aplicado ao domínio RAM . . . . .	74

5.4	Gráfico Acíclico Dirigido - DAG Bayesiano . . . . .	79
5.5	Regressão Linear - análise temporal do impacto do medicamento em exame	80
5.6	Estudo comparativo - Benchmark interno . . . . .	83
5.7	Shapley impacto global para RAM . . . . .	85
5.8	Impacto do valor SHAP - exemplo RAM: sangramento . . . . .	86
5.9	SHAP - Perspectiva do paciente . . . . .	87
5.10	SHAP Resumo da perspectiva do paciente . . . . .	88
5.11	Front end - ilustração de um diagnóstico de RAM . . . . .	89
6.1	Distribuição do Índice de Credibilidade . . . . .	90
6.2	Benchmark 1 - Modelo Base . . . . .	92
6.3	Benchmark 2 - Modelo Base + IC . . . . .	93
6.4	Benchmark 3 - Modelo Base + Índice de Credibilidade + componente inferência bayesiana . . . . .	94
6.5	Descrição do conjunto de dados Heart Disease- UCI Cleveland . . . . .	95
6.6	Estudo comparativo-evolução da precisão com o incremento gradual de variáveis do IC . . . . .	96

## Lista de Tabelas

3.1	Relação de estudos por categoria de extração de RAM . . . . .	27
3.2	Ferramentas e funções relevantes na extração de RAM . . . . .	29
3.3	A evolução da pesquisa RAM pelo modelo de aprendizado . . . . .	34
3.4	Experimentos Ensemble Learning para identificar RAM . . . . .	34
3.5	Dicionários e Conjuntos de Dados usados no domínio RAM . . . . .	39
5.1	Ferramentas bibliotecas utilizadas nessa tese . . . . .	72
5.2	DAG Probabilidades de Reação a Medicamentos . . . . .	78
5.3	DAG Probabilidades Doença-Sintoma . . . . .	78
6.1	Análise de Benchmark RAM . . . . .	91
6.2	Análise de Benchmark - Externa . . . . .	93
6.3	Análise de Benchmark aplicado ao domínio doenças do Coração . . . . .	95

## **Lista de Nomenclaturas**

<b>Sigla</b>	<b>Significado</b>
Entropia	O estudo que originou a entropia de Shannon e serviu de base para novos trabalhos, como o de [Gordon and Shortliffe, 1984], que institui conceito de ganho de informação atribuindo importância e prioridade ao tratamento de inconsistências e imperfeições nos dados. O trabalho de [Deng, 2016] demonstra que esforços recentes ainda são empenhados nessa linha de pesquisa.
Filtragem de partículas	técnica usada em fusão de dados envolvendo sensores, desde o início da pesquisa em fusão de dados, na década de 1980. O uso desse método cresceu muito com o advento de IOT e rede de sensores sem fio. A técnica trata amostras, pela necessidade de tempo real, otimiza a análise de grande volume de dados. Inspirou métodos como filtro de Kalman e foi aperfeiçoada com MCMM – cadeia de Markov.
Mg/Dl	Miligrama por decilitro de sangue. Medida utilizada em exames e testes laboratoriais.
OODA	consiste em um ciclo "observar-orientar-decidir-agir", desenvolvido pelo estrategista militar e coronel da Força Aérea dos Estados Unidos John Boyd . Boyd aplicou o conceito ao processo de operações de combate , muitas vezes no nível operacional durante as campanhas militares. Agora também é frequentemente aplicado para entender as operações comerciais e os processos de aprendizado.
OMNIBUS	Um modelo de fusão de dados que é composto por um fluxograma em torno do natureza cíclica do ciclo OODA de inteligência e o Boyd malha de controle, porém tem a capacidade de funcionamento de um loop dentro de outro loop. Usa as definições mais refinadas do Modelo JDL.
TBM	Transferable Belief Model - modelo que estimulou o uso da DST em processos de FIMS. Criado por Smets em 1994, a TBM é uma iniciativa mais ligada ao cálculo de pesos, não somente para estabelecer descontos pela incerteza, mas também para viabilizar a transformação pignística, importante aliada no uso da DST em inteligência artificial.

## **Lista de Símbolos e Unidades**

<b>Sigla</b>	<b>Significado</b>
AD	Árvore de decisão - um algoritmo de aprendizado de máquina supervisionado que é utilizado para classificação e para regressão, com raiz e folhas, em que o raiz é um dos atributos da base de dados e o nó-folha é a classe ou o valor que será gerado como resposta
BPA	do inglês - Basic probability assignment - atribuição de probabilidade básica Atribuir uma massa de probabilidade a um subconjunto, o significa que se coloca esse nível de confiança a esse subconjunto
DAG	Do inglês Directed Acyclic Graph - gráfico acíclico dirigido
DST	Dempster and Shafer Theory - Teoria Dempster and Shafer ou Teoria da Evidência
EHR	Do inglês Eletronic Healht Records - Registros eletrônicos de saúde(RES) - equivalente a prontuário eletrônico
FIMS	Fusão de Informação Múltiplos Sensores ou Multissensor
IC	Índice de Credibilidade - parte da contribuição científica dessa tese
IOT	Do inglês Internet of Things - Internet das coisas
JDL	Do inglês Join Defense Lybrary - método pioneiro de fusão de dados, criado para aplicações militares, que vem recebendo contribuições até os dias de hoje, permitindo o uso de modelos adaptados para o uso em gestão.
LSTM	Do inglês Long Short Long Short Term Memory - Redes de Memória de Curto e longo Prazo arquitetura de redes neurais utilizada em métodos de aprendizado profundo.
MCMM	Monte Carlo via Cadeias de Markov - O uso combinado de simulação de Monte Carlo com cadeias de Markov servem basicamente para gerar valores de uma distribuição, com valores correlacionados
NN	Do inglês Neural Network - rede neural(RN) - são sistemas de computação com nós interconectados que funcionam como os neurônios do cérebro humano. Usando algoritmos, elas podem reconhecer padrões escondidos e correlações em dados brutos, agrupá-los e classificá-los, e – com o tempo – aprender e melhorar continuamente.
PLN	Processamento de Linguagem Natural - técnica de inteligência artificial muito usada em extração e beneficiamento de dados textuais, de imagens, vídeos ,dentre outras
RAM	Reação adversa a medicamento qualquer resposta prejudicial ou indesejável, não intencional, a um medicamento, que ocorre nas doses usualmente empregadas no homem para profilaxia, diagnóstico, terapia da doença ou para a modificação de funções fisiológicas.
RES	Registros eletrônicos de saúde que formam um prontuário eletrônico, contendo dados pessoais e demográficos do paciente, medicamentos prescritos, exames realizados, tratamentos e diagnósticos em andamento, dentre outras informações relevantes ao estado do paciente.
SVM	Do inglês Support Vector Machine é um algoritmo de aprendizado de máquina supervisionado que pode ser usado para desafios de classificação ou regressão. Seu foco maior é no treinamento e classificação de um dataset.

# 1. Introdução

Este capítulo descreve o que motivou esta pesquisa, as questões a respeito do problema, os objetivos da investigação e a metodologia de pesquisa para alcançá-los.

## 1.1 Motivação

Quanto mais você souber sobre um problema, melhor decidirá o que fazer a respeito. A incerteza está presente em quase todos os ramos de atividade, produzindo fenômenos estudados em várias ciências e gerando inquietação. O filósofo [Bauman, 2013] define que vivemos em um mundo líquido, onde quase nada é certo.

A oferta de dados nos dias de hoje é muito farta e variada, sendo estes dados multi-fontes e multimodais. A Internet das Coisas (IoT) conecta mais de nove bilhões de dispositivos, incluindo a conectividade de coisas inteligentes possibilitando novas interações e interoperações entre coisas e pessoas [Gite and Agrawal, 2016].

Num mundo fortemente digitalizado em que vários sensores coletam dados do ambiente, é possível integrar uma grande quantidade de informações para gerar conhecimento. No entanto, é preciso lidar com dúvidas, imprecisões, eventos com baixas chances de ocorrência e outliers, independentemente do domínio. Juntos, esses aspectos representam incertezas, que devem ser adequadamente consideradas na hora de dar sentido aos dados para que a tomada de decisão possa ser confiável. Dependendo do contexto, a solução não é restringir tais dados imprecisos, pois cortar esse tipo de informação pode não apenas gerar erros ou perda de conhecimento, mas provocar um grave acidente ou até mesmo a perda de uma vida.

Todas essas questões estão no escopo da fusão de dados (data fusion), definida em [Castanedo, 2013] como a integração de dados e conhecimento de múltiplas fontes. Na fusão de informação multisensores (FIMS) do inglês multisensor information fusion (MSIF),

os dados de diferentes sensores são combinados para fornecer uma descrição completa de uma situação de interesse [Durrant-Whyte and Henderson, 2016]. A FIMS permite que decisões sejam tomadas com uma compreensão mais precisa do contexto e com menor incerteza [Gite and Agrawal, 2016]. De acordo com [Steinhauer and Karlsson, 2019], a FIMS permite uma melhor análise de dados ao possibilitar: a) o aumento da precisão das informações usando várias fontes de dados; b) a mitigação da incerteza para fornecer uma base aprimorada para a tomada de decisões e c) a redução da dimensionalidade de dados relevantes para fornecer uma visão facilmente compreensível de um domínio.

A FIMS enfrenta vários desafios para estruturar informações devido a problemas encontrados nos dados e sensores. Os dados podem sofrer com conflitos e correlações, e os sensores impõem dificuldades técnicas como imprecisões, frequência de captura e redundância [Esteban et al., 2005], [Castanedo, 2013], [Khaleghi et al., 2013]. De acordo com [Meng et al., 2020], o FIMS é um processo dinâmico. Isso significa que a complexidade da fusão de dados é causada não apenas pela diversidade de tipos de dados e pelo ambiente de coleta, mas também pelas *dependências temporais entre fatos e evidências*. Além disso, aplicações em tempo real envolvem dados que têm efeito significativo apenas por um período limitado. É o desafio da dinamicidade, onde eventuais ocorrências duram por um tempo limitado. Se os mecanismos de periodicidade de captura não estiverem bem configurados, pode-se perder informações relevantes. Para alguns sensores, a coleta de dados deve ser realizada de minuto em minuto. Para outros, de 10 em 10 minutos. Há outros sensores onde é necessário se calibrar a captura para que ocorra a cada hora. Distinguir a ordem correta dos fatos é crucial para extrair o significado preciso das informações [Meng et al., 2020]. Um sensor que analisa a droga injetada no paciente sob o ponto de vista de avaliar o comportamento de um determinado parâmetro sanguíneo deve estar sincronizado com o mecanismo que analisa efeitos temporais em tais marcadores sanguíneos, como glicose, ureia, creatinina dentre outros.

Para lidar com o desafio da resolução de incertezas durante a fusão de dados, as abordagens principais são a teoria da probabilidade Bayesiana, a teoria Dempster-Shafer (DST) e o raciocínio fuzzy [Verbert et al., 2017]. O raciocínio fuzzy, é uma abordagem baseada em computação que imita o raciocínio humano, definindo funções ou graus de pertinência que produzem possibilidades de valores intermediários. A lógica fuzzy pode ser embarcada em hardware, amplamente utilizada em sistemas de controle e inteligência artificial. A teoria da probabilidade contém duas interpretações de probabilidade, a bayesiana e a frequentista. A abordagem frequentista usa apenas dados, enquanto o raciocínio bayesiano usa dados para aumentar a crença inicial. Ao combinar dados e crença inicial, tem-se a probabilidade condicional, em que o modelo se beneficia e uma boa quantidade

de conhecimento prévio fica disponível [Yarnell et al., 2021]. A teoria DST foi criada para lidar com incertezas e informações incompletas ao permitir partições de combinação de evidências de diferentes fontes e computar um grau de crença chamado função de crença que leva em conta todas as evidências disponíveis [Shafer, 1976]. A DST é um raciocínio amplamente usado em diagnóstico de falhas, diagnóstico médico, sistemas de decisão e análise de risco. Ao contrário do framework Bayesiano, o DST permite a atribuição de crença a conjuntos de elementos no domínio. Por outro lado, o modelo Bayesiano atribui crença apenas a elementos individuais. Apesar de raciocínios distintos, são complementares e nesta tese, usou-se uma rede Bayesiana para realizar inferências que suportam crenças em um contexto DST.

## 1.2 Definição do Problema de Pesquisa

A integração de FIMS com ML incrementou o uso de classificadores conjuntos (ensemble learning) o que vem contribuindo para a melhora dos resultados em relação aos classificadores tomados de forma isolada. Mas [Florea and Bossé, 2009] reforçam a importância que deve ser dada também à análise de dados com qualidade, para o bom uso da DST. O estudo afirma ser precária utilização de análise de dados sob o ponto de vista temporal e causal, não reconhecendo eventuais dependências entre as fontes e sensores. [Pei et al., 2017] observa a importância de se analisar aspectos temporais no pré-processamento. O conhecimento do precário uso da análise de dados sobre a dependência entre fontes e sensores também foi observado por [Mao et al., 2020], que destacam a importância de analisar essas relações.

Para entender melhor a questão da incerteza, vai-se supor um exemplo fictício em que a incerteza está presente, mas não é prontamente perceptível. Maria foi ao hospital porque estava sentindo dores abdominais e reclamando de sangramento digestivo. Havia dois médicos de plantão, João e Paulo. Maria é uma paciente com câncer ginecológico em estágio avançado, que pode apresentar sangramento genital, e desde a última avaliação médica vem se queixando de dores abdominais. Ela está tomando medicamentos anti-inflamatórios fortes há um mês e voltou ao hospital com um pequeno sangramento digestivo. João e Paulo inicialmente discutiram o caso. João considera que o sangramento digestivo foi devido ao medicamento anti-inflamatório com 80% de chance, atribuindo os outros 20% de chance de ser outra coisa. Paulo discorda de João. Paulo acredita que, com 70% de chance, o sangramento digestivo foi causado pela evolução natural do câncer. Ele não eliminou a possibilidade de ser uma reação adversa a medicamentos (RAM) à opção do anti-inflamatório, mas apenas atribuiu a ela 20% de chance de ser verdadeira.

Além disso, ele considera 10% de chance de ser outra coisa. Essas são duas opiniões conflitantes para o diagnóstico. Esse tipo de situação é o dilema clássico da DST.

Ambos os especialistas concordaram que era necessária mais investigação e decidiram fazer um estudo mais aprofundado analisando notas clínicas, prescrições médicas mais recentes e estudo farmacológico detalhado considerando interações medicamentosas. Além disso, solicitaram aos técnicos que elaborassem (i) uma análise de todos os medicamentos prescritos ao paciente, (ii) a relação entre esses medicamentos e possíveis causas de RAM, (iii) outra análise com possibilidades de evolução da doença e (iv) análise de possíveis tendências nos parâmetros sanguíneos em testes de laboratório feitos no ano anterior. Por fim, solicitaram ajuda à equipe para fazerem inferências com probabilidades relacionadas aos sintomas e à doença do paciente que também permitiu análises estatísticas sobre medicamentos prescritos e possíveis RAM.

Algo chamou a atenção dos médicos especialistas ao analisarem a evolução de uma enzima hepática nos últimos três exames da paciente. Quando os examinaram com mais detalhes, identificaram um cenário de metástase no fígado. A variação apresentada na tendência de crescimento do valor dessa enzima estabelece uma relação de dependência com a evolução da doença. Um aumento nessa enzima provavelmente causa sangramento digestivo. Esse fenômeno é um sintoma da metástase da doença para o fígado, então praticamente elimina ou reduz substancialmente a crença no efeito adverso do anti-inflamatório.

Na prática, João e Paulo não tinham muito tempo e não havia como organizarem esses dados rapidamente. Os especialistas João e Paulo não conseguiram obter estudos comparativos com análise histórica, avaliando RAM e casos de evolução da doença em pacientes com aquele tipo de diagnóstico. Após todo o esforço, ambos os especialistas chegaram a um consenso sobre sangramento digestivo. No entanto, o anti-inflamatório possivelmente potencializador de tais sintomas teria como hipótese mais plausível o diagnóstico de progressão da doença para o fígado, reduzindo fortemente a crença em um episódio de RAM.

As hipóteses de João e Paulo mudaram: João para 10% de RAM anti-inflamatória, Paulo para 90% de evolução da doença e o outro com 5% para RAM e 95% para evolução da doença, o que, quando consolidado, gera um resultado de 92% de evolução da doença, a hipótese de RAM 5% e 3% de incerteza. Esse exemplo demonstra que, quando uma reação adversa ocorre num paciente, com um caso semelhante, nem sempre irá ocorrer em outro. Somos seres biopsicossociais diferentes [Karunamuni et al., 2021] e as evidências para identificar RAM são influenciadas por vários detalhes e aspectos dependentes.

Diante deste exemplo, percebe-se características temporais e causalmente dependentes podem ser detectadas e ajudar a obter melhores resultados de diagnóstico. Todavia, apesar da evolução dos processos FIMS, com o uso de ML concomitante com estimadores como DST para o tratamento da incerteza, poucos estudos utilizam de análise das relações entre dados e sensores antes do processo de FIMS.

A DST foi projetada para lidar com conflitos entre opiniões de especialistas. As opiniões devem ser independentes e vir de fontes diferentes sem qualquer dependência [Shafer, 1976]. Há um interesse crescente no uso da DST em FIMS, principalmente devido à integração de FIMS com ML. Essa integração reforça a importância de estimar crenças para usar como entrada para a função de atribuição de probabilidade básica (do inglês basic probability assignment (BPA)) a partir de conjuntos de dados. Conseqüentemente, um tratamento estatístico robusto é essencial no pré-processamento dos dados em modelos FIMS com DST para melhorar a qualidade da geração de BPA, seja no uso de métodos de contribuição probabilística [Denoeux, 2020], ou em abordagens com aprendizado de máquina [Denoeux, 2019]. Esta tese vai endereçar este problema conforme detalhado na seção seguinte.

### **1.3 Objetivos da Pesquisa**

Em situações de diagnóstico há várias fontes de dependência e variabilidade que afetam diretamente os conjuntos de dados e, portanto, reduzem a qualidade das evidências geradas a partir de tais conjuntos. O objeto de estudo desta pesquisa é obter um melhor forma de geração de BPA para o uso de DST em FIMS, por meio da identificação e saneamento de dependências entre evidências constatadas por análises de correlação, inferência causal e relações de dependência probabilística e temporal entre as fontes de dados. Assim, podem ser estabelecidas relações de causa e efeito e identificadas relações de dependência e causais identificadas a partir de análises de temporalidade. A hipótese dessa tese é que melhores resultados em termos de precisão e menor incerteza são alcançados quando os impactos das relações de dependência de dados em um ambiente FIMS são adequadamente identificados.

O objetivo principal dessa pesquisa é a identificação e saneamento de dependências entre dados e fontes de dados, melhorando a precisão e diminuindo a incerteza em decisões de diagnóstico.

Os objetivos secundários da pesquisa são: 1) a construção de um método que estabelece um IC projetado especificamente para uso na geração de evidências em FIMS com

DST. 2) a aplicação do método no domínio de RAM no intuito de melhorar a precisão e reduzir a incerteza nos resultados preditivos e de diagnóstico .

#### 1.4 Metodologia da Pesquisa

Para o desenvolvimento do modelo FIMS proposto nesta tese, seguiu-se modelo de arquitetura de FIMS proposto por [Blasch and Plano, 2005] no DFIG (Data Fusion Information Group) que privilegia o gerenciamento da informação e a participação do usuário gestor. O DFIG promove a evolução do modelo seminal do JDL (Joint Defense Laboratory) e defende a aplicação de FIMS em domínios de gestão do mundo real, não apenas nos domínios da robótica ou da automação militar, utilizando métodos com raciocínio baseado em evidências, como DST.

A estratégia de pesquisa passa pela construção de um modelo utilizando aspectos de inferência Bayesiana, correlação e análise temporal para mitigar os conflitos de dependência decorrentes do conjunto de dados, criando um método estruturado que define um índice de credibilidade(IC) projetado para gerar evidências com maior precisão em um ambiente FIMS com DST. Neste modelo, é realizada uma análise de dados robusta no pré-processamento, utilizando técnicas de correlação, inferência causal e análise de séries temporais. Para a geração da BPA com machine learning, são empregadas as técnicas *random forest*, rede bayesiana, correlação de Pearson e análise de regressão linear. Uma contribuição adicional será a utilização desse modelo para o estudo de caso de RAM em uma entidade de saúde pública, onde o modelo será aplicado.

A criação do IC reúne variáveis relacionadas à inferência Bayesiana, correlação e análise temporal, num contexto de fontes multissensoriais e interação entre fatores. As evidências que geram as probabilidades que são atribuídas às hipóteses para a formação da BPA são oriundas de vários eventos de leitura de sensores. Essas evidências compõem a geração do índice de credibilidade, que visa a uma melhor precisão na configuração do diagnóstico.

Os resultados obtidos mostram que o IC proposto aumenta significativamente a qualidade das evidências geradas a partir das técnicas clássicas. Os resultados foram medidos em diferentes cenários usando-se várias métricas e modelos de aprendizagem para mostrar a robustez do método proposto. Esse benchmark utilizou três conjuntos de dados: um com probabilidade frequentista (método base), outro com IC adicionado ao método base e um terceiro adicionando-se os resultados da rede Bayesiana ao segundo modelo. O segundo método permitiu um ganho de 15 % na precisão e o terceiro obteve uma pre-

cisão 21 % maior. Ganhos similares foram obtidos na métrica de cobertura. Foi realizado um teste com conjunto de dados externo, um tradicional de doenças cardíacas, obtendo resultados satisfatórios.

O modelo foi aplicado no domínio de RAM, onde prevalece a incerteza diagnóstica, alcançando uma precisão e revocação significativamente melhores em um experimento que utilizou dados de prontuários eletrônicos- registros eletrônicos de saúde (RES), do inglês Electronic Health records (EHR) coletados de um hospital público federal de câncer. Esses RES reuniram dados de mais de 50 mil pacientes ativos e 900 mil prontuários e 2 milhões de prescrições em 2019, das quais 63 mil possibilidades de RAM foram evidenciadas em notas clínicas envolvendo 5.937 pacientes.

## **1.5 Organização da Tese**

Uma vez que já se conheciam as vantagens e desvantagens entre as fontes de dados MSD e RES, precisaria-se decidir por qual caminho a pesquisa iria fluir. O desafio da limpeza, do pré-processamento, desafios de captura, limitações impostas cada vez mais por razões de privacidade foram apontados como limitadores. A utilização de RES se mostrou um caminho mais oportuno para a realização da pesquisa e havia um interesse da FIOCRUZ pelo projeto disponibilizando seus prontuários eletrônicos para a elaboração da pesquisa. Entretanto surgiu em março de 2020 a pandemia do corona vírus e a FIOCRUZ concentrou seus esforços de pesquisas sobre a COVID. O projeto foi então apresentado ao INCA que autorizou o projeto de pesquisa. O restante desta tese está organizado em mais 6 capítulos:

- Capítulo 2 – Fundamentação teórica: teoria da evidência, redes bayesianas e séries temporais. Fundamentos sobre o estudo de RAM.
- Capítulo 3 – Revisão de literatura - aborda os métodos de identificação automatizada de RAM e busca a resposta às questões dessa pesquisa.
- Capítulo 4 – Modelo - apresenta a proposta de contribuição da criação do índice de credibilidade para uso em modelo de FIMS com DST.
- Capítulo 5 – Método aplicado no domínio de RAM e no domínio de doenças cardíacas.
- Capítulo 6 – Resultados - Aborda os resultados obtidos no domínio de RAM e no domínio de doenças cardíacas (base de Cleveland).

- Capítulo 7 – Conclusão - Realiza o fechamento dessa tese, apontando limitações e trabalhos futuros.

## **2. Fundamentação Teórica**

Neste capítulo, são explorados os fundamentos teóricos e da aplicação que será posteriormente usada como modelo para comprovação.

### **2.1 Contextualização sobre RAM**

De acordo com o estudo de [Dev et al., 2017a] as reações adversas a medicamentos (RAM) correspondem a uma grande preocupação na área da saúde, estando entre as principais causas de morbimortalidade - Food and Drug Administration - (FDA). Apesar de ser conhecido o problema de subnotificação de RAM, o FDA revela que entre 2006 e 2014 houve um aumento de 230% na notificação de casos de RAM como causadora de mortalidade. O estudo de [Banda et al., 2016]) afirma que o custo anual de morbidade e mortalidade relacionada a medicamentos foi estimado em cerca de US\$ 170 bilhões e que as estimativas anuais de casos relacionados a RAM ultrapassam 1,9 milhão de visitas clínicas anuais nos EUA. A farmacovigilância é uma das áreas da vigilância em saúde e agrega processos que permitem o rastreamento de eventos relacionados à saúde humana [Calix et al., 2017]. Essa área monitora e rastreia o uso seguro de produtos farmacêuticos, monitorando RAM que podem ser causadas por medicamentos e outros materiais relacionados à saúde. Extensos ensaios e testes clínicos são necessários antes que um medicamento seja lançado no mercado e não é uma tarefa fácil descobrir todos os efeitos colaterais de um medicamento aprovado [Peng et al., 2016a]. A FDA monitora ativamente os medicamentos aprovados para identificar eventos adversos por meio de sistemas de notificação. O FDA Adverse Event Reporting System (FAERS), Sistema de Notificação de Eventos Adversos da FDA, é o principal banco de dados sobre RAM nos EUA. É alimentado por consumidores e fornecedores.

A subnotificação de RAM é um problema mundial. Se faz necessária a utilização de fontes alternativas para identificação de RAM. Os testes e ensaios clínicos para o lança-

mento de uma droga são caros e têm de ser realizados em curta duração. A ocorrência de RAM gera desconforto para pacientes e profissionais de saúde no que concerne a realizar o registro desse tipo de eventos. O estudo de [Liu and Chen, 2015] afirma que o problema da subnotificação de RAM é um fenômeno mundial que dificulta a avaliação efetiva do lançamento de novos medicamentos. [Huang et al., 2017] avaliaram como limitada a segurança dos medicamentos com apenas ensaios clínicos feitos por laboratórios, devido a amostra ser pequena, havendo pouco tempo para monitorar os efeitos nocivos. De acordo com [Mishra et al., 2015], fontes alternativas de registro de RAM precisam ser trabalhadas. Esse estudo de [Mishra et al., 2015] realizou a coleta, detecção e monitoramento de informações sobre RAM a partir de prontuários eletrônicos - registros eletrônicos de saúde, do inglês electronic health records (EHR). Os autores constataram o uso crescente de prontuários eletrônicos e concluíram que há também uma imensa variedade de blogs, comunidades e locais onde os pacientes compartilham dados sobre saúde, inclusive sobre RAM.

As mídias sociais são fontes onipresentes de informação para vários tópicos, compartilhando eventos diários, reclamações pessoais, consumo de drogas e RAM. São acessíveis e já estão inseridos no cotidiano da população, como no Twitter, DailyStrength e MedHelp [Mahata et al., 2018].

[Liu et al., 2016] definem que a mídia social digital (MSD) constitui fontes de dados pouco exploradas na identificação de RAM. No entanto, reconhecem ser uma difícil tarefa dada a alta dimensionalidade das informações não estruturadas. Nas notas clínicas, os profissionais de saúde em geral escrevem em formato livre, usam abreviaturas, expressões, siglas e termos de jargão. Os autores afirmam que características lexicais e semânticas podem aumentar a capacidade de identificação, melhorando a eficácia da extração de RAM.

Mesmo considerando se tratar de informações inseguras o estudo de [Comfort et al., 2018] argumenta haver interesse crescente em utilizar informações extraídas do MSD sobre RAM. E destaca como um dos desafios na identificação de RAM a necessidade de rotulagem e anotações de RAM, atividade que depende de especialistas. Devido à expansão de dados públicos massivos o estudo de [Peng et al., 2016a] define que redes sociais abrangentes como o Twitter permitem a captura de informações adicionais sobre RAM, por se poder agregar recursos computacionais, como análise de sentimentos, classificadores e técnicas de processamento de linguagem natural(PLN).

A necessidade de melhorias no pré-processamento das postagens nas redes sociais, foi enfatizada pelo estudo de [Liu and Chen, 2015] que observou que após a extração das

mensagens que relacionam medicamentos a sintomas, é necessário realizar um grande descarte de mensagens que se referem a estudos sobre RAM, notícias sobre RAM, propagandas sobre medicamentos que mitigam RAM, mas que efetivamente não estão relacionadas a comentários sobre RAM. A importância do pré-processamento para remoção desse tipo de ruído cresce com o volume de dados gerados em mídias sociais como o Twitter. [Calix et al., 2017].

## 2.2 Introdução e Histórico

O uso da FIMS vem se incrementando com a associação de técnicas de inteligência artificial. A FIMS é uma ferramenta para analisar e usar os conjuntos de dados disponíveis, tornando-se uma parte essencial dos processos de mineração de dados e aprendizado de máquina [Beddar-Wiesing and Bieshaar, 2020].

Um dos principais objetivos da FIMS é trabalhar dados com menor probabilidade de erros e maior confiabilidade, utilizando dados de múltiplas fontes [Blasch and Plano, 2005]. A partir de revisões de modelos de arquitetura de FIMS, os trabalhos de [Esteban et al., 2005], [Steinhauer and Karlsson, 2019] e [Blasch and Plano, 2005] ressaltam a importância da gestão e da estimação.

Para a resolução de conflitos, imperfeições e incertezas em FIMS, os estudos concentram os esforços em entropia, no uso de pesos e descontos e utilizam mapeamento de distâncias e correlações entre as evidências e sinais dos sensores.

O estudo de [Florea and Bossé, 2009] realizou uma revisão dos conceitos de critérios para estimar confiabilidade: estimativas, distâncias e similaridade. Abordou também a análise temporal no comportamento dos sensores, comparando diversas métricas. Classificou os sensores em simples, complexos e híbridos, alegando a relevância de tratá-los de forma distinta.

É notório o crescimento a partir de 2015 dos trabalhos de FI utilizando ML combinado com DST. O forte uso de ML e DST promove altas taxas de acurácia e amplia o leque de possibilidades de geração da BPA, sendo utilizadas várias técnicas e combinados de algoritmos(ensemble).

Nesta pesquisa serão utilizados fundamentos de FIMS, recursos de processamento de linguagem natural (PLN), séries temporais, inferência causal, técnicas de machine learning (cluster, SVM e redes neurais) e teoria da Evidência - DST.

## 2.3 Fusão de Informação

Atualmente os métodos comuns de fusão de dados estudam o valor dos dados sob diferentes perspectivas. O método de fusão de dados mais simples é a combinação direta de dois conjuntos de dados unidimensionais quando eles apresentam um significado comum. Com base na compreensão semântica existente nos dados de alta dimensionalidade disponíveis em formato de texto, a FIMS também pode ser conduzida a partir da semântica dos dados. Diferentes métodos de fusão de dados podem trazer resultados distintos de otimização para o modelo de aprendizado de máquina. O trabalho de [Liu et al., 2020] dividiu métodos de fusão de dados em três categorias como mostrado na figura 2.1. Nota-se uma evolução dos métodos de fusão ao longo do tempo. Os métodos baseados em características vem recebendo novas contribuições com o incremento de uso de aprendizado profundo. Os métodos baseados em estágios são os mais tradicionais, com a fusão sendo realizada em etapas, como a abordagem realizada neste trabalho. A fusão baseada em significado semântico requer avançadas técnicas de predição, com métodos estatísticos e probabilísticos, e recursos que possibilitam a transferência de aprendizado. A multimodalidade dos dados, em formatos de texto, gráficos, de imagens e vídeo permeiam essa evolução.

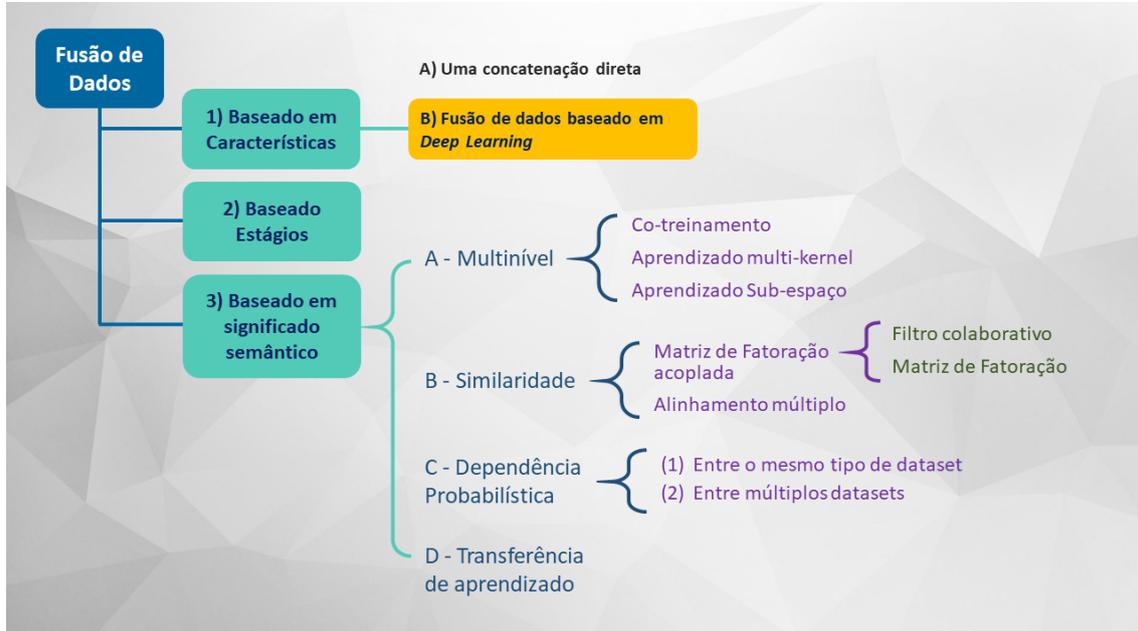


Figura 2.1: Métodos comuns de fusão de dados - inspirado em [Liu et al., 2020]

A recomendação de uso de métodos de probabilidade em FIMS pode ser vista em [Esteban et al., 2005] onde são referenciados os modelos de arquitetura de FIMS, desde o JDL até o modelo cascata, passando pelo modelo omnibus. Os autores propõem uma

abordagem baseada em sistema, num modelo que privilegia a estimação entre duas outras etapas, identificação e validação.

O estudo de [Liu et al., 2020] alerta que no método de combinação direta de recursos, sem os cuidados de verificação de compatibilidade para essa combinação, vários problemas precisam ser observados. Inicialmente ao se mesclar diretamente os dados, é necessário remover recursos duplicados. Após isso, deve-se cuidar de algumas características de diferentes dimensões, que fazem com que o modelo tenha bom desempenho, que podem ser perdas devido à fusão direta. Finalmente, a fusão direta de recursos pode resultar num viés no modelo (overfitting).

O método de fusão de dados baseado em etapas [Zhu et al., 2018] divide o problema em diferentes estágios. Por meio dos dados de cada estágio, as saídas de cada estágio do processo são mescladas.

[Liu et al., 2020] , [Kumar and Daumé, 2011] utilizam o método de fusão de dados baseado em similaridade [Liu et al., 2020], [Zheng, 2015] [Katz et al., 2019] e [Shang et al., 2014] aplicam o método de fusão de dados baseado em dependência probabilística.

O método de fusão de dados baseado em dependência probabilística utiliza estrutura de grafos. Esse método toma dados diferentes como nós em primeiro lugar, e então a relação entre os dados, como causalidade, é medida pela borda, que pode ser direcionada ou não. Isso permite a construção de um gráfico, e alguns métodos de estudo da estrutura do grafo podem ser usados para fundir os dados.

A FIMS é baseada num processo de níveis adaptados envolvendo agregação, mineração de dados e integração de informações de múltiplas fontes, resultando num produto de informação de maior valor do que qualquer parte isolada [Blasch and Plano, 2005]. O grupo DFIG introduziu o modelo de fusão de usuário, conforme apresentado na figura 2.2 afirmando a importância dos resultados do FIMS para representar o conhecimento a partir de uma fusão de estimativas, bem como o raciocínio do conhecimento para gerenciamento e controle.

A FIMS privilegia uma visão com referências cruzadas, não só nos dados, mas também na apresentação dos dados, no método visual empregado na interface com o usuário. O método de fusão de dados baseado em multi-visualização estuda um objeto a partir do conhecimento de diferentes pontos de vista [Zheng et al., 2015].

## CENTRO DE FUSÃO DE DADOS

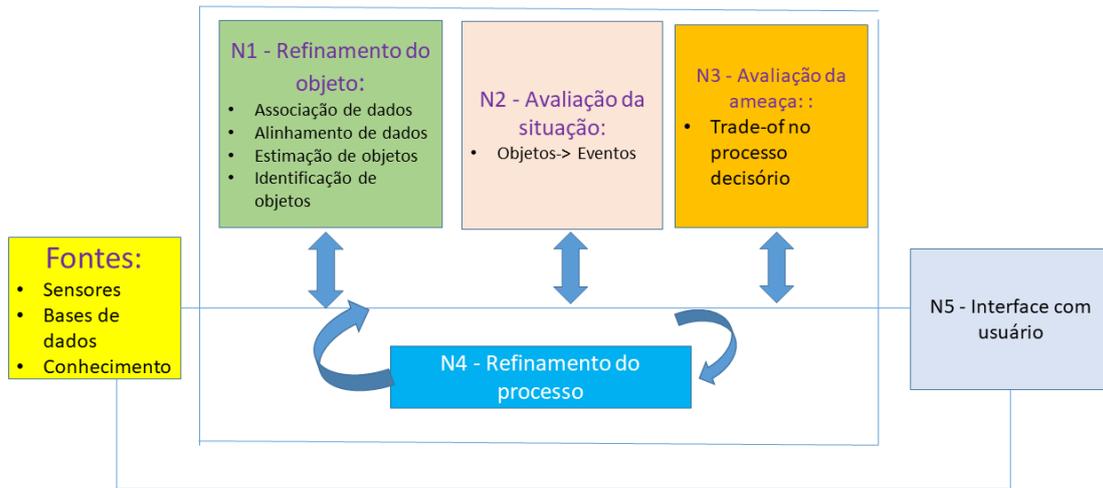


Figura 2.2: Arquitetura focada na gestão e em usuários- inspirado em DFIG [Blasch and Plano, 2005]

### 2.4 Rede Bayesiana

De acordo com a definição de [Russel and Norvig, 2003], as Redes Bayesianas (BN) são uma ferramenta eficaz de inferência Bayesiana capaz de representar achados incertos e incompletos, com base na probabilidade condicional e não no Teorema de Bayes. O raciocínio Bayesiano permite definir os efeitos de causalidade entre variáveis de um sistema. Uma relação ou causalidade entre variáveis é expressa por um grafo acíclico dirigido (do inglês directed acyclic graph (DAG)). Para variações discretas ou de conhecimento, se representa com o uso de tabelas de probabilidade condicional. Para variações contínuas, são usadas funções de densidade de probabilidade. Algumas das principais características de uma BN são satisfatórias para a propriedade de Markov local, na qual uma variável é condicionalmente dependente apenas de seus pais. Assumindo essa premissa, a cadeia de regras usada para calcular a probabilidade conjunta de uma variável pode ser simplificada para o seguinte:

$$P(X_1 \dots X_n) = \prod_{i=1}^n P(X_i | \text{parents}(X_i))$$

$X_i$  é uma variável aleatória numa rede Bayesiana e pais ( $X_i$ ) é um conjunto de todas aquelas que influenciam as variáveis.

## 2.5 Séries Temporais

Uma série temporal pode ser definida como uma coleção de observações feitas sequencialmente ao longo do tempo [Aminikhanghahi and Cook, 2017].

Em termos gerais, uma série temporal pode ser expressa como um conjunto possível de medidas infinitas ordenadas no tempo  $S$ , que consiste em pares  $(s_i, t_i)$  que se definem em pares medição do sensor  $s_i$  e instante do tempo  $t_i$ , de modo que:

$$S = \{(s_0, t_0), (s_1, t_1), \dots, (s_i, t_i) \dots\}, i, j \in \mathbb{N}^*$$

$$\forall t_i < t_j \text{ se } i < j$$

Sensores de medições  $s_i$  tomam valores num conjunto que depende das particularidades e características dos sensores. É comum que as séries temporais sejam definidas de forma mais simples por meio de medidas geralmente obtidas em intervalos de tempo iguais entre elas, sendo conhecidas como *série temporal regular*. Nesse caso, não é necessário referência à medida do tempo, sendo essa substituída por uma ordem de referência do índice regular, chegando-se a uma simplificação na expressão:

$$S = \{s_0, s_1, \dots, s_i, \dots\}, i \in \mathbb{N}^+$$

A expressão pode aparentar uma forma de equilíbrio estável, sendo nesse caso classificada como estacionária, ocasião em que a média é constante ao longo do tempo. Por outro lado, as séries temporais não estacionárias, não apresentam tal constância em relação à média ao longo do tempo. Logo, quando as propriedades estatísticas sofrem alterações ao longo do tempo, tem-se uma série não estacionária. O efeito não estacionário pode ser produzido em séries temporais por tendências, sazonalidade e correlações entre séries. A maioria dos métodos de previsão estatística baseia-se no pressuposto de que as séries temporais são aproximadamente estacionárias [Aminikhanghahi and Cook, 2017].

## 2.6 Teoria da Evidência - Dempster and Shafer Theory (DST)

A DST foi criada por Dempster em 1967 e aperfeiçoada por Shafer em 1976. A motivação está associada à dificuldade encontrada por esses pesquisadores na teoria da probabilidade para representarem ignorância e lidarem com necessidade das crenças atribuídas

a um evento e sua negação se complementarem, ou seja, somarem 1.

A DST promove um formalismo versátil e consiste numa extensão da teoria da probabilidade para descrever a incerteza na evidência. Na probabilidade clássica, pressupõe-se a existência de valores de probabilidade associados a determinados eventos, independentemente do observador poder saber o valor real da probabilidade. De forma distinta, a DST está centrada na credibilidade que se atribui ao fato de que um acontecimento ocorra sob determinado ponto de vista e conforme a experiência do observador.

Outro aspecto importante é o fato de a DST permitir que a evidência adquirida por meio de observações ou experimentos suporte várias conclusões mutuamente exclusivas de forma concomitante, entre elas nenhuma conclusão em particular, o que leva à atribuição de crenças ao componente de incerteza.

Na DST há um conjunto fixo de  $N$  elementos mutuamente exclusivos e exaustivos, chamado de quadro de discernimento. Seja  $\Theta$  um conjunto, indicado por  $\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_i, \dots, \Theta_N\}$ .

O conjunto de partes de  $\Theta$ , representado por  $P(\Theta)$ , é composto por todos os subconjuntos de  $\Theta$ . Isso envolve  $\emptyset$ , o conjunto vazio,  $P(\Theta)$ , e o próprio  $\Theta$ . o conjunto de composto de  $2^N$  elementos de  $\Theta$

$$P(\Theta) = \{\emptyset, \Theta_1, \dots, \Theta_N, \{\Theta_1, \Theta_2\}, \dots, \{\Theta_1, \Theta_2, \dots, \Theta_i\}, \dots, \Theta\}.$$

O elemento  $A$  representa qualquer um dos elementos que formam o conjunto de partes  $P(\Theta)$ . A massa  $m(A)$  representa o quão fortemente a evidência apoia  $A$ . Quando  $m(A) > 0$ ,  $A$  é chamado de elemento focal da função de massa.

Dada uma evidência, a massa atribuída a cada elemento de  $P(\Theta)$  equivale a um valor indicativo da crença a ele atribuído. A DST define essa função como massa  $m$ , que equivale a BPA e reúne as seguintes propriedades:

$m :$

$$P(\Theta) \rightarrow [0, 1], A \rightarrow [m(A)].$$

Tais propriedades satisfazem às seguintes condições:

$$m : 2^\Theta \rightarrow [0, 1] \quad (1)$$

$$m(\emptyset) = 0 \quad (2)$$

$$\sum_{A \in P(\Theta)} m(A) = 1 \quad (3)$$

- (1) Indica que a todos os subconjuntos de  $\Theta$  é atribuído um valor de crença entre 0 e 1;
- (2) Significa que a crença depositada no conjunto vazio é sempre zero; e
- (3) Que a soma de todos os valores atribuídos deve ser um.

Na DST uma função de massa também é chamada de atribuição de probabilidade básica (BPA). Supõe-se que existam duas BPA, operando em dois conjuntos de proposições B e C, indicados respectivamente por  $m_1, m_2$ ; eles serão então combinados pela soma ortogonal implementada na regra de Dempster.

A DST não requer que as probabilidades atribuídas a eventos sejam completas e precisas, e utiliza um conjunto de hipóteses em vez de hipóteses segregadas. A partir do acúmulo de evidências, a DST objetiva facilitar a atribuição da probabilidade de crença (BPA) às hipóteses.

É sobre esse quadro dinâmico, que acata mudanças, que se estabelece o raciocínio, sendo considerado o impacto das evidências coletadas.

O impacto da evidência deve ser observado não apenas nas hipóteses individuais originais, mas também nos agrupamentos dessas. Tais agrupamentos ainda são considerados hipóteses, pois são os subconjuntos de  $\Theta$ .

A função de massa  $m$  é semelhante a uma função de densidade de probabilidade da teoria da probabilidade. Mas tal função não contém a restrição Bayesiana de que a soma da crença atribuída aos subconjuntos unitários (hipóteses originais ou *singletons*) tem que somar um. Assim sendo, confirmar certa crença para um *singleton* não implica confirmar a crença restante para sua negação. Logo,  $m$  se comportaria como uma função de densidade de probabilidade, no caso de  $m$  atribuir valores diferentes de zero aos subconjuntos de unidades de  $\Theta$ .

Na DST, quando são obtidas novas evidências que suportam a crença numa hipótese, a crença que permanece até a unidade deve ser atribuída a  $\Theta$ , e não à negação da hipótese, como teria sido feito na teoria da probabilidade. Assim uma evidência sustenta parcialmente a crença numa hipótese. A evidência não precisa apoiar parcialmente a crença na negação da hipótese. Pelo contrário, a crença que não é atribuída como resultado de evidências deve ser compartilhada entre hipóteses mutuamente exclusivas.

O impacto de uma evidência suporta um subconjunto de hipóteses de  $\Theta$  que não é unitário. Isso equivale a afirmar que as evidências confirmam a crença nas hipóteses formando

subconjuntos. E não se faz necessário especificar o impacto na crença de cada uma delas. Assim é possível fortalecer a crença na incerteza e atribuir massa ao ambiente ( $\Theta$ ).

Com a BPA, a DST fornece uma forma de representar o impacto da evidência no Quadro de Discernimento. Logo, considerando as evidências disponíveis, a DST estabelece instrumentos para tentar determinar o grau de crença que pode ser depositado em cada hipótese. Isso permite que se possam ser especificadas várias métricas como os graus de crença, dúvida, verossimilhança, e intervalo de crença que será visto a seguir. Falhas ou omissão na identificação e saneamento de dependências entre essas evidências, impactarão em menor precisão e cobertura e num aumento da incerteza do diagnóstico. Um elemento focal é todo o subconjunto do quadro de discernimento cuja massa de atribuição de crença seja maior que zero, isto é  $A \in \Theta$  para o qual  $m(A) > 0$ . O grau de crença num elemento  $A$  de  $P(\Theta)$  é escrito como  $Bel(A)$  e representa a crença mínima na hipótese  $A$ , como resultado da evidência. Portanto, é definido como a soma das atribuições de probabilidade básicas feitas a todos os subconjuntos de  $A$ :

$$\forall A | A \in P(\Theta) : Bel(A) = \sum_{x \subseteq A} m(x)$$

É verdade que o grau de crença é sempre menor que o grau de plausibilidade. O intervalo entre o grau de crença e a plausibilidade de um elemento  $A$  de  $P(\Theta)$  se consiste no intervalo de crença em  $A$ . Esse intervalo é representado pelo par  $[Bel(A), Pl(A)]$  e representa o nível de incerteza sobre a hipótese  $A$ , como resultado de uma evidência.

Uma medida da incerteza permitida pela DST é estabelecida pela diferença entre  $Bel(A)$  e  $Pl(A)$ .

- Quando  $Bel(A)$  e  $Pl(A)$  apresentam valores iguais, tem-se absoluta certeza sobre o impacto da evidência na hipótese  $A$ ;
- A partir das evidências  $Bel(A)$  é 0 e  $Pl(A)$  é 1, a diferença entre as duas medidas é máxima e nada se sabe sobre o efeito em  $A$ ; e
- Quando os valores de  $Bel(A)$  e  $Pl(A)$  são diferentes, quanto maior a diferença entre os dois, maior é a incerteza sobre o impacto das evidências na hipótese  $A$ .

O quadro de discernimento da DST é formado por um processo iterativo para avaliar o impacto nas hipóteses (funções de massa) de evidências sucessivas. Nesse processo, a crença nas hipóteses adquiridas numa iteração ( $m_1$ ), em decorrência da consideração do

impacto das evidências, é combinada com a adquirida na iteração ( $m_2$ ), ao avaliar-se o impacto de novas evidências.

$$m_{12}(\emptyset) = 0$$

$$m_{12}(\Theta) = 1$$

$$\forall A | A \in P(\Theta) \wedge A \neq \emptyset \wedge A \neq \Theta :$$

A referida combinação ( $m_{12} = m_1 \oplus m_2$ ) é realizada usando-se a regra de combinação de Dempster:

$$m_{12}(A) = \frac{\sum_{B, C \in P(\Theta) \wedge B \cap C = A} m_1(B) * m_2(C)}{\sum_{B, C \in P(\Theta) \wedge B \cap C \neq \emptyset} m_1(B) * m_2(C)}$$

A referida combinação

( $m_{12} = m_1 \oplus m_2$ ) é realizada usando a regra de combinação de Dempster ou o que é equivalente para o último caso:

$$m_{12}(A) = \left( \frac{1}{1-k} \right) * \sum_{B, C \in P(\Theta) \wedge B \cap C = A} m_1(B) m_2(C)$$

Sendo o termo de conflito k

$$k = \sum_{B, C \in P(\Theta) \wedge B \cap C \neq \emptyset} m_1(B) m_2(C)$$

onde a constante k causa um efeito de normalização. Por tal efeito, a crença no conjunto vazio é eliminada, e é distribuída entre todos os outros elementos de  $P(\Theta)$ . A distribuição ocorre na proporção da crença depositada nos elementos.

Demonstra-se que a função obtida  $m_{12}$  é uma BPA como  $m_1 \oplus m_2$ . Independentemente da ordem em que as evidências ou funções sejam combinadas, a propriedade comutativa da multiplicação garante que essa regra gere os mesmos resultados.

A normalização das crenças se faz necessária somente quando a soma de todas as massas de crenças forem menor que 1. Isso ocorre quando existem elementos, com crença

atribuída, que não são comuns às evidências.

### **2.6.1 Benefícios da DST**

A DST permite atribuir crenças parciais quando não seria possível ou não seria prático conceder distribuições de probabilidade de forma confiável. O estudo de [Ferson et al., 2015]: imputa à DST a possibilidade de exprimir meios compreensivos e convenientes para a atenuação de vários problemas, como

- Especificações imprecisas de distribuições;
- Dependências desconhecidas ou fracamente conhecidas;
- Incerteza não descartável nas mensurações;
- Falta de detecção ou outro tipo de censura nas medições;
- Tamanho amostral pequeno.

### **2.6.2 Desafios da DST**

São três os principais desafios da DST:

- Resultado contraintuitivo;
- Complexidade computacional;
- Atribuição da probabilidade inicial(BPA).

Observa-se um comportamento contraintuitivo nas regras de combinação de evidências quando as concentrações de crença estão associadas a elementos disjuntos durante a combinação de evidências. Outra possibilidade aborda elementos comuns que reúnam baixos valores de crença e também quando ocorrem situações de pesos altos de conflito. Um risco alto é a utilização de subterfúgios que podem gerar o desprezo de evidências que apresentam essas características.

[Smets, 1988] alega que o comportamento contraintuitivo da regra de combinação de Dempster foi exemplificado inicialmente por [Zadeh, 1984]. O foco do problema é o critério de normalização da regra. Não há uma maneira de rebaixar de forma proporcional as crenças em relação ao volume de desconhecimento ou conflito entre as evidências. Então, é atribuído 100% de probabilidade, ou seja, a verdade absoluta, a um elemento

com baixa crença, o que é um paradoxo, gerando o resultado contraintuitivo. A literatura aborda um possível quarto desafio na abordagem DST que se refere ao tempo de processamento devido a quantidade de combinações possíveis do espaço do DST. Não obstante o incremento significativo na capacidade de processamento dos tempos atuais, a complexidade computacional endereça a questão de performance e foram propostos diversos trabalhos combatendo críticas à inexecutabilidade da DST.

### **2.6.3 Considerações Finais**

A teoria apontou um vasto caminho que pode ser utilizado num processo de FIMS com DST no intuito de melhor gerar a BPA para uma eficiente estimação com DST.

A DST difere dos métodos tradicionais de probabilidade. Ela não exige que a crença numa proposição implique que a descrença nela seja atribuída apenas à probabilidade de que o oposto ocorra. Esse aspecto a torna mais próxima da realidade, devido à incerteza, na abordagem de problemas.

A DST enriquece a teoria da probabilidade por meio de um arcabouço teórico mais generalizado, permitindo a distinção entre ignorância e incerteza e a introdução de elementos como a crença não atribuída, o grau de dúvida e o grau de plausibilidade.

Na DST também é importante enfatizar que os métodos numéricos de incerteza podem ser atribuídos a conjuntos sobrepostos de hipóteses, eventos ou proposições e não apenas a hipóteses individuais, como é feito tradicionalmente.

A escolha sempre foi realizada para ignorar a incerteza, eliminar dados ambíguos ou ausentes e considerar apenas as informações que são conhecidas com certeza. No entanto, a DST permite adotar uma perspectiva diferente, tentando aproveitar a falta de informação para um melhor entendimento da situação, melhorando a cobertura de diagnósticos.

O foco deste trabalho está no enriquecimento da credibilidade das evidências que participarão da BPA, identificando e saneando relações de dependência por ventura encontradas no domínio do problema.

## **3. Revisão de Literatura**

### **3.1 Metodologia da revisão**

Esta seção aborda o processo realizado para a revisão de literatura. As 4 questões de pesquisa utilizadas nesta revisão do sistema são: (1) quais as fontes de dados alternativas e métodos de automação e identificação de RAM?; (2) quais são as técnicas de aprendizado de máquina para extração e classificação de RAM e tarefas de diagnóstico?; (3) quais são os principais bancos de dados RAM e como eles são criados, mantidos e pesquisados?; (4) como atenuar a incerteza na formação dos bancos de dados ?

#### **3.1.1 Estratégia de Busca**

Das fontes de pesquisa inicialmente utilizadas na pesquisa, chama-se a atenção para a base SCOPUS, que apresenta muitos artigos de farmacovigilância e RAM, no entanto, via de regra não pertencem à área de sistemas de informação, frequentemente abordando a área de bioinformática ou biomedicina. Esta revisão de literatura considerou artigos publicados em inglês e os anos entre 2015 e 2020. Os motores de busca das quatro bases de dados foram configurados para buscar títulos, resumos e textos completos dos trabalhos. Os títulos e resumos dos artigos recuperados foram revisados por um pesquisador, sendo incluídos se mencionassem experimentos ou estudos sobre extração de RAM. Alguns deles foram recuperados porque no texto completo foram utilizadas algumas das palavras-chave de busca, mas o artigo estava fora do tópico, o que ocasionou sua exclusão sem a necessidade de leitura total do texto. Para estabelecer a estratégia de busca, foram realizadas consultas iniciais para avaliar o volume de estudos potencialmente relevantes. Os principais resultados de pesquisas preliminares, refinando os critérios, até chegar à melhor consulta. As perguntas de pesquisa foram: adversa ((reações a medicamentos usando mineração de opinião e aprendizado de máquina em mídias sociais digitais) OR (banco de dados e processamento de linguagem natural para reações adversas a medica-

mentos)) OR (uso de aprendizado profundo para reações adversas a medicamentos)). As fontes utilizadas foram 1) Biblioteca Digital IEEE, 2) Biblioteca Digital ACM, 3) Google Scholar e 4) SCOPUS.

Em seguida, os textos completos foram analisados. Eles foram recuperados e analisados individualmente, avaliando se o objetivo era determinar se os artigos propunham um experimento. Finalmente, os artigos foram revisados para determinar o método de análise de sentimento e classificação ou previsão de RAM. Foram excluídos os artigos que não continham pelo menos uma dessas características.

### 3.2 Triagem da revisão

Para ser aceito, o artigo precisa abordar o assunto de detecção e extração de RAM, utilizando algum classificador, seja ele tradicional (Support vector machines (SVM), Naive Bayes(NB) ou que utilize método com aprendizado profundo(deep learning). Foram rejeitados artigos da área de biomedicina, artigos fora do campo da farmacovigilância ou que não trouxessem qualquer contribuição. A revisão de literatura relatada neste artigo foi realizada entre março e junho de 2019. Os 98 artigos incluídos são artigos de periódicos, anais de congressos e capítulos de livros. Esta revisão de literatura foi elaborada com o auxílio da ferramenta textitPARSIF. Foram realizadas consultas diretas às bases de dados das fontes ACM, IEEE, Google Scholar, Scopus, Springer e Science Direct

BUSCA	filtro /parte do artigo	conector	palavras chave
1	abstract	and or	pharmacovigilance social media eletronic health records
2	abstract	and or	adverse drug reactions social media eletronic health records
3	todo o artigo	and or	adverse drug events social media eletronic health records
4	todo o artigo	conector and and or	adverse drug reactions machine learning social media eletronic health records
5	todo o artigo	and and or	adverse drug reactions sentiment analysis social media eletronic health records

Figura 3.1: Critério de busca nas bases de artigos

A figura 3.1 mostra as palavras chaves utilizadas como string de busca que foram realizadas em quatro bibliotecas - ACM, Google Scholar, Scopus e IEEE - e após o processo de triagem, por título e resumo(abstract), todos os trabalhos foram recuperados e lidos.

Quanto à elegibilidade, foram selecionados apenas os estudos que abordavam o tema farmacovigilância, principalmente na detecção de RAM e que se restringem ao domínio

de SI, pois existem muitos estudos em bioinformática e biomedicina abordando esse tema. Os artigos eleitos foram lidos na íntegra e os pesquisadores se concentraram em evidências explícitas de métodos, coleta de dados e mecanismos de entrega de informações. Vários métodos de avaliação foram aplicados e descritos nos estudos incluídos nesta revisão de literatura, 75% dos quais eram experimentos. Os estudos de caso são [Kim and Rhew, 2017] [Zolnoori et al., 2019a], [Liu and Chen, 2015], [Liu and Wang, 2018] e [Comfort et al., 2018].

### **3.2.1 Eixos da Revisão**

As respostas às perguntas de pesquisa desta revisão de literatura atingem quatro eixos de revisão:

- 1a Métodos de automação de descoberta de RAM;
- 2a Evolução da análise e classificação de sentimentos com modelo Supervisionado, melhorando o desempenho com ensemble learning;
- 3a Quais são os principais bancos de dados RAM e como eles são criados, mantidos e pesquisados?;
- 4a Quais estruturas de raciocínio podem ser aplicadas para mitigar a incerteza em um processo de fusão de dados?.

A seguir, se explicarão as respostas às questões de pesquisa desta revisão de literatura.

A tabela 3.1 mostra as fontes e técnicas utilizadas pelos artigos.

### **3.3 Resposta à 1ª questão Quais são as fontes alternativas e respectivos métodos para identificação de RAM?**

Ao responder essa questão chegou-se a uma série de formas distintas de captura, que são então sucedidas por técnicas de aprendizado de máquina, inicialmente utilizando análise de sentimentos baseada em polaridade, com destaque para o uso dos classificadores SVM e NB.

A revisão narrativa feita em [Wong et al., 2018] abordou os fundamentos do PLN e discutiu sua aplicação na identificação de RAM utilizando três fontes de dados: prontuários eletrônicos(EHR), sites e blogs publicados na internet. O principal benefício da PLN,

<b>Dataset</b>	<b>Tradicional</b>	<b>Lexical</b>	<b>Deep Learning</b>
TWITTER	[Peng et al., 2016b], [Wu et al., 2015]	[Masino et al., 2018], [Rajapaksha and Weerasinghe, 2015]	[Severyn and Moschitti, 2015a], [Huynh et al., 2016a], [Mahata et al., 2018], [Chen et al., 2018a] [Han et al., 2017]
CORPUS	[Kim et al., 2008] Corpus Diegolab [Song et al., 2017]		
DS - DaylyS-trengt	[Mishra et al., 2015] , [Sarker et al., 2015]	[Korkontzelos et al., 2015]	
Foruns	[Wong et al., 2018]	Diabetes Forums, [Zhang et al., 2019a]	
EHR	[Zhang and Geng, 2019]	[Dandala et al., 2018]	[Yang et al., 2018], [Santiso et al., 2018] [Song et al., 2017]
MedHelp	[Rajapaksha and Weerasinghe, 2015]	[Zhang et al., 2019a]	
Sider2	[Harpaz et al., 2017], [Thompson et al., 2018a]	[Rajapaksha and Weerasinghe, 2015]	[Liu and Wang, 2018]
Metamap	[Harpaz et al., 2017]	[Wunnava et al., 2017]	[Peng et al., 2016a] , [Rajapaksha and Weerasinghe, 2015]
Medline	[Huynh et al., 2016b] ,	[Dev et al., 2017a]	

Tabela 3.1: Relação de estudos por categoria de extração de RAM

além da economia associada à automação de tarefas, seria a extração de RAM comentadas em MSD que não seriam descobertas se não fosse por esse ferramental. Um trabalho que evoluiu o tratamento do léxico que partiu de vários MSD de comunidades de saúde é encontrado em [Mishra et al., 2015] . Este artigo usou opiniões sobre medicamentos usados no tratamento de câncer para identificar RAM comparando essas opiniões contra folhetos de medicamentos(bulas) aprovados pela FDA. O estudo mapeou as intervenções, contraindicações e sintomas presentes nos comentários dos pacientes, comparando-os com os rótulos da FDA, usando classificação baseada em SVM e análise de sentimentos com WordNet e SentiwordNet.

O estudo de [Peng et al., 2016b] propôs um pipeline para extrair RAM no Twitter. Esse estudo coletou durante quatro meses uma coleção de mensagens do Twitter. Esse método superou em cinco vezes o resultado de extração de RAM de um pipeline anterior, descobrindo 20% de RAM inéditas. Foram consideradas inéditas em relação à base FAERS via MetaMap. Este trabalho utilizou como benchmarking o trabalho de [Wu et al., 2015], ambos utilizaram o analisador de sentimento *TextBlob* para polaridade.

A figura 3.2, demonstra um pipeline típico para extrair RAM de *tweets*, atingindo precisão em torno de 70%, utilizando modelo de aprendizado não supervisionado e análise de sentimento de polaridade.

A vantagem principal do MSD está no volume e oferta muito diversificada. Em contrapartida, o desafio do pré-processamento é muito grande pois as mensagens não são

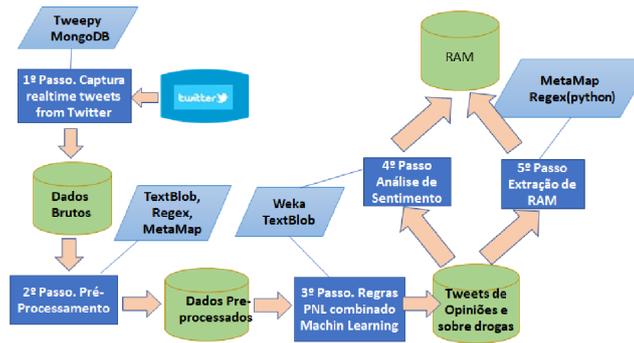


Figura 3.2: Análise de sentimento de polaridade X classificador SVM/NB adaptada de [Peng et al., 2016b]

padronizadas e apresentam um nível de ruído muito alto.

Quanto ao uso do EHR, os trabalhos de [Mishra and Diesner, 2018], [Zhang et al., 2019a], [Wunnava et al., 2018], [Dandala et al., 2018] e [Santiso et al., 2018] destacam a vantagem de reunir informações de observação do paciente, realizadas por profissionais de saúde, por equipes multidisciplinares (médicos, enfermeiros, farmacêuticos). Apresenta a desvantagem, em menor proporção do que em MSD, das características de alta dimensionalidade. Em especial nas notas clínicas são usadas muitas siglas, termos de jargões, abreviaturas e demais aspectos que demandam a criação de procedimentos PLN.

O desafio da privacidade de dados sensíveis do paciente em EHR requer cuidado especial, inclusive em nível de comitê de ética da instituição de saúde. A oferta de dados é razoável, mas não existe um padrão, pois há diversos sistemas de gestão hospitalar que utilizam estruturas e bancos de dados diferentes. Assim, deve-se ter cuidado na criação de modelos proprietários, ou seja, que funcionam apenas para um sistema de gestão específico, vinculado a um determinado fornecedor. É importante abordar a redundância de fontes de dados, pois não é incomum existir mais de um sistema de gestão hospitalar na instituição, havendo a necessidade de trabalhar a redundância de dados. Um processo de fusão de informações é recomendado para organizar a extração e identificação de eventos em fases.

Podemos concluir da análise dessa questão algumas informações relevantes:

- MSD traz análise de sentimento e necessidade de supervisão, inclusive para atividade de anotação e rotulação de dados para o processo de classificação
- Seja EHR ou MSD deve-se cuidar da alta dimensionalidade
- EHR é mais organizado e requer menor pré-processamento para tratamento de ruído

- MSD apresenta volume imenso, o que é bom, mas requer muito esforço para tratamento de ruído em rotinas de pré-processamento
- EHR traz necessidades de fusão de informação a partir de bases despadronizadas, redundantes e imprecisas, demandando cuidados para tratamento de incerteza

### 3.4 Resposta à 2a questão - Qual a evolução das técnicas de classificação e análise de sentimentos para a extração de RAM?

Entre os anos de 2010 a 2015 configura-se uma primeira geração de experimentos baseados em análise de sentimentos conjugada com técnicas de PLN utilizando classificadores tradicionais como SVM e NB. Entre 2016 e 2020, a comunidade científica começa a buscar experimentos mais incrementados tendo em vista melhor precisão nos experimentos. Os métodos de análise de sentimentos mais simples do início são aprimorados com o uso de recursos de redes neurais e aprendizado profundo. Muitos experimentos realizam combinações de métodos - ensemble learning - alcançando resultados com métricas mais favoráveis. Os destaques em algoritmos são CNN, RNN e LSTM são as técnicas predominantes, com uso crescente, em geral, combinado, de NER e CRF. O PLN evoluiu fortemente, com destaque para técnicas de *embedding* para o processamento de palavras, destacando-se *Word2Vec* e *Glove*, e representação vetorial de emoções.

FERRAMENTA	FUNÇÃO	REFERÊNCIAS
TextBlob – Biblioteca Python	Biblioteca Python Python Análise de Sentimentos subjetividade e polaridade.	[Wu et al., 2015], [Peng et al., 2016c]
Ctakes	Ferramenta livre de PLN no campo médico processa o texto livre clínico de <i>corpus</i> ou EHR permitindo representações ontológicas via PLN	
Word2vec	Usam redes neurais de duas camadas que são treinadas para reconstruir contextos de linguagem de palavras. Vetores de palavras são posicionados em espaço vetorial, então palavras que compartilham contextos comuns são muito próximas.	[Lin et al., 2015]
SentiWordNet	É um recurso léxico publicamente disponível para mineração de opinião. três pontuações de sentimento: positividade, negatividade, objetividade	[Hsu et al., 2017]
GLOVE-Global	Vetores para representação de palavras, utiliza algoritmo com aprendizado não supervisionado para obter representações vetoriais para palavras utilizando subestruturas de espaço vetorial, do inglês <i>vector space</i>	[Lin et al., 2015]

Tabela 3.2: Ferramentas e funções relevantes na extração de RAM

### 3.4.1 Modelos profundos e combinações - Uma evolução

A evolução dos métodos utilizando *ensemble* incrementa o uso de composições envolvendo aprendizado profundo. A análise de sentimentos inicialmente focada na polaridade trouxe ao longo do tempo avanços significativos do PLN nos últimos anos. É notório o uso de processamento vetorial, seja texto ou imagens – algo útil para as emoções dos *tweets*, que inseriram definitivamente modelos de redes neurais na pesquisa de extração de RAM de MSD. O *embedding* está para o tratamento de palavras assim como o tratamento de imagens de *emotions* está na evolução da análise de sentimento para expressões. O aprendizado profundo funciona seja com o modelo de rede neural recorrente (RNN) ou a partir de rede neural convolucional (CNN). Pode haver combinações de ambos, especialmente RNN com modelos LSTM e CRF.

A Técnica CNN de aprendizado profundo teve suas portas abertas no estudo de RAM por meio do uso de camadas profundas de processamento, convoluções, agrupamento e uma camada de classificação. Além do processamento de imagens para emoções, a CNN foi aplicada com sucesso em tarefas de reconhecimento de similaridade semântica na representação de palavras.

Um estudo pioneiro em farmacovigilância usando CNN que criou um experimento que utilizou aprendizado profundo para analisar sentimentos de tweets está em [Severyn and Moschitti, 2015b]. A principal contribuição deste trabalho é um novo modelo para inicializar pesos de parâmetros para configurar a CNN, usando um modelo neural não supervisionado para treinar a incorporação de palavras iniciais. Em seguida utiliza um corpus obtido por método supervisionado para aprimorar o modelo como um todo. Essa mescla de aprendizado supervisionado e não supervisionado é denominada aprendizado semi-supervisionado.

O estudo [Huynh et al., 2016b] propõe uma composição de dois modelos de redes neurais, uma rede neural convolucional recorrente (CRNN) pela concatenação de uma CNN com redes neurais recorrentes e a rede neural convolucional com atenção (CNNA) que adiciona pesos de atenção a uma CNN. Este estudo avaliou várias arquiteturas de modelos neurais em um conjunto de dados do Twitter contendo linguagem informal e um conjunto de dados construído por amostragem de casos MEDLINE. O estudo conclui que a CNNA é mais eficaz na base do Twitter do que a CRNN, que, por sua vez, apresenta melhores resultados que a primeira no caso da base extraída do MEDLINE.

Um avanço na computação afetiva para triagem de humor, emoções e sentimentos dos pacientes, além de combater a automedicação, foi alcançado no estudo de [Mahata

et al., 2018] que criou um experimento classificador, denominado "InfyNLP". Esse método identifica menções de automedicação confessadas por meio de tweets. Um modelo agregado de uma rede CNN foi treinado em um conjunto de dados rotulado e anotado. O modelo utilizou uma busca aleatória para ajustar os hiperparâmetros dos modelos CNN, apresentou um conjunto de melhores modelos para a tarefa de previsão e produz resultados com uma média de 0,693 em F-score.

Uma composição de CNN e LSTM foi apresentada numa competição pelo trabalho de [Han et al., 2017] que se divide em três tarefas: a primeira seria identificar *tweets* contendo menções a RAM; a segunda objetiva a classificação automática de mensagens de consumo de drogas no Twitter e a terceira tarefa seria identificar a RAM na base de dados MedDRA. Eles propuseram abordagens CNN e TLM para a primeira e a segunda tarefa, obtendo, respectivamente 40,2% e 68,9% de f-measure. A terceira tarefa foi a mais bem-sucedida e utilizou redes LSTM recorrentes para a detecção de RAM e alcançou 87,2% de f-measure.

Um método para mitigar o problema de anotação e rotulação de RAM, objetivando redução do esforço para rotular tweets em larga escala, é definido no estudo de [Masino et al., 2018]. Esse modelo desenvolveu uma rede CNN que processa grandes conjuntos de *tweets* representados por vetores de palavras criados por um modelo de aprendizado não supervisionado. No ambiente de teste atingiu métrica F1 de 0,46 e sensibilidade de 0,78 para classificação, superando dois modelos de *embedding* usando *n-grams* e recursos de léxico. Um ponto forte desse modelo foi o uso de reconhecimento de entidades nomeadas (REN) para extração de termos relevantes no domínio.

### 3.4.2 Tarefa PLN em destaque - reconhecimento de entidades nomeadas (REN)

Entre as aplicações da PLN, o REN vem sendo bem empregado na pesquisa de RAM. A REN pretende captar o significado de mensagens ou textos das quais se pretende extrair sentimentos, identificando entidades e efetuando a sua classificação e atribuindo uma categoria semântica a essas entidades. No campo da análise de sentimentos, a abstração do termo linguístico equivale a termos específicos. Eles podem se referir a opiniões ou expressões que denotam desacordo ou acordo, paixão, ódio ou repulsa. O objetivo é que todas as sentenças contenham uma opinião sobre RAM, fazendo com que o conteúdo do corpus e a classificação que se deseja obter dele especifiquem as entidades nomeadas que representam bem o domínio, como, por exemplo no contexto de RAM, as entidades drogas, reações adversas, sintomas e doenças.

Uma importante contribuição em tarefas específicas de *embedding* para tratar palavras

ocultas, propondo uma forma de farmacovigilância em tempo real para reforçar a segurança de medicamentos é apresentada no estudo de [Dandala et al., 2018]. Esta pesquisa empregou três tarefas diferentes de PLN em EHR. O método realiza a identificação de REN e relacionamento em conjunto nas extrações de entidade e nas tarefas de identificação de relacionamento. Em segundo lugar, um processo de duas etapas foi usado para extrair entidades usando os modelos Bi-LSTM-CRF e a conclusão é alcançada com um método LSTM Attention-Bi.

O estudo de [Yang et al., 2018] teve como objetivo a REN para detectar o nome do medicamento e seus atributos - dosagem, frequência, via, duração e menções de RAM. O modelo combinou várias arquiteturas de aprendizado profundo. Uma segunda tarefa tratou de indicações de outros sinais e sintomas de anotações clínicas. Então utilizou essas duas tarefas com diferentes estratégias de treino, utilizando LSTM-CRF, comprovando a eficiência do REN para o estudo de RAM.

### **3.4.3O uso de *embedding* no LSTM e associação com CRF**

Com base em vetores de representação de palavras, o trabalho de [Nikfarjam et al., 2015] gerou um experimento chamado "ADRMINE" para modelar palavras com semelhanças semânticas encontradas nas postagens. Esse experimento utiliza um método não supervisionado e trabalha a extração de conceitos complexos por meio de aprendizado profundo usando CRF e obteve F-measure de 0,82.

No intuito de classificar RAM como graves e não graves, um modelo de rede neural recorrente LSTM foi trabalhado no estudo de [Dev et al., 2017b] atingindo F1-Score médio de 95% .

Com base no uso de *embeddings* gerados por lemas criados a partir de EHR, como fonte de dados para a descoberta de RAM, o estudo de [Santiso et al., 2018] implementou uma subclasse de uma rede neural bidirecional conjunta AB-LSTM. Esse estudo obteve melhores resultados do que as abordagens de extração supervisionada. Eles concluíram que usando "lemas" como características-chaves em vez de formas de palavras, o desempenho do AB-LSTM melhora, apresentando bom desempenho na identificação de RAM em EHR.

Uma proposta de incorporar emoções em um modelo bilateral a partir de *tweets* foi feita por [Chen et al., 2018b] que inicialmente avalia a polaridade dos *tweets* como sentimentos positivos ou negativos. Feito isso, eles treinaram um classificador de sentimentos lidando com as emoções incorporadas usando um AB-LSTM, melhorando os níveis de

compreensão semântica e sentimentos.

Utilizando MSD e texto médico obtido em notas clínicas EHR para formar a base para a realização da análise e classificação de RAM, o estudo de [Song et al., 2017] utilizou CRF. Esse trabalho realizou uma série de operações de padronização, objetivando o monitoramento de RAM a baixo custo.

A redução de ruído em texto clínico foi obtida com o uso de técnicas de tokenização no estudo [Wunnava et al., 2018]. O método utilizou uma arquitetura de três camadas, combinando RNN, BI-LSTM e CRF, utilizando como base os prontuários eletrônicos (EHR). Trabalhou em três camadas, a primeira com um modelo RNN combinado com AB-LSTM para representação de palavras em nível de caractere. Na segunda camada utilizou um modelo AB-LSTM para representação de contexto e na terceira camada usou um modelo CRF para a previsão da saída final.

Uma combinação de LSTM-CRF na identificação de RAM está no estudo de [Tang et al., 2018] que realizou um benchmark com o “corpus” CADEC e concluiu que o resultado da combinação LSTM-CRF supera o CRF isolado e que *embedding* são úteis no reconhecimento de RAM.

O *embedding* foi o foco do estudo de [Lin et al., 2015], que examinou métodos de representação de diferentes palavras para a tarefa de reconhecimento de RAM, incluindo a normalização de "tokens", e dois métodos de última geração para incorporação de palavras, o *Word2vec* e *Glove*. Após análise manual dos clusters gerados a partir desses dois métodos, observou-se que o *Word2vec* apresentou tendências de cluster em relação ao *Glove*.

No próximo tópico serão vistos alguns experimentos e técnicas que auxiliaram na produção de léxicos para promover pesquisas em RAM.

A tabela 3.3 aborda as características da pesquisa de RAM quanto ao uso da abordagem do modelo de aprendizado. No início dos tempos, prevaleciam os métodos não supervisionados na expectativa de que a análise de sentimentos e técnicas de classificação fossem o suficiente para identificar RAM em fontes alternativas, sejam MDS ou EHR. O modelo supervisionado alcança melhores métricas, mas o trabalho de anotação de RAM requer a alocação de profissionais especializados e muito tempo para avaliação e criação de rótulos. Surge a necessidade de algo híbrido, concretizado pelo uso do semi-supervisionado, inicialmente com ensemble learning(EL) e mais recentemente incrementado com deep learning(DL). As métricas evoluem ao longo do tempo, e a necessidade de bases anotadas se faz menor no modelo semi-supervisionado.

MODELO NÃO SUPERVISIONADO	MODELO SUPERVISIONADO	MODELO SEMI-SUPERVISIONADO(EL)	MODELO SEMI-SUPERVISIONADO(DL)	
MÉTODO GERAL	Captura, polaridade, ML	Captura, polaridade, léxico, ML	Captura, polaridade, léxico, ML, EL	Captura, polaridade, léxico, ML, DL
INTEGRAÇÃO	MEDHELP SIDER2	MEDHELP SIDER2 , Metamap	MEDHELP SIDER2 , Metamap	MEDHELP SIDER2 , Metamap
ANÁLISE DE SENTIMENTOS	Polaridade	Polaridade	Polaridade, Emotions	Polaridade, Emotions(textuais e gráficos)
REPRESENTAÇÃO DE PALAVRAS	bag of words, cesta de palavras	bag of words, cesta de palavras	bag of words, cesta de palavras	Embedding, PLN redes LSTM
ACURÁCIA MÉDIA	72,50%	74,50%	82,50%	74%
ALGORITMOS	NB, SVM	NB, SVM, AD	NB, SVM, AD, EL	RN, CNN, RNN
CREDIBILIDADE	baixa	baixa	baixa	moderada
FORMAÇÃO DE DATASET	difícil	difícil	difícil	difícil
ANOTAÇÕES	Necessária	Necessária	Desejada	Complementar
COMPARTILHAMENTO	difícil	difícil	difícil	difícil
ESTUDOS	[Sarker et al., 2015] , [Peng et al., 2016c],	[Azzouza et al., 2017], [Dritsas et al., 2018], [Liu et al., 2018a],	[Sintsova and Pu, 2016a], [Dandala et al., 2018],	[Masino et al., 2018], [Huynh et al., 2016b], [Liu and Wang, 2018], [Masino et al., 2018],

Tabela 3.3: A evolução da pesquisa RAM pelo modelo de aprendizado

EXPERIMENTO	BAG OF WORDS	CRITÉRIO DE VOTAÇÃO	TÉCNICA
[Liu et al., 2018b]	N-grams; FAERS , VigiBase	voto majoritário	Bagging, Boosting SVM e Random Space
[Liu and Wang, 2018]) -SRSS	FAERS e CHV with Meta-map	Seleção de características	SVM , NB, Bagging subspace
[Sintsova and Pu, 2016b]	32 mm <i>tweets</i> , sentimentos	Voto ponderado	Multinomial NB and Logistic

Tabela 3.4: Experimentos Ensemble Learning para identificar RAM

A tabela 3.4 ilustra alguns experimentos que usam ensemble learning. [Cai et al., 2019].

O estudo de [Hsu et al., 2017] inovou no tratamento da mineração de frequência de RAM ao melhorar as versões anteriores dos métodos de extração de *tweets*. O método trabalhou um aspecto longitudinal, no tempo, identificando os primeiros *tweets* que continham RAM, tentando estabelecer uma dinâmica no tempo enquanto tentava obter uma frequência em que esses efeitos colaterais ocorreriam. Foram usados quatro léxicos: AFINN [Nielsen, 2011] , SentiWordNet [Baccianella et al., 2010], Bing.Liu [Hu and Liu, 2004] e MPQA [Wilson et al., 2005].

Um esforço para melhorar a integração do uso de consultas aos bancos de dados FAERS e VIGIBASE, além de identificar RAM foi trabalhado em [Liu et al., 2018b]. Esse estudo auxiliou as autoridades regulatórias a organizar ações de retirada de mercado e recalls de medicamentos, reduzindo custos e perdas financeiras para organizações da indústria farmacêutica. O artigo propõe um experimento chamado "SSEL-ADE" que emprega aprendizado não supervisionado e aprendizado contínuo integrado *ensemble* para

extrair RAM do MSD combinando várias formas de recursos e técnicas. O experimento criado por [Liu and Wang, 2018] criou um método inovador para identificação de RAM, implementando a extração de mensagens de fóruns de doenças cardíacas, FAERS e CHV, usando métodos UMLS e Stanford e MetaMap. Esse experimento criou associações por meio da exploração de características linguísticas e semânticas, que obteve bons resultados.

Observa-se uma importante utilização da DST em experimentos ensemble learning. A combinação de crenças oferecidas pela DST é agregada à técnica de ensemble. Por exemplo, o estudo [Wang et al., 2020] criou um classificador de conjuntos baseado em DST para reconhecimento de falhas, alegando que os classificadores individuais têm um desempenho ruim nesse sentido.

Entre os muitos métodos de *ensemble* disponíveis, um dos mais usados é o *Boosting*, que é originalmente aplicado para construir um classificador forte a partir de uma coleção de classificadores fracos.

A fórmula de resolução de conflitos da DST é uma possibilidade de uso que integra a combinação de classificadores individuais em processos FIMS. Essas combinações podem ser baseadas na regra clássica de combinação Dempster. Por trabalhar com informações incertas e incompletas, a DST tem sido amplamente utilizado em aplicações de diagnóstico de falhas.

A maioria dos trabalhos utilizou métodos diferentes para detectar falhas nos sensores, mas o diferencial da DST em relação a esses trabalhos é a capacidade de modelar e levar em conta ambiguidades, incertezas e conflitos sobre os dados. A DST permite o gerenciamento parcial do conhecimento, ao possibilitar atribuição de crença à incerteza.

Existe uma diferença entre métodos *boosting* como *AdaBoost* e DST em relação à combinação de classificadores. *Boosting* atribui um peso ponderado a cada classificador individual com base no erro de treinamento, enquanto a DST atribui a BPA - geralmente no sentido de peso - gerado pela saída de cada classificador individual.

### **3.5 Resposta à terceira questão - Quais são as principais bases de dados sobre RAM e como são criadas, mantidas e pesquisadas?**

Respondendo à terceira questão de pesquisa vai abordar-se a geração de dados e o fornecimento de conjuntos de dados para desenvolvimentos de pesquisa em RAM.

Apesar de os enfoques de extração serem diferenciados quanto à fonte de dados, sejam MSD ou EHR, há recursos computacionais de PLN e formação de léxicos que auxiliam em ambas opções de fontes de dados. Sejam oriundos de MSD ou EHR, os registros apresentam alta dimensionalidade, enquanto as fontes MSD contam com corpora e dicionários tradicionais. As fontes EHR normalmente vão demandar procedimentos de captura que exigem uma estruturação e organização de um processo com uso de técnica de fusão de dados. Os processos de fusão de dados são empregados em projetos EHR dada a variedade de fontes, bancos de dados com formatos distintos, redundâncias e outros aspectos que serão vistos nas próximas subseções.

Essa subseção é subdividida em quatro partes: i) dicionários e abordagens baseadas em regras; ii) corpus e ferramentas auxiliares para a formação de léxicos; iii) conjuntos de dados tradicionais e iv) fusão de dados.

### **3.5.1 Dicionários e abordagens baseadas em regras.**

Interpretando o linguajar da internet e utilizando seleção de características para a classificação de RAM, o estudo de [Rajapaksha and Weerasinghe, 2015] realizou a extração de RAM a partir do Twitter filtrando mensagens relacionadas a drogas usando abordagens baseadas em regras de dicionário para identificar termos médicos, emoções e gírias.

O estudo de [Wunnava et al., 2017] implementou uma sistema de extração de categorias estruturadas de grandes narrativas denominado *meta extração framework (MEFA)*. O MEFA trabalha a partir das características sintáticas, semânticas e morfológicas da base FAERS. Com essa abordagem de extração de regras geraram informações demográficas que melhoraram a qualidade do processo de identificação de RAM.

Uma categorização de RAM foi criada no estudo de [Kim et al., 2008] que classificou postagens como possíveis, prováveis e improváveis, com base nas relações causais especificadas nas regras do dicionário WHO-UMC. Usou o classificador NB combinado com técnicas de mineração de texto, gerando um léxico com termos frequentes nas opiniões de especialistas, extraídos de nuvens de palavras e de relatórios de casos de RAM disponíveis nos dados do sistema de notificação coreano.

Outro caso de categorização de RAM em relação à sua gravidade que também usou dicionário WHO-UMC foi definido em [Zhang et al., 2018b] que objetivou explorar as características de ocorrência de RAM para medicamentos hipolipemiantes, drogas usadas para controle do colesterol. Um repositório aberto de dados RAM contendo 579 notificações de eventos na China foi rotulado manualmente. Tais eventos foram mapeados no

dicionário Who-art, com 171 termos e um padrão de descrição.

### **3.5.2 “Corpus” e ferramentas auxiliares para a formação de léxicos, integrando a bases preexistentes.**

Um dos desafios na pesquisa de RAM é o trabalho de anotação, que é artesanal, feito por especialista, consumindo um esforço considerável. Essa limitação gerou uma evolução natural dos esforços de pesquisa concentrada na formação de corpus, gerando léxicos que consolidam dados sobre nichos, tratam de problemas de sintaxe, jargão ou sotaque, o que é combatido com dicionários e semânticas que trabalham regras específicas. Além de mitigar a precária disponibilidade de novos conjuntos de dados, a formação de corpus traz agilidade às pesquisas pela economia de esforço de captura e possibilita bases para estabelecer comparação, possivelmente economizando atividade de anotação e rotulação das opiniões. O corpus pode ser usado para fins de linguagem ou emoção.

Além da importância do compartilhamento de dados, a pesquisa de referência como conjuntos de dados tradicionais é uma boa prática para validar potenciais RAM descobertas nos estudos.

O trabalho de [Thompson et al., 2018b] reúne textos rotulados da base MEDLINE, gerando e disponibilizando o corpus "Phaedra and Himeria" [Thompson et al., 2016]. Esse trabalho gerou uma contribuição para bases preexistentes ao codificar informações importantes sobre RAM e interações medicamentosas. As notas mostraram altos níveis de concordância entre diferentes especialistas e trabalharam efeitos de múltiplas drogas ou RAM. As RAM foram categorizadas de acordo com sua intensidade, diferenciando efeitos experimentais de especulativos.

Uma integração que permitiu o mapeamento UMLS para o SNOMED-CT foi obtida com a criação do corpus "AEOLUS no estudo de [Banda et al., 2016], uma iniciativa pioneira que passou a permitir que pesquisadores o vinculem a outras ontologias como a Classificação Internacional de Doenças(CID International Disease Classification IDC nos EUA). Esse estudo partiu das bases FAERS e MedDRA, com mapeamento de dicionário OHDSI [Hripcsak et al., 2015], usou recursos do dicionário RxNorm e gerou uma versão padronizada dos eventos no FAERS, removendo duplicidades e fornecendo métricas de RAM.

Relacionando termos correspondentes em vocabulários controlados, como termos clínicos SNOMED e MedDRA, o corpus "CSERO Adverse Drug Event Corpus"(CADEC), foi criado no trabalho de [Karimi et al., 2015a]. O CADEC contém anotações como no-

mes de medicamentos, RAM, doenças e sintomas obtidos de postagens no fórum médico "AskaPatient". Apresentam texto em linguagem coloquial, no qual regras formais de gramática e pontuação em inglês geralmente não são usadas. As anotações contêm menções a entidades como drogas, RAM, sintomas e doenças relacionados a essas entidades.

Um corpus de nicho psiquiátrico foi criado pelo estudo de [Zolnoori et al., 2019a] que extraiu efeitos farmacológicos a partir de comentários de pacientes em MSD. Esse experimento desenvolveu o corpus *PsyTAR - Psytrary Treatment Adverse Reactions* a partir de rótulos que utilizam dados narrativos de pacientes com síndrome Serotonin Reuptake Inhibitor (SSRI) e Serotonin Norepinephrine Reuptake Inhibitor (ISRI). O trabalho utilizou a linguagem UMLS consultando o SNOMED-CT. Um componente de classificação de frases do *PsyTAR* foi usado para treinar automaticamente uma série de classificadores, obtendo 0,90 F-score com SVM. Esse estudo comprovou que RAM psicológicas oriundas de problemas funcionais cognitivos têm maior variabilidade semântica quando comparadas a RAM fisiológicas.

Nota-se um fenômeno de distinção de termos usados entre ramos medicamentosos, o que faz com que drogas relacionadas à oncologia possam gerar termos diferentes daqueles encontrados em opiniões sobre drogas cardiológicas, por exemplo. Portanto, usar o léxico implementado para outro propósito provavelmente fará com que um classificador usado numa base distinta gere uma precisão diferente. Não adiantaria muito, portanto, seria como formar uma base em RAM a partir de opiniões sobre eleições ou filmes. Essa característica torna as linguagens do corpus e os léxicos uma barreira na interoperabilidade.

### 3.5.3 Conjunto de dados tradicionais

Será visto como o uso de ferramentas e bases tradicionais possibilitam a extração de regras auxiliam no desenvolvimento dos estudos. O trabalho de [Zhang and Geng, 2019] define que os dados obtidos de EHR do FAERS possuem uma estrutura temporal bem estruturada, mas falta clareza na definição de RAM, apresenta métricas confusas e fere a privacidade tornando os dados do paciente expostos. Respeitando a importância de combinar múltiplas fontes, o estudo realizou rigorosas análises de segurança, contemplando 12 drogas antimaníacas. Investigando três fontes de dados: sistema de notificação de eventos adversos da FDA, banco de dados de citações MEDLINE e registros dos principais mecanismos de busca da web, o trabalho de [Harpaz et al., 2017] propôs uma avaliação do tempo de espera até a detecção de RAM por cada uma dessas fontes.

A Tabela 3.5 mostra diversos recursos de PLN e alguns “corpus” e léxicos disponibi-

DICIONÁRIO	DESCRIÇÃO	CARACTERÍSTICAS	OBSERVAÇÃO	REFERÊNCIAS
Conjunto de dados FAERS	Os sistemas de notificação de eventos adversos da FDA	Termos do dicionário, atividades regulamentares (MedDRA)	útil para profissionais de saúde e consumidores	[Zhang and Geng, 2019], [Harpaz et al., 2017], [Wunnava et al., 2018], [Liu et al., 2018a]
Efeito colateral do conjunto de dados SIDER	Biblioteca de Medicina	Útil para profissionais de saúde	Biblioteca Nacional de Medicina dos EUA	[Harpaz et al., 2017], [Thompson et al., 2018a]
Conjunto de dados SNOMED	Nomenclatura Sistemática de Medicina Biblioteca de Medicina dos EUA	Banco de dados com uma coleção de termos médicos sistematicamente organizados e	Fornecer códigos, termos, sinônimos e definições usados na documentação clínica e relatórios	[Banda et al., 2016], [Karimi et al., 2015b], [Zolnoori et al., 2019b]
OMS-ART Organização Mundial da Saúde	UMLS básico Dicionário de terminologia usado para termos de RAM	Dicionário usado para RAM (descontinuado)	Centro Colaborador ERA OMS para Monitoramento	[Kim and Rhew, 2017], [Zhang et al., 2018b]
MEDDRA Médico	Dicionário para Atividades Regulatórias	Usar efeitos de saúde no formato UMLS	Medicamentos Internacionais	[Banda et al., 2016], [Karimi et al., 2015b]
CHV-Consumidor Saúde Vocabulário	Base UMLS Completa a UMLS ajuda o consumidor em saúde	Recupera e analisa de dados sobre medicamentos RAM.		
METAMAP	base UMLS	Configurável para mapear texto biomédico para Metatheasurus UMLS	Utiliza PNL e linguística computacional, sendo aplicada em mineração de dados	
RxNorm	Permita a troca de informações entre sistemas de forma eficiente	Nomenclaturas próprias RxNorm propõe padronizações		[Banda et al., 2016]

Tabela 3.5: Dicionários e Conjuntos de Dados usados no domínio RAM

lizados, contendo nichos específicos que podem auxiliar a pesquisa de RAM.

### 3.5.4 Bancos de dados próprios utilizando processo de fusão de dados

Um estudo que utilizou recursos de fusão de dados aplicado em farmacovigilância foi um modelo baseado em classificação local (LCM) para prever interações medicamentosas. O método define uma nova regra de mesclagem supervisionada para agregar as saídas de vários classificadores numa instância não marcada ao mesclar DST com LCM, criando o método (LCM-DS) que agrega os resultados de vários LCM. Para fazer a previsão final, o LCM-DS integra três aspectos de vários classificadores: a saída de probabilidades posteriores para classificadores individuais, a proximidade entre os perfis de decisão de determinadas instâncias e os perfis de referência [Shi et al., 2018]. Esse estudo utiliza a potencialidade da regra de combinação de Dempster para reduzir a incerteza dos diagnósticos de interação medicamentosa.

Esse estudo despertou a necessidade de entender melhor o que seria o conceito de fusão de dados e como a DST colaboraria nesse processo. Nesse ínterim, abstrai-se o domínio do problema RAM, e realiza-se a uma investigação sobre fusão de dados em si, mais precisamente fusão de informação multissensores.

### 3.5.5 Fusão de informação multisensores FIMS

De acordo com [Steinhauer and Karlsson, 2019], a FIMS permite uma melhor análise de dados ao proporcionar:

- o aumento da precisão das informações usando várias fontes de dados;
- a mitigação da incerteza a fim de fornecer melhores bases para a tomada de decisão;
- a redução da dimensionalidade dos dados relevantes com o intuito de proporcionar uma visão facilmente compreensível de um domínio.

O trabalho de [Khaleghi et al., 2013] aborda a estruturação dos dados, realizando uma classificação em incerteza, imperfeição, conflitantes e correlacionados.

Para selecionar uma técnica de fusão de dados, o primeiro aspecto a focar é a análise da constituição dos dados. Esta etapa de aquisição requer a análise de anomalias nos dados capturados [Steinhauer and Karlsson, 2019], que destaca os seguintes aspectos:

**(a) Imperfeição nos dados** que pode ser atenuada com:

- fusão probabilística: A ideia principal é representar a imperfeição dos dados por meio da incerteza na forma de distribuições de probabilidade
- raciocínio baseado em evidências: A DST define possíveis medidas de hipóteses obtidas a partir de correspondentes de crenças e plausibilidade, contemplando as imperfeições.

**(b) Inconsistência** pode ser categorizada como:

- dados espúrios. Os dados espúrios são aqueles desordenados devido a falhas permanentes ou atrasos na transmissão dos sensores. Normalmente devem ser removidos.
- validade e ordem (sequência) - Os dados devem ser atualizados em intervalos e sequência apropriados a fim de garantir um resultado de fusão válido quando o histórico de medição for importante para o processo de fusão [Kaugerand et al., 2018].
- conflitantes - O problema dos dados conflitantes é discutido nas extensões da DST desenvolvidas especialmente para resolver o problema dos dados inconsistentes mencionados em [Steinhauer and Karlsson, 2019].

Segundo [Meng et al., 2020] o processo de FIMS é considerado dinâmico. Pode-se observar que a complexidade da fusão de dados é causada não apenas pelo tipo de dados e ambiente de coleta, mas também pela *temporalidade*. Para estimar um estado do sistema, especialmente em um contexto de variação de tempo, os dados podem ser significativos apenas por um intervalo limitado. Esse desafio deve ser bem mapeado em um ambiente de aplicação em tempo real. Os centros de fusão devem ser capazes de distinguir a validade dos dados e a correta ordem deles. Vai-se explicar sobre a natureza dos dados, inclusive em seus aspectos espaço-temporais.

Num ambiente corporativo da atualidade, seja de qualquer indústria ou segmento, é gerada uma grande quantidade de dados, em diferentes estruturas e formatos. No tocante à natureza dos dados, tais dados podem ser divididos de acordo com as dimensões espaciais e temporais como dados com atributo temporal, dados com atributo espacial e dados com atributos temporais e espaciais(dados espaço-temporais), conforme a figura 3.3.

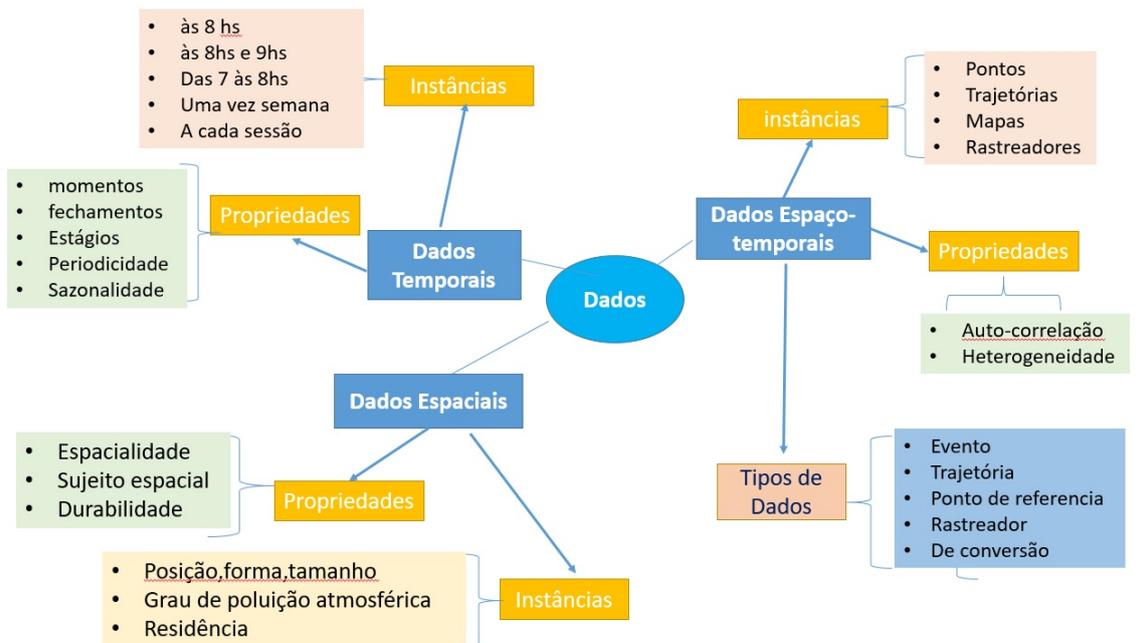


Figura 3.3: Dados espaço-temporais adaptado de [Liu et al., 2020]

A revisão feita em [Khaleghi et al., 2013] aborda as vantagens da fusão de dados, em especial na autenticidade de dados, o que permite melhora da detecção, aumentando a confiança e confiabilidade, reduzindo a ambiguidade nos conjuntos de dados. No estudo de [Liu et al., 2020] a avaliação espacial e temporal é determinante para se obter disponibilidade. Os dados com propriedade espacial descrevem um objeto estático, que geralmente não muda com o tempo em um curto período, como um endereço, um edifício, uma rua. Já a propriedade temporal dos dados é representada por unidades de tempo, como dia, hora, minuto ou segundo. As propriedades desses dados são momento, proxi-

midade, periodicidade, sazonalidade e estágio. Assim, os dados das seis horas mostram um momento. Caso haja uma relação entre os dados de seis horas com os dados das oito horas, ocorre uma proximidade. A periodicidade está relacionada com a frequência, como uma vez por dia, hora, semana. Os dados das seis às nove horas mostram o estágio. Os atributos temporais são fundamentais para estudar a evolução e desenvolvimento de fenômenos. Dependendo do caso, os dados de cada estação mostram sazonalidade.

O desenvolvimento de objetos em diferentes campos é essencialmente temporal e espacial. Dessa forma, a maioria dos dados é gerada no processo de criação dos objetos desenvolvidos com propriedades temporais e espaciais, denominadas espaço-temporais, que reúnem as propriedades da correlação e heterogeneidade [Atluri et al., 2018].

A correlação é definida em [Liu et al., 2020] como observações feitas no espaço e no tempo. Quando correlacionadas entre si, é caracterizada uma situação de dependência, que pode gerar fenômenos de causalidade. Assim, é fundamental explicar a estrutura da auto-correlação entre observações ao analisar dados espaço-temporais. A heterogeneidade é comprovada pela alta variabilidade dos tipos e formatos de dados. Os conjuntos de dados usados para estudar um ambiente hospitalar, por exemplo, reúnem dados de emergência, numa série de diferentes tipos e formatos, e provavelmente oriundos de diferentes fontes.

A FIMS é baseada em um processo de níveis adaptados envolvendo agregação, mineração de dados e integração de informações de múltiplas fontes, resultando num produto de informação de maior valor do que qualquer parte isolada [Blasch and Plano, 2005]. O grupo DFIG introduziu o modelo de fusão de usuário, conforme apresentado na figura 2.2 afirmando a importância de apresentar os resultados do FIMS para representar o conhecimento a partir de uma estimativa de fusão, bem como o raciocínio do conhecimento para gerenciamento e controle.

A recomendação de uso de métodos de probabilidade em FIMS pode ser vista em [Esteban et al., 2005] onde são referenciados os modelos de arquitetura de FIMS, desde o JDL até o modelo cascata, passando pelo modelo omnibus. Os autores propõem uma abordagem baseada em sistema, num modelo que privilegia a estimação entre duas outras etapas, identificação e validação.

O estudo de [Liu et al., 2020] alerta que no método de combinação direta de recursos, vários problemas precisam ser observados. Inicialmente, ao se mesclar diretamente os dados, é necessário remover recursos duplicados. Após isso, deve-se cuidar de algumas

características de diferentes dimensões, que fazem com que o modelo tenha bom desempenho, podem ser perdidas devido à fusão direta. Finalmente, a fusão direta de recursos pode resultar num viés no modelo (overfitting).

O método de fusão de dados baseado em etapas [Zhu et al., 2018] divide o problema em diferentes estágios. A cada etapa, por meio dos dados deste estágio e, finalmente, as saídas de cada estágio do processo são mescladas.

Os trabalhos [Liu et al., 2020], [Zheng, 2015] e [Kumar and Daumé, 2011] utilizaram método de fusão de dados baseado em similaridade porquanto [Shang et al., 2014], [Katz et al., 2019], [Shang et al., 2014] utilizaram método de fusão de dados baseado em dependência probabilística. O método de fusão de dados baseado em dependência probabilística utiliza estrutura de grafos. Ele toma dados diferentes como nós em primeiro lugar, e então a relação entre os dados, como causalidade, é medida pela borda, que pode ser direcionada ou não. Isso permite a construção de um gráfico, e alguns métodos de estudo da estrutura do grafo podem ser usados para fundir os dados.

A FIMS privilegia uma visão cruzada, não só dos dados, mas também na apresentação, no método visual. O método de fusão de dados baseado em multi-visualização estuda um objeto a partir do conhecimento de diferentes pontos de vista [Zheng et al., 2015].

Se na segunda questão de pesquisa dessa revisão de literatura analisou-se a questão de multidimensionalidade e ruídos como dificuldade para captura e reforço de técnicas para análise de sentimento e emoções e representação de palavras, com uso intenso de PLN, a resposta à terceira questão dessa revisão trouxe a questão do processo de fusão de informação. A FIMS trabalha a incerteza, utiliza raciocínio probabilístico para lidar com os problemas de imprecisão, redundância e ambiguidade. Interessante o potencial de reunir elementos gráficos para realizar inferências e fusão de dados. A DST pode ser um aliada na geração de evidências para a formação de hipóteses baseadas em métodos de aprendizado de máquina. Isso traz mais um tópico de pesquisa, o processo de geração de evidências que equivalem a BPA, que será explanado na próxima seção.

### **3.6 Resposta à quarta questão - Que estruturas de raciocínio podem ser aplicadas para mitigar a incerteza pelas relações de dependência na formação das hipóteses em processos de fusão de dados?**

Se as três primeiras questões de pesquisa desta revisão de literatura cobriram exclusivamente o domínio de RAM, nesta quarta questão o olhar direcionou-se mais para questão da incerteza de uma forma geral, pois o estudo desse tema ficaria muito limitado se fosse

restrito ao domínio de RAM.

Esse tópico de estudo surgiu na pesquisa da resposta à segunda questão, onde, em especial nos métodos para EHR a formação de conjuntos de dados a partir de um processo de fusão de dados, propicia a necessidade de tratamento de incerteza.

O estudo de [Florea and Bossé, 2009] realizou uma revisão dos conceitos de critérios para estimar a confiabilidade: estimativas, distâncias e similaridade. Também abordou a análise temporal do comportamento dos sensores, comparando diferentes métricas. Ele classificou os sensores em simples, complexos e híbridos, alegando a relevância de tratá-los de forma diferenciada.

O crescimento a partir de 2015 de empregos de FIMS usando ML combinado com DST é notório. O forte uso de ML e DST promove melhores métricas de precisão e amplia o leque de possibilidades de geração de BPA, utilizando diversas técnicas e algoritmos combinados pelo uso de *ensemble learning*.

O gráfico mostrado na figura 3.4 demonstra o uso de métodos probabilísticos e métodos ML com DST. Os métodos com ML utilizam métodos de cluster, SVM, neurais e ensemble e vêm superando os estudos probabilísticos sobre o tema. O uso de cluster e SVM aparece em maior uso do que métodos neurais, como as principais técnicas de inteligência artificial. Também é possível observar o forte uso de *ensemble learning* com DST, em que são combinados resultados de diferentes classificadores. Há uma tendência de uso de classificação binária e agrupamentos na indústria de diagnóstico de falhas, o que justifica o alto uso de SVM e *clustering* neste segmento.

..

### **3.6.1 Iniciativas para reforçar a confiança e credibilidade**

Surgem iniciativas para atribuição de confiança e credibilidade às evidências baseadas em dados, como abordado no trabalho de [Schubert, 2011], que criou um experimento que determina um *grau de falsidade*, a fim de mitigar conflitos por meio de uma abordagem que avalia sinais falsos entre fontes e sensores, ora pela distância, ora pela qualidade do quadro recebido.

Também utilizando uma abordagem de discordâncias por distância, [Yang et al., 2013] utilizaram a distância de Jusselme e propuseram um novo método, aprimorando as ideias do *grau de falsidade de Schubert*.

[Tabassian et al., 2012] introduziram uma nova abordagem de classificação supervi-

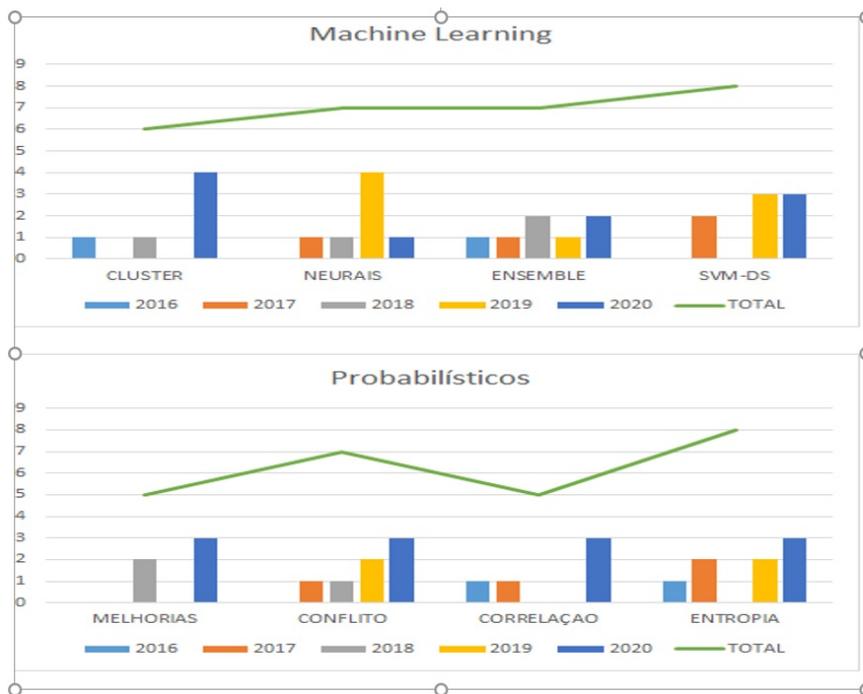


Figura 3.4: Evolução de publicações FIMS com DST

sionada usando *ensemble learning* e DST para lidar com rótulos de dados imperfeitos. O experimento avalia a incerteza em cada rótulo de amostra de aprendizagem, aplicando um procedimento de re-rotulagem a partir de protótipos de classe pré-definidos.

Existem iniciativas baseadas na resolução de conflitos por meio de critérios de otimização [Chen et al., 2014]. A redução de dimensionalidade com base numa proposta de agrupamento de análise de componentes principais recursiva (R-PCA) que agrega dados de sensores redundantes e ajuda a detectar discrepâncias é encontrada em [Yu et al., 2017].

Uma nova proposição para o cálculo da incerteza é proposta em [Yang and Han, 2016], calculando a BPA por meio da distância média entre o intervalo de crença de cada elemento focal em relação ao elemento com maior incerteza.

O uso de descontos para resolução de conflitos pode ser visto em [Zhao et al., 2016]. Neste trabalho, é proposto um método de combinação aprimorado para evidências conflitantes com base em medidas inconsistentes. Inicialmente, uma nova abordagem é proposta para medir o conflito entre duas evidências. A evidência conflitante é então revisada selecionando-se os coeficientes de desconto (coeficientes de conflito combinados com a DST) calculados pelo peso médio da evidência.

### **3.6.2 Associação, combinação na análise de dados - trabalhando a condicionalidade e temporalidade**

Existem estudos que trabalham causalidade e temporalidade a partir das associações e combinações de dados. Por exemplo uma estimativa de BPA criada a partir de um algoritmo de otimização para seleção de modelos foi proposta por [Wang et al., 2017]. A otimização teve como objetivo selecionar modelos de previsão de deslizamento. Os aspectos relevantes avaliados para o modelo foram paisagismo, confiabilidade, segurança, economia razoável e viabilidade técnica. Esses aspectos foram combinados usando pesos como fatores de avaliação. O estudo então determina graus de incerteza correspondentes a diferentes níveis usando um modelo de otimização baseado em decisões multicritério. Em seguida, as funções de massa (BPA) de diferentes índices.

A condicionalidade e a dependência foram abordadas por [Khakzad, 2019], que utilizou redes Bayesianas com DST criando uma rede de evidências. Esse trabalho se transforma em conhecimento *a priori*, usa condicionalidade com redes Bayesianas, calcula probabilidades, para evitar imprecisões em funções de massa e propagar a atualização para gerar as funções de crença. Num veículo conectado inteligente (*veículo conectado inteligente-ICV*) o algoritmo Dynamic Bayesian Network (DBN) foi usado em [Zhang et al., 2018a] para detecção precisa de falhas ao adquirir as correlações espaciais e temporais dos dados do veículo. No ano seguinte a esse estudo, parte desse grupo de pesquisa acrescentou aspectos temporais e ampliou o método DBN, criando um esquema de detecção e reparo de falhas baseado em limites [Zhang et al., 2019b].

Dois estudos despertaram para a importância de investigar temporalidade. O estudo de [Jiang et al., 2017] apresentou uma estrutura de evidências para reduzir a incerteza baseada em fusão para a condição de avaliação com o objetivo de melhorar a precisão das decisões de avaliação. O método mescla recursos extraídos de diferentes fontes de evidências e constrói FOD a partir de séries temporais. O trabalho de [Pippa et al., 2018] apresentou um modelo de fusão de dados no nível característico combinado com o uso de séries temporais multidimensionais. Eles usaram análise espectral para extração de características e três esquemas de fusão para combinar informações entre sensores para avaliar convulsões e eventos semelhantes, classificando os registros de sinais elétricos emitidos pelo cérebro em um exame encefalograma.

### **3.6.3 Combinação de técnicas Bayesianas, Fuzzy e DST para melhora de precisão**

Abordando um cenário de imprecisão sobre o conhecimento da dependência entre as diferentes variáveis que compõem um modelo probabilístico, o trabalho de [Zhang et al.,

2017] observa escassez de dados devido a essa imprecisão ou falta de conhecimento.

No intuito de angariar melhor precisão pela redução de incerteza, surgiram estudos que utilizam os raciocínios Bayesiano, Fuzzy e baseado em DST conforme descrito no trabalho de revisão de [Verbert et al., 2017]. Na subseção anterior, viu-se como os métodos que realizam aproximação para geração de BPA. Há várias abordagens possíveis para a geração da BPA a partir de conjunto de dados utilizando técnicas de aprendizado de máquina. Em geral, a técnica é utilizada para a geração das evidências, que vão gerar a BPA e formar o quadro de discernimento DST. Entretanto, faz-se necessário todo o cuidado no pré-processamento, em especial para não incorrer no erro de considerar as fontes e variáveis independentes.

### **3.7 Críticas a BPA e origem das limitações de Dependência**

Os estudos [Monney and Chan, 2007] e [Monney et al., 2011] estabelecem uma relação reflexiva das fontes de informação com o BPA para definir a dependência interna. O estudo de [Guralnik et al., 2006]) propôs um método para dividir a evidência entre dependência de partes independentes e dependentes, levando as estruturas de crença a reduzir a dupla contagem ao mesclar informações. A pesquisa de [Fu and Yang, 2014] descontou a elaboração de funções de crença utilizando pesos positivos e negativos para minimizar a influência da dependência.

Esses métodos têm limitações para [Mao et al., 2020], que definem qual regra de correspondência proposta por Cattaneo precisa essencialmente analisar conflitos de evidências em vez de dependência. A Regra Cautelosa e a Regra Disjuntiva Ousada, representadas por [Destercke and Dubois, 2011], não podem lidar com evidências que são incapazes de se desintegrar. A desvantagem de [Monney et al., 2011] é que é difícil representar informações de incerteza na hierarquia das fontes de informação. Enquanto isso, o método de [Yager, 2009] não calcula o grau de dependência e o cálculo é muito complexo quando se trata de grandes volumes de evidências. O método de [Su et al., 2014] efetivamente equivale à dependência, mas os elementos são necessários nas estruturas. Foi definido em [Mao et al., 2020] que a regra de combinação de Dempster é amplamente utilizada no sistema de fusão de dados, alegando que é difícil satisfazer a suposição de independência entre diferentes evidências, pois existem muitos fatores influenciadores. Em [Su et al., 2018], discute-se que as evidências coletadas de diferentes fontes de informação para serem usadas num sistema de fusão de informações geralmente possuem alguma relação de dependência. No entanto, a regra clássica de combinação de Dempster

só pode ser aplicada a casos em que a evidência é independente. Essa suposição é muitas vezes irreal e pode levar a decisões irracionais. Ele afirma que a dependência entre as evidências é ignorada e a regra de correspondência de Dempster é aplicada diretamente ao resultado da mesclagem, sob o risco de uma decisão irracional e equivocada.

O trabalho de [Xiao and Qin, 2018] mostrou um método de combinação ponderada para evidências conflitantes em FIMS ao considerar a interação entre a medição de evidências. Isso foi feito usando o cosseno de similaridade para estabelecer um grau de confiabilidade para a geração da BPA.

De acordo com [Zhang and Shields, 2020], as entradas do sistema são frequentemente consideradas mutuamente independentes ou correlacionadas por uma *distribuição gaussiana multivariada*. No entanto, essa suposição subjetiva pode influenciar a estimativa de resposta se a estrutura de dependência real se desviar dessa suposição. Eles a propuseram, como método de quantificação e propagação da incerteza associada à modelagem de dependência, principalmente quando os dados para caracterizar os modelos de probabilidade são limitados.

### **3.7.1 Dependências- causalidade - usando Bayesiano conjugado com DST**

Em [Song et al., 2020], baseado numa estrutura de rede Bayesiana, é proposto um método de análise de dependência, que combina aritmética afim (linear) e método P-box para lidar com incerteza híbrida e dependência não determinística.

Coefficientes de desconto podem ser gerados com base em grau de evidências dependentes. Esse recurso foi utilizado no estudo de [Su et al., 2018] que propôs um método para combinar evidências dependentes baseado na técnica de *informação mútua*. O método finaliza aplicando um modelo de agregação baseado no coeficiente de desconto total e a combinação da regra de Dempster.

Vários métodos tratam dependências de algumas forma, com a criação de redes de evidências dependentes [Lee, 2019], através de mecanismos de inferência Bayesiana [Song et al., 2020], por meio de cálculo de distâncias, por transformação pignística de probabilidade [Mao et al., 2020], combinando Bayesiano com DST [Verbert et al., 2017].

O estudo de [Wang et al., 2019] usa uma estrutura de rede para reduzir a incerteza das opiniões de especialistas por meio de um método de implementação em vez de metas. Executa uma análise de sensibilidade para verificar as regras de agregação. Cria um modelo de propagação de confiança que integra diferentes tipos de inferências. Define as métricas de confiança de segurança gerando a BPA, em seguida, propaga a confiança

usando a regra de Dempster. O método aborda o problema de reduzir a incerteza num ambiente de fusão de informações com base na opinião de especialistas. Ele implementa um método de implementação utilizando ensemble baseado em fuzzy, criando uma métrica de credibilidade com base em pesos de evidência atingindo precisão superior a 95%.

Os estudos que abordam mais adequadamente as relações de dependência podem ser organizados em duas categorias principais. Um primeiro grupo de trabalhos que utilizam DST e rede bayesiana abordando dependências evidentes e um segundo grupo que aborda a criação de processos de atribuição de confiança às evidências, implementando descontos na geração da BPA.

### **3.8 Discutindo os resultados desta revisão de literatura**

A revisão de literatura sobre FIMS chegou a quatro revisões anteriores sobre RAM: [Khaleghi et al., 2013], [Meng et al., 2020], [Steinhauer and Karlsson, 2019] e [Liu et al., 2020]. Quanto ao uso de fontes alternativas para a detecção de RAM, o ponto de partida desse estudo foi o conjunto de três revisões de literatura anteriores. A primeira revisão [Sarker et al., 2015] investigou métodos automatizados de extração de RAM a partir de MSD. A revisão investigou 69 artigos nas bases de dados Medline, Emase, Scopus e Web of Science e Google Scholar. Este estudo concluiu que sites e blogs relacionados à saúde contêm dados mais relevantes do que informações postadas em redes sociais. Por sua vez, o volume de dados de MSD é significativamente maior. Eles observaram que apenas seis estudos ( 9,5%) disponibilizaram suas bases rotuladas publicamente. O estudo destaca que a maioria das abordagens de extração de RAM conta com o desafio de usar léxicos para identificar ou extrair menções de RAM em textos obtidos em MSD. Ele destacou a importância do uso de PLN para mitigar a dificuldade de captura no Twitter, a criação de *corpus* e algoritmos de aprendizado de máquina, ressaltando, porém, que a maioria dos classificadores pesquisados utiliza apenas linguísticas básicas como *bag of words (BoW)* . Outra fragilidade apontada por essa revisão, ocorre em mais de 80% dos estudos observados e consiste na ausência de análises estatísticas nos estudos que extraíram RAM.

A segunda revisão [Tricco et al., 2018] apresentou duas questões de pesquisa abordando quais plataformas de MSD existem para detectar RAM e qual a validade e confiabilidade dos dados em MSD para identificação de RAM. Foram revisados 70 artigos publicados entre 2001 e 2016 nas bases de dados MEDLINE, EMBASE e Cochrane Library. Desses 70 artigos, 46 (66%) descreveram um sistema de informação automatizado para detectar RAM a partir de MSD. Existem muitos sistemas e bancos de dados preexis-

tentes sobre RAM, como o FAERS, mas apenas um quinto dos trabalhos checam as RAM contra registros cadastrados nesses bancos de dados. Isso compromete a credibilidade e a validação dos resultados obtidos. Esta revisão observou que os métodos são muito incipientes, com resultados muito variados e que os estudos não abordam a validade e confiabilidade do método como um todo.

A terceira revisão [Wong et al., 2018] realizou uma narrativa respondendo à pergunta sobre quais bancos de dados estão disponíveis para extração de RAM. O estudo aborda a disponibilidade de dados dessas fontes, uma vez que há um importante uso de algoritmos computacionais para ajudar a detectar automaticamente associações entre medicamentos e RAM, destacando recursos de PLN.

Foi atingido o objetivo da revisão de literatura desta tese, de se conhecer os métodos automatizados de extração de RAM abordando fontes de dados, seja em mídia social digital (MSD) seja a partir de registros eletrônicos de saúde(EHR) sob o prisma da evolução de algoritmos e técnicas de classificação e análise de sentimentos. Identificaram-se desafios nesse tipo de pesquisa, como: a alta dimensionalidade e dispersão desse tipo de fonte de dados; a precária disponibilidade e compartilhamento de bancos de dados produzidos por pesquisas sobre o tema; dificuldades de comunicação entre os sistemas de registro de países que prejudicam a interoperabilidade entre suas agências reguladoras; a necessidade de tratar os grupos de drogas de forma específica na construção do corpus. Esses fatores tornam necessária a integração de classificadores com léxicos, dicionários e "corpus". É preciso investigar sobre a atualização dos métodos usados para extração de RAM e como os dados e os diferentes léxicos são usados neste campo de pesquisa. Também é importante observar que com o aumento do volume de pesquisas utilizando dados obtidos em MSD e EHR é importante cuidar da ética e privacidade quanto ao uso de dados das pessoas que postam suas opiniões ou dos pacientes e dados sensíveis envolvidos.

Foram constatados os resultados gerados por revisões anteriores, que apontavam a falta de padronização nos resultados, várias possibilidades de experimentos, falta de compartilhamento de bancos de dados, dificuldade na geração desses conjuntos de dados e uma análise estatística insuficiente.

Os artigos abordam o MSD ou EHR como fonte de dados e há uma predominância de trabalhos que utilizam o Twitter e o uso da análise de sentimentos tem sido ampliado com base no uso de léxicos e tratamento de emoções e até sinais. No início dos tempos, os esforços eram no aprendizado não supervisionado, com análise de sentimento avaliando a polaridade de uma mensagem ou sentença(*post*) como positiva, negativa ou neutra, mas permite uma alta ocorrência de falsos positivos.

A maioria dos trabalhos investigados cria suas próprias bases de estudo, geralmente apoiadas em léxicos específicos, e menos da metade utiliza as bases disponíveis. Apenas 20% dos trabalhos verificam suas extrações candidatas contra RAM contra bases pré-existentes. As comunidades têm uma credibilidade mais aguda do que as redes sociais, pois, em geral, há uma organização com curadores e especialistas. A classificação de postagens e emoções em tweets permite a análise de hashtags e emoticons que podem gerar termos para a representação de palavras (embedding).

Usando um modelo não supervisionado, o artigo de [Nikfarjam et al., 2015] afirma que a técnica de extração automática de RAM é escalável e adequada para mineração de mídia social. Tal técnica poderia atuar em grandes volumes de dados não rotulados, o que diminuiria a necessidade de grandes conjuntos de dados de treinamento anotados.

Mas os resultados dos estudos mostram que para obter melhor acurácia, é necessário utilizar o processo no modelo Supervisionado. Assim, as notações de rótulos são feitas por especialistas, o que costuma ser caro num contexto de grande volume como o do MSD. Isso torna os conjuntos de dados rotulados em grande escala escassos, demorados e caros para obter. Sistemas de crowdsourcing são uma boa alternativa para obter anotações de forma distribuída, pois são uma forma eficiente de agilizar o processo de coleta de dados de treinamento por diversos anotadores, formando uma base de anotações de volume a um custo acessível. Nessa linha, destaca-se o trabalho de [Sintsova and Pu, 2016b] que cria o “Dystemo“, método supervisionado à distância e que atribui ao reequilíbrio dos dados inicialmente rotulado como um passo essencial para o sucesso do método proposto, que contribui com a introdução de heurísticas para encontrar automaticamente tweets neutros, demonstrando a importância de incluí-los no processo de aprendizagem. A pesquisa de RAM vem evoluindo com o uso de dados rotulados ou alternativamente por meio de ferramentas que permitem o processo semi-supervisionado baseado em léxicos que melhoram a precisão dos resultados. É uma tendência a uma forte necessidade de revisão e pré-processamento, para uma melhor análise de sentimentos, composições (“ensembles”) de algoritmos e léxicos, muitas vezes agregando aprendizado profundo.

É relevante que o trabalho no modelo supervisionado inclua exemplos neutros e negativos na formação dos dados, como mostra o trabalho de [Masino et al., 2018] que visou o reconhecimento de RAM rara. Quando usado com um texto de entrada que é conhecido por conter RAM, foi obtido um bom desempenho, mas os resultados diminuem com altas taxas de falsos positivos quando as anotações não continham exemplos de RAM.

Outro desafio na pesquisa de RAM a partir de MSD está no campo semântico no nível da linguagem e agrupamento no que diz respeito à família da droga envolvida. O

estudo de caso em diabetes feito no trabalho de [Liu and Chen, 2015] que observou que a relevância dos tweets para a farmacovigilância depende muito do medicamento considerado. Nesse estudo, metade das potenciais RAM detectadas refere-se a uma única droga, "vareniclina", e nenhum tweet descreveu potenciais RAM para 12 outras drogas selecionadas para o trabalho. Um bom classificador para medicamentos do grupo opioide pode não levar a bons resultados para um conjunto de dados formado por medicamentos do grupo cardíaco. Um exemplo de que RAM psicológicas de problemas funcionais cognitivos têm maior variabilidade semântica quando comparadas a RAM fisiológicas está no estudo de [Zolnoori et al., 2019b].

Ao ampliar essa questão de especificidades sobre drogas, idiomas e costumes, pode-se perceber que isso causa um desafio ao problema da interoperabilidade, apontado na pesquisa de [Wong et al., 2018]. Uma iniciativa criada para mitigar esse problema pode ser atribuída a "AEOLUS" no artigo de [Banda et al., 2016], uma iniciativa pioneira que começou a permitir que pesquisadores a vinculassem a outras ontologias como a Classificação Internacional de Doenças (CID) e FAERS, depurando inconsistências, integrando com SNOMED-CT e MedDRA utilizando RxNorm, enfim, utilizou uma gama de ferramentas comprovando que é possível evoluir na padronização de bases. Lembrando que existe uma barreira mais difícil que é a do idioma já que os medicamentos são usados em todo o mundo. No entanto, as agências reguladoras são de cada país, o uso de recursos que facilitam a padronização como UMLS, Metamap e o uso de bases oficiais se torna ainda mais relevante.

Uma importante avaliação de aspectos relacionados às questões de legislação em farmacovigilância no mundo, assim como questões relacionadas à privacidade dos dados dos autores MSD, pode ser vista em [Pappa and Stergioulas, 2019]. Essas são questões relevantes, a primeira, sobre a regulação influencia a interoperabilidade, a integração de *datasets*. E a privacidade é uma questão que envolve vários aspectos, inclusive a ética. E isso afeta a pesquisa à medida que mais MSD são usados, o que é uma tendência. Vale ressaltar que o RES também oferece riscos ainda maiores no que diz respeito ao possível desrespeito à confidencialidade dos dados do paciente.

O maior desafio na pesquisa de RAM com MSD esteja no discurso informal das mídias sociais, sem preocupações sintáticas, muitas vezes cometendo exageros ou descumprimento de pontuação. Assim como costumam envolver aspectos emocionais na fala, abordando desafios à análise de sentimentos, que vem recebendo investimentos em estudos sobre sarcasmo, tensão e relaxamento. O sarcasmo é uma forma especial de sentimento que o define como "uma forma sutil de linguagem em que os indivíduos dizem o contrário do que está implícito", conforme definido no estudo realizado por [Samonte et al., 2018].

Esforços de pesquisa na linha de reconhecimento de imagens postadas *emojis e emoticons* também têm sido explorados. Tais estudos analisam o comportamento contextual do usuário, semelhante ao papel de *tweets* anteriores de um usuário ou aos níveis gerais de sarcasmo de uma determinada discussão. Essa forma de apelo pode ser vista nos estudos de [Dandala et al., 2018] e [Tang et al., 2018], que reúnem léxicos criados a partir dessas técnicas.

Mesmo a capacidade de detectar sinais de estresse e relaxamento humano na condução das postagens foi considerada fundamental para o diagnóstico oportuno de doenças relacionadas ao estresse no trabalho [Gopalakrishna Pillai, 2018]. O estudo afirma que os métodos tradicionais empregam escalas de medição de estresse ou monitoramento fisiológico que podem ser intrusivos e inconvenientes. Por outro lado, a natureza onipresente das mídias sociais pode ser aproveitada para identificar sinais de estresse e relaxamento. Pode funcionar para avaliar esse estudo para a pesquisa de RAM em MSD.

Algoritmos tradicionais de classificação ainda são amplamente utilizados na busca de RAM, como NB e SVM. Mas é notório o crescimento do uso de *ensemble*, aumentando a publicação de trabalhos utilizando a combinação de léxicos específicos com métodos de aprendizagem normal, geralmente recorrentes ou profundos. A característica dos “embeddings” usando CRF e REN combinados e somados ao poder dos classificadores neurais, sejam eles recorrentes ou convolucionais, com destaque para o LSTM, bidirecional ou não [Masino et al., 2018], [Chen et al., 2018a], [Wunnavu et al., 2018].

Em resumo, viu-se que o processo semi-supervisionado utilizando MSD por meio de plataformas ao máximo colaborativo, a importância do corpus específico de acordo com o grupo de medicamentos, a interoperabilidade como objetivo incentivando a disponibilização de conjuntos de dados originados em estudos e o registro de descobertas em sites oficiais e principais bases de dados o respeito à ética e à privacidade com a tendência de maior utilização de MSD e RES são os pilares que ajudarão a combater a sub-notificação de RAM.

### **3.9 Trabalhos relacionados**

Vai ser apresentada uma visão geral dos principais estudos desenvolvidos para lidar com incertezas decorrentes de dependências. Nota-se uma ênfase no uso da rede bayesiana. Vários métodos teóricos foram utilizados, como abordagens probabilísticas, fuzzy e bayesiana. Nossa pesquisa contribui para a criação de um método que utiliza relações de dependência temporal e probabilística num contexto multisensores e de interação entre os

elementos. Os seis trabalhos a seguir serviram de inspiração para esse método.

O estudo de [Martí et al., 2015] define que, para se encontrar padrões de frequência, o objetivo primário de segmentação de séries temporais é a redução de dimensionalidade. Esse trabalho realiza método de detecção de anomalia em turbo-máquinas empregando SVM e séries temporais. Em [Tian et al., 2018] é usado um mapa de probabilidades e feito um estudo longitudinal com um modelo de detecção de alterações em imagens utilizando séries temporais de imagens e DST. O objetivo é avaliar mudanças em construções em 3D.

O trabalho de [Zhang et al., 2019b] utilizou séries-temporais e mencionou redes Bayesianas para tratar de relações de dependências nos dados.

Outro estudo que utilizou séries temporais é o de [Khan and Anwar, 2019], no qual um algoritmo de oito passos com nova função entropia é proposto. Usou-se rede neural convolucional CNN para BPA, com critérios temporais para resolução de conflitos. Dados conflitantes de etapas de tempo recebem pesos mais baixos em comparação com dados de etapas de tempo em que há menos discordância entre os sensores.

O estudo de [Khakzad, 2019] identificou condicionalidade e dependência entre as fontes de dados, que utilizaram a rede Bayesiana com DST, criando uma rede de evidências. Esse método utilizou o conhecimento a priori possibilitado pelo uso de redes Bayesianas. As vantagens deste método são identificar condicionalidades e evitar imprecisões nas funções de massa e conseqüentemente nas funções de crença. Esse estudo não trabalhou com requisitos temporais e de interações entre fatores.

Um novo método para lidar com a dependência de evidências em DST e dividido em duas etapas é proposto em [Mao et al., 2020]. Inicialmente, são quantificados os pesos relativos dos componentes do sistema, considerando suas interações. Numa segunda etapa, a dependência externa é descrita, utilizando fractais utilizando recursos de *transformação pignística de probabilidade*. Por meio da análise de sensibilidade, são demonstradas as vantagens da proposta em lidar com evidências dependentes, melhorando a avaliação da perda de informações. A desvantagem é que não lida com dependências de BPA entre hipóteses, apenas entre fontes.

O estudo de [Su et al., 2018] propõe um método para combinar evidências dependentes baseadas em *informações mútuas* entre fontes. Medir o grau de dependência entre as evidências, gerando uma matriz de coeficiente de desconto para ajustar as incertezas das hipóteses para corresponder à regra de Dempster. Os resultados obtidos são de acurácia entre 92 e 94% com o conjunto de dados IRIS, utilizando os métodos de Pearson e Ken-

dall estreita correlação. A vantagem do método proposto é que a correlação pode lidar com casos lineares e não lineares que são mais comuns na realidade. A desvantagem é que não implementa inferência causal.

O trabalho de [Song et al., 2020] usa uma estrutura de rede Bayesiana e propõe um *método de análise de dependência não-determinística* que usa *aritmética afim*. A vantagem pode produzir *p-box* de confiabilidade resultante mais restrita do que as obtidas por *desigualdades de Frechet* e é mais econômica do que o *2-estágios método Monte Carlo*. A desvantagem é que não considera causalidade, temporalidade, correlação e interações, restringindo o uso a problemas mais complexos.

Criando um método de combinação retrátil que atua numa rede de evidências que forma uma estrutura de discernimento DST, a pesquisa de [Lee, 2019] produz uma matriz de coeficientes e ajusta as antigas funções de crença. A vantagem é ajustar os valores de crença que eventualmente apresentem valores negativos para reverter o efeito da combinação indesejada de evidências. Quando a evidência regular é combinada, os valores de crença negativos são usados para influenciar os valores de crença de forma oposta. As desvantagens são que não trabalha com causalidade, temporalidade, interações. Não determina o grau de dependência do BPA.

Uma abordagem multi-modelo hierárquica bayesiana é proposta em [Zhang and Shields, 2020] para quantificar a incerteza na forma de modelo de dependência e parâmetros de modelo que resultam de conjuntos de dados sm3all. Essa abordagem identifica um conjunto de modelos candidatos marginais e suas probabilidades de modelo correspondentes usando o modelo de inferência Bayesiana. Em seguida, estima a incerteza numa estrutura de dependência baseada na técnica de cópula condicional sobre as marginais e seus parâmetros. O modelo cria uma nova abordagem de reponderação de amostragem para propagar incertezas de forma eficiente. Ele começa com um modelo probabilístico baseado em *p-box*. Em seguida, realiza uma abordagem linear para uma rede Bayesiana de dependência não determinística e compara os resultados com a desigualdade de Frechet. A vantagem é estimar a incerteza na quantidade de interesse dadas as distribuições de entrada do modelo múltiplo candidatas a um baixo custo computacional em comparação com as típicas simulações de Monte Carlo aninhadas. A desvantagem é a complexidade computacional.

Os trabalhos identificam a dependência de diferentes formas, mas não integram as dimensões causal, correlacional, temporal e interação entre os fatores. O uso da inferência bayesiana em conjunto com a DST é um aspecto relevante. A correlação entre os fatores é explorada com maior profundidade em poucos estudos e, em particular, a dependência

temporal é pouco investigada, possivelmente pela necessidade de um estudo longitudinal. O modelo proposto nessa tese aborda os aspectos de interação, correlação, dependência causal e temporal, integrando-os em um índice de credibilidade para gerar evidências de forma mais precisa.

### 3.10 Comentários finais sobre a revisão de literatura

A pesquisa de RAM percorreu um longo caminho nos últimos cinco anos, mas ainda há muito a avançar. A evolução do processo não supervisionado para o semi-supervisionado alavancou o uso de um ensemble de classificadores. Essa combinação de algoritmos de classificação normalmente atuam em “corpus”, léxicos e “datasets” para tarefas de análise de sentimentos. Este trabalho faz uma atualização sobre o formulário de pesquisa, os algoritmos e categorias de combinação de classificadores utilizados na extração automática de RAM.

A oferta de novos experimentos e ferramentas, a disponibilidade de *datasets* e a formação de *corpus* continuam sendo um desafio. A transição do processo não supervisionado para o supervisionado encontrou um equilíbrio entre o dispêndio de esforço e o benefício de melhores resultados no processo semi-supervisionado.

Apesar da notável evolução do PLN e da consequente geração de léxicos, observou-se que classificadores de RAM atribuídos a um determinado grupo de medicamentos podem gerar um resultado de precisão diferente se submetidos a outro “dataset” referente a outro grupo de medicamentos. Isso porque os sintomas, sentimentos, reações são diferentes, havendo a necessidade de compor léxicos e modelos específicos para melhores resultados.

O grande benefício do volume do MSD, que permite o enriquecimento do léxico e consequentemente uma análise mais precisa dos sentimentos, possibilita a produção de pesquisas com melhores resultados. Esse é um grande recurso no combate ao problema da dimensionalidade.

O problema da autenticidade e credibilidade das opiniões no MSD e EHR, a precária disponibilidade de “datasets” e a falta de análises cronológicas de interação medicamentosa continuam sendo um campo a ser explorado, abrindo oportunidades de pesquisa para trabalhos futuros.

## **4. DEP-DST: Método de geração de evidências com ajuste de dependências**

Esta seção descreve a abordagem DEP-DST em um contexto genérico para a geração de evidências DST e em seguida demonstrar a aplicação desse modelo no domínio de RAM, em um caso específico de um hospital público federal que trata câncer. Primeiro, descrevemos o modelo de arquitetura e as variáveis de componentes para integrar um IC. Em seguida, descrevemos os materiais e métodos experimentais que implementaram o modelo no domínio de RAM e, por fim, detalhamos o processo FIMS. Esta proposta apresenta um modelo que aborda o problema dos relacionamentos de dependência, identificando-os por meio de técnicas de análise correlacional, causal e temporal e integrando-os na proposição de um método que gere um índice de credibilidade para uso na geração de crenças em um contexto de DST em FIMS.

### **4.1 Modelo Arquitetônico**

A figura 4.1 descreve o modelo criado com base em um processo FIMS. As etapas do processo de fusão de dados contribuem para criar o índice de credibilidade. O nível zero do modelo de fusão DFIG cuida do pré-processamento e organização da captura. A identificação do objeto de nível 1 gera o centro de fusão 1, com a associação das fontes de dados primários e sensores, permitindo conhecer as interações entre os elementos do domínio que permitirão a geração da variável tamanho-efeito. A análise de correlação também é viável - o que habilita a variável DepCorrelation.

Com base no centro de fusão criado no nível 1, no nível 2 é necessária a participação de especialistas para uma avaliação da situação (SA) quanto à identificação do objeto, refinando e ajustando detalhes da integração de primeiro nível. Esses ajustes baseados em opiniões de especialistas geram evidências em formato de linguagem natural, exigindo procedimentos de PLN e procedimentos de agregação de dados que possibilitem

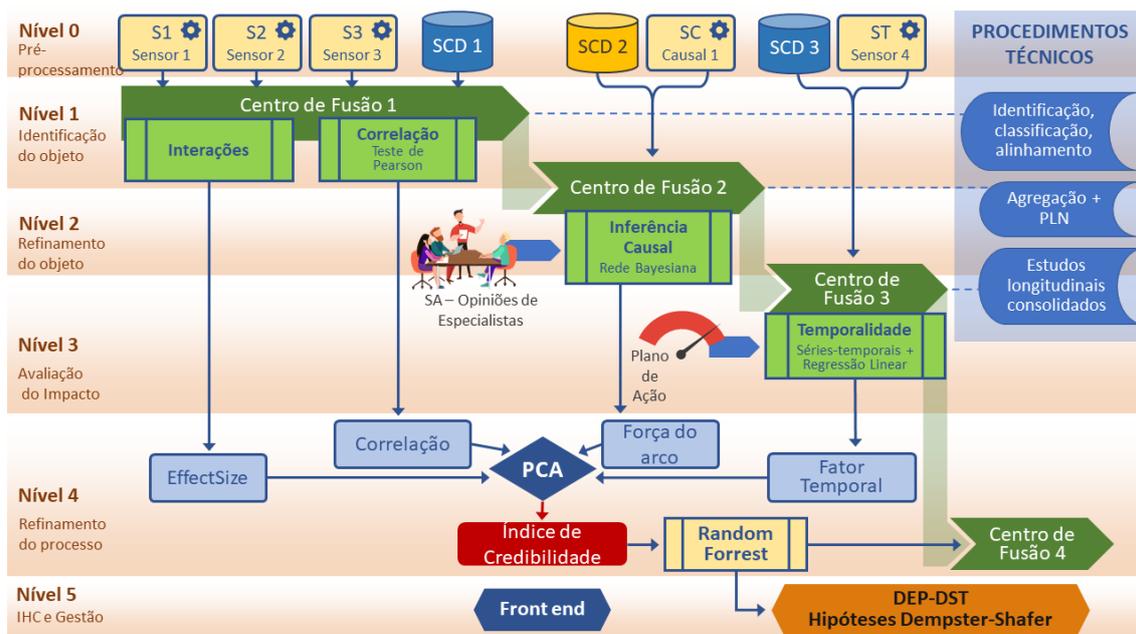


Figura 4.1: DEP-DST Modelo de geração de evidências DST com ajuste de dependências

novas fontes de informação. Objetos potenciais de interesse, descobertos com o apoio de procedimentos de inferência causal, são identificados, chegando ao centro de fusão 2.

O nível 3 trata da avaliação de impacto, que traz novas possibilidades de identificação de objetos de interesse com o apoio de procedimentos planejados num plano de ação. Nesse ponto do modelo, os processos que determinam as dependências esclarecidas pelos efeitos temporais são executados e a variável temporalidade é gerada para compor o IC.

O nível 4 realiza o refinamento do processo como um todo. A estimativa da probabilidade de DST é nesse nível, em que é gerado o centro de fusão 4, que é a base mais robusta do processo. Há uma interação com os níveis anteriores nesse nível, permitindo uma revisão de procedimentos. A geração de BPA pelo método de inteligência artificial possibilita a geração de hipóteses para a tomada de decisão do modelo DEP-DST. O nível 5 realiza o gerenciamento de novos requisitos e objetivos do processo FIMS e cuida da interface com usuário por métodos visuais.

A figura 4.1 mostra o fluxo para o processo de formação do índice de credibilidade de forma genérica. Essa abordagem pode ser aplicada a qualquer domínio. O IC é formado por quatro componentes. O primeiro é gerado pelas relações de dependência identificadas por um método de inferência causal. O segundo componente decorre do teste de correlação entre as entidades em estudo. O terceiro componente representa interação entre os elementos que produzirão o efeito tamanho. O quarto membro é formado pela dependência observada num fenômeno temporal.

## 4.2 Componentes do Índice de credibilidade para o tratamento das dependências

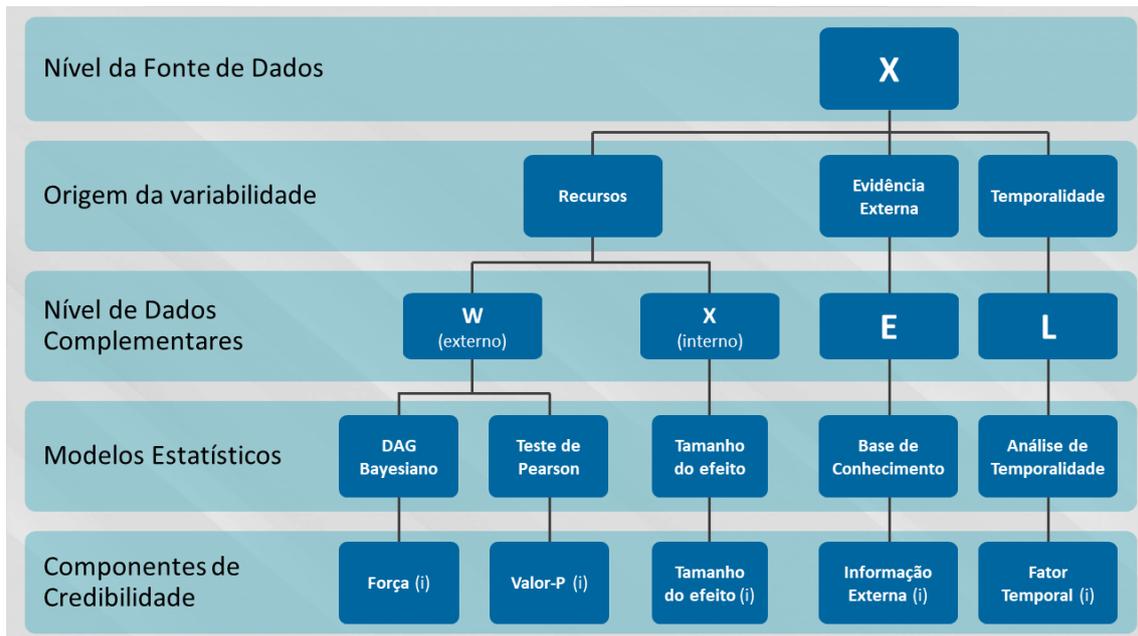


Figura 4.2: Esquema do Índice de Credibilidade

A figura 4.2 mostra as diferentes camadas para o tratamento de dependências com as quais o índice de credibilidade é formado. Inicialmente, as características dos dados são incrementadas a partir de recursos internos (x) e externos (W) criados a partir de modelos estatísticos. A letra E significa conhecimento de especialistas e bases de conhecimento. A letra L representa uma complementariedade de dados a partir de análise temporal. São geradas 4 variáveis nesse processo, a DepCorrelation, DepCausal, DepTemporal e Effect-Size, todas definidas formalmente na próximas subseções 4.2.1 a 4.2.4. A dependência por correlação é analisada com o apoio do modelo estatístico Teste de Pearson Test e gera a variável componente DepCorrelation. A dependência causal é mais elaborada e utiliza o modelo estatístico de inferência Bayesiana, produzindo a variável força do arco que forma a variável DepCausal. Características internas como o efeito de tamanho produzido pelas interações entre os fatores, permitem a formação do componente de efeito de tamanho. A evidência externa vem das bases de conhecimento. Por fim, a variável índice de credibilidade componente DepTemporal é formada a partir de um modelo estatístico que elabora uma análise de temporalidade no nível complementar dos dados L.

Considere uma estrutura de geração de evidências na qual uma estrutura de relacionamento nos dados consiste num subconjunto de covariáveis X, que explicam o fenômeno de interesse y. Suponha ainda que um conjunto de fatores externos W afete o fenômeno de interesse y da mesma maneira. Tomando esse quadro de referência, serão explicados a seguir os principais elementos que compõem o índice de credibilidade do PCA proposto.

### 4.2.1 DAG Bayesiano - Probabilidades Condicionais

Seja  $y = (y_1, \dots, y_n)$  a variável de destino no modelo (destino). A força do arco para o indivíduo  $x_i$  é definida como a seguinte probabilidade condicional

$$Strength(x_i) = P(y_i | x_{i1}, \dots, x_{im})$$

Essa probabilidade é estimada a partir do DAG seguindo uma abordagem baseada em simulações de amostra, conforme ilustrado no processo iterativo pela figura 4.3. Seja  $S$  o número total de iterações.

---

**Algorithm 1** Amostragem no DAG - Bayesiano.

---

Seja  $S$  o total de iterações.

$y_{x_i} = array(1, S)$

**for**  $s$  **in**  $(1, \dots, S)$  **do**

$y_s(x_i) = y_i | sample(x_{s1}, \dots, x_{sm})$

$y_{x_i}.append(y_s(x_i))$

$Strength(x_i) = \sum_{s=1}^S y_{x_i}(s)$

**end for**

---

Figura 4.3: Amostragem no DAG Bayesiano

Sob esse esquema da figura 4.2, o conjunto de fatores  $W$  representa aquelas variáveis que podem afetar  $y$ , mas cujo efeito deseja-se isolar. Uma maneira de fazer isso é primeiramente identificar se a variável  $y$  é causada por  $X$  ou por  $W$ . Tem-se então que estudar o comportamento das probabilidades condicionais:

- 1)  $P(y|X)$
- 2)  $P(y|W)$
- 3)  $P(y|X, W)$

O comportamento dessas probabilidades condicionais forma o contexto de interação

mais simples entre esses conjuntos de variáveis, podendo-se modelar estruturas mais complexas por meio de um grafo (DAG).

O objetivo dessa representação é determinar a causalidade da variável  $y$  para descartar relações na base que possam gerar evidências falsas. Esse processo constitui o primeiro pilar do nosso modelo. A partir dele, é possível estimar a estrutura do DAG sob uma abordagem Bayesiana - cadeia de Markov Monte Carlo (MCMC) e determinar em quais casos se poderia identificar uma relação de dependência por se tratar de uma fonte não causal nos dados.

Uma das saídas dentro de um processo de estimação Bayesiano é a força do arco. Essa variável permite resumir, em termos percentuais, a frequência com que percorreu cada relação (aresta) em cada simulação do algoritmo MCMC.  $DepCausal(z)$  é a variável que representa a força do arco em cada relacionamento do DAG para cada elemento "z" na população, então  $DepCausal(z)$  é a primeira variável que será parte do índice de credibilidade proposto.

#### 4.2.2 Análise de Correlação com Teste de Pearson

Embora correlação e causalidade não sejam conceitos equivalentes, e embora a causalidade possa desempenhar um papel mais relevante, ambas as propriedades são altamente desejáveis para uma boa análise de dados. Correlação descreve comportamentos em que a ocorrência de um evento está diretamente relacionada à ocorrência de outro. Esses aspectos são essenciais em questões de modelagem preditiva e aprendizado automatizado, pois o desempenho dos modelos depende diretamente dessa característica nos dados.

O teste de Pearson é um método estatístico não paramétrico utilizado para testar hipóteses nas quais se deseja saber se existe uma diferença significativa entre quaisquer duas distribuições. Por essa razão, é útil determinar quando há um efeito substancial de uma variável externa ou não. Para tanto, a distribuição sob o impacto dessa variável é estudada e comparada com um "placebo". Se a hipótese nula for rejeitada, o que caracteriza uma homogeneidade na distribuição, obtém-se evidência de um efeito significativo.

Interessa-se observar relações nas quais a variável  $y$  é causada e fortemente correlacionada com  $X$  e não com  $W$ . A ideia é usar o teste de Pearson para determinar as relações que parecem ter um forte efeito no conjunto de covariáveis  $W$  e ajustá-las.

O seguinte teste é proposto:

- $H_0: y = y|W$  : O W causa o y.
- $H_1: y \neq y|W$  : O W não causa o y

E o valor p é calculado da seguinte forma

$$p\text{-valor}(x_i) = 2P(T > t),$$

em que  $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

$n = \text{número de observações}$

$r = \text{coeficiente de correlação de Pearson}$

Se o  $p\text{-valor}$  obtido pelo teste for  $< 0,05$ , então é rejeitado  $H_0$  em favor de  $H_1$ . Interessante em preservar aqueles casos em que a hipótese nula não é rejeitada. Uma vez que sob  $H_0$  não há evidência de um efeito de correlação de  $y$  com  $W$ , então  $DepCorrelation(z)$  mede a credibilidade no nível do registro em termos da correlação sendo o segundo componente do nosso índice.

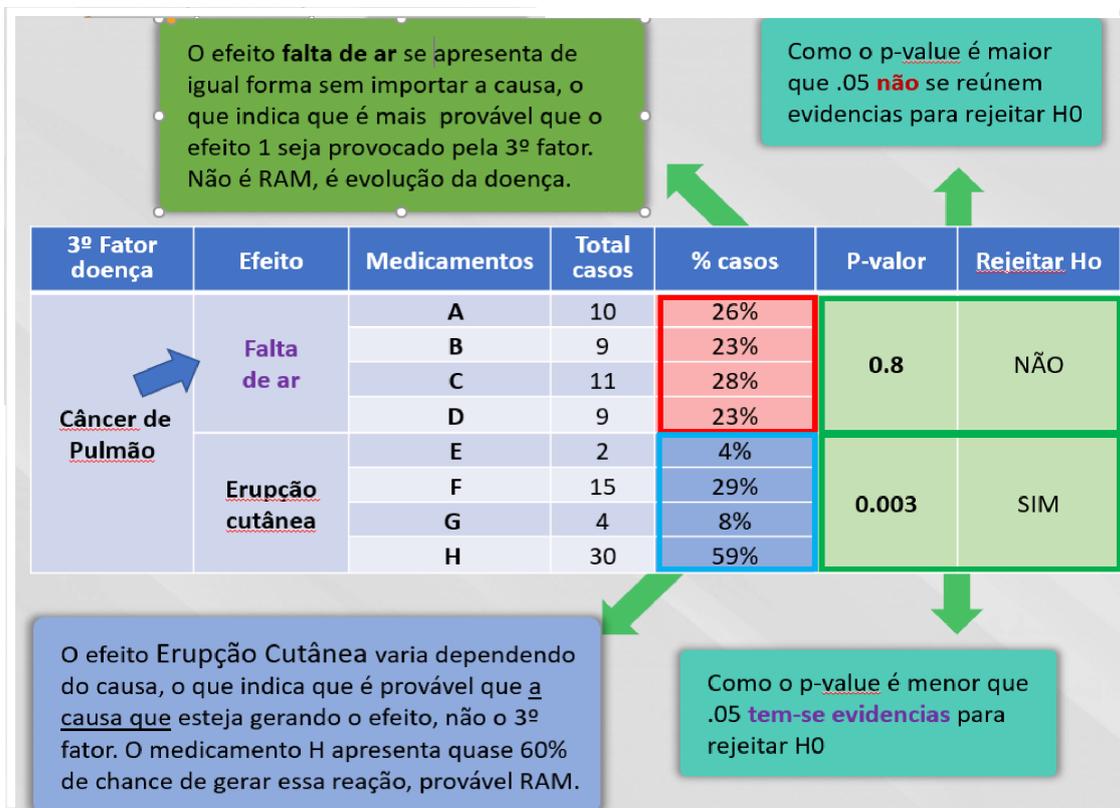


Figura 4.4: Simulação de teste de Pearson para análise de correlação

A figura 4.4 mostra um exemplo hipotético de um teste de Pearson envolvendo medicamentos usados por um paciente, sintomas e reações relacionados. Para um determinado sintoma, as frequências de distribuição são avaliadas e uma correlação é analisada.

### 4.2.3 Tamanho do Efeito- análise de interações

Normalmente há contextos em que a variável objetiva  $y$  pode ser explicada por meio de uma série de fatores, embora não se saiba se foi causada por um deles, por todos eles ou por uma combinação entre eles. À medida que cresce o número de interações que afetam a mesma variável, a relação torna-se menos evidente e as conclusões obtidas por meio dela perdem qualidade.

Para medir o grau de interações em  $X$  na variável alvo  $y$ , propõe-se o desenho deste componente que resume essas interações da seguinte forma. Suponha que  $X_i$  seja uma variável categórica com " $c$ "categorias em  $X$  e  $I_j$  seja a variável indicadora da categoria  $j$ , quer dizer:

- $I_j(Z) = 1$  se o indivíduo  $z$  pertencer à categoria  $j$
- $I_j(Z) = 0$  se o  $z$  individual não pertencer à categoria  $j$

Em seguida, definimos o tamanho do efeito para  $z$  individual como a taxa de incidência de  $z$  em cada categoria, ou seja:

$$EffectSize(z) = \frac{1}{c} \sum_{j=1}^c I_j(z)$$

### 4.2.4 Análise Temporal

Muitas vezes há a possibilidade de identificar e obter insights de tendências temporais nas relações de dados. Essas tendências podem conter padrões lineares ou não lineares com o passar do tempo. Estas observações permitem obter conclusões sobre a variável de interesse a partir de tendências temporais associadas ao conjunto de covariáveis  $X$  que podem levar a decisões sobre a variável alvo  $y$ .

Essas tendências podem ser modeladas por meio de técnicas de regressão linear para avaliar seu nível de significância estatística ( $\alpha$ ) e poder concluí-las.

Seja  $L = (L_1, \dots, L_k)$  um conjunto de covariáveis com informações complementares a “X”, que contém informações temporais sobre o subconjunto de observações (ou indivíduos) no estudo. Uma análise de regressão linear simples é proposta para determinar a tendência de cada covariável  $L_j$ ,  $\forall j$  associada a cada indivíduo através de “L”. Então temos um modelo da forma:

$$L_j = \beta_0 + \beta_1 * t + \varepsilon_j \quad ; \quad j = 1, \dots, k$$

$$\varepsilon_j \sim \text{Gaussiana}(0, \sigma^2) \quad ; \quad j = 1, \dots, k$$

Este é o modelo de regressão linear mais comum, a partir do qual as seguintes hipóteses são de interesse:

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

Se  $\beta_1 > 0$ , o efeito é crescente e decrescente se  $\beta_1 < 0$ . Procura-se então rejeitar a hipótese nula  $H_0 : \beta_1 = 0$  para ter significância estatística na tendência.

Uma vez determinado um fator de tendência para cada indivíduo para as variáveis  $L = (L_1, \dots, L_k)$ , é possível validar se um aumento ou diminuição nos valores dessas variáveis pode ser associado com a variável de destino “y”. Isso pode ser determinado a partir de uma análise complementar ou de uma base de conhecimento previamente capturada seguindo os princípios mencionados no ponto 3.4 (Fontes de Conhecimento Externas).

Formalmente, seja  $(L, y_L)$  o par de evidências obtido da análise temporal. O fator de temporalidade para cada indivíduo  $z$  dentro do conjunto de covariáveis  $L = (L_1, \dots, L_k)$  é então definido como:

$$Dep - Temporal(z) = \begin{cases} 1 & \text{se ação } (X_z, y_z) \text{ está contida} \\ em (L, y_L) & \\ 0 & \text{em outro caso} \end{cases}$$

### 4.3 Método PCA - Integração dos Componentes do Índice de Credibilidade

O método proposto é baseado em uma análise de componentes principais dos índices de credibilidade individuais. PCA (Principal Component Analysis) é uma técnica de redução de dimensão inventada em 1901 pelo estatístico Karl Pearson. Esse algoritmo consiste em uma projeção ortogonal que busca a direção da máxima variabilidade nos componentes e projeta com base nele para obter uma versão que sumariza o comportamento original e os índices, preservando o máximo de informações possível.

Para cada registro  $z$  em  $X_{m \times n}$  considere o vetor de índices individuais:

$$V(z) = (Dep - Causal(z), Dep - Correlation(z), EffectSize(z), Dep - Temporal(z))$$

Este vetor define um espaço de 5 dimensões sobre o conjunto de dados original  $X_{m \times n}$ , este espaço é definido como *espaço de credibilidade*. Este espaço determina uma nova representação dos dados a partir dos quais buscamos obter uma única medida para cada registro  $z$ . Primeiro rodamos o espaço de credibilidade, procurando uma nova representação  $y$  no sentido de máxima variabilidade dos índices. Temos então que  $y$  é dado por:

$$y_{1 \times 5}(z) = A_{5 \times 5} V^t(z)$$

De onde  $A_{5 \times 5}$  é uma matriz ortogonal de rotação. Dado que se busca a rotação na direção de máxima variabilidade, o novo vetor de variáveis  $y_{1 \times 5}(z)$  deve ser não correlacionado, isso implica que sua matriz de variâncias e covariâncias deva ser diagonal, isto é:

$$Var(y) = \begin{bmatrix} \sigma_{y1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{y5} \end{bmatrix}$$

Isso implica que:

$$\text{Var}(AV^t) = A \text{Var}(V^t) A^t = \begin{bmatrix} \sigma_{y_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{y_5} \end{bmatrix} = \text{diag}(\sigma_{y_1}, \dots, \sigma_{y_5})$$

De maneira geral se tem o seguinte resultado elementar de álgebra linear:

*Proposição:* Seja  $S$  uma matriz quadrada  $p$ -dimensional definida não negativa e  $C$  uma matriz ortogonal. Suponha que  $C^t S C = \text{diag}(\lambda_1, \dots, \lambda_p)$ , então o vetor  $(\lambda_1, \dots, \lambda_p)$  corresponde aos autovalores próprios de  $S$  e as colunas da matriz ortonormal  $C$  correspondem aos eigenvetores normalizados de  $S$ . Quer dizer:

Se  $C^t S C = \text{diag}(\lambda_1, \dots, \lambda_p)$  então  $(\lambda_1, \dots, \lambda_p) = \text{autovalores}(S)$  y  $C = \text{autovetores}(S)$

Da proposição anterior segue-se que a matriz de rotação  $A^t$  tem como colunas os autovetores próprios da matriz de variâncias e covariâncias do vetor de índices individuais  $S = \text{Var}(V^t)$  e que o vector de variâncias  $(\sigma_{y_1}, \dots, \sigma_{y_4})$  do novo espaço obtido da rotação  $y(z) = AV^t(z)$  está dado por seus próprios valores.

Em particular, tomando o primeiro autovetor da matriz  $S = \text{Var}(V^t)$  e projetando-se sobre si mesmo, obtém-se uma representação unidimensional no espaço  $\mathbb{R}$  cuja proporção de variabilidade é dada pela taxa de contribuição do primeiro autovalor de  $S$ :

$$\% \text{ de Variabilidade Explicada} = \frac{\sigma_{y_1}}{\sigma_{y_1} + \cdots + \sigma_{y_5}} = \frac{\lambda_{y_1}}{\lambda_{y_1} + \cdots + \lambda_{y_4}}$$

Os principais pontos do algoritmo com o método PCA para cálculo do índice de credibilidade proposto estão resumidos a seguir.:

- Obtenha o vetor de índices individuais  $V^t$  para cada registro  $z$  em  $X$  segundo a metodologia descrita na seção anterior.
- Aplique o método PCA obtendo o primeiro autovalor e o vetor de autovetores da matriz de variâncias e covariâncias de  $V^t$  para obter a projeção  $y(z) = AV^t(z)$ .

- Por fim, o IC é definido como a projeção  $y(z)$  padronizada para o intervalo unitário  $[0, 1]$  e a taxa de Variabilidade Explicada da projeção em para ter uma medida da qualidade do índice.

#### 4.4 Algoritmos

A primeira parte do algoritmo ilustrada na figura 4.5 aborda a geração das variáveis que compõe o IC, quais sejam a DepCorrelation, EffectSize, DepCausal e DepTemporal e o cálculo do IC. A função concatenar empilha os vetores para formar uma matriz com eles. O axis=0 concatena horizontalmente e seu axis=1 concatena verticalmente. O X new seria a matriz completa, aquela que inclui os dados originais, as variáveis individuais que formam o IC.

##### PASSO 1: Gerar variáveis que vão compor o IC

seja  $X_{n,m}$  o conjunto de dados original.

```

V = array(n, 4)
Para i em (1, ..., n)
    v1(xi) = Depcausal(xi)
    v2(xi) = DepCorrelation(xi)
    v3(xi) = EffectSize(xi)
    v4(xi) = DepTemporal(xi)
    V(xi) = (v1(xi), v2(xi), v3(xi), v4(xi))
V = concatenar(V, V(xi), axis = 1)

```

##### PASSO 2: Calcula o índice de credibilidade.

```

S = Var(Vt)
At = eigenvectors(S)
IC - PCA = array(n, 1)
Para k em (1, ..., n):
    IC - PCAk = A · V(k)t
    IC - PCA = (IC - PCAk, axis = 0)
Xnew = concatenar(X, V, IC - PCAk, axis = 0)

```

Figura 4.5: Gerar variáveis componentes do IC

A parte final do algoritmo ilustrada na figura 4.6 demonstra a execução das instâncias de treinamento e testes utilizando a técnica RF.

**PASSO 3: Gerar a evidência pela técnica Random Forest.**

```
 $X_{train}, y_{train}, X_{test}, y_{test} = \text{sampling}(X_{new}, y, 75\%)$   
 $Grid = \{p_1, \dots, p_l\}$  (espaço de parâmetros)  
 $performance = \text{array}(n, 1)$   
  
Para  $i \in \{1, \dots, l\}$ :  
     $model = \text{RandomForest.Train}(X_{train}, y_{train}, p_i)$   
     $performance(i) = model.evaluation(X_{val}, y_{val})$   
     $p_{optimal} = p_{\max\{performance\}}$   
 $best - model$   
     $= \text{RandomForest.Train}(X_{train}, y_{train}, p_{optimal})$   
 $Evidence = best - model.predict(X_{new}, y_{new})$ 
```

Figura 4.6: Gerar a evidência baseada na técnica Random Forest

#### 4.5 Ameaças à validade do modelo

A utilização do método PCA para a combinação linear dos componentes que geram o índice de credibilidade pode ser substituída por algum outro instrumento que se mostre mais eficaz.

#### 4.6 Considerações Finais sobre o modelo

O modelo DEP-DST a partir a criação do IC propõe o estabelecimento de um diagnóstico a partir de uma situação causa e efeito, considerando relações de dependência. Na próxima seção esse modelo será instanciado em um domínio para que seja possível acolher resultados práticos.

## 5. Método DEP-DST aplicado no domínio de RAM

Neste capítulo será abordada a aplicação do modelo genérico apresentado na seção 4.1. O modelo é formado a partir de um processo de fusão de informações que integra dados de pacientes, notas clínicas, prescrições médicas e prescrições obtidas de pacientes internados. Nas prescrições médicas são registrados os medicamentos aplicados ao paciente. O CID, Código Internacional de Doenças, informa uma codificação da doença pela qual o paciente é acometido. Os pacientes são expostos a diferentes medicamentos, lembrando que em caso de comorbidades (diabetes, hipertensão arterial, insuficiência renal, convulsões, neuropatia e depressão), a quantidade de medicamentos aumenta, gerando interações medicamentosas. Durante uma consulta clínica, o paciente expõe algumas queixas ao médico. Este, por sua vez, observa sintomas que o exame clínico permite avaliar. Essas queixas e observações são registradas numa nota clínica. Existe ainda uma fonte adicional de dados, necessária para alguns casos, em exames laboratoriais do paciente.

Um paciente em tratamento de câncer passa por uma série de exames ao longo do tempo, pelo menos um exame por mês. Se ele passa por períodos de internação, essa frequência de exames é mais intensa. A coleta de dados de testes laboratoriais permite a geração de séries temporais. Em alguns casos, podem ser observadas tendências na variação de alguns parâmetros sanguíneos (como glicose, ureia, creatina, leucócitos, entre outros). Algumas dessas variações nos parâmetros laboratoriais podem indicar RAM. Vários sintomas e queixas podem ser RAM ou podem surgir de uma evolução natural da doença em tratamento. A aplicação do modelo no domínio RAM determina as relações de dependências entre as diferentes fontes (receitas, prescrições, avaliações clínicas e queixas dos pacientes) e as confronta.

A proposta de criação do IC permite avaliar dependências causais e temporais por inferência causal da relação entre doença e sintomas e em relação entre RAM e medicamentos, relatada na variável força obtida pela inferência da rede Bayesiana. A associa-

ção entre medicamentos e reações/sintomas é mapeada pela variável DepCorrelation, que relaciona medicamentos e sintomas, por meio de uma análise multivariada de um teste qui-quadrado de Pearson, avaliando a correlação que avalia a proporcionalidade entre os fatores da distribuição, disponibilizando o (*p-value*) desse teste de hipótese.

Por fim, o IC é formado por um componente temporal, extraído de uma regressão linear obtida pelas séries temporais de exames, quando esses permitem a formação de tendências altas ou baixas nos parâmetros sanguíneos, levando, em alguns casos, a possíveis RAM intensificadas pelo uso de medicamentos ao longo do tempo. Esta função é desempenhada pela variável DepTemporal que cuida da dependência temporal e integra o IC.

Tendo um cenário com uma decisão diagnóstica num ambiente de incerteza, o modelo permite a geração de evidências com um algoritmo de floresta aleatória. Foram utilizados cenários distintos de conjuntos de dados, inserindo paulatinamente os componentes do IC e permitindo avaliar a eficácia evolutiva que a aplicação dessas variáveis melhoraria no modelo.

O experimento contou com 5.937 pacientes que foram tratados para 85 doenças diferentes. Havia 55% de mulheres e 45% de homens, com faixa etária média de 45 anos, 10% crianças e 85% idosos (> 65 anos).

Inicialmente, foram recuperados 63 mil notas clínicas registradas ao longo de 2019. Dessas notas clínicas, aproximadamente 8,7% referiram medicamentos ou sintomas, com 55 mil registros selecionados nesta fase. Mais de 10% desses registros foram gerados após a homogeneização para gerar pares de medicamentos e sintomas, gerando 5.600 registros de possíveis candidatos a RAM. Em seguida, realizou-se a combinação entre as notas clínicas e as prescrições aplicadas ao paciente, fossem elas ambulatorial ou utilizadas durante tratamento ou internações, que geraram mais de 6300 prescrições combinadas com os registros candidatos a RAM a partir das evoluções registradas nas notas clínicas.

Os exames laboratoriais colaboraram com 40.414 registros gerados por 911 dos 960 pacientes selecionados na etapa anterior, ou seja, 95% dos pacientes realizaram algum exame e com uma frequência razoável, com média superior a quatro exames anuais por paciente. Sabe-se que pacientes internados ou em fase crítica de tratamento coletam exames com maior frequência. Em seguida, foi realizado um procedimento para avaliar tendências nos parâmetros do exame, gerando 4.415 registros de tendência de alta e 3.333 de tendência de queda.

## 5.1 Formação do Banco de dados - Conhecimento, Informações Contextuais e Semiestruturadas

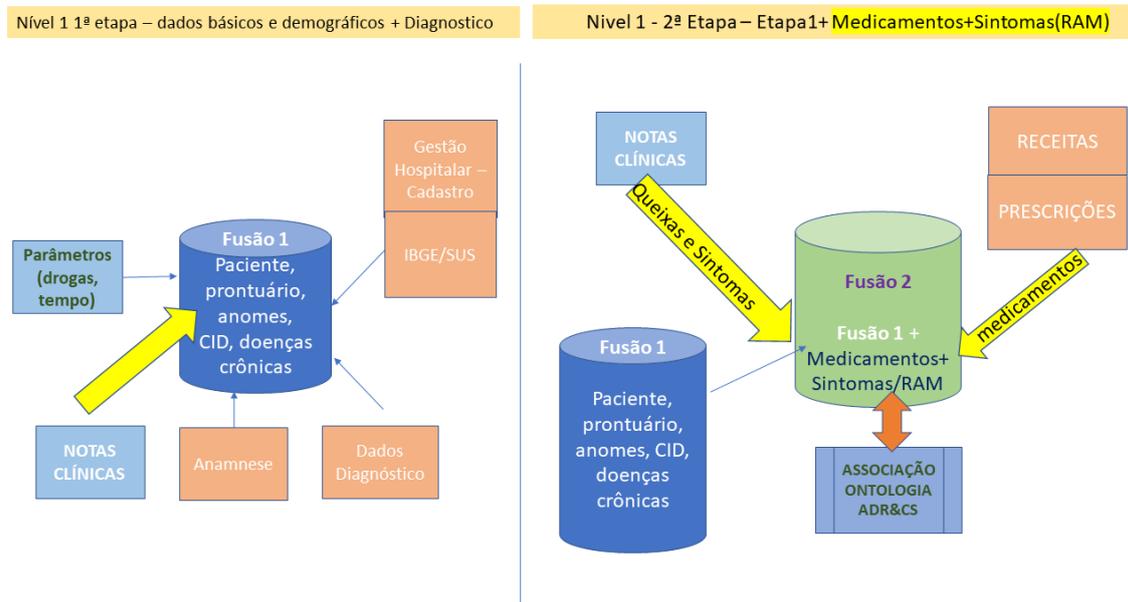


Figura 5.1: Formação do banco de dados

A figura 5.1 ilustra a coleta de dados realizada no primeiro nível de fusão de dados. Os principais sensores e fontes de dados são a nota clínica e as receitas médicas. A partir da nota clínica são extraídas as queixas e sintomas relatadas pelo paciente ou observadas pelo profissional de saúde, utilizando-se técnicas de PLN para esse fim. Esses dados são associados a dados demográficos e pessoais do paciente e geram um primeiro centro de fusão de dados. Já no nível posterior de fusão os eventos candidatos a RAM são gerados por meio da combinação dos sintomas e queixas apontadas pelos pacientes relacionados aos medicamentos prescritos.

Os materiais utilizados no processo de fusão de informação são: Um prontuário eletrônico armazena as informações clínicas do paciente e forma a fonte primária de dados do registros eletrônicos de saúde(RES). Os dados que compõem o RES podem ser categorizados em observacionais, contextuais e de conhecimento. Os dados observacionais são extraídos de bancos de dados e sistemas eletrônicos que constroem o RES. O registro mais importante é a nota clínica. Os medicamentos podem ser indicados para o paciente em tratamento ambulatorial, no consultório, mediante a emissão de receitas. Caso o paciente esteja internado, em tratamento ou recebendo novas dosagens de medicamentos, frequência, volume, tipo de administração e dosagem costumam ser alterados, caracterizando informações contextuais e risco aumentado de RAM. Os dados obtidos por meio do acesso ao RES em bancos de dados geralmente são estruturados, mas existem outros

tipos de representação.

Uma nota clínica é formada por dados não estruturados, onde o profissional de saúde registra, em formato livre, informações diversas, caracterizando uma alta dimensionalidade das informações. Alguns dados incorporam uma regra de negócios agregada, conhecida como dados semiestruturados. Essa situação ocorre com testes laboratoriais e testes determinados por informações de conhecimento de domínio. Por exemplo, os valores normais de referência para o parâmetro de glicose no sangue estão entre 70 e 99mg/dl. Quando a glicemia está abaixo de 70 mg/dL, a hipoglicemia pode afetar o paciente. Quando a glicemia está acima de 99 mg/dL, o paciente está na faixa de pré-diabetes, e o paciente é considerado diabético se esta medida for superior a 120 mg/dL.

A tabela 5.1 retrata as ferramentas utilizadas no desenvolvimento dessa aplicação.

## 5.2 Materiais utilizados na implementação dessa tese

FUNCIONALIDADE MÉTODO	LINGUAGEM	BIBLIOTECA
Análise de Componentes Principais PCA	Python	Sklearn.decomposition
Modelo Random Forest RandomForestClassifier	Python	Sklearn.ensemble
Matriz de Confusão <i>confusion<sub>m</sub>matrix</i>	Python	Sklearn.metrics
Teste de Pearson chisquare	Python	Sklearn.stats
Curva ROC e AUC <i>roc<sub>c</sub>curve, auc</i>	Python	Sklearn.metrics
Validação cruzada <i>train<sub>t</sub>est<sub>s</sub>split</i>	Python	Sklearn.model <sub>s</sub> election
DAG Bayesiano <i>parameter<sub>t</sub>learning.fit</i>	Python	bnlearn
Regressão linear	Python	<i>Linear<sub>r</sub>egression().fit</i>
Testes de Hipóteses OLS(Ordinary Least Squares)	Python	statsmodels.api
Shapley Additive Values shap.TreeExplainer	Python	shap

Tabela 5.1: Ferramentas bibliotecas utilizadas nessa tese

A figura 5.2 demonstra o mapa mental do domínio de RAM. Os dados demográficos dos pacientes são agregados com dados de diagnóstico, como a doença que está sendo tratada e comorbidades. As principais entidades do modelo RES são o cadastro do paciente e os dados demográficos referentes a medicamentos, reações, sintomas, doenças, exames e prescrições de acordo com a condição do paciente (internado ou em tratamento ambulatorial). A coleta de exames laboratoriais pode determinar tendências nos parâmetros sanguíneos, que podem indicar RAM.

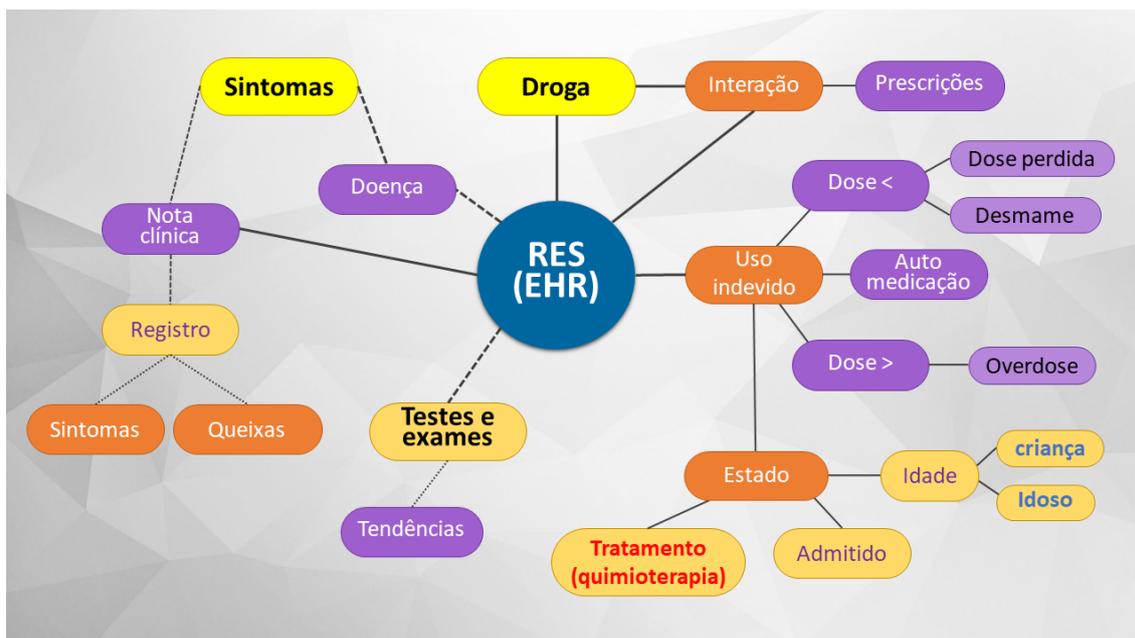


Figura 5.2: Mapa Mental - RAM a partir de RES

A heterogeneidade e a multimodalidade das fontes de dados trazem consigo um alto grau de incerteza, incluindo fontes de informação redundantes e ambíguas. Por exemplo, duplicidade - onde os mesmos dados são provenientes de duas fontes diferentes, apresentando valores divergentes - exige a escolha de uma fonte predominante.

### Bases de Conhecimento

A estrutura para análise de RAM é formada a partir de um modelo de domínio, que relaciona doenças, medicamentos e reações adversas. Nosso modelo usa duas bases de conhecimento, a Classificação Internacional de Doenças (*CID*) e uma ontologia de efeitos adversos [Cai et al., 2015]), para buscar reações conhecidas que possam ser usadas como um catálogo de medicamentos [Cai et al., 2015], contendo informações do tipo de conhecimento, como uma espécie de "bula" que reúne uma grande relação de medicamentos. As RAM registradas representam relações droga-reação para as quais já existe um precedente clínico, testado, reconhecido e aceito, significando conhecimento de domínio.

### Recursos e volumes gerais da base de RAM

As características gerais do banco de dados são descritas a seguir.

- a) O banco de dados contém registros que relacionam reações e medicamentos ad-

ministrados aos pacientes desde 2019.

- b) O banco de dados é composto por 81.740 registros desse tipo e 11 colunas que descrevem, entre outras coisas, a identificação do paciente, data de nascimento, sexo, setor clínico, reação e medicamentos administrados.
- c) Os registros informados no banco de dados correspondem a 5.937 pacientes atendidos nesse período. As variáveis mais importantes no banco de dados são a clínica médica de origem, os medicamentos administrados e os sintomas ou reações relatadas e observadas. A partir dessas variáveis, constrói-se o quadro de discernimento DST para estimar a crença de RAM a partir de medicamentos aplicados ou a evolução natural da doença.

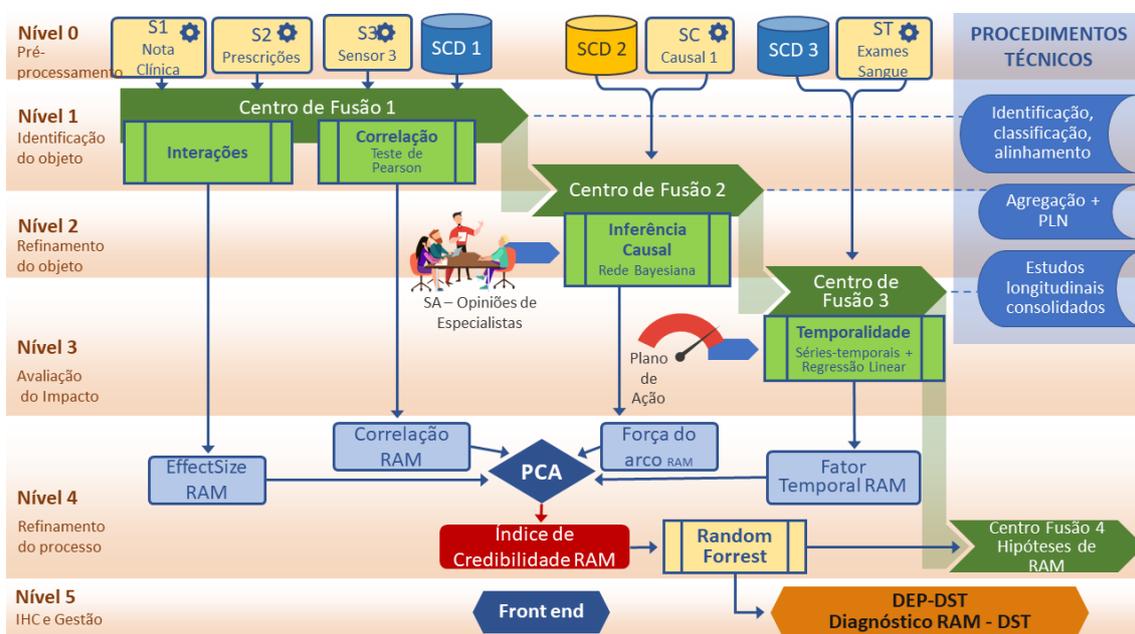


Figura 5.3: Modelo DEP-DST aplicado ao domínio RAM

A figura 5.3 mostra os elementos drogas e as entidades em estudo são RAM ou podem ser sintomas ou reações. É a instância do modelo contemplado na figura 4.1 instanciada para o problema de RAM. Dessa forma, os sensores genéricos desta figura ganham nomenclatura, como nota clínica, receitas, prescrições, como representado nos sensores da captura na primeira linha da figura, equivalente ao nível zero do FIMS.

A relação de dependências será ponderada pelas variáveis de correlação, inferência causal e análise temporal. Essa figura inclui os sensores e fontes de dados detalhados abaixo:

- Nota Clínica - a partir de uma fonte de exame do paciente, apresenta-se um formato

não estruturado que requer o uso de técnicas de PLN para a extração de medicamentos e sintomas que possam indicar RAM. Uma nota clínica relata uma avaliação da condição de um paciente por um profissional de saúde. Contém dados sobre as queixas descritas pelo paciente, o estado clínico e geral do paciente observado pelo profissional de saúde, dados registrados sobre a evolução da doença e resultados de marcadores e exames;

- Prescrição - uma fonte de prescrições ambulatoriais que contém as dispensações de medicamentos para pacientes que são atendidos e em domicílio e em ambulatório, que não estão internados;
- Prescrição médica - contém as prescrições de medicamentos para pacientes internados. É informação contextual; geralmente, as aplicações de medicamentos ocorrem por via intravenosa em doses e demandas mais potentes, aumentando a incidência de RAM;
- Quimioterapia - a partir do sensor que reúne dados de pacientes que fazem tratamento quimioterápico;
- Exames Laboratoriais - com frequência variada, os dados são obtidos através dos sensores dos exames laboratoriais, sendo os mais comuns os exames de sangue e urina. Eles relatam informações estruturadas sobre os valores de referência do resultado dos marcadores. O uso de informações do domínio de conhecimento médico se faz necessário para avaliar desvios de comportamento e detectar tendências.

### **5.3 FIMS - Processo de Fusão de Dados**

A seguir, analisam-se as informações obtidas e os processos realizados em cada nível do FIMS proposto. A figura 5.3 mostra os níveis dos estágios de fusão utilizados no modelo proposto. Cada nível tem sua função, contribuição, interação e o feedback entre os níveis que ocorrem, proporcionando dinamismo ao processo.

#### **5.3.1 Nível 0 - Sinal e Pré-processamento**

Os sensores que contribuem para esse processo de fusão reúnem as informações básicas do cadastro do paciente, evolução anotada na nota clínica, prescrições ambulatoriais, internação e tratamento quimioterápico. Há também um sensor que capta os testes de laboratório. Os dados são recuperados de hora em hora, duas vezes por dia ou uma vez

por dia no processamento noturno. Após configurar a captura de dados dos sensores, é realizada a análise, avaliação e previsão dos estados observáveis dessas informações. Essa avaliação permite a eliminação de ruídos e redundâncias nas alocações e medições, otimizando o processamento dos seguintes níveis. Essa configuração de pré-processamento atende às condições de pré-processamento e prepara o ambiente para o próximo nível de FIMS.

### 5.3.2 Nível 1 - Refinamento de Objeto

O nível 1 enfatiza a associação entre os dados, possibilitando identificar e localizar RAM a partir de propriedades e atributos. A combinação dos dados ocorre para obter a estimativa mais confiável e precisa da RAM. As referências de espaço e tempo são alinhadas, adquirindo o interesse dos objetos RAM por meio dos múltiplos sensores do domínio. As fontes obtidas no nível 0 são associadas, necessitando de processos de homogeneização para possibilitar a correta associação entre o prontuário do paciente e as informações cadastrais presentes no RES a ele relacionado. Este nível fornece o primeiro centro de fusão de processos, fornecendo a infraestrutura essencial para detecção de RAM a partir das informações recém-fundidas. O domínio de conhecimento de RAM descreve as possíveis RAM decorrentes da ingestão e interações medicamentosas. O modelo pode obter as informações fornecidas por meio de procedimentos de pesquisa, incluindo técnicas específicas de PLN da infraestrutura de informações associada a esse nível. O foco é identificar RAM. Dos textos não estruturados de alta dimensão que ocorrem na descrição de uma NC, extraem-se os elementos que constituem uma RAM: medicamentos e reações. A combinação de possíveis reações de RAM de medicamentos de ingestão e interações capturadas são verificadas em relação à base de conhecimento. As interações entre os medicamentos aplicados a um paciente são realizadas através da implementação da variável *EffetcSize* que irá integrar o IC. No nível 1 ocorre a primeira análise de dependência do modelo, por meio da análise de correlação.

**Dependência - Análise de Correlação** A partir dos medicamentos e sintomas atribuídos ao paciente, é possível fazer uma análise de correlação. É aplicado um teste de Pearson, que permite a análise de correlação dessa possível relação de dependência. Assim, é produzida a variável *DepCorrelation* onde uma potencial RAM falsa receberá um peso menor. A figura 4.4 mostra um exemplo simulado de um teste de Pearson envolvendo medicamentos usados por um paciente e sintomas e reações relacionados. Para um determinado sintoma, as frequências de distribuição são avaliadas e uma correlação é analisada, conforme definido na seção **Análise de Correlação com Teste de Pearson**.

### 5.3.3 Nível 2 - Refinamento da situação

O nível 2 usa técnicas de agregação de dados e PLN para extrair dados de interesse do domínio. As identificações foram realizadas a partir de fichas de candidatos a RAM, em que as reações e medicamentos verificados estavam nos laudos presentes numa nota clínica.

[Blasch and Plano, 2005] utilizaram a consciência situacional (SA) no nível 2. Esse nível envolve a avaliação da RAM identificada preliminarmente. A partir do centro de fusão<sup>1</sup>, podem ser convocados especialistas para discutir os resultados preliminares. Os resultados da rede bayesiana são analisados para novos ajustes. Anotações clínicas são geradas em formatos especiais e novos procedimentos de extração são aplicados usando-se PLN e anotações de especialistas após reuniões de comitê para avaliação do processo. Este é um passo crucial para os demais níveis do FIMS, relacionando-se as informações contextuais adquiridas em RAM. A RAM neste momento é um registro potencial, que precisará ser avaliado e passar pelo processo de estimativa das etapas seguintes.

As prescrições e sensores de quimioterapia fornecem as fontes de medicamentos normalmente prescritos para um paciente com câncer. Os dados de prescrição, que são medicamentos prescritos regularmente para os pacientes, são incorporados. Normalmente, o tratamento quimioterápico ocorre em ciclos, nos quais os mesmos medicamentos são aplicados por várias sessões consecutivas. São aplicados via injeção, dissolvidos em soluções e, na maioria dos casos, aplicados por injeção com solventes e outros tipos de drogas auxiliares que geralmente não estão na zona de influência de RAM. Os dados precisam ser consolidados, com a limpeza de dados redundantes. Assim, a granularidade dos registros candidatos de RAM envolvendo a combinação de medicamentos e reações adversas extraídas a partir de notas clínicas na etapa anterior é adequada.

**Dependência - Análise de Causalidade - Inferência Bayesiana** A própria evolução natural da doença pode ser a causa real de uma reação adversa anteriormente suposta. Na realidade, seria um sintoma, um sinal da doença.

As tabelas 5.2 e 5.3 demonstram as matrizes de probabilidade condicional obtidas da rede Bayesiana. A tabela 5.3 menciona as probabilidades condicionais obtidas pela rede bayesiana que aborda os sintomas das doenças. A tabela 5.2 aborda RAM em relação a medicamentos. A doença pode ser o fator que leva a sintomas que podem ser confundidos com reações medicamentosas. Essa relação de dependência deve ser avaliada para obterem-se resultados mais precisos.

Tabela 5.2: DAG Probabilidades de Reação a Medicamentos

DROGA	Reação Não	Reação Sim	RAM
Dactomicina	49%	51%	Edema
Dactomicina	82%	18%	Hipoteca
Dactomicina	95%	5%	Diarréia
Morfina	81%	19%	Hipoteca
Morfina	61%	19%	Edema
Morfina	98%	2%	Diarréia
Fludarabina	84%	16%	Hipoteca
Fludarabina	70%	30%	Edema
Fludarabina	87%	13%	Diarréia

Tabela 5.3: DAG Probabilidades Doença-Sintoma

DOENÇA	Sintoma Não	Sintoma Sim	Sintoma
c498	46%	54%	Edema
c498	89%	11%	Hipoteca
c498	92%	8%	Diarréia
A250	61%	39%	Edema
A250	81%	19%	Hipoteca
A250	98%	2%	Diarréia
A320	61%	39%	Edema
A320	81%	19%	Hipoteca
A320	98%	2%	Diarréia

A figura 5.4 ilustra o DAG Bayesiano para medir com maior precisão a probabilidade de apresentar-se um sintoma ou RAM quando da ingestão de um medicamento. A partir da análise, é necessário identificar situações onde o efeito não corresponde a uma RAM, mas é causado pela evolução da doença e equivale a um sintoma.

Observe que na tabela 5.3 a doença c498 tem uma probabilidade condicional de 54% para o sintoma de edema, que por sua vez mantém uma probabilidade condicional de 51% ser RAM relacionada ao medicamento Dactomicina.

### 5.3.4 Nível 3 - Avaliação de Impacto

O conjunto completo de medicamentos e registros candidatos a RAM está disponível nesta etapa e permite o refinamento do modelo nesta terceira etapa da fusão. O conhecimento a priori estabelecido nos níveis anteriores é interpretado para prever estimativas e efeitos. Em seguida, é realizado o refinamento da análise de ameaças de RAM, modelando a situação atual para permitir inferências futuras.

Especialistas realizam um plano de ação no qual alguns marcadores são especificados em exames patológicos, possibilitando a avaliação de ameaças. Os sensores capturam

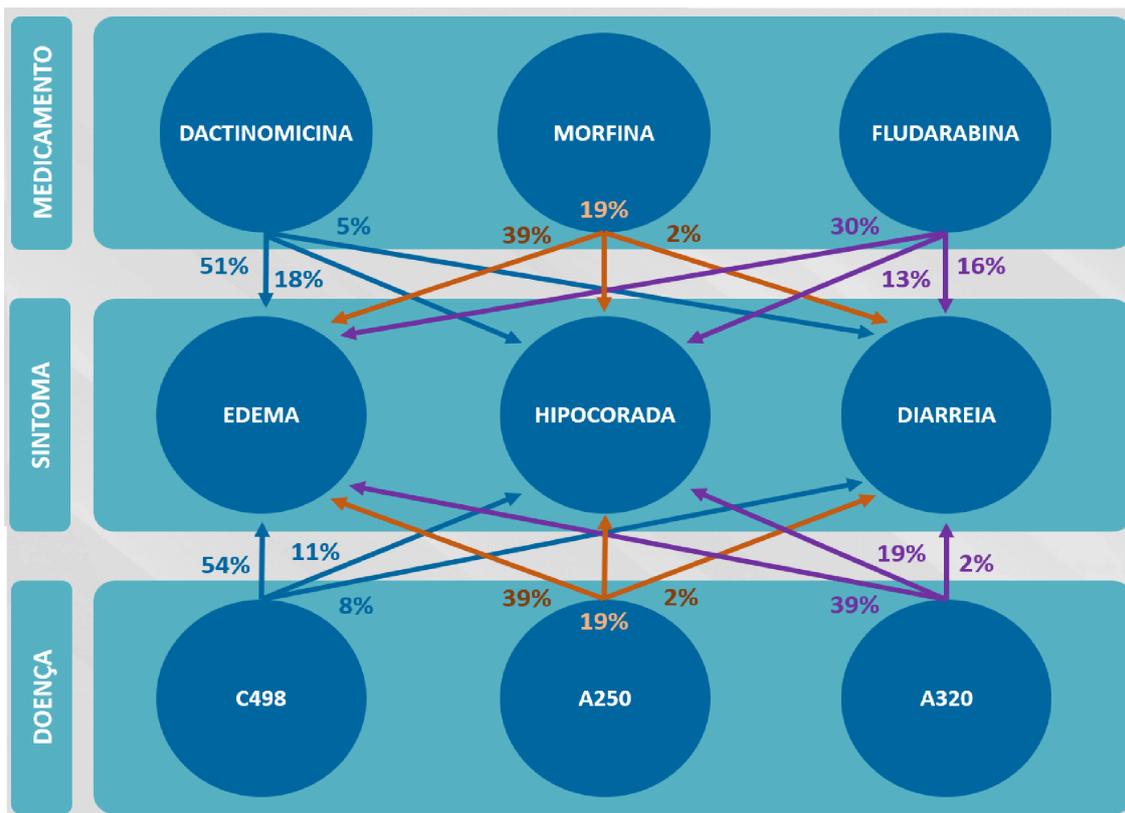


Figura 5.4: Gráfico Acíclico Dirigido - DAG Bayesiano

esses marcadores, como glicose, ureia, creatinina, hemócitos e leucócitos. Com base nas evidências dos resultados desses marcadores ao longo do tempo, verifica-se a ocorrência de uma tendência para um marcador, seja de subida ou de descida. Essa tendência pode estar relacionada à RAM desencadeada por um determinado medicamento. Por exemplo, se o paciente apresenta tendência de aumento de ureia nos últimos exames, talvez algum medicamento tenha causado essa variação. Este centro de fusão engloba a formação de um conjunto de dados longitudinal que permite a avaliação da temporalidade vista a seguir. Em alguns casos, uma relação de dependência do tempo pode ser estabelecida, como visto na seção **Análise Temporal**.

**Modelo para Análise de Exames Patológicos** A metodologia utilizada baseia-se em modelos de regressão linear e testes de hipóteses, utilizando o F-Test. A figura 5.5 mostra o gráfico de regressão linear. O modelo proposto considera as informações obtidas nos exames de sangue do paciente, conforme segue:

- a) A base de dados utilizada nesta etapa é composta por 2.338 prontuários e 11 variáveis.

- b) O banco de dados contém os resultados correspondentes aos valores de creatinina, glicose, ureia e hemograma total.
- c) 78 pacientes foram considerados nesta base.

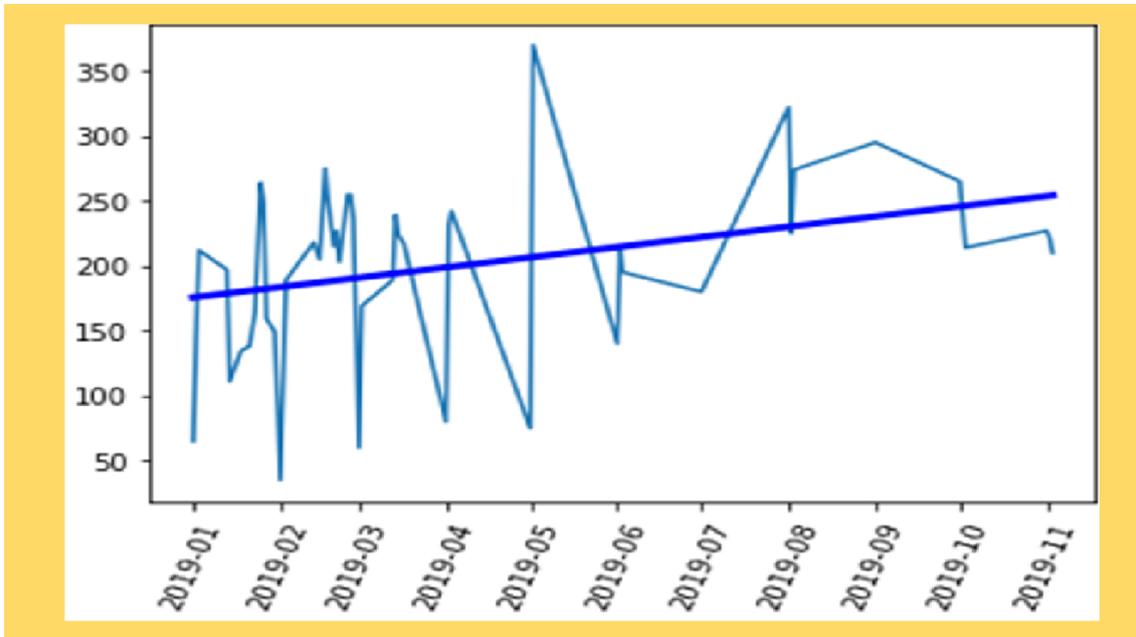


Figura 5.5: Regressão Linear - análise temporal do impacto do medicamento em exame

**Dependência Temporal - Análise de Regressão Linear** A figura 5.5 demonstra o gráfico de regressão linear que permite a análise temporal do impacto do medicamento em exame.

Como exemplo particular, considere o caso do paciente que recebeu morfina. O nível de plaquetas desse paciente foi contado ao longo período de janeiro a outubro de 2019.

- 1.- Um modelo de regressão linear foi ajustado para determinar a tendência do valor monitorado durante a aplicação da droga ao longo de 360 dias; Neste caso, pode-se observar uma tendência decrescente no valor da contagem de plaquetas ao longo do tempo. De acordo com a experiência médica, esta diminuição pode causar os sintomas: sangramento, edema, astenia e neuropatia.
- 2.- Para testar o efeito da droga sobre o valor do marcador, foi aplicado um teste de hipótese.  $H_0$ : A droga não afeta o marcador sanguíneo  $H_1$ : A droga afeta o marcador.
- 3.- Para cada exame de sangue, determina-se a probabilidade de apresentarem-se variações para cima ou para baixo como efeito da aplicação de um medicamento.

A significância estatística da tendência anterior foi validada por uma Análise de Regressão Linear da série temporal do paciente. Os resultados mostram evidências para rejeitar a hipótese nula  $H_0\beta_1 = 0$  em favor de  $H_1\beta_1 \neq 0$  ao obterem um *p-value* com uma significância menor do que de  $\alpha = 0,05$ . A partir dessa análise, conclui-se que a droga afetou o marcador de plaqueta. Efetivamente, ocorre uma tendência decrescente no volume das plaquetas em estudo.

### 5.3.5 Nível 4- Refinamento do Processo - Calcula a Probabilidade de RAM e Aplica IC

No nível 4, o objetivo dos processos de refinamento é fechar o ciclo do modelo. Um modelo de estimativa de geração de evidências para RAM é implementado, caracterizando a incidência dessa relação entre um medicamento e uma determinada reação ou sintoma. Esse nível permite fontes potenciais de informações aprimoradas, como rastreadores, novos medicamentos, novas RAM e alocação de sensores. Os parâmetros usados em diferentes níveis, como granularidade, frequência de captura e fontes de notícias, são otimizados e ajustados. Uma técnica eficaz é incluir fontes de dados pré-processadas de uma interface. Análises de causalidade mais precisas podem ser inseridas aqui, ajustando-se parâmetros para inferência com redes Bayesianas para melhorar o modelo.

Os níveis anteriores têm seus desempenhos monitorados. Há um fechamento de metas nesse momento. A identificação e a estimativa de RAM servem de base para a formação do IC, que orienta a geração de evidências para construir o quadro de discernimento do DST. O processo de geração do índice de credibilidade aplicado no domínio RAM, o processo de geração de evidências DST e a aplicação de um benchmark serão abordados a seguir.

A modelagem de um contexto de RAM é um desafio se o sintoma pode vir da doença e não for uma RAM causada pelo uso de medicamentos, ou pode ser resultado de uma interação de dois ou mais medicamentos. Este estudo estabelece critérios que foram analisados e revisados ao longo da pesquisa para atribuir medidas para estabelecer a credibilidade de uma RAM. As funções destinadas a quantificarem esses parâmetros são formalmente detalhadas.

- 1.- DepCausalRAM: o valor referente à força do arco a partir da inferência da rede Bayesiana:

$$DepCausalRAM(z) = streghnt(z)$$

Conforme descrito na seção DAG Bayesiano - Probabilidades Condicionais e visto nas tabelas 5.2 e 5.3 as probabilidades condicionais são calculadas, gerando a variável força e o gráfico que permite inferência Bayesiana como na figura 5.4.

- 2.- DepCorrelationRAM: O valor complementar de  $p$ -value obtido com o teste de Pearson foi utilizado para quantificar o efeito da doença na RAM. Para cada paciente, um valor de  $(1 - p\text{-value})$  foi calculado para medir o componente de correlação da credibilidade de RAM.

Para tanto, foi utilizado o teste do qui-quadrado de Pearson, modelando as hipóteses:

- $H_0$  : A doença causa a reação ou sintoma.
- $H_1$  : A doença não causa a reação ou sintoma

- 3.- EffectSizeRAM :o tamanho do efeito foi medido para considerar que uma RAM em que várias drogas interagiram é menos crível do que uma em que o efeito é isolado. Se  $n$  é o número de medicamentos que o paciente  $i$  tomou, então:

$$EffectSizeRAM(z) = \frac{1}{c} \sum_{j=1}^c L_j(z)$$

- 4.- DepTemporalRAM: se a potencial RAM está contido no *TemporalityBasis*, o registro é marcado com o valor booleano um. Caso contrário, é zero. Esta função é essencialmente um indicador de variável booleana.

Implementou-se um método PCA para representar as variáveis numa dimensão menor. Essa técnica resumiu os três fatores de crença (efeito de tamanho, efeito de conhecimento e efeito causal) em um único indicador. O método define uma combinação linear ponderada de fatores de crença atribuindo diferentes pesos a cada um deles.

O método busca a direção da máxima variabilidade nos dados para resumir a quantidade mais significativa de informações disponíveis nos primeiros componentes (novas dimensões). Assim, a PCA 1ª dimensão pode ser utilizada como um indicador linear que resume o comportamento dos fatores de crença.

Usou-se RF para a técnica de geração de evidências de RAM. O RF é um modelo de ML que ajusta uma série de árvores de decisão em subamostras do conjunto de dados e usa médias para melhorar as métricas de previsão. RAM envolve um contexto de alta

dimensão. Nesse tipo de cenário, é fundamental controlar a complexidade e o tamanho das árvores. Uma forma de controle é definir os valores de precisão preditiva e overfitting de controle. As variáveis são sempre permutadas aleatoriamente em cada divisão. Portanto, a melhor divisão encontrada pode variar, mesmo com os mesmos dados de treinamento.

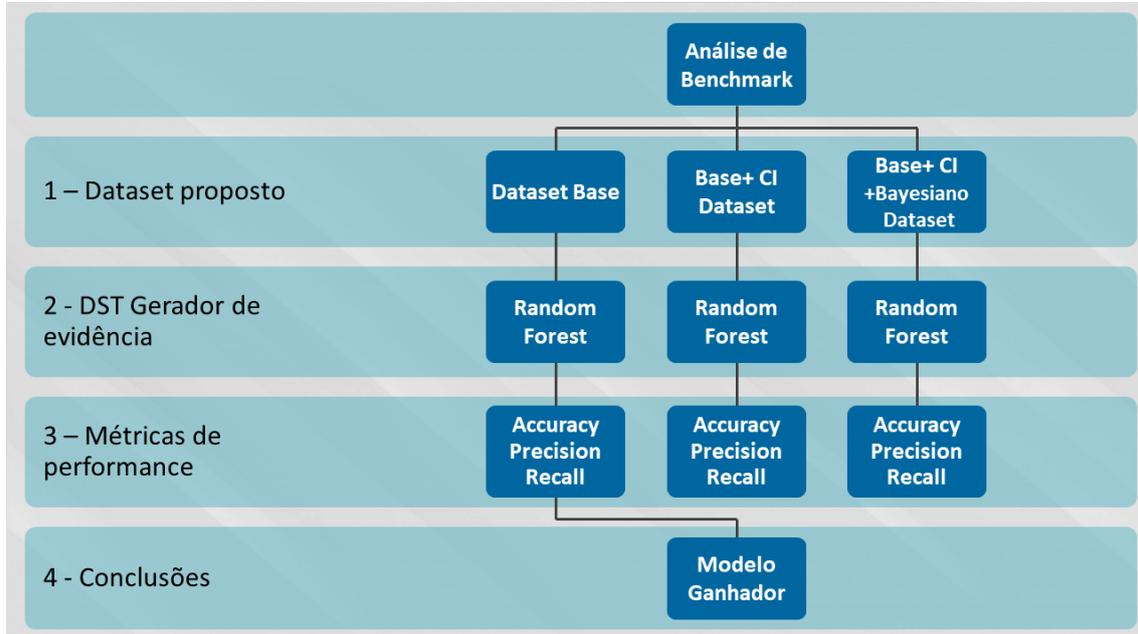


Figura 5.6: Estudo comparativo - Benchmark interno

As informações obtidas pela análise de séries temporais a partir de evidências temporais obtidas no domínio são agregadas, indicando relações de dependência entre as variáveis. Por fim, há um bloco referente à inferência causal que utiliza uma rede Bayesiana, na qual a força dos arcos é um indicador de peso para a formação do IC. Em cada bloco, é formado um BPA (formação de hipóteses) que corresponde às funções de massa.

Observe que há uma conjectura de hipóteses correspondentes a BPA criada a partir de cada nível e evolui à medida que as etapas do processo passam. São fatores de interação que combinam e produzem BPA com valor de massa  $m$  para cada hipótese.

A figura 5.6 demonstra o método de benchmark.

- Geração de Evidencias DST: a evidência é gerada para um contexto de DST por meio de um modelo de aprendizado Random Forest treinado em três diferentes cenários de conjuntos de dados.
- Cenários propostos: O primeiro cenário inclui uma base de dados com probabilidade frequentista. O segundo cenário contém um conjunto de dados que incorpora

o índice de credibilidade do PCA sem a variável Bayesiana. O terceiro cenário que também considera as variáveis da inferência Bayesiana.

- Métricas de desempenho: três métricas de desempenho foram utilizadas para quantificar diferentes aspectos e ter uma visão objetiva do desempenho dos modelos: exatidão, precisão, recall.

### 5.3.6 Nível 5 Gerenciamento de Processos e Interface Homem-Computador (IHC)

Por fim, o Nível 5 permite o gerenciamento do modelo e da IHC, a seleção de medicamentos e reações adversas e a especificação de métodos visuais de apresentação dos resultados.

Neste último nível do FIMS, são determinados os recursos de interação entre o ser humano e o sistema (HCI) e revistos os aspectos de gestão. O refinamento cognitivo é realizado. Iniciativas relacionadas à gestão de processos acontecem. Os métodos visuais são especificados para exibir os formulários e as informações de saída dos resultados do processo de fusão. As informações são utilizadas como conhecimento para especialistas que trabalham na seleção de medicamentos e reações relacionadas que são o foco de atenção para o gerenciamento de problemas.

Utilizou-se a ferramenta SHAP para interpretação em alguns resultados do modelo. As figuras a seguir mostram explicações visuais para prever RAM usando a técnica SHAP [Lundberg and Lee, 2017]. Assim, na parte vermelha do gráfico, temos as características que mais influenciam na predição de RAM e, na região azul, as características que menos influenciam no diagnóstico de RAM. Tomando a média do valor absoluto dos valores de Shapley, é possível estimar a contribuição global de cada variável para o modelo. Após o IC, percebe-se que a força do arco avaliada pelo DAG Bayesiano é a que mais contribui para o modelo, seguida do *p-value* do teste de correlação e do tamanho do efeito.

A figura 5.7 demonstra que o IC e seus componentes atuam como um primeiro filtro no processo decisório, avaliando o nível de crença em cada registro. Em seguida, temos as variáveis associadas ao efeito de cada medicamento, que acabam ajustando a probabilidade no processo de decisão. É essencial observar a flutuação de medicamentos e doenças (representadas pela CID) à medida que os cenários variam.

Considerando os valores de Shapley para cada saída na figura 5.8, é possível estimar a contribuição de cada variável para o nível de sintoma. Nesse exemplo, observa-se que valores mais altos de IC contribuem positivamente para a probabilidade do sintoma SANGRAMENTO, enquanto valores mais baixos diminuem a probabilidade do sintoma.

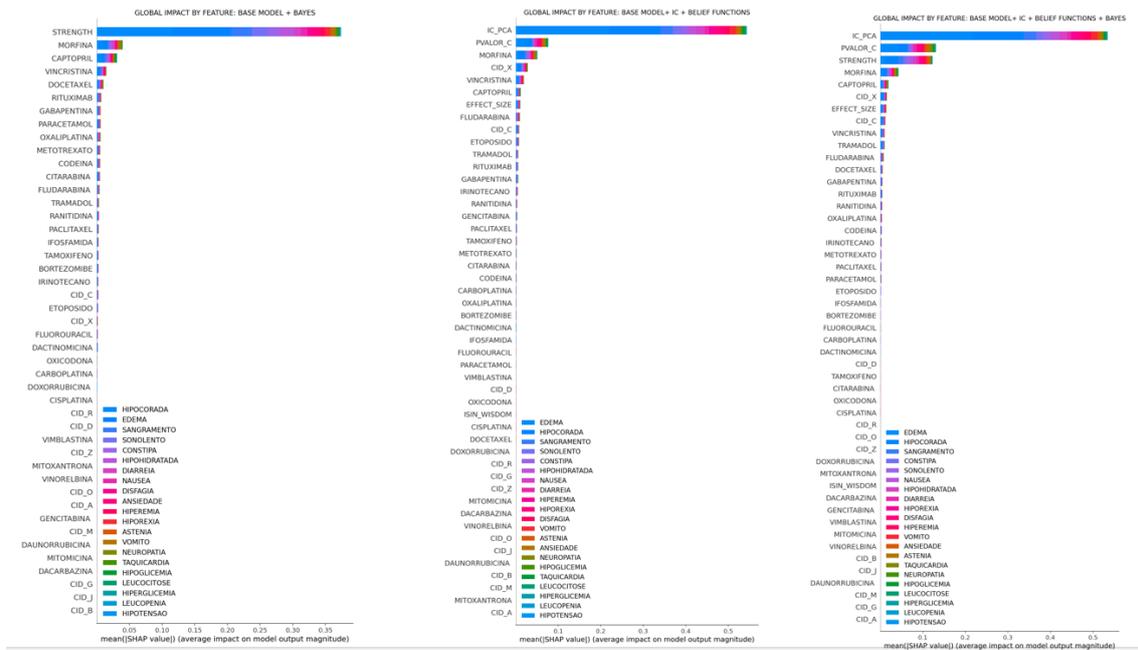


Figura 5.7: Shapley impacto global para RAM

Analisaram-se alguns medicamentos pelo SHAP e observou-se que alguns deles como DOCETAXEL, PACLITAXEL e IRINOTECAN são correlacionados positivamente. Esses medicamentos apresentam maior probabilidade de SANGRAMENTO. Tal fato pode indicar que o efeito desses medicamentos pode afetar a presença desse sintoma ou RAM.

A figura 5.9 ao nível do paciente, o SHAP oferece uma visualização que permite identificar as contribuições de cada variável na probabilidade atribuída a cada observação de paciente. Por exemplo, nesse caso, o modelo parte do valor esperado do sintoma NÁUSEA (39,5%). Observa-se que essa probabilidade diminui ou aumenta de acordo com o efeito de algumas drogas como citharabina, vincristina, morfina, gabapentina e captopril.

Por fim, na figura 5.10 permite uma análise sobre o impacto provocado por cada uma das variáveis componentes do IC, effectSize, DepCausal (Strength) e DepCorrelation, acabando por aumentar a probabilidade do sintoma em até 62,4 %.

## 5.4 Explicações sobre o a geração de BPA

Um modelo Random Forest é treinado a partir do conjunto de dados. No caso de RAM, o modelo considera o contexto de medicação atual do paciente e prevê as reações mais prováveis. O modelo foi enriquecido com os fatores de crenças individuais que geraram o IC durante a fase de treinamento, para melhorar o desempenho. As características

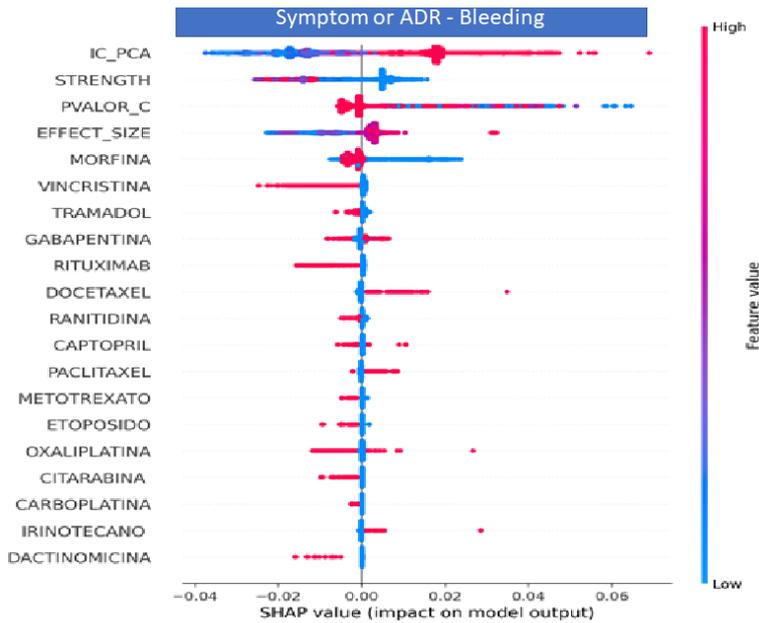


Figura 5.8: Impacto do valor SHAP - exemplo RAM: sangramento

gerais dos pacientes são consideradas como covariáveis e são utilizadas durante a fase de treinamento para estimar a probabilidade de ocorrência de RAM. O BPA é construído usando o método Random Forest a partir de covariáveis, IC-PCA e variáveis de crenças individuais (Depcausalstrength, Depcorrelational, Depeffectsize e Deptemporal). Depois disso, o BPA gerado pode ser combinado com as informações externas, como da base de conhecimento que contem as bulas dos medicamentos, utilizando a regra Dempster.

A figura 5.11 exibe a funcionalidade do front end da aplicação para o domínio RAM. O usuário escolhe o medicamento, sintoma, doença e paciente. O modelo apresenta a probabilidade de RAM, as variáveis que compõem o cálculo do IC. O IC e a incerteza são exibidos, juntamente com a crença obtida das bases de conhecimento, no caso da embalagem da bula do medicamento. O BPA corresponde ao campo "ADR PROBABILITY" no DASH e é a probabilidade estimada diretamente pelo *Random forest*. No exemplo a seguir está dentro da caixa vermelha e corresponde ao valor de 97,22%.

Como o IC é uma medida da confiança da RAM, seu complemento ( 1 - IC) representaria a incerteza como desconfiança da evidência em um contexto de mundo fechado.

A figura 5.11 exibe a funcionalidade do front end da aplicação RAM. O usuário escolhe o medicamento, sintoma, doença e paciente. O modelo apresenta a probabilidade de RAM, as variáveis que compõem o cálculo do IC. O IC e a incerteza são exibidos, juntamente com a crença obtida das bases de conhecimento, no caso oriundo da base de conhecimento que contém as informações de bula do medicamento.

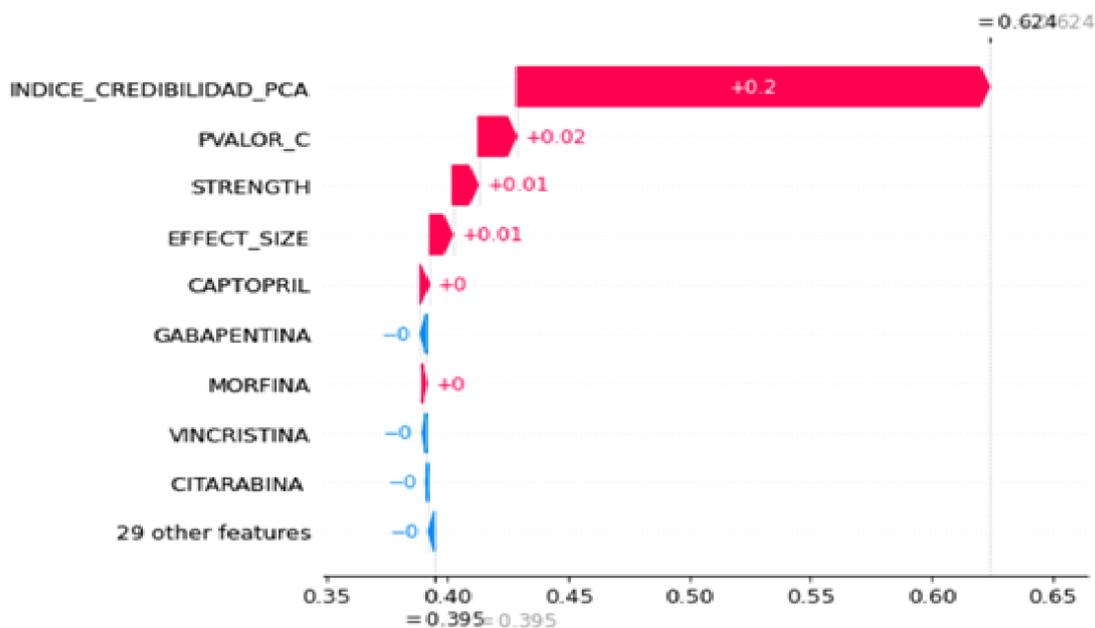


Figura 5.9: SHAP - Perspectiva do paciente

O paciente consumiu o medicamento Ranitidina e apresentou Edema. Ressalta-se que o sistema buscou, além da Ranitidina, outros dois medicamentos - Morfina e Docetaxel - que são prescritos para esse paciente. A probabilidade de morfina gerar edema é de 92,7%. Este é o BPA calculado pelo sistema, com base no IC e suas variáveis componentes. Existe informação externa, obtida a partir da bula do medicamento que pode ser combinada. A Figura 5.11 mostra o modelo realizando a combinação Dempster de BPA e informações externas e o modelo atribui a crença de 99,5% à RAM.

## 5.5 Ameaças à validade

As limitações mais importantes deste estudo estão relacionadas às ameaças de validade interna e do artefato relativo ao experimento. Não foi estabelecido um processo de revisão de código. Também não foram analisados casos de teste para a aplicação.

A obtenção de base de dados sobre RAM é muito difícil, especialmente por motivos relacionados à ética e privacidade. Ficou inviável avaliar a qualidade e desempenho do modelo em outros conjuntos de dados sobre RAM. Para se realizar um estudo comparativo, demonstrado na tabela 6.2, foram mencionados métricas, técnicas e bases de dados utilizadas em outros estudos da literatura.

Não foi possível a validação por médicos. Profissionais de farmácia auxiliaram a validação da ontologia utilizada [Cai et al., 2015]. Foram utilizadas 30 drogas no experi-

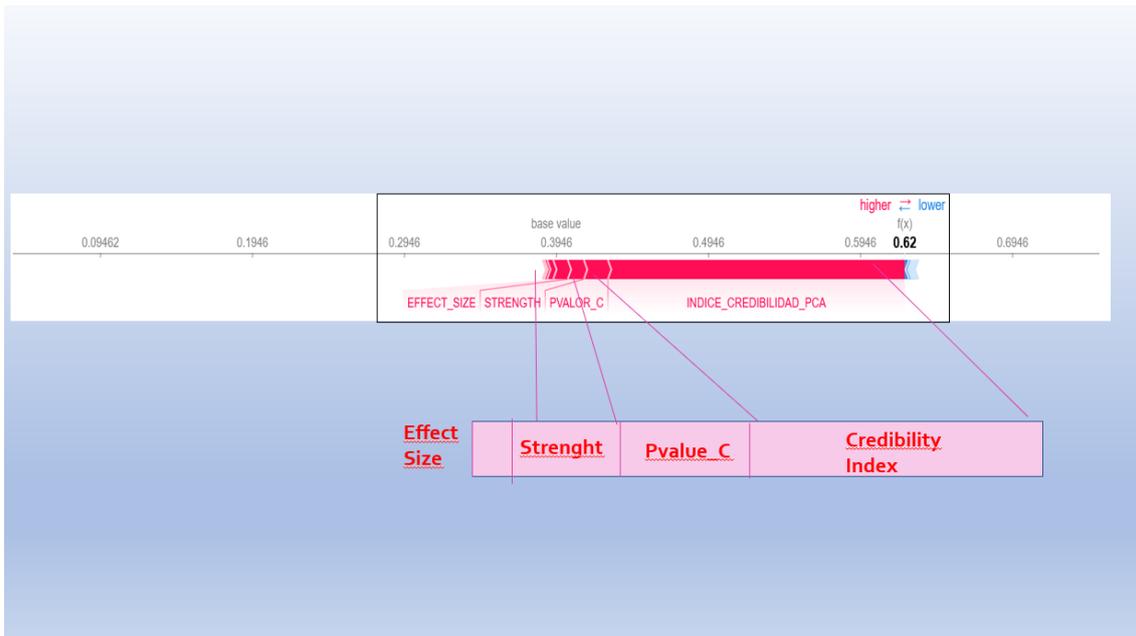


Figura 5.10: SHAP Resumo da perspectiva do paciente

mento, escolhidas com o auxílio de profissionais de farmácia.

Apesar da utilização do método PCA linear atingiu resultados satisfatórios, uma possibilidade de evolução para um método não linear pode alcançar métricas mais efetivas.

A taxonomia de RAM normalmente empregada classifica os eventos em grave ou não grave, o que leva a interpretação dos resultados, que podem influenciar os achados. O DEP-DST foi projetado para acolher as opiniões de especialistas farmacêuticos, que não foram coletadas nessa instânciação e não tratam, de forma automática, a categorização da RAM em relação à gravidade.

Do ponto de vista de ameaças de validade externa, um outro domínio foi explorado e os resultados serão apresentados no capítulo seguinte. Essa é uma tentativa para avaliar capacidade de generalização do modelo e, apesar de chegarmos a inferências relevantes sobre a influencia da frequência cardíaca na evolução de doenças, não foi possível contar com um especialista em cardiologista para validação.

## 5.6 Considerações Finais sobre a implantação do modelo no domínio RAM

Esse modelo foi parcialmente implantado num hospital oncológico e poderá ser aplicado em outras instituições de saúde. Adicionalmente foi aplicado em uma base externa que avalia doenças do coração. O *transfer learning* pode ser uma excelente oportunidade, principalmente para aplicar o modelo proposto em outras instituições de saúde interessa-

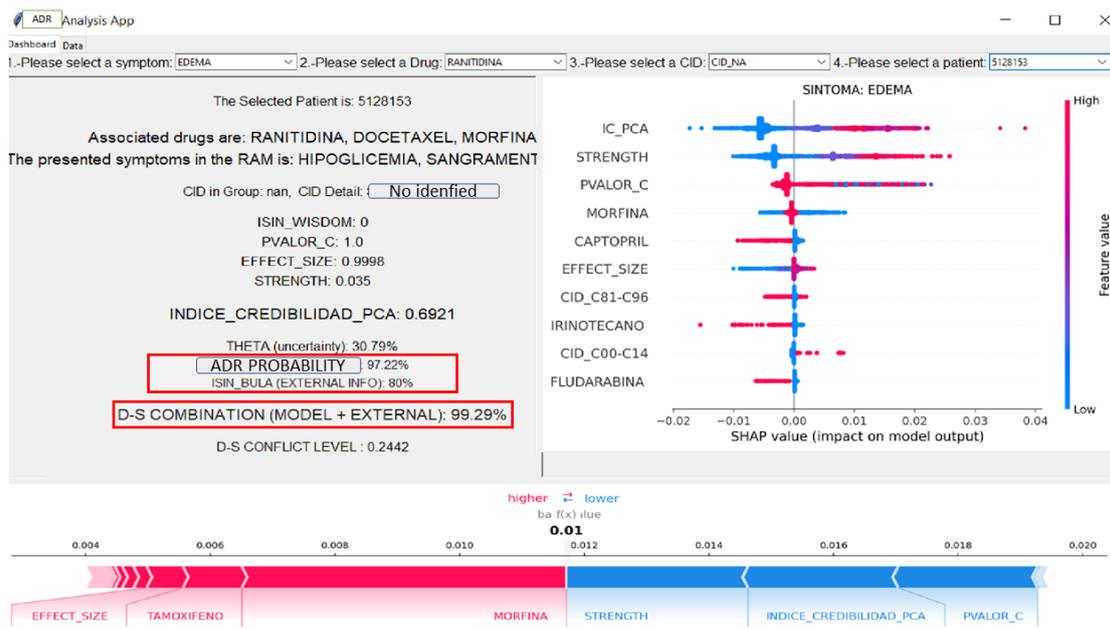


Figura 5.11: Front end - ilustração de um diagnóstico de RAM

das numa instância em seu ambiente. Dois estudos relevantes para a aplicação da aprendizagem por transferência são [Zheng et al., 2020] e [Hao and Chen, 2021]. Eles usam uma técnica de aprendizado que pode otimizar significativamente o tempo de treinamento e a necessidade de atividades de notificação e rotulagem de casos.

O modelo FIMS usando DST com aquisição de credibilidade foi alcançado. O IC gerado pela combinação de fatores de correlação, causais e temporais funcionou para a geração de hipóteses em um quadro de discernimento DST.

A análise dos dados ao longo do tempo por meio de regressão linear funcionou para detectar tendências na variação dos parâmetros sanguíneos obtidos por meio dos exames realizados pelos pacientes.

O benchmark proposto evidenciou a melhora das métricas de desempenho através da incorporação paulatina e sucessiva de variáveis componentes do índice de credibilidade, com destaque para a inferência bayesiana, comprovando que as relações de dependência, por inferência causal e correlação, mostram-se importantes para aumentar a precisão do modelo.

## 6. Resultados

Este capítulo apresenta dois estudos comparativos na aplicação do DEP-DST. Os resultados obtidos pelo método no domínio de RAM, conforme aplicação detalhada no capítulo anterior, e no domínio de doenças cardíacas. Os resultados mostram métricas que indicam que o modelo agrega valor em tarefas de diagnóstico.

### 6.1 Resultados no domínio de RAM

A figura 6.1 mostra como o IC foi padronizado para se enquadrar em uma escala percentual entre 0 e 1. Um possível ponto de corte para o índice com o limite superior da zona de baixa credibilidade passa a ser considerado o primeiro valor interquartil em que o intervalo (0, 0,21). Seria o equivalente a RAM com menos do que 21% de credibilidade.

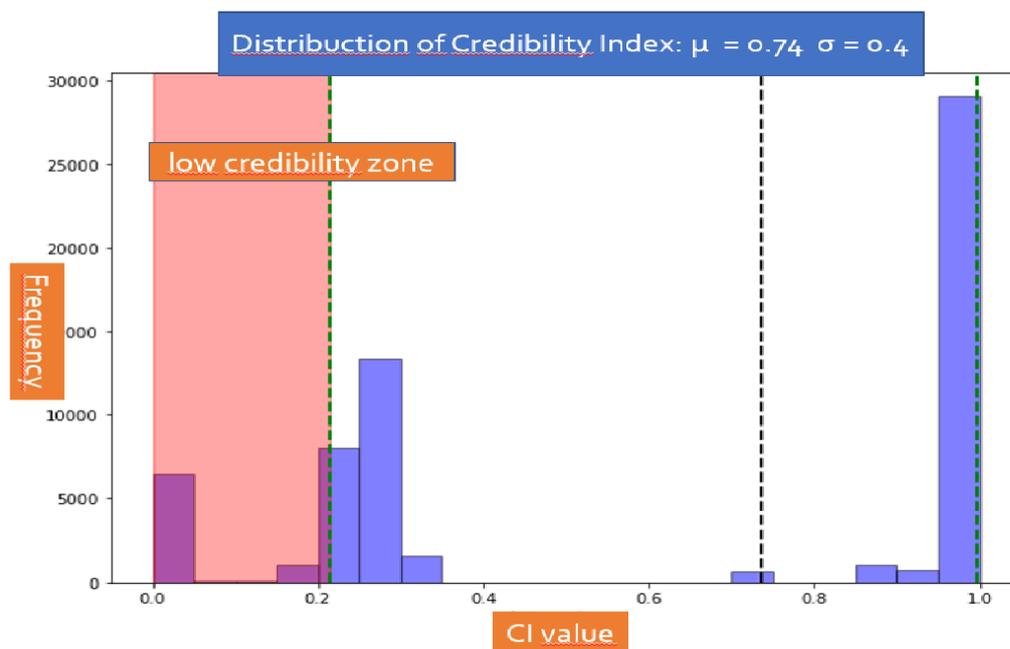


Figura 6.1: Distribuição do Índice de Credibilidade

Vamos explicar os resultados do índice de credibilidade a partir do histograma exibido na figura 6.1 e em seguida vamos explorar os resultados obtidos em cenários com o uso da técnica *Random Forest* a partir do conjunto de dados de RAM em três versões com incremento paulatino dos atributos que compõe o IC. Dessa forma, se utiliza a mesma técnica *Random Forest* avaliando três conjunto de dados formados a partir da mesma base, com o incremento de variáveis que integram o IC, demonstrando, através da obtenção de melhores resultados de métricas precisão, acurácia e supervisão.

A tabela 6.1 consolida os resultados do benchmark entre as versões do conjunto de dados de RAM incrementadas pelos componentes de IC e finalmente aumentadas com a variável bayesiana *strength*.

Tabela 6.1: Análise de Benchmark RAM

CENÁRIO	DESCRIÇÃO	ACURÁCIA	PRECISÃO	REVOCAÇÃO	GANHO
Conjunto de dados de base RAM	Inclui as variáveis geradas no contexto de interações múltiplas, aquelas variáveis binárias que indicam a presença ou ausência do medicamento no período de observação do sintoma	81%	59%	61%	'n / D'
Conjunto de dados de base RAM + CI	Inclui as variáveis do conjunto de dados Base mais o percentual de credibilidade obtido através do índice PCA (sem bayesiano)	93%	85%	86%	21%
Conjunto de dados de base RAM + IC+ Bayes	Inclui as variáveis do conjunto de dados Base mais o percentual de credibilidade obtido através do índice PCA mais o modelo de rede Bayesiana.	95%	90%	91%	25%

A figura 6.2 mostra a matriz de confusão obtida pela execução da técnica *Random Forest* aplicada sobre o conjunto de dados sem os componentes do índice de credibilidade. A figura 6.3 demonstra a matriz de confusão obtida a partir da *Random Forest* aplicada ao conjunto de dados base incrementado com os componentes IC de associação e temporalidade. A figura 6.4 começa a partir do conjunto de dados com os componentes de correlação, interação e temporal do índice de credibilidade e adiciona os indicadores de inferência bayesiana do componente de força. Esse terceiro cenário, portanto, equivale à configuração final das variáveis componentes do índice de credibilidade .

Os valores da diagonal da matriz da figura (6.4 são maiores que da figura 6.3, o que demonstra a evolução das métricas, o que pode ser observado quando a precisão aumenta de 81% para 95% no segundo caso (modelo base + IC+ componente bayesiano).

No intuito de aferir a capacidade do modelo em outros ramos de atividade, a seguir

PRECISÃO		Geração de Evidências com RANDOM FOREST - Conjunto de dados Base																		61%			
81%		EDEMA	HIPOCORADA	SANGRAMENTO	CONSTIPA	NAUSEA	SONOLENTO	HIPOREXIA	HIPOHIDRATADA	DIARREIA	HIPEREMIA	DISFAGIA	VOMITO	ASTENIA	ANSIEDADE	TAQUICARDIA	NEUROPATIA	HIPOGLICEMIA	LEUCOCITOSE	HIPERGLICEMIA	LEUCOPENIA	HIPOTENSÃO	RECALL
TARGET	EDEMA	23.008	795	196	75	21	4	24	0	103	28	13	11	33	5	8	28	0	8	0	0	0	94%
	HIPOCORADA	9.250	1.302	59	45	17	0	21	0	34	9	7	13	10	5	1	10	0	0	0	0	0	85%
	SANGRAMENTO	5.414	162	334	11	9	0	1	0	14	7	3	0	7	1	3	7	0	2	0	0	0	84%
	CONSTIPA	3.029	64	62	197	17	1	19	0	21	1	7	2	14	4	0	9	0	0	0	0	0	56%
	NAUSEA	1.592	132	37	6	69	0	21	0	35	0	2	17	5	3	0	9	0	0	0	0	0	43%
	SONOLENTO	1.767	6	5	2	0	6	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	60%
	HIPOREXIA	1.478	24	17	27	18	2	64	0	21	1	2	2	2	3	0	0	6	0	2	0	0	52%
	HIPOHIDRATADA	1.517	126	9	7	2	0	0	0	0	3	2	1	0	0	0	1	0	0	0	0	0	0%
	DIARREIA	1.067	65	30	13	8	0	10	0	296	1	4	0	14	0	0	3	57	2	0	0	0	75%
	HIPEREMIA	1.407	28	10	9	1	0	0	0	11	65	0	0	0	0	2	5	8	0	0	0	0	71%
	DISFAGIA	1.352	39	28	14	0	0	4	0	4	2	50	0	3	0	0	0	0	0	0	0	0	79%
	VOMITO	1.209	37	24	9	17	0	8	0	32	0	4	26	1	0	0	4	0	0	0	0	0	28%
	ASTENIA	723	31	12	7	11	0	6	0	37	0	6	0	90	0	1	2	0	0	0	0	0	97%
	ANSIEDADE	670	6	8	6	4	0	0	0	1	0	0	0	2	16	0	0	0	0	0	0	0	89%
	TAQUICARDIA	590	35	4	0	4	0	0	0	15	0	0	0	6	0	36	0	4	0	0	0	0	78%
	NEUROPATIA	383	26	5	2	2	0	4	0	11	0	1	0	3	0	0	78	0	0	0	0	0	96%
	HIPOGLICEMIA	335	6	9	2	0	0	0	0	48	0	0	0	1	0	0	0	57	0	0	0	0	98%
	LEUCOCITOSE	340	18	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	23	0	0	0	100%
	HIPERGLICEMIA	213	3	13	0	0	0	1	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0%
	LEUCOPENIA	55	0	0	1	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0%
HIPOTENSÃO	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	
59%	ACURÁCIA	42%	62%	55%	63%	51%	75%	54%	0%	57%	87%	70%	57%	88%	100%	97%	92%	100%	85%	0%	0%	0%	67%

Figura 6.2: Benchmark 1 - Modelo Base

vão-se analisar comparativamente os resultados do DEP-DST em relação a outros trabalhos no domínio de RAM e em um domínio distinto.

## 6.2 Benchmark com outros trabalhos no domínio de RAM

No domínio de RAM, comparou-se o DEP-DST contra 5 outros trabalhos, que implementaram técnicas diferentes de aprendizado de máquina.

A tabela 6.2 espelha o resultado comparativo, em que o DEP-DST apresenta resultados melhores que outros cinco trabalhos que utilizaram técnicas diversas como SVM, DT e de aprendizado profundo.

## 6.3 Benchmark com base de domínio de deonças cardíacas

No intuito de aferir a capacidade de generalização e desempenho do modelo realizou-se um teste comparativo com uma base pública muito utilizada em pesquisas científicas de saúde e de ciência da computação. Tentou-se utilizar bases de câncer, onde poderia ser buscada a ajuda em análise de dados por especialistas, mas as variáveis eram características das imagens de tumores, não contendo características demográficas ou da doença, nem de exames laboratoriais que permitissem análises causais e temporais que são a ênfase de nosso modelo.

PRECISÃO		Geração de Evidências com RANDOM FOREST - Conjunto de dados Base + IC (sem Bayesiano)																				92%	
89%		EDEMA	HIPOCORADA	SANGRAMENTO	CONSTIPA	NAUSEA	SONOLENTO	HIPOREXIA	HIPOHIDRATADA	DIARREIA	HIPEREMIA	DISFAGIA	VOMITO	ASTENIA	ANSIEDADE	TAQUICARDIA	NEUROPATIA	HIPOGLUCEMIA	LEUCOCITOSE	HIPERGLUCEMIA	LEUCOPENIA	HIPOTENSAO	RECALL
TARGET	EDEMA	23.709	143	193	79	40	19	31	1	7	30	17	24	17	0	8	25	6	10	1	0	0	97%
	HIPOCORADA	1.002	9.427	185	18	41	4	8	12	2	11	46	7	3	0	2	1	8	6	0	0	0	96%
	SANGRAMENTO	4.062	48	1.657	53	32	4	15	0	5	27	27	3	7	0	3	5	8	15	4	0	0	89%
	CONSTIPA	2.196	21	86	1.019	12	0	31	0	13	15	13	10	12	0	0	2	6	11	0	0	0	83%
	NAUSEA	991	19	132	39	612	0	39	1	8	14	7	47	6	0	2	3	8	0	0	0	0	82%
	SONOLENTO	107	4	15	2	0	1.616	0	0	0	13	0	1	9	22	0	0	0	1	0	0	0	97%
	HIPOREXIA	1.034	9	69	31	35	4	441	0	9	9	7	6	0	0	1	0	4	9	0	0	0	84%
	HIPOHIDRATADA	108	9	30	3	14	0	1	1.491	0	2	4	0	0	0	0	0	3	0	3	0	0	98%
	DIARREIA	214	17	59	19	6	0	20	0	1.181	2	3	5	17	0	1	0	25	1	0	0	0	94%
	HIPEREMIA	792	30	47	13	5	16	6	0	3	597	24	3	1	1	2	0	6	0	0	0	0	90%
	DISFAGIA	690	6	76	24	11	0	5	0	0	7	663	5	1	0	2	0	3	3	0	0	0	95%
	VOMITO	831	8	57	27	27	3	16	0	2	12	12	357	5	0	3	2	1	8	0	0	0	80%
	ASTENIA	173	22	25	13	2	80	6	0	12	1	0	6	570	9	1	0	5	1	0	0	0	97%
	ANSIEDADE	7	0	0	0	0	211	0	0	0	0	0	0	5	487	0	1	0	0	0	0	2	98%
	TAQUICARDIA	321	12	19	1	1	0	4	1	0	4	2	3	5	0	319	0	1	0	0	1	0	98%
	NEUROPATIA	97	11	12	6	0	12	2	0	0	0	1	0	3	7	0	361	0	3	0	0	0	97%
	HIPOGLUCEMIA	151	0	10	2	5	0	2	0	16	8	2	5	0	0	0	0	254	3	0	0	0	99%
	LEUCOCITOSE	198	1	9	0	2	2	6	0	0	2	3	2	2	0	0	0	2	155	0	0	0	97%
	HIPERGLUCEMIA	101	2	26	5	2	0	1	1	1	1	1	3	0	0	0	10	0	1	80	0	0	88%
	LEUCOPENIA	4	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	2	0	0	50	0	96%
HIPOTENSAO	0	0	0	0	0	14	0	0	0	0	0	0	0	3	0	0	0	0	0	0	8	73%	
96%	ACURÁCIA	64%	98%	71%	81%	85%	83%	80%	100%	96%	89%	91%	80%	97%	96%	100%	97%	96%	95%	100%	98%	80%	95%

Figura 6.3: Benchmark 2 - Modelo Base + IC

Tabela 6.2: Análise de Benchmark - Externa

MÉTRICA	FLORESTA ALEATÓRIA	DEEP LEARNING bi-LSTM	Neural LSTM com CRF	SVM	ÁRVORE DE DECISÃO	Este trabalho DEP-DST
Acurácia	92,3	86% [Santiso et al., 2018]	84,5% [Dandala et al., 2018]	90,7% [Cassillas et al., 2016]	87,3% [Imai et al., 2017]	90%
Precisão	88% [Cassillas et al., 2016]	86% [Santiso et al., 2018]	84% [Dandala et al., 2018],	70,6% [Cassillas et al., 2016]	87,3% [Imai et al., 2017]	95%
Revocação	85,7% [Cassillas et al., 2016]	86% [Santiso et al., 2018]	71,9% [Dandala et al., 2018],	59,9% [Cassillas et al., 2016]	87,3% [Imai et al., 2017]	91%
Número de hipóteses	mantém	mantém	mantém	mantém	mantém	reduz
Grau de certeza	Não mudou	aumenta				
Base usada	EHR Hospital Galdakao-Usansolo	corpus ixa e MedGS MedCH	pacientes com câncer EHR	EHR Hospital Galdakao-Usansolo	EHR okaido University Hospital	EHR INCA - Hospital do Câncer brasileiro
Características da Base	alta dimensionalidade descritiva					

Utilizou-se a base *UCI heart disease* da Universidade de Cleveland para experimentos sobre doenças cardíacas. Esse conjunto de dados reúne atributos gerais do paciente e campos contendo valores de diferentes medições de sangue e informações sobre doenças ou condições de risco do paciente, conforme descrito na figura 6.5.

PRECISÃO		Geração de Evidências com RANDOM FOREST - Conjunto de dados Base + IC + Bayesiano																			95%		
91%		EDEMA	HIPOCORADA	SANGRAMENTO	CONSTIPA	NAUSEA	SONOLENTO	HIPOREXIA	HIPOHIDRATADA	DIARREIA	HIPEREMIA	DISFAGIA	VOMITO	ASTENIA	ANSIEDADE	TAQUICARDIA	NEUROPATIA	HIPOGLICEMIA	LEUCOCITOSE	HIPERGLICEMIA	LEUCOPENIA	HIPOTENSÃO	RECALL
TARGET	EDEMA	24.073	47	153	19	15	11	7	2	11	4	1	5	3	0	0	6	3	0	0	0	0	99%
	HIPOCORADA	1.045	9.490	162	3	18	0	4	5	2	1	43	1	2	1	0	0	6	0	0	0	0	97%
	SANGRAMENTO	4.108	31	1.769	3	11	2	10	1	12	9	12	1	0	0	0	0	4	2	0	0	0	96%
	CONSTIPA	2.254	12	162	964	15	0	8	0	7	6	8	2	5	0	0	0	4	0	0	0	0	82%
	NAUSEA	1.011	12	176	16	633	0	33	1	7	3	6	17	4	0	0	2	6	0	1	0	0	89%
	SONOLENTO	100	1	22	0	0	1.649	0	0	1	2	0	0	0	12	1	0	2	0	0	0	0	99%
	HIPOREXIA	964	3	109	19	30	0	515	0	7	6	6	2	3	0	0	0	2	2	0	0	0	90%
	HIPOHIDRATADA	112	4	40	0	10	0	3	1.495	0	1	2	0	0	0	0	2	6	0	1	0	0	99%
	DIARREIA	176	12	68	10	6	0	1	0	1.273	2	0	0	6	0	0	0	14	1	0	1	0	98%
	HIPEREMIA	806	10	99	13	10	12	3	0	0	570	18	0	0	1	0	0	4	0	0	0	0	92%
	DISFAGIA	706	2	132	5	10	0	4	0	0	631	1	0	0	0	0	0	2	0	0	0	0	97%
	VOMITO	849	3	110	15	31	0	15	1	1	4	8	325	2	0	0	2	2	3	0	0	0	82%
	ASTENIA	174	6	32	8	3	94	5	0	11	2	3	2	570	11	0	1	4	0	0	0	0	97%
	ANSIEDADE	4	0	0	0	0	215	0	0	0	0	0	0	4	488	0	0	0	0	0	0	2	99%
	TAQUICARDIA	343	5	38	3	8	0	2	0	0	3	5	1	3	0	282	0	0	1	0	0	0	99%
	NEUROPATIA	94	6	12	3	0	20	0	0	1	1	2	0	3	6	0	367	0	0	0	0	0	98%
	HIPOGLICEMIA	157	0	37	8	6	0	5	0	11	0	3	2	0	0	0	0	227	2	0	0	0	99%
	LEUCOCITOSE	200	1	36	3	7	0	2	0	0	1	1	0	0	0	0	0	0	133	0	0	0	100%
	HIPERGLICEMIA	58	0	9	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	164	0	100%
	LEUCOPENIA	4	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	51	0	98%
HIPOTENSÃO	0	0	0	0	0	16	0	0	0	0	0	0	0	1	0	0	0	0	0	0	8	89%	
96%	ACURÁCIA	65%	99%	62%	90%	84%	82%	87%	100%	97%	95%	92%	93%	98%	96%	100%	100%	98%	98%	100%	100%	80%	97%

Figura 6.4: Benchmark 3 - Modelo Base + Índice de Credibilidade + componente inferência bayesiana

Utilizou-se como comparativo o estudo de [Latha and Jeeva, 2019], que obteve uma acurácia de 85% em seu melhor método relacionando diversas pesquisas que alcançaram uma média de 80% de acerto. O DEP-DST cobriu esse estudo, atingindo 95% de acurácia. Nessa aplicação do modelo, mais uma vez ficou notória a importância da inferência causal. Utilizou-se a estratégia comparativa com e sem as variáveis que forma o IC, chegando a decisão análoga. A importância da inferência causal é decisória para o modelo. O modelo puro, sem nenhum componente do IC chega a 70%, e nos três cenários seguintes se incrementam a variável do teste do quiquadrado de Pearson, variável de correlação, o resultado melhora para 76%. Com o incremento do fator causal a acurácia chega aos 95%. Esses cenários estão na figura 6.6, que apresenta as curvas AUC relativos aos resultados demonstrados na tabela 6.3 evoluindo os cenários de cima pra baixo da esquerda para a direita.

Demonstrando que os componentes do IC agregam valor às métricas, passou de 70% para 86% com a inferência Bayesiana e atingiu 95% de precisão na AUC (curva ROC) com os componentes do IC agregados à inferência bayesiana.

No caso da aplicação em doenças cardíacas os resultados da inferência probabilística mostram o comportamento da probabilidade de sofrer de uma doença cardíaca (num>0) quando há frequência cardíaca normal (thal=3), ou há defeitos na frequência (thal = 6 ou 7). Pode-se notar que a probabilidade de sofrer de doenças cardíacas aumenta à medida que são identificados defeitos na frequência cardíaca do paciente. Com frequência

<ul style="list-style-type: none"> <li>• 1.-age: idade em anos</li> <li>• 2.-sex: sexo (1 = masculino; 0 = feminino)</li> <li>• 3.-cp: tipos de dor no peito <ul style="list-style-type: none"> <li>-- Valor 1: angina típica</li> <li>-- Valor 2: angina típica</li> <li>-- Valor 3: dor não-anginosa</li> <li>-- Valor 4: assintomática</li> </ul> </li> <li>• 4.-trestbps: pressão arterial em repouso (em mm Hg no momento de admissão no hospital)</li> <li>• 5.-chol: (colesterol) soro do colesterol em mg/dl</li> <li>• 6.-fbs: (glicose) (glicemia no sangue em jejum &gt; 120 mg/dl) (1 = verdadeiro; 0 = falso)</li> <li>• 7.-restecg: (Fator Temporal) resultado do eletrocardiograma em repouso <ul style="list-style-type: none"> <li>-- Valor 0: normal</li> <li>-- Valor 1: com anormalidade da onda ST-T (inversões da onda T e/ou elevação ou depressão de ST &gt; 0,05 mV)</li> <li>-- Valor 2: mostrando provável ou definitiva hipertrofia ventricular esquerda pelos critérios de Estes</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• 8.-thalach: frequência cardíaca máxima atingida</li> <li>• 9.-exang: angina induzida por exercício (1 = sim; 0 = não)</li> <li>• 10.-oldpeak = depressão ST induzida por exercício em relação ao repouso</li> <li>• 11.-slope: a inclinação do segmento ST em exercício de pico <ul style="list-style-type: none"> <li>-- Valor 1: ascendente</li> <li>-- Valor 2: plano</li> <li>-- Valor 3: descendente</li> </ul> </li> <li>• 12.-ca: número de vasos principais(0-3) coloridos por fluoroscopia</li> <li>• 13.-thal: 3 = normal; 6 = defeito corrigido; 7 = defeito reversível</li> <li>• 14.-num: diagnóstico de doença cardíaca(estado angiográfico da doença) <ul style="list-style-type: none"> <li>-- Valor 0: estreitamento do diâmetro &lt; 50%</li> <li>-- Valor 1: estreitamento do diâmetro &gt; 50%</li> </ul> </li> </ul> <p>(em qualquer vaso principal: atributos 59 a 68 são vasos)</p>
<ul style="list-style-type: none"> <li>• <b>Fator Sanguíneo:</b> Variáveis usadas para realizar a análise sanguínea temporal do IC (trestbps, chol, thalach)</li> <li>• <b>Doenças (CID):</b> Variable usada para incluir el fator de la diabetes como enfermedad del paciente (fbs)</li> <li>• <b>Fator Temporal:</b> Resultados dos eletrocardiograma resumidos em uma análise de regressão linear refletindo a tendencia(restecg)</li> </ul>	

Figura 6.5: Descrição do conjunto de dados Heart Disease- UCI Cleveland

Tabela 6.3: Análise de Benchmark aplicado ao domínio doenças do Coração

CENÁRIO	DESCRIÇÃO	ACURÁCIA	PRECISÃO	REVOCAÇÃO	AUC
Conjunto de dados de base Coração	dados puros do dataset Heart disease UCI. sem as variáveis geradas no contexto do índice de credibilidade	69,2%	65,2%	68,2%	70 %
Conjunto de dados de base Coração + IC	Inclui as variáveis do conjunto de dados Base mais o percentual de credibilidade obtido através do índice PCA (sem bayesiano)	76%	73,1%	74,2%	76 %
Conjunto de dados de base Coração + DAG bayesiano	Inclui as variáveis do conjunto de dados Base mais o percentual de credibilidade obtido através do índice PCA mais o modelo de rede Bayesiana.	84,8%	77,7%	94,2%	86 %
Conjunto de dados de base Coração + IC+ Bayes	Inclui as variáveis do conjunto de dados Base mais o percentual de credibilidade obtido através do índice PCA mais o modelo de rede Bayesiana.	94,9%	95,5%	93,4%	95 %

alterada (thal=7) aumenta em mais de 20% o agravamento de quadro.

## 6.4 Considerações Finais sobre os resultados

Os resultados obtidos pelo DEP-DST tanto no domínio RAM quanto no domínio de doenças cardíacas demonstram que o IC proposto reúne com eficácia a variabilidade provocada por dependências causais e temporais, além de considerar as interações de efeitos

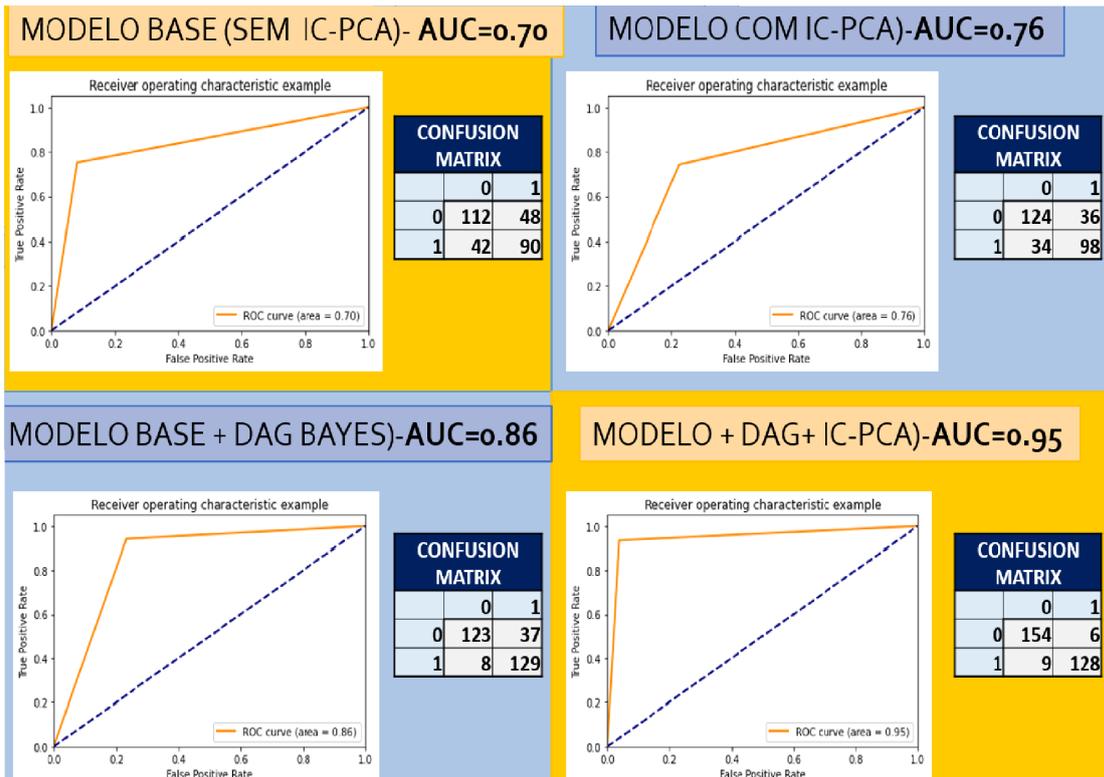


Figura 6.6: Estudo comparativo-evolução da precisão com o incremento gradual de variáveis do IC

e associação.

A capacidade de generalização do modelo com os bons resultados obtidos na base de doenças cardíacas demonstra que o método pode ser útil em tarefas de diagnóstico.

A inovação propiciada pela agregação de variáveis causais e temporais, considerando aspectos de correlação e interação, pode ser aplicada em domínios de diagnóstico, seja para a indústria de saúde, financeira ou de indústria. A teoria da evidência tem no IC proposto um aliado para melhor representar hipóteses em um ambiente de incerteza.

## 7. Conclusão

Neste capítulo, apresentam-se as conclusões e principais contribuições dessa investigação. Adicionalmente, se apresentam as limitações e questões em aberto não tratadas nesse trabalho. Por fim são indicados possíveis caminhos para trabalhos futuros a partir dessa pesquisa.

O objetivo de criar um modelo FIMS usando DST com aquisição de credibilidade foi alcançado. O IC gerado pela combinação de fatores de correlação, causais e temporais mostrou uma melhora na geração de hipóteses em um quadro de discernimento DST. O modelo proposto foi aplicado no domínio RAM em formato de estudo de caso. E para mostrar a capacidade de generalização do modelo, instanciou-se uma aplicação na base de doenças cardíacas de Cleveland.

A detecção de dependência causal se mostrou relevante para redução de incerteza na geração de hipóteses, tanto no domínio de RAM quanto no domínio de doenças cardíacas. Essa análise causal implementada a partir de inferência bayesiana, demonstra que várias das queixas obtidas nas notas clínicas dos pacientes são oriundas da evolução da doença e não são um efeito da ingestão de drogas, ou seja, não são RAM.

No caso da aplicação em doenças cardíacas os resultados da inferência probabilística realça o impacto na probabilidade de sofrer de uma doença cardíaca em relação a variação no comportamento da frequência cardíaca normal, condicionada a outras variáveis importantes como a presença de dor no peito e o gênero do paciente. A probabilidade de sofrer de doenças cardíacas graves aumenta mais de 20% à medida que são identificados defeitos na frequência cardíaca do paciente.

A análise temporal de dados por meio de regressão linear se mostrou como um componente eficiente do modelo, demonstrando eficácia na detecção de tendências na variação dos parâmetros sanguíneos obtidos por meio dos exames realizados pelos pacientes. Essas tendências temporais demonstram um efeito potencial causado pelo uso de um de-

terminado medicamento ao longo do tempo. Os resultados preliminares mostram que a apresentação dos resultados obtidos com o uso do DST comprovaram eficácia no gerenciamento de RAM.

O IC proposto contribui para a redução da incerteza, pois permite uma categorização inferior dos registros com baixa credibilidade de serem RAM. A doença é o fator  $C \rightarrow B$ , no domínio em que o resultado de interesse é  $A \rightarrow B$ , drogas que possam causar a RAM.

Os resultados do benchmarking mostram que é possível melhorar as métricas de desempenho através da incorporação do índice de credibilidade e da inferência bayesiana. Os resultados mostram uma melhoria média das métricas usadas para medir o desempenho de 21% no modelo com a utilização do Índice de Credibilidade e de até 25% quando se adiciona o efeito da inferência bayesiana na aplicação do domínio RAM. Esse mesmo comportamento de melhoria evolutiva à medida que se incorporam os componentes do IC ocorreu na aplicação do domínio de doenças cardíacas.

Isso garante que algumas reações sejam causadas por sintomas normais da doença, são sintomas naturais que não devem ser confundidos com reações causadas pelo uso de medicamentos. O estudo das relações de dependência, por inferência causal e correlação, mostrou-se importante para aumentar a precisão do modelo.

A escolha da ferramenta SHAP no auxílio de avaliação do modelo detectou aspectos relevantes em relação à influência, não somente das variáveis componentes do IC, como também provocada pelos medicamentos utilizados no ambiente de avaliação para tomada de decisão de diagnóstico de RAM, ressaltando:

- 1) Cada paciente é um caso, porque tem sua doença, sua natureza imunológica, toma medicamentos específicos e comorbidades, levando a reações e sintomas específicos
- 2) É possível ter um SHAP por sintoma/RAM ou por medicamentos. Essa flexibilização melhora a explicação do modelo e auxilia na gestão de RAM, uma vez que os especialistas consideram uma interface amigável gerada por essa ferramenta.
- 3) A temporalidade contribui em cerca de 10% dos casos e, quando ocorre, é decisiva.

As métricas alcançadas de 92% de precisão, 96% de acurácia e 95% de recall são bem satisfatórias para um domínio com tanta incerteza. O fato de o modelo ter sido implantado em um domínio diferente da RAM, no caso para diagnóstico de doenças cardíacas,

atingindo 95% de cobertura do caso benchmark com um período relativamente curto de treinamento, leva-se a crer em um bom modelo para aplicação em tarefas de diagnósticos, em especial, onde predomina incerteza.

## 7.1 Contribuições

Ao longo dessa jornada, essa pesquisa fomentou algumas publicações:

- 2019 - Desenho de Pesquisa SBSC2019 - Publicação de desenho de pesquisa sobre detecção de RAM. Á época ainda pensando em utilizar MDS e análise de sentimentos, [Ribeiro and Garcia, 2019]
- 2020 - CSCWD artigo . publicação apresentado em abril de 2021 na conferência CSCWD 2021, esse artigo aborda o processo de preparação do dataset a partir de MSD [Ribeiro et al., 2021a]
- 2021 - FLAIRS artigo apresentado em maio de 2021 na conferência FLAIRS34 com o projeto inicial do índice de credibilidade para atuar no domínio de RAM, [Ribeiro et al., 2021b]
- 2022 - Journal Sensors artigo publicado em março de 2022, contemplando o projeto do IC sendo formado por meio de um processo de FIMS [Ribeiro et al., 2022]

## 7.2 Trabalhos futuros

O modelo aplicado ao domínio RAM o modelo PCA linear pode ser substituído pela versão Kernel PCA [Schölkopf et al., 1997]. A ideia básica do kernel PCA é usar uma função de kernel não linear ( $k$ ) em vez do produto escalar padrão, implicitamente executa o PCA em um espaço de alta dimensão que não está linearmente relacionado ao espaço de entrada. Esta mudança de espaço permite a obtenção de espaços lineares onde o tradicional PCA é beneficiado, aumentando a possibilidade de capturar uma variação maior nos primeiros componentes e consequentemente obtendo um melhor rank.

O modelo pode ser adequado para ser utilizado em um ambiente de tempo real de monitoramento, onde o diagnóstico gerado pelo DEP-DST pode ser integrado a um ambiente de tempo real, sendo as hipóteses geradas pelo modelo podendo servir de base para geração de alertas e envios de aviso.

No domínio de RAM, podem ser acrescentadas fontes de dados obtidas a partir de sensores de imagens, como ressonância magnética ou tomografia computadorizada. Seria necessário o uso de métodos de visão computacional que podem extrair coleções de imagens formando uma série temporal semelhante a feita com os exames laboratoriais. Também podem ser agregadas sob essa técnica imagens obtidas a partir de exames histopatológicos obtidos de materiais oriundos de biópsia.

No domínio de doenças do coração, é possível aprimorar a extração de dados para melhor modelar a inferência bayesiana, o que provavelmente servirá para melhorar as métricas obtidas em relação ao benchmark, o qual o DEP-DST cobriu 95%. As comorbidades - hipertensão e diabetes - talvez possam ser inferidas o que além de melhorar o motor de inferência causal vai permitir melhora do

Esse modelo foi implantado em um hospital oncológico e poderá ser aplicado em outras instituições de saúde. A técnica de Transfer Learning pode ser uma excelente oportunidade, principalmente para aplicar o modelo proposto em outras instituições de saúde interessadas em uma instância em seu ambiente. Dois estudos relevantes para a aplicação da aprendizagem por transferência são [Zheng et al., 2020] e [Hao and Chen, 2021]. Eles usam uma técnica de aprendizado que pode otimizar significativamente o tempo de treinamento e a necessidade de atividades de notificação e rotulagem de casos.

Essa pesquisa poderia ainda endereçar os seguintes tópicos:

- a cobertura do cenário e aplicação em relação a RAM no contexto do Instituto, assim como a aplicação da teoria de evidências ao longo de todo o cenário de aplicação no Instituto, discutindo e justificando os graus de confiabilidade.
- identificar o grau de aceitação de profissionais de saúde e eventualmente dos pacientes que possam utilizar os resultados dessa pesquisa. Há uma reunião chamada de "mesa redonda" onde pode-se pleitear uma fatia de tempo para discussão desse assunto.

### **7.3 Limitações**

O tema dependências em incerteza é muito amplo, não pôde ser totalmente coberto, bem como não seria possível esgotar a utilização de relações de causalidade nessa análise. Não encontramos na literatura o uso da DST para diagnosticar RAM, mas constatamos o ambiente incerto que ocorre na tomada de decisão clínica no geral, no diagnóstico de RAM e doenças cardíacas, como abordou-se nesse trabalho.

A obtenção de bases de dados sobre RAM não é simples, pois envolve privacidade e dados sensíveis. O processo de FIMS é intenso em volume e heterogeneidade de fontes e dados. Essa combinação de fatores torna necessária uma análise rigorosa sobre o domínio, em especial para utilização em novas instanciações.

Alguns exames importantes não puderam ser coletados por estarem em plataforma de difícil captura, ou protegidos por métodos criptografados. Uma limitação se refere a drogas utilizados em processo de radioterapia, como contrastes, que não foram contemplados no trabalho por terem suas doses e prescrições armazenadas em softwares embarcados em servidores de terceiros, inviabilizando tal captura. Essas medicações poderiam influenciar o ambiente que propicie RAM.

É necessário trabalhar um grande volume de dados, sendo fundamental o prontuário eletrônico e dados sensíveis de paciente, como medicações, exames e tratamento. Isso trouxe a necessidade de se realizar um processo de anonimização para atender exigências da legislação sobre a Lei Geral de Proteção de Dados e ao comitê de ética da instituição, que permitiram a coleta de dados com a abrangência de um ano. Como o modelo trabalha com análise temporal, provavelmente se o conjunto de dados trabalhasse com um período de tempo de dois anos, poderiam surgir oportunidades de melhoria no modelo assim como identificação de novas RAM.

A escolha de drogas demanda apoio de especialistas, pois foi revelado pela revisão de literatura, que um modelo bom para um tipo de medicamentos pode não ser bom para outro grupo de drogas. Na instanciação para RAM, foram utilizados trinta medicamentos de vários grupos, abrangendo não somente tratamento do câncer, mas também dos grupos relacionados às comorbidades dos pacientes. Provavelmente, se o conjunto de dados contivesse um número maior de medicamentos, seriam identificadas novas RAM e surgiriam oportunidades de melhoria do modelo.

Realizou-se uma abordagem linear, tanto na utilização do PCA para gerar o IC como na avaliação das séries temporais que realizam a análise de tendências para estabelecer dependência temporal. Provavelmente métodos não lineares podem evoluir ambos.

Embora a geração da BPA por meio da RF tenha obtidos boas métricas nesse trabalho, poderiam ser experimentados outros arranjos que permitiriam uma avaliação mais ampla do modelo.

Não foi possível realizar uma validação por meio de uma pesquisa qualitativa de aceitação do modelo e sobre o funcionamento da aplicação.

## Referências Bibliográficas

- [Aminikhanghahi and Cook, 2017] Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367.
- [Atluri et al., 2018] Atluri, G., Karpatne, A., and Kumar, V. (2018). Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, 51(4):1–41.
- [Azzouza et al., 2017] Azzouza, N., Akli-Astouati, K., Oussalah, A., and Bachir, S. A. (2017). A Real-time Twitter Sentiment Analysis Using an Unsupervised Method. In *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics, WIMS '17*, New York, NY, USA. ACM.
- [Baccianella et al., 2010] Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, pages 2200–2204.
- [Banda et al., 2016] Banda, J. M., Evans, L., Vanguri, R. S., Tatonetti, N. P., Ryan, P. B., and Shah, N. H. (2016). A curated and standardized adverse drug event resource to accelerate drug safety research. *Scientific data*, 3:160026.
- [Bauman, 2013] Bauman, Z. (2013). *Liquid times: Living in an age of uncertainty*. John Wiley & Sons.
- [Beddar-Wiesing and Bieshaar, 2020] Beddar-Wiesing, S. and Bieshaar, M. (2020). Multi-sensor data and knowledge fusion—a proposal for a terminology definition. *arXiv preprint arXiv:2001.04171*.
- [Blasch and Plano, 2005] Blasch, E. and Plano, S. (2005). Dfig level 5 (user refinement) issues supporting situational assessment reasoning. In *2005 7th International Conference on Information Fusion*, volume 1, pages xxxv–xlili.

- [Cai et al., 2015] Cai, M.-C., Xu, Q., Pan, Y.-J., Pan, W., Ji, N., Li, Y.-B., Jin, H.-J., Liu, K., and Ji, Z.-L. (2015). Adres: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic acids research*, 43(D1):D907–D913.
- [Cai et al., 2019] Cai, Q., Wang, H., Li, Z., and Liu, X. (2019). A survey on multimodal data-driven smart healthcare systems: Approaches and applications. *IEEE Access*, 7:133583–133599.
- [Calix et al., 2017] Calix, R. A., Gupta, R., Gupta, M., and Jiang, K. (2017). Deep grammar: Improving precision in the classification of personal health-experience tweets with deep learning. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1154–1159.
- [Casillas et al., 2016] Casillas, A., Pérez, A., Oronoz, M., Gojenola, K., and Santiso, S. (2016). Learning to extract adverse drug reaction events from electronic health records in spanish. *Expert Systems with Applications*, 61:235–245.
- [Castanedo, 2013] Castanedo, F. (2013). A review of data fusion techniques. *The Scientific World Journal*, 2013.
- [Chen et al., 2014] Chen, Y., Xia, X., Ge, S., and An, Y. (2014). An approach to conflict evidence combination based on two criteria optimization. *J. Comput. Inf. Syst*, 10(1):2727–2734.
- [Chen et al., 2018a] Chen, Y., Yuan, J., You, Q., and Luo, J. (2018a). Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM. In *Proceedings of the 26th ACM International Conference on Multimedia, MM '18*, pages 117–125, New York, NY, USA. ACM.
- [Chen et al., 2018b] Chen, Y., Yuan, J., You, Q., and Luo, J. (2018b). Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM. In *Proceedings of the 26th ACM International Conference on Multimedia, MM '18*, pages 117–125, New York, NY, USA. ACM. event-place: Seoul, Republic of Korea.
- [Comfort et al., 2018] Comfort, S., Perera, S., Hudson, Z., Dorrell, D., Meireis, S., Nagarajan, M., Ramakrishnan, C., and Fine, J. (2018). Sorting through the safety data haystack: using machine learning to identify individual case safety reports in social-digital media. *Drug safety*, 41(6):579–590.
- [Dandala et al., 2018] Dandala, B., Joopudi, V., and Devarakonda, M. (2018). IBM Research System at MADE 2018: Detecting adverse drug events from electronic health

- records. In *International Workshop on Medication and Adverse Drug Event Detection*, pages 39–47.
- [Deng, 2016] Deng, Y. (2016). Deng entropy. *Chaos, Solitons Fractals*, 91:549 – 553.
- [Denœux, 2019] Denœux, T. (2019). Logistic regression, neural networks and dempster–shafer theory: A new perspective. *Knowledge-Based Systems*, 176:54–67.
- [Denœux, 2020] Denœux, T. (2020). Distributed combination of belief functions. *Information Fusion*, 65:179–191.
- [Destercke and Dubois, 2011] Destercke, S. and Dubois, D. (2011). Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. *Information Sciences*, 181(18):3925–3945.
- [Dev et al., 2017a] Dev, S., Zhang, S., Voyles, J., and Rao, A. S. (2017a). Automated classification of adverse events in pharmacovigilance. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 905–909. IEEE.
- [Dev et al., 2017b] Dev, S., Zhang, S., Voyles, J., and Rao, A. S. (2017b). Automated classification of adverse events in pharmacovigilance. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 905–909. IEEE.
- [Dritsas et al., 2018] Dritsas, E., Livieris, I. E., Giotopoulos, K., and Theodorakopoulos, L. (2018). An Apache Spark Implementation for Graph-based Hashtag Sentiment Classification on Twitter. In *Proceedings of the 22Nd Pan-Hellenic Conference on Informatics, PCI '18*, pages 255–260, New York, NY, USA. ACM. event-place: Athens, Greece.
- [Durrant-Whyte and Henderson, 2016] Durrant-Whyte, H. and Henderson, T. C. (2016). Multisensor data fusion. In *Springer handbook of robotics*, pages 867–896. Springer.
- [Esteban et al., 2005] Esteban, J., Starr, A., Willetts, R., Hannah, P., and Bryanston-Cross, P. (2005). A review of data fusion models and architectures: towards engineering guidelines. *Neural Computing & Applications*, 14(4):273–281.
- [Ferson et al., 2015] Ferson, S., Kreinovich, V., Grinzburg, L., Myers, D., and Sentz, K. (2015). Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
- [Florea and Bossé, 2009] Florea, M. C. and Bossé, É. (2009). Crisis management using dempster shafer theory: Using dissimilarity measures to characterize sources’ reliability. *C3I for Crisis, Emergency and Consequence Management*.

- [Fu and Yang, 2014] Fu, C. and Yang, S. (2014). Conjunctive combination of belief functions from dependent sources using positive and negative weight functions. *Expert systems with applications*, 41(4):1964–1972.
- [Gite and Agrawal, 2016] Gite, S. and Agrawal, H. (2016). On context awareness for multisensor data fusion in iot. In *Proceedings of the Second International Conference on Computer and Communication Technologies*, pages 85–93. Springer.
- [Gopalakrishna Pillai, 2018] Gopalakrishna Pillai, R. (2018). Detection of Strength and Causal Agents of Stress and Relaxation for Tweets. In *Companion Proceedings of the The Web Conference 2018, WWW '18*, pages 837–841, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- [Gordon and Shortliffe, 1984] Gordon, J. and Shortliffe, E. H. (1984). The Dempster-Shafer theory of evidence. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, 3:832–838.
- [Guralnik et al., 2006] Guralnik, V., Mylaraswamy, D., and Voges, H. (2006). On handling dependent evidence and multiple faults in knowledge fusion for engine health management. In *2006 IEEE aerospace conference*, pages 9–pp. IEEE.
- [Han et al., 2017] Han, S., Tran, T., Rios, A., and Kavuluru, R. (2017). Team UKNLP: Detecting ADRs, Classifying Medication Intake Messages, and Normalizing ADR Mentions on Twitter. In *SMM4H@AMIA*.
- [Hao and Chen, 2021] Hao, C. and Chen, D. (2021). Software/hardware co-design for multi-modal multi-task learning in autonomous systems. In *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pages 1–5. IEEE.
- [Harpaz et al., 2017] Harpaz, R., DuMouchel, W., Schuemie, M., Bodenreider, O., Friedman, C., Horvitz, E., Ripple, A., Sorbello, A., White, R. W., Winnenburg, R., and Shah, N. H. (2017). Toward multimodal signal detection of adverse drug reactions. *Journal of Biomedical Informatics*, 76:41 – 49.
- [Hripcsak et al., 2015] Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., Suchard, M. A., Park, R. W., Wong, I. C. K., Rijnbeek, P. R., et al. (2015). Observational health data sciences and informatics (ohdsi): opportunities for observational researchers. *Studies in health technology and informatics*, 216:574.
- [Hsu et al., 2017] Hsu, D., Moh, M., and Moh, T.-S. (2017). Mining Frequency of Drug Side Effects over a Large Twitter Dataset Using Apache Spark. In *Proceedings of the*

2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, pages 915–924, New York, NY, USA. ACM. event-place: Sydney, Australia.

- [Hu and Liu, 2004] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- [Huang et al., 2017] Huang, J., Zhang, X., du, J., Duan, R., Yang, L., H Moore, J., Chen, Y., and Tao, C. (2017). Comparing Different Adverse Effects Among Multiple Drugs Using FAERS Data. *Studies in health technology and informatics*, 245:1268.
- [Huynh et al., 2016a] Huynh, T., He, Y., Willis, A., and Rüger, S. (2016a). Adverse Drug Reaction Classification With Deep Neural Networks. In *Proceedings of COLING 2016: Technical Papers*, pages 877–887. COLING.
- [Huynh et al., 2016b] Huynh, T., He, Y., Willis, A., and Rüger, S. (2016b). Adverse Drug Reaction Classification With Deep Neural Networks. In *Proceedings of COLING 2016: Technical Papers*, pages 877–887. COLING.
- [Imai et al., 2017] Imai, S., Yamada, T., Kasashi, K., Kobayashi, M., and Iseki, K. (2017). Usefulness of a decision tree model for the analysis of adverse drug reactions: Evaluation of a risk prediction model of vancomycin-associated nephrotoxicity constructed using a data mining procedure. *Journal of evaluation in clinical practice*, 23(6):1240–1246.
- [Jiang et al., 2017] Jiang, H., Wang, R., Gao, J., Gao, Z., and Gao, X. (2017). Evidence fusion-based framework for condition evaluation of complex electromechanical system in process industry. *Knowledge-Based Systems*, 124:176–187.
- [Karimi et al., 2015a] Karimi, S., Metke-Jimenez, A., Kemp, M., and Wang, C. (2015a). CadeC: A corpus of adverse drug event annotations. *Journal of Biomedical Informatics*, 55:73 – 81.
- [Karimi et al., 2015b] Karimi, S., Metke-Jimenez, A., Kemp, M., and Wang, C. (2015b). CadeC: A corpus of adverse drug event annotations. *Journal of Biomedical Informatics*, 55:73 – 81.
- [Karunamuni et al., 2021] Karunamuni, N., Imayama, I., and Goonetilleke, D. (2021). Pathways to well-being: Untangling the causal relationships among biopsychosocial variables. *Social Science & Medicine*, 272:112846.

- [Katz et al., 2019] Katz, O., Talmon, R., Lo, Y.-L., and Wu, H.-T. (2019). Alternating diffusion maps for multimodal data fusion. *Information Fusion*, 45:346–360.
- [Kaugerand et al., 2018] Kaugerand, J., Ehala, J., Motus, L., and Preden, J.-S. (2018). Time-selective data fusion for in-network processing in ad hoc wireless sensor networks. *International Journal of Distributed Sensor Networks*, 14:155014771881130.
- [Khakzad, 2019] Khakzad, N. (2019). System safety assessment under epistemic uncertainty: Using imprecise probabilities in bayesian network. *Safety science*, 116:149–160.
- [Khaleghi et al., 2013] Khaleghi, B., Khamis, A., Karray, F. O., and Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1):28–44.
- [Khan and Anwar, 2019] Khan, M. N. and Anwar, S. (2019). Time-domain data fusion using weighted evidence and dempster–shafer combination rule: Application in object classification. *Sensors*, 19(23):5187.
- [Kim and Rhew, 2017] Kim, H. H. and Rhew, K. Y. (2017). A Machine Learning Approach to Classification of Case Reports on Adverse Drug Reactions Using Text Mining of Expert Opinions. In *Advances in Computer Science and Ubiquitous Computing*, pages 1072–1077. Springer.
- [Kim et al., 2008] Kim, J. . D., Ohta, T., and Tsujii, J. (2008). Corpus annotation for mining biomedical events from literature. *BMC Bioinform*, 9.
- [Korkontzelos et al., 2015] Korkontzelos, I., Piliouras, D., Dowsey, A. W., and Ananiadou, S. (2015). Boosting drug named entity recognition using an aggregate classifier. *Artif Intell Med*, 65.
- [Kumar and Daumé, 2011] Kumar, A. and Daumé, H. (2011). A co-training approach for multi-view spectral clustering. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 393–400.
- [Latha and Jeeva, 2019] Latha, C. B. C. and Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, 16:100203.
- [Lee, 2019] Lee, G. S. (2019). An evidence retraction scheme on evidence dependency network. *International journal of advanced smart convergence*, 8(1):133–140.

- [Lin et al., 2015] Lin, W., Dai, H., Jonnagaddala, J., Chang, N., Jue, T. R., Iqbal, U., Shao, J. Y., Chiang, I., and Li, Y. (2015). Utilizing different word representation methods for twitter data in adverse drug reactions extraction. In *2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 260–265.
- [Liu et al., 2020] Liu, J., Li, T., Xie, P., Du, S., Teng, F., and Yang, X. (2020). Urban big data fusion based on deep learning: An overview. *Information Fusion*, 53:123–133.
- [Liu and Wang, 2018] Liu, J. and Wang, G. (2018). Pharmacovigilance from social media: An improved random subspace method for identifying adverse drug events. *International Journal of Medical Informatics*, 117:33 – 43.
- [Liu et al., 2018a] Liu, J., Zhao, S., and Wang, G. (2018a). Ssel-ade: A semi-supervised ensemble learning framework for extracting adverse drug events from social media. *Artificial Intelligence in Medicine*, 84:34 – 49.
- [Liu et al., 2018b] Liu, J., Zhao, S., and Wang, G. (2018b). SSEL-ADE: A semi-supervised ensemble learning framework for extracting adverse drug events from social media. *Artificial Intelligence in Medicine*, 84:34 – 49.
- [Liu et al., 2016] Liu, J., Zhao, S., and Zhang, X. (2016). An ensemble method for extracting adverse drug events from social media. *Artificial Intelligence in Medicine*, 70:62 – 76.
- [Liu and Chen, 2015] Liu, X. and Chen, H. (2015). Identifying adverse drug events from patient social media: a case study for diabetes. *IEEE intelligent systems*, 30(3):44–51.
- [Lundberg and Lee, 2017] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777.
- [Mahata et al., 2018] Mahata, D., Friedrichs, J., Shah, R., and Jiang, J. (2018). Detecting Personal Intake of Medicine from Twitter. *IEEE Intelligent Systems*, 33(04):87–95.
- [Mao et al., 2020] Mao, S., Han, Y., Deng, Y., and Pelusi, D. (2020). A hybrid dematel-fractal method of handling dependent evidences. *Engineering Applications of Artificial Intelligence*, 91:103543.
- [Martí et al., 2015] Martí, L., Sanchez-Pi, N., Molina, J. M., and Garcia, A. C. B. (2015). Anomaly detection based on sensor data in petroleum industry applications. *Sensors*, 15(2):2774–2797.

- [Masino et al., 2018] Masino, A. J., Forsyth, D., and Fiks, A. G. (2018). Detecting Adverse Drug Reactions on Twitter with Convolutional Neural Networks and Word Embedding Features. *Journal of Healthcare Informatics Research*, 2(1-2):25–43.
- [Meng et al., 2020] Meng, T., Jing, X., Yan, Z., and Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, 57:115–129.
- [Mishra et al., 2015] Mishra, A., Malviya, A., and Aggarwal, S. (2015). Towards Automatic Pharmacovigilance: Analysing Patient Reviews and Sentiment on Oncological Drugs. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1402–1409.
- [Mishra and Diesner, 2018] Mishra, S. and Diesner, J. (2018). Detecting the Correlation Between Sentiment and User-level As Well As Text-Level Meta-data from Benchmark Corpora. In *Proceedings of the 29th on Hypertext and Social Media, HT '18*, pages 2–10, New York, NY, USA. ACM.
- [Monney and Chan, 2007] Monney, P.-A. and Chan, M. (2007). Modelling dependence in dempster-shafer theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(01):93–114.
- [Monney et al., 2011] Monney, P.-A., Chan, M., and Romberg, P. (2011). A belief function classifier based on information provided by noisy and dependent features. *International journal of approximate reasoning*, 52(3):335–352.
- [Nielsen, 2011] Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.
- [Nikfarjam et al., 2015] Nikfarjam, A., Sarker, A., O’Connor, K., Ginn, R., and Gonzalez, G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association : JAMIA*, 22(3):671–681.
- [Pappa and Stergioulas, 2019] Pappa, D. and Stergioulas, L. K. (2019). Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions. *International Journal of Data Science and Analytics*, 8(2):113–135.
- [Pei et al., 2017] Pei, F.-Q., Li, D.-B., Tong, Y.-F., and He, F. (2017). Process service quality evaluation based on dempster-shafer theory and support vector machine. *Plos one*, 12(12):e0189189.

- [Peng et al., 2016a] Peng, Y., Moh, M., and Moh, T.-S. (2016a). Efficient Adverse Drug Event Extraction Using Twitter Sentiment Analysis. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '16, pages 1011–1018, Piscataway, NJ, USA. IEEE Press.
- [Peng et al., 2016b] Peng, Y., Moh, M., and Moh, T.-S. (2016b). Efficient Adverse Drug Event Extraction Using Twitter Sentiment Analysis. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '16, pages 1011–1018, Piscataway, NJ, USA. IEEE Press.
- [Peng et al., 2016c] Peng, Y., Moh, M., and Moh, T.-S. (2016c). Efficient adverse drug event extraction using twitter sentiment analysis. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1011–1018. IEEE.
- [Pippa et al., 2018] Pippa, E., Zacharaki, E. I., Özdemir, A. T., Barshan, B., and Megalooikonomou, V. (2018). Global vs local classification models for multi-sensor data fusion. In *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, pages 1–5.
- [Rajapaksha and Weerasinghe, 2015] Rajapaksha, P. and Weerasinghe, R. (2015). Identifying Adverse Drug Reactions by analyzing Twitter messages. pages 37–42.
- [Ribeiro and Garcia, 2019] Ribeiro, L. and Garcia, A. C. (2019). O uso de análise de sentimentos sobre mídia social digital para extração de reações adversas a medicamentos devido à interação medicamentosa durante um intervalo de tempo. In *Anais Estendidos do XV Simpósio Brasileiro de Sistemas Colaborativos*, pages 7–12, Porto Alegre, RS, Brasil. SBC.
- [Ribeiro et al., 2021a] Ribeiro, L. A., Cinalli, D., and Garcia, A. C. B. (2021a). Discovering adverse drug reactions from twitter: A sentiment analysis perspective. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 1172–1177. IEEE.
- [Ribeiro et al., 2021b] Ribeiro, L. A. P. A., Garcia, A. C. B., and dos Santos, P. S. M. (2021b). Temporal and causal relations on evidence theory: an application on adverse drug reactions. In *The International FLAIRS Conference Proceedings*, volume 34.
- [Ribeiro et al., 2022] Ribeiro, L. A. P. A., Garcia, A. C. B., and dos Santos, P. S. M. (2022). Dependency factors in evidence theory: An analysis in an information fusion scenario applied in adverse drug reactions. *Sensors*, 22(6).

- [Russel and Norvig, 2003] Russel, S. and Norvig, P. (2003). Artificial intelligence: A modern approach, 2003. *EUA: Prentice Hall*.
- [Samonte et al., 2018] Samonte, M. J. C., Dollete, C. J. T., Capanas, P. M. M., Flores, M. L. C., and Soriano, C. B. (2018). Sentence-Level Sarcasm Detection in English and Filipino Tweets. In *Proceedings of the 4th International Conference on Industrial and Business Engineering, ICIBE' 18*, pages 181–186, New York, NY, USA. ACM.
- [Santiso et al., 2018] Santiso, S., Perez, A., and Casillas, A. (2018). Exploring Joint AB-LSTM with embedded lemmas for Adverse Drug Reaction discovery. *IEEE Journal of Biomedical and Health Informatics*, pages 1–1.
- [Sarker et al., 2015] Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T., and Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance: A review. *Journal of Biomedical Informatics*, 54:202 – 212.
- [Schölkopf et al., 1997] Schölkopf, B., Smola, A., and Müller, K.-R. (1997). Kernel principal component analysis. In *International conference on artificial neural networks*, pages 583–588. Springer.
- [Schubert, 2011] Schubert, J. (2011). Conflict management in dempster–shafer theory using the degree of falsity. *International Journal of Approximate Reasoning*, 52(3):449–460.
- [Severyn and Moschitti, 2015a] Severyn, A. and Moschitti, A. (2015a). Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, pages 959–962, New York, NY, USA. ACM.
- [Severyn and Moschitti, 2015b] Severyn, A. and Moschitti, A. (2015b). Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, pages 959–962, New York, NY, USA. ACM.
- [Shafer, 1976] Shafer, G. (1976). *A mathematical theory of evidence*, volume 42. Princeton university press.
- [Shang et al., 2014] Shang, J., Zheng, Y., Tong, W., Chang, E., and Yu, Y. (2014). Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1027–1036.

- [Shi et al., 2018] Shi, J.-Y., Shang, X.-Q., Gao, K., Zhang, S.-W., and Yiu, S.-M. (2018). An integrated local classification model of predicting drug-drug interactions via dempster-shafer theory of evidence. *Scientific Reports*, 8(1):11829.
- [Sintsova and Pu, 2016a] Sintsova, V. and Pu, P. (2016a). Dystemo: Distant Supervision Method for Multi-Category Emotion Recognition in Tweets. *ACM Trans. Intell. Syst. Technol.*, 8(1):13:1–13:22.
- [Sintsova and Pu, 2016b] Sintsova, V. and Pu, P. (2016b). Dystemo: Distant Supervision Method for Multi-Category Emotion Recognition in Tweets. *ACM Trans. Intell. Syst. Technol.*, 8(1):13:1–13:22.
- [Smets, 1988] Smets, P. (1988). Non-standard logics for automated reasoning.
- [Song et al., 2017] Song, Q., Li, B., and Xu, Y. (2017). Research on Adverse Drug Reaction Recognitions Based on Conditional Random Field. pages 97–101.
- [Song et al., 2020] Song, Y., Mi, J., Cheng, Y., Bai, L., and Chen, K. (2020). A dependency bounds analysis method for reliability assessment of complex system with hybrid uncertainty. *Reliability Engineering & System Safety*, 204:107119.
- [Steinhauer and Karlsson, 2019] Steinhauer, H. and Karlsson, A. (2019). *Information Fusion*, pages 61–78.
- [Su et al., 2018] Su, X., Li, L., Shi, F., and Qian, H. (2018). Research on the fusion of dependent evidence based on mutual information. *IEEE Access*, 6:71839–71845.
- [Su et al., 2014] Su, X., Xu, P., Mahadevan, S., and Deng, Y. (2014). On consideration of dependence and reliability of evidence in dempster-shafer theory. *JOURNAL OF INFORMATION & COMPUTATIONAL SCIENCE*, 11(14):4901–4910.
- [Tabassian et al., 2012] Tabassian, M., Ghaderi, R., and Ebrahimpour, R. (2012). Combining complementary information sources in the dempster-shafer framework for solving classification problems with imperfect labels. *Knowledge-Based Systems*, 27:92–102.
- [Tang et al., 2018] Tang, B., Hu, J., Wang, X., and Chen, Q. (2018). Recognizing continuous and discontinuous adverse drug reaction mentions from social media using LSTM-CRF. *Wireless Communications and Mobile Computing*, 2018:1–8.
- [Thompson et al., 2016] Thompson, P., Batista-Navarro, R. T., Kontonatsios, G., Carter, J., Toon, E., McNaught, J., Timmermann, C., Worboys, M., and Ananiadou, S. (2016). Text mining the history of medicine. *PloS one*, 11(1):e0144717.

- [Thompson et al., 2018a] Thompson, P., Daikou, S., Ueno, K., Batista-Navarro, R., Tsujii, J., and Ananiadou, S. (2018a). Annotation and detection of drug effects in text for pharmacovigilance. *Journal of Cheminformatics*, 10(1):37.
- [Thompson et al., 2018b] Thompson, P., Daikou, S., Ueno, K., Batista-Navarro, R., Tsujii, J., and Ananiadou, S. (2018b). Annotation and detection of drug effects in text for pharmacovigilance. *Journal of Cheminformatics*, 10(1):37.
- [Tian et al., 2018] Tian, J., Dezert, J., and Qin, R. (2018). Time-series 3d building change detection based on belief functions. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 1–5. IEEE.
- [Tricco et al., 2018] Tricco, A. C., Zarin, W., Lillie, E., Jeeblee, S., Warren, R., Khan, P. A., Robson, R., Hirst, G., Straus, S. E., and others (2018). Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. *BMC medical informatics and decision making*, 18(1):38.
- [Verbert et al., 2017] Verbert, K., Babuška, R., and De Schutter, B. (2017). Bayesian and dempster–shafer reasoning for knowledge-based fault diagnosis—a comparative study. *Engineering Applications of Artificial Intelligence*, 60:136–150.
- [Wang et al., 2017] Wang, F., Wang, Z., and Su, J. (2017). Study on optimization of landslide treatment schemes based on entropy weight grey incidence and ds evidence theory. *The Open Civil Engineering Journal*, 11(1).
- [Wang et al., 2019] Wang, Y., Liu, F., and Zhu, A. (2019). Bearing fault diagnosis based on a hybrid classifier ensemble approach and the improved dempster-shafer theory. *Sensors*, 19(9):2097.
- [Wang et al., 2020] Wang, Z., Wang, R., Gao, J., Gao, Z., and Liang, Y. (2020). Fault recognition using an ensemble classifier based on dempster–shafer theory. *Pattern Recognition*, 99:107079.
- [Wilson et al., 2005] Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 347–354.
- [Wong et al., 2018] Wong, A., Plasek, J. M., Montecalvo, S. P., and Zhou, L. (2018). Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 38(8):822–841.

- [Wu et al., 2015] Wu, L., Moh, T., and Khuri, N. (2015). Twitter opinion mining for adverse drug reactions. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 1570–1574.
- [Wunnava et al., 2018] Wunnava, S., Qin, X., Kakar, T., Rundensteiner, E. A., and Kong, X. (2018). Bidirectional LSTM-CRF for adverse drug event tagging in electronic health records. In *International Workshop on Medication and Adverse Drug Event Detection*, pages 48–56.
- [Wunnava et al., 2017] Wunnava, S., Qin, X., Kakar, T., Socrates, V., Wallace, A., and Rundensteiner, E. (2017). Towards Transforming FDA Adverse Event Narratives into Actionable Structured Data for Improved Pharmacovigilance. In *Proceedings of the Symposium on Applied Computing, SAC '17*, pages 777–782, New York, NY, USA. ACM.
- [Xiao and Qin, 2018] Xiao, F. and Qin, B. (2018). A weighted combination method for conflicting evidence in multi-sensor data fusion. *Sensors*, 18(5):1487.
- [Yager, 2009] Yager, R. R. (2009). On the fusion of non-independent belief structures. *International journal of general systems*, 38(5):505–531.
- [Yang et al., 2018] Yang, X., Bian, J., and Wu, Y. (2018). Detecting medications and adverse drug events in clinical notes using recurrent neural networks. In *International workshop on medication and adverse drug event detection*, pages 1–6.
- [Yang and Han, 2016] Yang, Y. and Han, D. (2016). A new distance-based total uncertainty measure in the theory of belief functions. *Knowledge-Based Systems*, 94:114–123.
- [Yang et al., 2013] Yang, Y., Han, D., and Han, C. (2013). Discounted combination of unreliable evidence using degree of disagreement. *International Journal of Approximate Reasoning*, 54(8):1197–1216.
- [Yarnell et al., 2021] Yarnell, C. J., Abrams, D., Baldwin, M. R., Brodie, D., Fan, E., Ferguson, N. D., Hua, M., Madahar, P., McAuley, D. F., Munshi, L., Perkins, G. D., Rubinfeld, G., Slutsky, A. S., Wunsch, H., Fowler, R. A., Tomlinson, G., Beitler, J. R., and Goligher, E. C. (2021). Clinical trials in critical care: can a bayesian approach enhance clinical and scientific decision making? *The Lancet Respiratory Medicine*, 9(2):207–216.

- [Yu et al., 2017] Yu, T., Wang, X., and Shami, A. (2017). Recursive principal component analysis-based data outlier detection and sensor data aggregation in iot systems. *IEEE Internet of Things Journal*, 4(6):2207–2216.
- [Zadeh, 1984] Zadeh, L. (1984). A mathematical theory of evidence (book review). *AI magazine*, 5:81–83.
- [Zhang et al., 2017] Zhang, H., Ha, L., Li, Q., and Beer, M. (2017). Imprecise probability analysis of steel structures subject to atmospheric corrosion. *Structural Safety*, 67:62–69.
- [Zhang et al., 2018a] Zhang, H., Zhang, Q., Liu, J., and Guo, H. (2018a). Fault detection and repairing for intelligent connected vehicles based on dynamic bayesian network model. *IEEE Internet of Things Journal*, 5(4):2431–2440.
- [Zhang and Shields, 2020] Zhang, J. and Shields, M. (2020). On the quantification and efficient propagation of imprecise probabilities with copula dependence. *International Journal of Approximate Reasoning*, 122:24–46.
- [Zhang and Geng, 2019] Zhang, M. and Geng, G. (2019). Adverse drug event detection using a weakly supervised convolutional neural network and recurrent neural network model. *Information*, 10(9):276.
- [Zhang et al., 2019a] Zhang, M., Zhang, M., Ge, C., Liu, Q., Wang, J., Wei, J., and Zhu, K. Q. (2019a). Automatic discovery of adverse reactions through Chinese social media. *Data Mining and Knowledge Discovery*.
- [Zhang et al., 2019b] Zhang, Y., Jiang, W., and Deng, X. (2019b). Fault diagnosis method based on time domain weighted data aggregation and information fusion. *International Journal of Distributed Sensor Networks*, 15(9):1550147719875629.
- [Zhang et al., 2018b] Zhang, Y., Wang, X., Shen, L., Hou, Z., Guo, Z., and Li, J. (2018b). Identifying Adverse Drug Reactions of Hypolipidemic Drugs from Chinese Adverse Event Reports. In *2018 IEEE International Conference on Healthcare Informatics Workshop (ICHI-W)*, pages 72–73.
- [Zhao et al., 2016] Zhao, Y., Jia, R., and Shi, P. (2016). A novel combination method for conflicting evidence based on inconsistent measurements. *Information Sciences*, 367:125–142.
- [Zheng et al., 2020] Zheng, J., Lu, C., Hao, C., Chen, D., and Guo, D. (2020). Improving the generalization ability of deep neural networks for cross-domain visual recognition. *IEEE Transactions on Cognitive and Developmental Systems*.

- [Zheng, 2015] Zheng, Y. (2015). Methodologies for cross-domain data fusion: An overview. *IEEE transactions on big data*, 1(1):16–34.
- [Zheng et al., 2015] Zheng, Y., Zhang, H., and Yu, Y. (2015). Detecting collective anomalies from multiple spatio-temporal datasets across different domains. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, pages 1–10.
- [Zhu et al., 2018] Zhu, L., Guo, F., Polak, J. W., and Krishnan, R. (2018). Urban link travel time estimation using traffic states-based data fusion. *IET Intelligent Transport Systems*, 12(7):651–663.
- [Zolnoori et al., 2019a] Zolnoori, M., Fung, K. W., Patrick, T. B., Fontelo, P., Kharrazi, H., Faiola, A., Wu, Y. S. S., Eldredge, C. E., Luo, J., Conway, M., Zhu, J., Park, S. K., Xu, K., Moayyed, H., and Goudarzvand, S. (2019a). A systematic approach for developing a corpus of patient reported adverse drug events: A case study for SSRI and SNRI medications. *Journal of Biomedical Informatics*, 90:103091.
- [Zolnoori et al., 2019b] Zolnoori, M., Fung, K. W., Patrick, T. B., Fontelo, P., Kharrazi, H., Faiola, A., Wu, Y. S. S., Eldredge, C. E., Luo, J., Conway, M., Zhu, J., Park, S. K., Xu, K., Moayyed, H., and Goudarzvand, S. (2019b). A systematic approach for developing a corpus of patient reported adverse drug events: A case study for ssri and snri medications. *Journal of Biomedical Informatics*, 90:103091.

[authordate,bibencoding=auto,strict,backend=biber,natbib]biblatex-chicago