

PGCOMP - Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal da Bahia (UFBA)  
Av. Milton Santos, s/n - Ondina  
Salvador, BA, Brasil, 40170-110

<https://pgcomp.ufba.br>  
[pgcomp@ufba.br](mailto:pgcomp@ufba.br)

A doença falciforme é uma condição genética que afeta a hemoglobina, resultando em complicações sistêmicas que influenciam a qualidade e sobrevivência dos pacientes. A cromatografia líquida de alta performance (HPLC) é apontada como a melhor técnica de escolha para pesquisa de hemoglobinas variantes, já a eletroforese de hemoglobinas constitui a técnica padrão ouro para diagnóstico da doença, sendo de difícil acesso em localidades remotas do Brasil e em vários países em desenvolvimento, tais como do continente africano. A heterogeneidade clínica e a estratificação da gravidade da doença têm sido desafiadoras na prática médica, de modo que a predição de desfechos envolve análises de uma série de marcadores clínicos e laboratoriais.

Vários estudos discutiram a aplicação de modelos de inteligência artificial no diagnóstico da doença falciforme e seu monitoramento, tendo os autores se concentrado na classificação tradicional, onde um único desfecho é analisado. Essa condição limita alguns problemas complexos de saúde, onde existe mais de um desfecho por paciente. A proposta deste estudo foi utilizar de algoritmos multirrótulo para observar as relações entre os desfechos da doença falciforme e avaliar a hipótese de que o uso de modelos de inteligência artificial pode permitir uma melhor estratificação de pacientes com a doença, considerando os múltiplos desfechos clínicos.

Foi utilizada uma base de dados com 609 pacientes e 63 variáveis. A base foi obtida de estudos experimentais de coorte transversal gerados pelo Laboratório de Pesquisa em Anemias da Faculdade de Farmácia (UFBA) e do Laboratório de Investigação em Genética e Hematologia Translacional (LIGHT - FIOCRUZ-BA). Foram utilizados os algoritmos multirrótulo Binary Relevance (BR), Classifier Chains (CC), Label Powerset (LP) e Random k-labelsets (RAKEL), para predição dos desfechos de vaso-oclusão, crises de dor, hospitalização e infecções. Esses algoritmos são utilizados em problemas de transformação multirrótulo, que transformam o problema multirrótulo em uma série de problemas de classificação binária independentes, multiclasse e ensemble.

Os resultados apontaram que quando a ordem dos desfechos é relevante, o modelo RAKELo com Random Forest se destaca, apresentando um Label Ranking Loss de 0,52. Por outro lado, quando a ordenação não é um fator clínico importante, os modelos baseados em SVM e Random Forest, em todas as abordagens (BR, CC, LP, RAKELo), mostraram uma combinação ideal de baixo Hamming Loss (0,20) e alto Micro-averaged F1-score (0,89).

Palavras-chave: Doença Falciforme, Classificação multirrótulo, Binary Relevance, Classifier Chains, Label Powerset, Random k-labelsets

# Análise de Classificação Multirrótulo em Desfechos da Doença Falciforme

Tiago Fernandes Machado

Tese de Doutorado

Universidade Federal da Bahia

Programa de Pós-Graduação em  
Ciência da Computação

Dezembro | 2024

DSC | 54 | 2024

Análise de Classificação Multirrótulo em Desfechos da Doença Falciforme

Tiago Fernandes Machado

UFBA







Universidade Federal da Bahia  
Instituto de Computação

Programa de Pós-Graduação em Ciência da Computação

**ANÁLISE DE CLASSIFICAÇÃO  
MULTIRRÓTULO EM DESFECHOS DA  
DOENÇA FALCIFORME**

Tiago Fernandes Machado

TESE DE DOUTORADO

Salvador  
19 de dezembro de 2024

TIAGO FERNANDES MACHADO

**ANÁLISE DE CLASSIFICAÇÃO MULTIRRÓTULO EM  
DESFECHOS DA DOENÇA FALCIFORME**

Esta Tese de Doutorado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

Orientador: Prof. Dr. Marcos Ennes Barreto  
Co-orientadora: Profa. Dra. Cynara Gomes Barbosa

Salvador  
19 de dezembro de 2024

Ficha catalográfica elaborada pela Biblioteca Universitária de Ciências e Tecnologias  
Prof. Omar Catunda, SIBI – UFBA.

M149 Machado, Tiago Fernandes.  
Análise de Classificação Multirrótulo em Desfechos da Doença Falciforme  
/ Tiago Fernandes Machado – Salvador, 2024.  
69p.: il.

Orientador: Prof. Dr. Marcos Ennes Barreto.

Co-orientadora: Profa. Dra. Cynara Gomes Barbosa.

Tese (Doutorado) – Universidade Federal da Bahia, Instituto de Computação, 2024.

1. Computação. 2. Doença Falciforme. 3. Classificação multirrótulo.  
I. Barreto, Marcos Ennes. II. Barbosa, Cynara Gomes. III. Universidade  
Federal da Bahia. Instituto de Computação. IV. Título.

CDU – 616.15

**Termo de Aprovação**  
**TIAGO FERNANDES MACHADO**  
**ANÁLISE DE CLASSIFICAÇÃO MULTIRRÓTULO EM**  
**DESFECHOS DA DOENÇA FALCIFORME**

Esta tese foi julgada adequada à obtenção do título de Doutor em Ciência da Computação e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação da UFBA.

Salvador, 19 de dezembro de 2014

*Marcos Ennes Barreto*

box SIGN 42L8J2YW-1R9263KX

Prof. Dr. Marcos Ennes Barreto  
(Orientador - UFBA)

*Cynara Gomes Barbosa*

box SIGN 467V6J68-1R9263KX

Profa. Dra. Cynara Gomes Barbosa  
(Co-orientadora - UFBA)

*Gecynalda Gomes*

box SIGN 4P72VJ3P-1R9263KX

Profa. Dra. Gecynalda Soares da Silva Gomes (UFBA)

*Robespierre Dantas da Rocha Pita*

box SIGN 4Q2XQ8J8-1R9263KX

Prof. Dr. Robespierre Dantas da Rocha Pita (UFBA)

*Argenton Ramos*

box SIGN 4LRV5W99-1R9263KX

Prof. Dr. Argenton Ramos (UNIVASF)

*Elisângela Vitória Adorno*

box SIGN 4VW6ZYZZ-1R9263KX

Profa. Dra. Elisângela Vitória Adorno (UFBA)

## RESUMO

A doença falciforme é uma condição genética que afeta a hemoglobina, resultando em complicações sistêmicas que influenciam a qualidade e sobrevida dos pacientes. A cromatografia líquida de alta performance (HPLC) é apontada como a melhor técnica de escolha para pesquisa de hemoglobinas variantes, já a eletroforese de hemoglobinas constitui a técnica padrão ouro para diagnóstico da doença, sendo de difícil acesso em localidades remotas do Brasil e em vários países em desenvolvimento, tais como do continente africano. A heterogeneidade clínica e a estratificação da gravidade da doença têm sido desafiadoras na prática médica, de modo que a predição de desfechos envolve análises de uma série de marcadores clínicos e laboratoriais.

Vários estudos discutiram a aplicação de modelos de inteligência artificial no diagnóstico da doença falciforme e seu monitoramento, tendo os autores se concentrado na classificação tradicional, onde um único desfecho é analisado. Essa condição limita alguns problemas complexos de saúde, onde existe mais de um desfecho por paciente. A proposta deste estudo foi utilizar de algoritmos multirrótulo para observar as relações entre os desfechos da doença falciforme e avaliar a hipótese de que o uso de modelos de inteligência artificial pode permitir uma melhor estratificação de pacientes com a doença, considerando os múltiplos desfechos clínicos.

Foi utilizada uma base de dados com 609 pacientes e 63 variáveis. A base foi obtida de estudos experimentais de coorte transversal gerados pelo Laboratório de Pesquisa em Anemias da Faculdade de Farmácia (UFBA) e do Laboratório de Investigação em Genética e Hematologia Translacional (LIGHT - FIOCRUZ-BA). Foram utilizados os algoritmos multirrótulo Binary Relevance (BR), Classifier Chains (CC), Label Powerset (LP) e Random k-labelsets (RAKEL), para predição dos desfechos de vaso-oclusão, crises de dor, hospitalização e infecções. Esses algoritmos são utilizados em problemas de transformação multirrótulo, que transformam o problema multirrótulo em uma série de problemas de classificação binária independentes, multiclasse e ensemble.

Os resultados apontaram que quando a ordem dos desfechos é relevante, o modelo RAKELo com Random Forest se destaca, apresentando um Label Ranking Loss de 0,52. Por outro lado, quando a ordenação não é um fator clínico importante, os modelos baseados em SVM e Random Forest, em todas as abordagens (BR, CC, LP, RAKELo), mostraram uma combinação ideal de baixo Hamming Loss (0,20) e alto Micro-averaged F1-score (0,89).

**Palavras-chave:** Doença Falciforme; Classificação multirrótulo; Binary Relevance; Classifier Chains; Label Powerset; Random k-labelsets

## ABSTRACT

Sickle cell disease is a genetic condition that affects hemoglobin, resulting in systemic consequences that influence patients' quality of life and survival. Hemoglobin electrophoresis is gold standard technique for diagnosing the disease and is difficult to access in remote areas of Brazil and several developing countries, such as the African continent. Clinical heterogeneity and disease severity stratification have been tested in medical practice, so predicting results involves analyzing a series of clinical and laboratory markers.

Several studies have discussed the application of artificial intelligence models in diagnosing sickle cell disease and its monitoring, with the authors focusing on the traditional classification, where a single stage is developed. This condition limits some health problems where there is more than one stage per patient. The purpose of this study was to use multi-label algorithms to observe the relationships between the outcomes of sickle cell disease and to evaluate the hypothesis that using artificial intelligence models can allow for better stratification of patients with the disease, considering the multiple clinical outcomes.

A database with 609 patients and 63 variables was used. The database was obtained from experimental cross-sectional cohort studies generated by the Anemia Research Laboratory of the School of Pharmacy (UFBA) and the Translational Genetics and Hematology Research Laboratory (LIGHT - FIOCRUZ-BA). Multi-label algorithms Binary Relevance (BR), Classifier Chains (CC), Label Powerset (LP), and Random k-label sets (RAKEL) were used to predict the development of vaso-occlusion, pain crises, hospitalization, and infections. These algorithms are used in multi-label transformation problems, which transform the multi-label problem into a series of independent, multi-class, and ensemble binary classification problems.

The results showed that when the order of the stages is relevant, the RAKELo model with Random Forest stands out, presenting a Label Ranking Loss of 0.52. On the other hand, when ordering is not an important clinical factor, the SVM- and Random Forest-based models, in all approaches (BR, CC, LP, RAKELo), presented an optimal combination of low Hamming Loss (0.20) and high microaverage F1-score (0.89).

**Keywords:** Sickle Cell Disease; Multi-label Classification; Binary Relevance; Classifier Chains; Label Powerset; Random k-labelsets



# SUMÁRIO

<b>Lista de Figuras</b>	vii
<b>Lista de Tabelas</b>	viii
<b>Lista de Siglas</b>	ix
<b>Capítulo 1—Introdução</b>	1
1.1 Contexto e motivação . . . . .	3
1.2 Objetivos e questões de pesquisa . . . . .	4
1.3 Organização da tese . . . . .	5
<b>Capítulo 2—Background</b>	8
2.1 Doença falciforme . . . . .	8
2.1.1 Epidemiologia . . . . .	8
2.1.2 Aspectos clínicos . . . . .	9
2.1.3 Tratamento . . . . .	11
2.2 Algoritmos multirrótulo . . . . .	12
2.2.1 Binary Relevance . . . . .	14
2.2.2 Classifier Chains . . . . .	15
2.2.3 Label Powerset . . . . .	16
2.2.4 Random k-labelsets (RAKEL) . . . . .	17

SUMÁRIO	vii
2.3 Interface entre computação e saúde . . . . .	18
<b>Capítulo 3—Trabalhos relacionados</b>	<b>21</b>
3.1 Abordagens no diagnóstico . . . . .	21
3.2 Abordagens no monitoramento . . . . .	26
<b>Capítulo 4—Material e métodos</b>	<b>35</b>
4.1 Coleta e descrição dos dados . . . . .	35
4.2 Definição e categorização das variáveis . . . . .	35
4.3 Treinamento do modelo . . . . .	38
<b>Capítulo 5—Resultados</b>	<b>43</b>
<b>Capítulo 6—Discussão</b>	<b>54</b>
<b>Capítulo 7—Conclusões</b>	<b>58</b>
<b>Referências Bibliográficas</b>	<b>60</b>
<b>Apêndice A—Resultados experimentais</b>	<b>65</b>

## LISTA DE FIGURAS

1.1	Visão geral da estrutura da tese . . . . .	6
2.1	Opções para resolver problemas de classificação multirrótulo. . . . .	14
2.2	Atendimento ao paciente sem uso de tecnologia. . . . .	19
2.3	Atendimento ao paciente com uso de tecnologia. . . . .	20
4.1	Pré-processamento dos dados. . . . .	39
4.2	Features selecionadas na função Recursive Feature Elimination e sua importância para o modelo. . . . .	40
4.3	Frequência dos rótulos do conjunto de dados. . . . .	41
5.1	Matriz de confusão do rótulo Vaso-oclusão. . . . .	45
5.2	Matriz de confusão do rótulo Crises de Dor. . . . .	46
5.3	Matriz de confusão do rótulo Hospitalização. . . . .	47
5.4	Matriz de confusão do rótulo Infecções. . . . .	48
5.5	Correlações entre os desfechos. . . . .	53

## LISTA DE TABELAS

3.1	Principais algoritmos utilizados no diagnóstico da doença falciforme, tendo como entrada imagens de esfregaço de sangue. . . . .	22
3.2	Principais algoritmos utilizados no monitoramento da doença falciforme. . . . .	28
4.1	Relatório de estatística descritiva da base de dados . . . . .	36
5.1	Relatório de classificação do Random Forest para Vaso-oclusão. . . . .	44
5.2	Relatório de classificação do Random Forest para Crises de Dor. . . . .	46
5.3	Relatório de classificação do Random Forest para Hospitalização. . . . .	48
5.4	Relatório de classificação do Random Forest para Infecções. . . . .	49
5.5	Ocorrência dos Atributos selecionados para os desfechos de vaso-oclusão, crises de dor, hospitalização e infecções. . . . .	50
5.6	Desempenho dos algoritmos com diferentes métricas. . . . .	51

## LISTA DE SIGLAS

<b>DF</b>	Doença Falciforme . . . . .	1
<b>HbS</b>	Hemoglobina S . . . . .	1
<b>HM</b>	Hemácias (milhões/uL) . . . . .	44
<b>HT</b>	Hematócrito (%) . . . . .	44
<b>VCM</b>	Volume Corpuscular Médio (fL) . . . . .	44
<b>HCM</b>	Hemoglobina Corpuscular Média (pg) . . . . .	47
<b>CHCM</b>	Concentração de Hemoglobina Corpuscular Média (g/dL) . . . . .	44
<b>LEUCO</b>	Contagem Global Leucócitos (mil/uL) . . . . .	45
<b>BAST</b>	Bastonetes Neutrófilos (mil/uL) . . . . .	45
<b>SEG</b>	Segmentados Neutrófilos (mil/uL) . . . . .	44
<b>EOS</b>	Eosinófilos (mil/uL) . . . . .	47
<b>BAS</b>	Basófilos (mil/uL) . . . . .	44
<b>LIT</b>	Linfócitos (mil/uL) . . . . .	47
<b>LIA</b>	Linfócitos Atípicos (mil/uL) . . . . .	45
<b>MON</b>	Monócitos (mil/uL) . . . . .	47
<b>PLA</b>	Plaquetas (mil/uL) . . . . .	45
<b>RET</b>	Reticulócitos (%) . . . . .	44
<b>GLI</b>	Glicemia jejum (mg/dL) . . . . .	44
<b>COLHDL</b>	Colesterol HDL (mg/dL) . . . . .	44

<b>COLLDL</b>	Colesterol LDL (mg/dL) . . . . .	44
<b>COLVLDL</b>	Colesterol VLDL (mg/dL) . . . . .	45
<b>TRIG</b>	Triglicérides Colesterol HDL (mg/dL) . . . . .	45
<b>AURICO</b>	Ácido úrico (mg/dL) . . . . .	44
<b>UREIA</b>	Ureia (mg/dL) . . . . .	44
<b>CREA</b>	Creatinina (mg/dL) . . . . .	44
<b>FERRO</b>	Ferro sérico (mg/dL) . . . . .	44
<b>AST</b>	Aspartato aminotransferase (U/L) . . . . .	47
<b>ALT</b>	Alanina aminotransferase (U/L) . . . . .	44
<b>GGT</b>	Gama Glutamil Transferase (U/L) . . . . .	45
<b>BD</b>	Bilirrubina Direta (mg/dL) . . . . .	44
<b>FAL</b>	Fosfatase Alcalina (U/L) . . . . .	48
<b>PT</b>	Proteínas Totais (g/dL) . . . . .	44
<b>ALBUM</b>	Albumina (g/dL) . . . . .	45
<b>GLOB</b>	Globulinas (g/dL) . . . . .	47
<b>RELAG</b>	Relação Albumina/globulinas (g/dL) . . . . .	44
<b>LDH</b>	Lactato Desidrogenase (U/L) . . . . .	45
<b>FERRIT</b>	Ferritina (ng/mL) . . . . .	45
<b>ASLO</b>	Antiestreptolisina O (UI/mL) . . . . .	44
<b>PCR</b>	Proteína C Reativa (mg/L) . . . . .	44
<b>HAPTO</b>	Haptoglobina (mg/dL) . . . . .	44
<b>A1ATS</b>	Alfa 1 antitripsina (mg/dL) . . . . .	48
<b>HBS</b>	Hemoglobina S (%) . . . . .	44
<b>HBA2</b>	Hemoglobina A2 (%) . . . . .	44

<b>HBF</b>	Hemoglobina Fetal (%) . . . . .	44
<b>HBA1</b>	Hemoglobina A1 (%) . . . . .	47
<b>HBC</b>	Hemoglobina C (%) . . . . .	45
<b>HAPL</b>	Haplótipos ligados ao gene da globina beta . . . . .	44
<b>TAL</b>	Talassemia alfa . . . . .	45
<b>QCRISDOR</b>	Quantidade de ocorrências de crises de dor nos últimos 6 meses . .	45
<b>QVO</b>	Quantidade de ocorrências de vaso-oclusão . . . . .	44

## Capítulo

# 1

## INTRODUÇÃO

A Doença Falciforme (DF) é uma condição genética que afeta a hemoglobina, uma proteína encontrada nas hemácias, responsáveis pelo transporte de oxigênio no organismo. A presença de Hemoglobina S (HbS), seja em homozigose ou heterozigose em combinação com outras variantes de hemoglobina, é a principal característica da doença falciforme. Essa condição é marcada pela falcização das hemácias quando expostas a condições de desoxigenação (NEUMAYR; HOPPE; BROWN, 2019). Essas hemácias falcizadas causam oclusão do fluxo sanguíneo, principalmente no nível da microcirculação, acarretando uma variedade de complicações sistêmicas, que variam desde eventos dolorosos a danos a diversos órgãos (KATO et al., 2018).

Das hemoglobinopatias, a DF é reconhecida como uma das condições hereditárias mais comuns, afetando cerca de 300.000 crianças anualmente, com aproximadamente 75% dos casos ocorrendo na África Subsaariana (BARROS et al., 2024). Nos Estados Unidos, embora menos prevalente, estima-se que aproximadamente 100.000 pessoas convivam com a doença, em contraste com taxas de 43 por 100.000 nascidos vivos em regiões da Europa (CANÇADO et al., 2023). No Brasil, acredita-se que existam 60.000 a 100.000 pessoas com essa condição genética; na Bahia, estudos apontam uma taxa de incidência de 1:650 nascidos vivos a cada ano, um número superior ao de outras regiões do país (ARAÚJO et al., 2024).



A mortalidade relacionada à DF no Brasil revela a gravidade da condição e os desafios enfrentados pelo Sistema Único de Saúde (SUS) no cuidado aos pacientes. Entre 2000 e 2018, foram registradas 9.817 mortes associadas à doença, com uma mediana de idade ao óbito de apenas 29,5 anos. Isso significa uma redução na sobrevida em comparação com a população geral, cuja mediana é de 69 anos. As principais causas de morte são septicemias (32,1%), pneumonias (19,4%) e insuficiência respiratória (18,2%), complicações que refletem a agressividade da doença (SANTO, 2022). Além disso, crianças entre 1 e 9 anos apresentam um risco 32 vezes maior de morte, e jovens entre 10 e 39 anos têm 13 vezes mais chances de óbito quando comparados à população geral, reforçando a necessidade de intervenções preventivas e melhorias na assistência médica (CANÇADO et al., 2023).

Entre os exames laboratoriais, o hemograma elucida a morfologia das hemácias e aponta uma possível hemoglobinopatia, sendo insuficiente para conclusão diagnóstica. A cromatografia líquida de alta performance (HPLC) é apontada como a melhor técnica de escolha para pesquisa de hemoglobinas variantes, e a eletroforese de hemoglobinas constitui a técnica padrão ouro, sendo capaz de avaliar o perfil quantitativo dos diferentes tipos de hemoglobinas. Esses procedimentos em localidades mais remotas do Brasil e em diversos países do continente africano necessitam de profissionais e sistemas analíticos laboratoriais especializados, mas o acesso nem sempre é possível (HAAN et al., 2020).

Embora o transplante de células-tronco hematopoéticas (TCTH) seja considerado uma alternativa terapêutica curativa, seu uso é limitado pela falta de doadores compatíveis e possibilidade de eventos pós-transplante, tal como doença do enxerto versus hospedeiro (HOUWING et al., 2019). Dessa forma, o acompanhamento clínico e laboratorial regular é a alternativa disponível mais adequada para a sobrevida, qualidade de vida e redução dos eventos. O uso de hidroxiureia, transfusão de hemácias e opioides são os tratamentos comumente usados para controlar os principais sintomas, tendo ainda a terapia gênica como alternativas em desenvolvimento (VICENT; SIMON; YONASI, 2022).

A detecção precoce da doença falciforme pode ser crucial para o tratamento e manejo da condição clínica do paciente. Dessa forma, a tecnologia de aprendizado de máquina pode ser usada em diversos aspectos da doença, tanto para ajudar no suporte ao diag-

nóstico precoce, como para prever seus diversos desfechos ou monitorar a eficácia dos tratamentos existentes, permitindo que os profissionais de saúde decidam sobre as opções terapêuticas mais adequadas para cada paciente, reduzindo assim a probabilidade de desfechos negativos.

## 1.1 CONTEXTO E MOTIVAÇÃO

Na abordagem tradicional de aprendizado de máquina de rótulo único, cada desfecho clínico é analisado de forma isolada, ignorando as relações interdependentes entre os diferentes desfechos que muitas doenças apresentam. Essa limitação é particularmente relevante na doença falciforme, já que a condição é caracterizada por uma ampla heterogeneidade de manifestações clínicas que frequentemente ocorrem de forma simultânea. Nesse contexto, a classificação multirrótulo surge como uma abordagem mais alinhada com a realidade clínica, permitindo modelar a interdependência entre desfechos e oferecendo uma compreensão mais integrada da condição dos pacientes. Isso é importante para lidar com os desafios impostos pela complexidade dessa doença, onde tratamentos dependem de uma análise abrangente dos múltiplos fatores que influenciam o curso clínico.

No recorte dos estudos revisados, grandes avanços no uso de algoritmos de aprendizado de máquina para o diagnóstico e monitoramento de pacientes com doença falciforme foram observados. Técnicas como Random Forest, Extra Trees e Gradient Boosting (PETROVIĆ et al., 2020), além de redes neurais convolucionais (CNN) (HAAN et al., 2020; TENGSHE et al., 2021; ABDULHAY; ALLOW; AL-JALOULY, 2021; SONI; THAKKAR; SINGH, 2022; PRASHANTHI; SINGH, 2023), ResNet, VGG e Inception (SEN et al., 2021; GOSWAMI et al., 2023; JENNIFER et al., 2023; GOSWAMI et al., 2024), alcançaram altos níveis de precisão em tarefas específicas com imagens de esfregaços de sangue e dados hematológicos. Apesar desses avanços, desafios persistem, incluindo a dificuldade de generalização dos modelos em diferentes contextos clínicos, a detecção de células sobrepostas, o tamanho limitado das bases de dados disponíveis e a dependência de dados sintéticos. Além disso, a abordagem de desfechos simultâneos e a identificação

de biomarcadores relevantes para múltiplos desfechos clínicos representam lacunas que devem ser exploradas, apontando caminhos para esta tese e para pesquisas futuras.

Nesse contexto, a aplicação de classificadores como Binary Relevance (BR), Classifier Chains (CC), Label Powerset (LP) e Random k-labelsets (RAkEL) têm sido avaliados em diferentes cenários (BOGATINOVSKI et al., 2022; GARCÍA-PEDRAJAS et al., 2024), porém nenhuma abordagem considerou pacientes com a doença falciforme como objeto de estudo até a escrita da tese. Esses algoritmos não apenas fornecem suporte ao diagnóstico médico, como também possibilitam a identificação precoce de complicações simultâneas, favorecendo o suporte a intervenções preventivas e mais assertivas.

A escassez de estudos que explorem a aplicação da classificação multirrótulo ressalta o caráter inovador desta tese. Ao introduzir essa abordagem de computação aplicada no contexto clínico, o trabalho avança o conhecimento científico e fornece uma ferramenta promissora de suporte ao diagnóstico médico. Tendo potencial de viabilizar estratégias mais eficazes de tratamento e cuidado, promovendo benefícios diretos aos pacientes e contribuindo para a evolução e entendimento das relações entre os desfechos e seus principais biomarcadores.

## 1.2 OBJETIVOS E QUESTÕES DE PESQUISA

Partindo da premissa de que o uso de algoritmos de aprendizado de máquina multirrótulo oferece suporte ao diagnóstico dos diversos desfechos da doença falciforme de maneira mais adequada do que os métodos tradicionais de rótulo único, nossa investigação busca explorar as vantagens dessa abordagem. Isso ocorre porque os algoritmos multirrótulo são capazes de prever múltiplos desfechos simultaneamente, refletindo a complexidade e a natureza da doença. Diferente dos métodos convencionais, que tratam cada desfecho de forma isolada, os algoritmos multirrótulo têm a capacidade de capturar a interdependência entre os desfechos, proporcionando uma visão mais abrangente da condição clínica do paciente. Este estudo tem como objetivo principal analisar o uso desses algoritmos na previsão de desfechos da doença falciforme, levando em consideração a variabilidade e a imprevisibilidade das relações clínicas e laboratoriais.

O objetivo geral de nossa investigação é o seguinte:

**Analisar como os algoritmos de classificação multirrótulo performam na previsão dos desfechos da doença falciforme.**

Além disso, com o intuito de aprofundar nossa compreensão sobre a aplicação dos algoritmos no contexto da doença falciforme, procuramos responder a um conjunto de questões de pesquisa que abordam aspectos relacionados ao desempenho dos modelos, à seleção de variáveis e ao processamento de dados. As questões de pesquisa que orientam nossa análise são as seguintes:

**QP1.** Quais são os parâmetros clínicos e laboratoriais mais relevantes para a previsão de múltiplos desfechos simultâneos na doença falciforme, e como esses fatores interagem para fornecer uma previsão?

**QP2.** Quais estratégias de pré-processamento de dados, como normalização, imputação de valores ausentes e redução de dimensionalidade, demonstram ter um impacto positivo nos modelos multirrótulo, contribuindo para a melhoria das métricas de desempenho?

**QP3.** Os algoritmos de aprendizado de máquina multirrótulo são capazes de identificar padrões preditivos exclusivos e complexos que não seriam facilmente detectados por métodos tradicionais de classificação, como os algoritmos de rótulo único?

Essas questões são fundamentais para entender as potencialidades dos algoritmos multirrótulo na análise de dados de pacientes com a doença falciforme e visam fornecer uma base sólida para futuras pesquisas e aplicações clínicas na área.

### 1.3 ORGANIZAÇÃO DA TESE

A Figura 1.1 apresenta uma visão abrangente da estrutura da tese, que é composta por uma introdução, seguida de seis capítulos que exploram de forma detalhada os diversos aspectos da pesquisa. Cada capítulo foi cuidadosamente elaborado para fornecer uma compreensão aprofundada dos temas abordados, estabelecendo uma base sólida para os objetivos e as questões de pesquisa propostas.

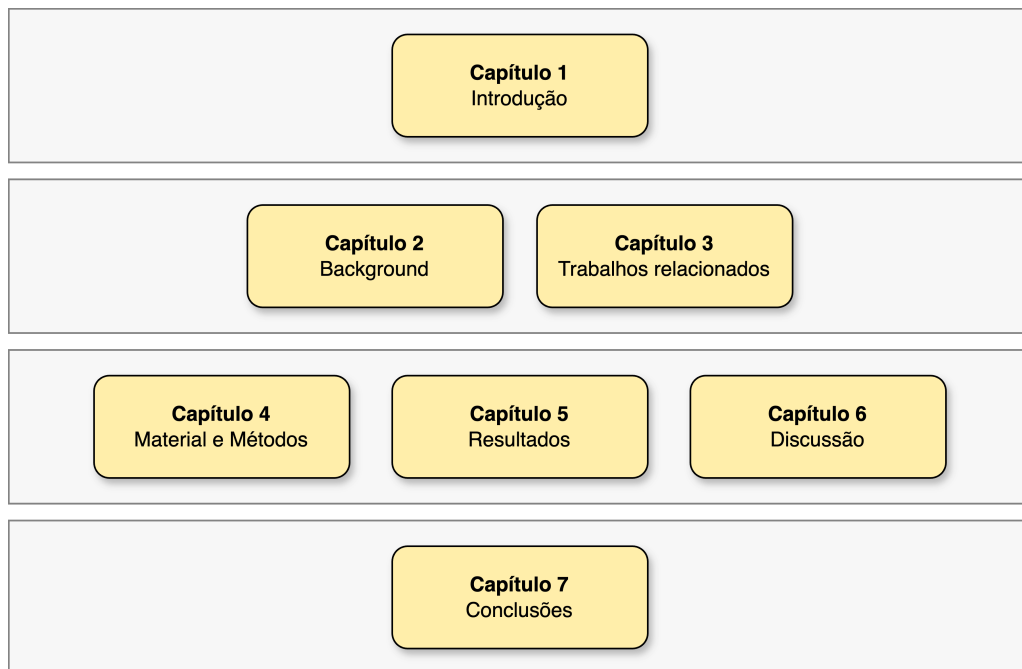


Figura 1.1: Visão geral da estrutura da tese

O Capítulo 1, intitulado **Introdução**, oferece um panorama geral sobre a doença falciforme, a relevância do estudo e os objetivos desta pesquisa. Além disso, são introduzidas as motivações para a utilização de algoritmos de aprendizado de máquina, especialmente os algoritmos de classificação multirrótulo, como uma abordagem inovadora para enfrentar os desafios relacionados a múltiplos desfechos clínicos.

O Capítulo 2, denominado **Background**, fornece uma visão detalhada sobre a doença falciforme, abordando sua epidemiologia, as manifestações clínicas que caracterizam a condição e as opções de tratamento atualmente disponíveis. Esse capítulo também apresenta os conceitos fundamentais de classificação multirrótulo, explicando como esses algoritmos podem ser aplicados a problemas complexos como o da doença falciforme. São discutidos, ainda, os métodos de transformação de problemas para adaptá-los à abordagem multirrótulo e são descritos os principais algoritmos utilizados nesse contexto.

O Capítulo 3, denominado **Trabalhos Relacionados**, revisa o estado da arte sobre o uso de algoritmos de inteligência artificial em dados de pacientes com a doença falciforme. Este capítulo apresenta uma visão geral das pesquisas existentes, identificando as limitações desses estudos e obtendo as oportunidades de investigação que ainda não fo-

ram suficientemente exploradas. As lacunas encontradas serviram como ponto de partida para os experimentos realizados ao longo da tese.

O Capítulo 4, intitulado **Material e Métodos**, descreve de forma detalhada os procedimentos seguidos durante os experimentos, incluindo a descrição e o pré-processamento dos dados, a escolha dos modelos de aprendizado e as métricas utilizadas para avaliar o desempenho dos algoritmos. Esse capítulo fornece o contexto necessário para a compreensão dos experimentos realizados.

O Capítulo 5, denominado **Resultados**, apresenta os resultados obtidos a partir da aplicação de classificadores tradicionais de rótulo único e multirrótulo na previsão de desfechos simultâneos da doença falciforme. Neste capítulo, são relatados os resultados obtidos para cada um dos modelos de aprendizado de máquina empregados, incluindo uma análise detalhada de seus desempenhos.

O Capítulo 6, intitulado **Discussão**, oferece uma análise crítica dos resultados encontrados, comparando-os com os achados da literatura existente e destacando as contribuições originais desta pesquisa. A partir dessa comparação, são discutidas as implicações dos resultados para o avanço do conhecimento na área da doença falciforme e o potencial das abordagens de aprendizado de máquina para melhorar o suporte ao diagnóstico da doença.

Por fim, o Capítulo 7, denominado **Conclusões**, resume as principais contribuições alcançadas por meio desta tese e apresenta as conclusões gerais. Este capítulo também discute as perspectivas futuras, sugerindo direções para novas pesquisas que possam expandir o entendimento sobre a doença falciforme e explorar ainda mais as possibilidades do uso de algoritmos de aprendizado de máquina.

## **BACKGROUND**

### **2.1 DOENÇA FALCIFORME**

#### **2.1.1 Epidemiologia**

Globalmente, cerca de 300 mil crianças nascem anualmente com a doença, a maioria na África Subsaariana, onde as taxas de portadores podem ultrapassar 30% em algumas comunidades (KAVANAGH; FASIPE; WUN, 2022). Além disso, a migração humana e a globalização têm ampliado a distribuição da doença, tornando-a um problema de saúde pública significativo em regiões como Europa, América do Norte e Oriente Médio.

Na África Subsaariana, a doença falciforme é uma das principais causas de mortalidade infantil. Estudos indicam que até 80% das crianças afetadas morrem antes de completar cinco anos devido à falta de diagnóstico precoce e tratamento adequado (ARAÚJO et al., 2024). A pobreza, a infraestrutura médica limitada e o acesso insuficiente a medicamentos essenciais, como antibióticos e vacinas, agravam o cenário. Em contraste, países de alta renda têm reduzido a taxa de mortalidade associada à doença por meio de programas de triagem neonatal, intervenções preventivas e manejo clínico avançado.

Nos Estados Unidos, estima-se que 100 mil pessoas convivam com a doença, predominantemente entre descendentes de africanos (TEBBI, 2022). Graças à triagem neonatal obrigatória e à melhoria nos cuidados médicos, a expectativa de vida dos pacientes

aumentou para cerca de 40 a 60 anos. No Reino Unido, onde a triagem também é amplamente implementada, o suporte às famílias e os avanços no tratamento têm permitido uma qualidade de vida significativamente melhor. Essas intervenções incluem o uso de hidroxiureia e transfusões regulares para prevenir complicações graves, como acidentes vasculares cerebrais e crises vaso-oclusivas.

As iniciativas globais para combater as desigualdades no manejo da doença incluem esforços da Organização Mundial da Saúde (OMS) para aumentar a conscientização, melhorar a triagem e garantir o acesso a cuidados de saúde (CORDOVIL et al., 2023). Países como Índia e Arábia Saudita, onde a prevalência também é significativa, têm implementado programas regionais para diagnóstico e prevenção, com base em abordagens de saúde pública. No entanto, esses esforços muitas vezes enfrentam desafios culturais, econômicos e estruturais que dificultam sua efetividade.

No Brasil, dados indicam uma queda na taxa de natalidade nas últimas décadas, em contrapartida ao aumento no número de casos diagnosticados no mesmo período (ARAÚJO et al., 2024). Os dados da triagem neonatal da doença falciforme brasileira de 2012 apontam que a maior incidência de casos ocorre na Bahia, com uma taxa de 1:650, contra 1:13.500 indivíduos nascidos a cada ano nos estados da Região Sul (SILVA-PINTO et al., 2019). Dados sobre a mortalidade na população brasileira com a doença indicam que a idade média ao óbito é de 32 anos, inferior à média da população geral, de 69 anos. A sobrevida mediana estimada para pacientes com doença falciforme é de 40 anos, enquanto para a população em geral é de 80 anos (CANÇADO et al., 2023). O risco de mortalidade entre crianças com a doença é reportado como 30 vezes maior em comparação com aquelas sem a condição.

### **2.1.2 Aspectos clínicos**

Os aspectos clínicos da doença falciforme são caracterizados por uma ampla gama de manifestações, variando de sintomas leves a complicações graves que impactam significativamente a qualidade e a expectativa de vida dos pacientes. A base fisiopatológica da doença é uma mutação no gene HBB, responsável por codificar a hemoglobina. Essa



mutação resulta na produção de hemoglobina S, que se polimeriza sob condições de baixa oxigenação, levando à deformação das hemácias em formato de foice (ROSSI et al., 2024). Essas células falciformes possuem uma vida útil reduzida e podem obstruir os vasos sanguíneos, resultando em crises vaso-oclusivas.

As crises vaso-oclusivas são a manifestação clínica mais comum e debilitante da doença falciforme. Elas são desencadeadas pela obstrução de pequenos vasos sanguíneos pelas hemácias falciformes, resultando em dor severa e isquemia tecidual. Essas crises podem ocorrer em qualquer órgão ou tecido, mas frequentemente afetam ossos, articulações, pulmões e abdômen. Episódios recorrentes de dor são uma das principais razões para hospitalizações e têm um impacto profundo na qualidade de vida do paciente (PADHEE et al., 2021; JI et al., 2021; PADHEE et al., 2022).

Outra complicação importante é a anemia hemolítica crônica, que resulta da destruição acelerada das hemácias falciformes, cuja vida útil média é de apenas 10 a 20 dias, em comparação aos 120 dias das hemácias normais (IDRIS; BOTCHWEY; HYACINTH, 2022). A anemia contribui para sintomas como fadiga, fraqueza, dispneia e palidez. Em crianças, a anemia severa pode prejudicar o crescimento e o desenvolvimento, aumentando a necessidade de intervenções médicas frequentes.

O sistema imunológico dos pacientes com doença falciforme também é comprometido. A disfunção da função esplênica, muitas vezes devido a autoinfartos repetidos no baço, aumenta a suscetibilidade a infecções bacterianas graves, como septicemia e meningite. Organismos encapsulados, como *Streptococcus pneumoniae* e *Haemophilus influenzae*, representam um risco significativo, especialmente em crianças (KATO et al., 2018). Vacinação e profilaxia com antibióticos são estratégias essenciais para reduzir a mortalidade associada a essas infecções.

A síndrome torácica aguda é uma complicação potencialmente fatal que afeta cerca de 30% dos pacientes com doença falciforme ao longo da vida (KATO et al., 2018). Ela é caracterizada por dor torácica, febre, tosse e hipoxemia, muitas vezes associada a infecções ou crises vaso-oclusivas nos pulmões. A intervenção imediata com antibióticos, oxigênio suplementar e, em casos graves, transfusões sanguíneas, é crucial para evitar

complicações adicionais, como insuficiência respiratória (BHASIN; SARODE, 2023).

No sistema nervoso central, a doença falciforme pode levar a acidentes vasculares cerebrais (AVCs) isquêmicos ou hemorrágicos. Estima-se que 11% das crianças com a condição sofram um AVC antes dos 20 anos, com um risco maior em idades mais jovens (TEBBI, 2022). O uso de exames de doppler transcraniano para monitoramento e transfusões sanguíneas regulares têm se mostrado eficazes na redução da incidência de AVC em crianças de alto risco.

Lesões renais crônicas também são comuns, com até 30% dos pacientes adultos desenvolvendo algum grau de nefropatia falciforme (GÜNTÜRKÜN et al., 2021). Essa condição inclui proteinúria, hematúria e, em casos avançados, insuficiência renal. Além disso, a falcização crônica em pequenos vasos renais pode comprometer a capacidade de concentração urinária, levando à desidratação.

### 2.1.3 Tratamento

Embora seja uma condição genética incurável em muitos casos, avanços terapêuticos têm proporcionado melhorias significativas na expectativa e na qualidade de vida dos pacientes. O manejo clínico engloba desde estratégias de suporte até terapias avançadas que visam corrigir ou mitigar os efeitos da doença.

O tratamento de suporte inclui transfusões sanguíneas regulares, que têm como objetivo principal prevenir complicações graves, como os acidentes vasculares cerebrais (AVCs), e corrigir a anemia severa (DEBAUN et al., 2020). A suplementação de ácido fólico é amplamente recomendada para apoiar a produção de hemácias, essencial frente à destruição acelerada das células falciformes (AGBOR et al., 2024). Aliado a isso, o controle rigoroso da dor, especialmente em crises vaso-oclusivas, demanda o uso escalonado de analgésicos, desde anti-inflamatórios não esteroides (AINEs) até opioides em casos mais graves.

A hidroxiureia consolidou-se como uma das mais importantes intervenções terapêuticas para pacientes com a condição (VICENT; SIMON; YONASI, 2022). Este fármaco age aumentando os níveis de hemoglobina fetal (HbF), que, por sua vez, inibe a polimerização

da hemoglobina S e reduz as crises vaso-oclusivas e a síndrome torácica aguda. Seu uso a longo prazo tem demonstrado segurança e impacto positivo, com redução expressiva na frequência de hospitalizações.

As infecções, uma das principais causas de morbidade e mortalidade, são prevenidas por meio de vacinação precoce e profilaxia antibiótica. Crianças pequenas recebem penicilina desde o diagnóstico, enquanto pacientes de todas as idades são vacinados contra patógenos encapsulados, como *Streptococcus pneumoniae* e *Haemophilus influenzae* (KATO et al., 2018). Em conjunto, essas medidas têm reduzido drasticamente as taxas de infecções graves, sobretudo em regiões com acesso limitado a serviços médicos.

O transplante de células-tronco hematopoiéticas (TCTH) é, atualmente, a única abordagem curativa disponível. Apesar de promissor, este procedimento é restrito a pacientes com doadores totalmente compatíveis e apresenta riscos significativos, como rejeição do enxerto e infecções pós-transplante (HOUWING et al., 2019). Ainda assim, avanços nas técnicas de condicionamento e imunossupressão têm ampliado as perspectivas para este tratamento.

Além das intervenções clássicas, novas terapias, como a terapia gênica, estão emergindo como alternativas revolucionárias. Essas abordagens visam corrigir diretamente a mutação genética ou modular a expressão de genes relacionados, como o que regula a produção de HbF (RAGHURAMAN et al., 2024).

O manejo adequado da doença falciforme não se limita ao tratamento médico. Intervenções educacionais e suporte psicológico são fundamentais para capacitar pacientes e cuidadores, ajudando-os a lidar com as limitações impostas pela condição. Programas de educação podem reduzir hospitalizações e melhorar a adesão às terapias propostas.

## 2.2 ALGORITMOS MULTIRRÓTULO

Os algoritmos de classificação multirrótulo surgiram como uma resposta às limitações dos métodos tradicionais de classificação de rótulo único, que não capturam a complexidade dos problemas em que uma instância pode estar associada a múltiplas categorias simultaneamente. Essa necessidade tornou-se evidente em áreas como bioinformática,

processamento de linguagem natural e visão computacional, onde a interdependência entre classes são importantes nas predições.

Formalmente, os classificadores multirrótulo seguem a seguinte definição matemática: seja  $X = \mathbb{R}^d$  (ou  $\mathbb{Z}^d$ ) o espaço de instâncias  $d$ -dimensionais e  $Y = \{y_1, y_2, \dots, y_q\}$  o espaço de rótulos com  $q$  possíveis valores. O objetivo do aprendizado multirrótulo é determinar uma função  $h : X \rightarrow 2^Y$  a partir de um conjunto de treinamento  $D = \{(x_i, Y_i) \mid 1 \leq i \leq m\}$ , onde cada exemplo  $(x_i, Y_i)$  é formado por um vetor de características  $d$ -dimensional  $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^\top$  e um conjunto de rótulos associados  $Y_i \subseteq Y$ . Para uma instância  $x \in X$  não vista, o classificador  $h(\cdot)$  retorna  $h(x) \subseteq Y$  como o conjunto de rótulos mais apropriados (ZHANG; ZHOU, 2013).

A trajetória dos algoritmos multirrótulo datam do início dos anos 2000, com o desenvolvimento de abordagens de transformação de problema. O Binary Relevance (BR), um dos primeiros métodos, separava o problema em classificadores binários independentes, cada um focado em um rótulo específico. Embora simples, essa abordagem ignora as relações entre rótulos, limitando sua eficácia em cenários complexos. Paralelamente, o Label Powerset (LP) foi introduzido (BOGATINOVSKI et al., 2022), transformando combinações de rótulos em classes únicas. Apesar de capturar interdependências, sua aplicabilidade é prejudicada em problemas com um número elevado de rótulos, devido à combinatória exponencial.

Os avanços subsequentes concentraram-se em métodos que modelam as dependências entre rótulos. O Classifier Chains (CC), por exemplo, ordena os rótulos sequencialmente, permitindo que as predições de um rótulo influenciassem as de outro (GARCÍA-PEDRAJAS et al., 2024). Essa abordagem trouxe ganhos consideráveis em precisão, ao mesmo tempo que preserva a simplicidade computacional. Além disso, algoritmos baseados em aprendizado por instância, como o ML-kNN (ZHANG; ZHOU, 2013), utilizam estatísticas locais para prever múltiplos rótulos, mostrando-se particularmente úteis em conjuntos de dados escassos.

Com o avanço das redes neurais, os algoritmos multirrótulo passaram por uma revolução significativa. Redes convolucionais (CNNs) foram adaptadas para tarefas mul-

tirrótulo, especialmente em visão computacional, permitindo a classificação simultânea de várias características em imagens (PRIYA; PETER, 2022). No processamento de linguagem natural, modelos baseados em arquiteturas Transformer, como o BERT, abriram novas possibilidades para capturar interdependências entre rótulos em textos extensos e complexos (WANG et al., 2024).

A Figura 2.1 resume como os problemas de classificação multirrótulo são tratados na literatura, destacando os algoritmos utilizados na tese.

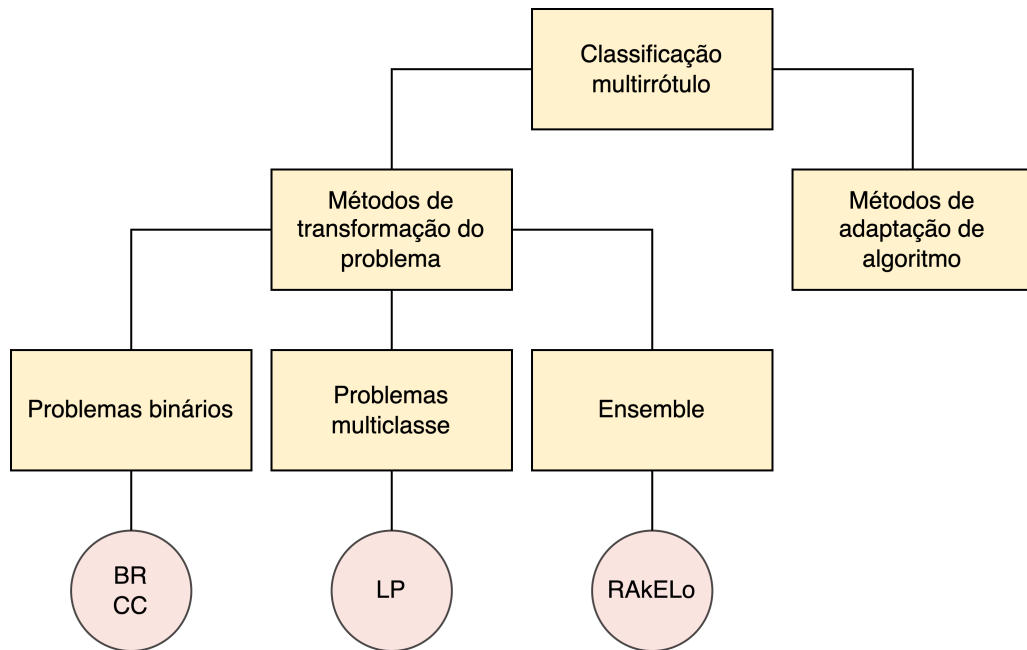


Figura 2.1: Opções para resolver problemas de classificação multirrótulo.

### 2.2.1 Binary Relevance

O método *Binary Relevance* é amplamente utilizado na classificação multirrótulo, conhecida por sua simplicidade, eficiência computacional e ampla aplicação em diversas áreas. Ele transforma o problema multirrótulo em uma série de problemas binários independentes, tratando cada rótulo como uma tarefa de classificação separada. Embora seja frequentemente criticado por desconsiderar as interdependências entre os rótulos, o BR continua sendo utilizado como uma baseline para avaliar métodos mais novos e como componente em algoritmos híbridos. Além disso, devido à sua escalabilidade e compati-

bilidade com classificadores tradicionais, o BR é útil em cenários com conjuntos de dados escassos ou de alta dimensionalidade, onde métodos mais complexos podem ser inviáveis computacionalmente (ZHANG; ZHOU, 2013).

Formalmente, no BR, o aprendizado de múltiplos rótulos é transformado em  $q$  problemas de classificação binária independentes, onde cada problema corresponde a um rótulo do espaço de rótulos. Para cada rótulo  $l_i$ , um classificador binário é treinado para prever se uma instância pertence ou não a esse rótulo, de forma independente dos demais. Assim, o conjunto de treinamento original  $D$ , formado por instâncias  $(x, Y)$ , onde  $Y$  é o conjunto de rótulos associados, é decomposto em  $q$  conjuntos binários  $D_i$ . Cada  $D_i$  contém pares  $(x, y_i)$ , onde  $y_i = 1$  se  $l_i \in Y$  e  $y_i = 0$  caso contrário.

Para superar as limitações do método BR, os pesquisadores desenvolveram versões aprimoradas, como o *Classifier Chains* (READ et al., 2009). Essas variações incorporam técnicas que modelam as dependências entre os rótulos, resultando em melhorias no desempenho preditivo.

### 2.2.2 Classifier Chains

O *Classifier Chains* é uma abordagem poderosa para problemas de classificação multirrótulo que supera a limitação do *Binary Relevance* ao modelar explicitamente as dependências entre os rótulos. No BR, os rótulos são tratados de forma independente, enquanto o CC cria uma cadeia de classificadores, onde cada classificador depende das previsões anteriores na sequência. Isso permite que o modelo capture interações entre os rótulos, potencialmente melhorando a acurácia nas tarefas em que as dependências entre rótulos são significativas (READ et al., 2009).

A ideia central do *Classifier Chains* é que, ao invés de tratar cada rótulo de forma isolada, cada rótulo  $l_i$  em um problema multirrótulo depende não apenas das características da instância  $x$ , mas também das previsões anteriores de rótulos. Em um processo sequencial, o primeiro classificador  $h_1$  é treinado da maneira tradicional, com as instâncias  $(x, y_1)$ , onde  $y_1$  é o rótulo binário. O segundo classificador  $h_2$  é então treinado com as instâncias  $(x, y_1, y_2)$ , ou seja, inclui a previsão de  $y_1$  como uma característica adicional.

Esse processo continua até o último rótulo.

A principal vantagem do CC é a sua capacidade de modelar dependências entre rótulos, o que pode ser especialmente útil em domínios onde essas dependências são naturais ou críticas, como em biologia ou análise de texto. Entretanto, a dependência de uma sequência de predições pode fazer com que a ordem dos classificadores na cadeia tenha um impacto no desempenho, sendo necessário um processo cuidadoso para escolher essa ordem de forma adequada.

A limitação do CC reside na ordem dos classificadores na cadeia. Como as predições anteriores influenciam os classificadores subsequentes, uma escolha subótima da ordem pode impactar negativamente o desempenho. Várias abordagens têm sido propostas para mitigar esse problema, como o uso de ordens aleatórias ou baseadas em relevância dos rótulos. Outra desvantagem é a computação mais intensiva. O processo de encadeamento de predições aumenta a complexidade do modelo, tornando-o mais custoso em termos de tempo de treinamento e memória, especialmente em grandes conjuntos de dados.

### 2.2.3 Label Powerset

O *Label Powerset* (LP) é uma abordagem singular dentro da classificação multirrótulo, diferenciando-se substancialmente de métodos como *Binary Relevance* (BR) e *Classifier Chains* (CC). Em vez de tratar cada rótulo individualmente (como o BR) ou em sequência (como o CC), o LP transforma o problema multirrótulo em uma tarefa de classificação multiclasse. Ele faz isso ao tratar cada combinação única de rótulos presentes no conjunto de treinamento como uma classe distinta, permitindo que o modelo preveja diretamente as combinações de rótulos para cada instância (TSOUMAKAS; KATAKIS, 2007).

A principal característica do LP está nessa conversão: todas as combinações únicas de rótulos do conjunto de treinamento tornam-se classes. Por exemplo, dado um conjunto de rótulos  $l_1, l_2, l_3$ , as combinações possíveis são tratadas como classes individuais, como  $\{l_1\}, \{l_2\}, \{l_3\}, \{l_1, l_2\}, \{l_1, l_3\}, \{l_2, l_3\}$  e  $\{l_1, l_2, l_3\}$ . Dessa forma, o classificador aprende a prever essas classes combinadas, capturando explicitamente as interações e dependências entre os rótulos.

O LP é particularmente eficaz em cenários onde o número de combinações únicas de rótulos é relativamente pequeno e os rótulos apresentam forte interdependência. Essa abordagem é valiosa quando as combinações de rótulos têm significados específicos e não devem ser tratadas de forma independente. Sua capacidade de modelar as interações entre rótulos o torna poderoso para problemas onde a co-ocorrência de rótulos carrega informações importantes (READ et al., 2009).

Entretanto, o método apresenta limitações significativas relacionadas à escalabilidade. À medida que o número de rótulos cresce, o número de combinações possíveis aumenta exponencialmente, resultando em um grande espaço de classes. Esse crescimento exponencial pode levar a problemas de sobreajuste, especialmente quando algumas combinações de rótulos aparecem raramente no conjunto de dados. Além disso, o treinamento e a inferência podem se tornar computacionalmente caros, impactando diretamente a eficiência do modelo.

Portanto, embora o LP seja poderoso para capturar dependências complexas entre rótulos, ele é mais adequado para cenários com um número limitado de rótulos e combinações. Em situações com muitos rótulos ou interações menos significativas entre eles, abordagens como *Classifier Chains* podem oferecer maior escalabilidade e eficiência, sem sacrificar a capacidade de modelar interdependências entre rótulos (ZHANG; ZHOU, 2013).

#### **2.2.4 Random k-labelsets (RAKEL)**

O RAKEL (RANdom k-labELsets) é uma abordagem baseada em conjuntos para a classificação multirrótulo, projetada para lidar de maneira mais eficiente com a complexidade e limitações associadas a problemas desse tipo. Diferentemente do Classifier Chains (CC), o RAKEL utiliza uma combinação de conjuntos aleatórios de rótulos por meio do Label Powerset (LP). Essa abordagem permite criar subconjuntos menores de rótulos, reduzindo a complexidade computacional e o risco de overfitting, problemas comuns em tarefas com muitos rótulos (TSOUMAKAS; VLAHAVAS, 2007).

A principal inovação do RAKEL está na divisão dos rótulos em subconjuntos aleató-



rios, permitindo que o modelo aplique o conceito do Label Powerset em cada subconjunto. Essa estratégia reduz o número de combinações de rótulos, evitando o crescimento exponencial associado ao LP puro. A randomização também contribui para uma maior robustez, minimizando o impacto de dependências específicas entre os rótulos e aumentando a capacidade de generalização do modelo.

Além disso, o RAKEL adota técnicas para lidar com o desequilíbrio dos rótulos. Ao distribuir os rótulos em subconjuntos menores e independentes, o método mitiga problemas associados a rótulos pouco representados no conjunto de dados. Essa abordagem é especialmente relevante em domínios como a biomedicina, onde as combinações de doenças ou sintomas podem ser muito variadas, mas algumas aparecem com menor frequência.

Contudo, como qualquer abordagem baseada em combinações de rótulos, o RAKEL ainda enfrenta desafios. A criação de múltiplos subconjuntos e o treinamento de vários classificadores aumentam os custos computacionais, tanto em tempo quanto em recursos. Além disso, o desempenho do modelo pode ser impactado pela escolha inadequada do tamanho dos subconjuntos ou pela falha em capturar corretamente as dependências importantes entre os rótulos.

## 2.3 INTERFACE ENTRE COMPUTAÇÃO E SAÚDE

Embora o aprendizado de máquina esteja amplamente difundido, buscamos esclarecer o posicionamento da nossa pesquisa. Para isso, ilustramos dois cenários que destacam o uso da tecnologia no diagnóstico e monitoramento da doença falciforme. No primeiro cenário (Figura 2.2), apresentamos como esses processos ocorrem sem suporte tecnológico. Já no segundo (Figura 2.3), exploramos como os algoritmos de inteligência artificial multirrótulo podem dar suporte a prática clínica.

No cenário tradicional, o diagnóstico e monitoramento da doença falciforme dependem de análises manuais, como a observação de esfregaços sanguíneos e interpretação de dados hematológicos por especialistas. Esse método, embora amplamente utilizado, apresenta limitações significativas: maior susceptibilidade a falhas humanas, dependência da experiência do profissional, maior tempo de interpretação e dificuldade na identificação

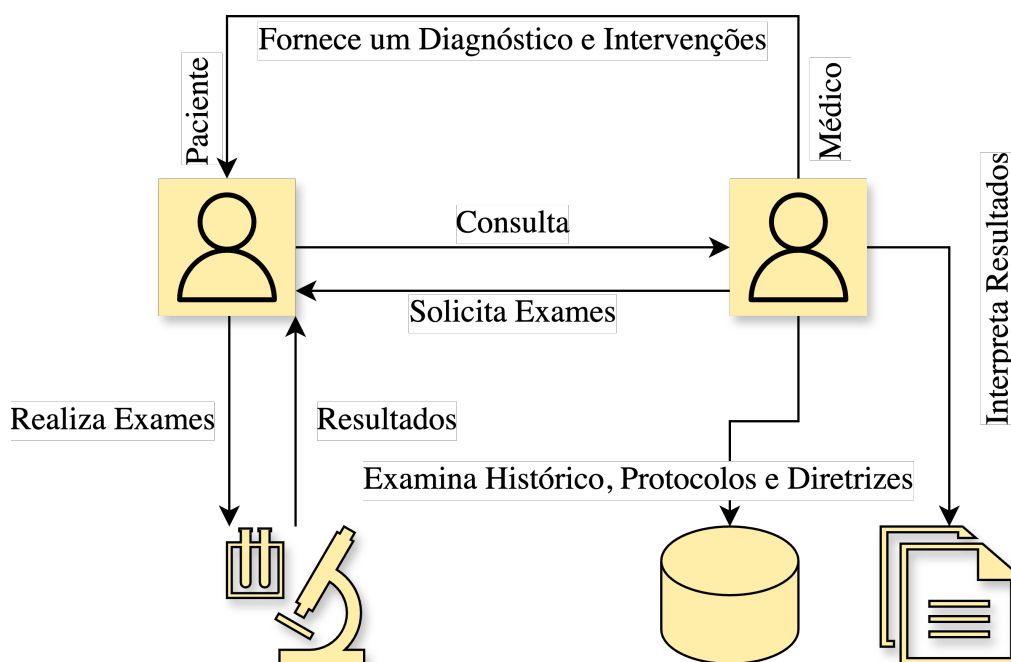


Figura 2.2: Atendimento ao paciente sem uso de tecnologia.

de padrões complexos entre os biomarcadores. Consequentemente, o tratamento personalizado é comprometido, especialmente em casos que necessitem de urgência.

Em contraste, o cenário assistido por inteligência artificial utiliza algoritmos tradicionais de rótulo único, de redes neurais convolucionais e classificadores multirrótulo para realizar análises automatizadas de grandes volumes de dados clínicos e imagens. Essas tecnologias permitem não apenas diagnósticos mais rápidos e precisos, mas também previsões de possíveis complicações. Essa abordagem integrada promove maior eficiência no uso de recursos médicos, além de oferecer suporte para decisões clínicas personalizadas, contribuindo diretamente no cuidado aos pacientes.

Entretanto, alguns desafios precisam ser superados para garantir a qualidade da resposta dos modelos. Um dos mais frequentemente relatados na literatura é a necessidade de uma quantidade suficiente de dados para permitir uma boa generalização. Modelos de aprendizado de máquina são altamente dependentes de grandes volumes de dados de qualidade para evitar problemas como *overfitting* ou viés nos resultados. Além disso, a escolha dos dados de entrada desempenha um papel muito importante no processo, in-

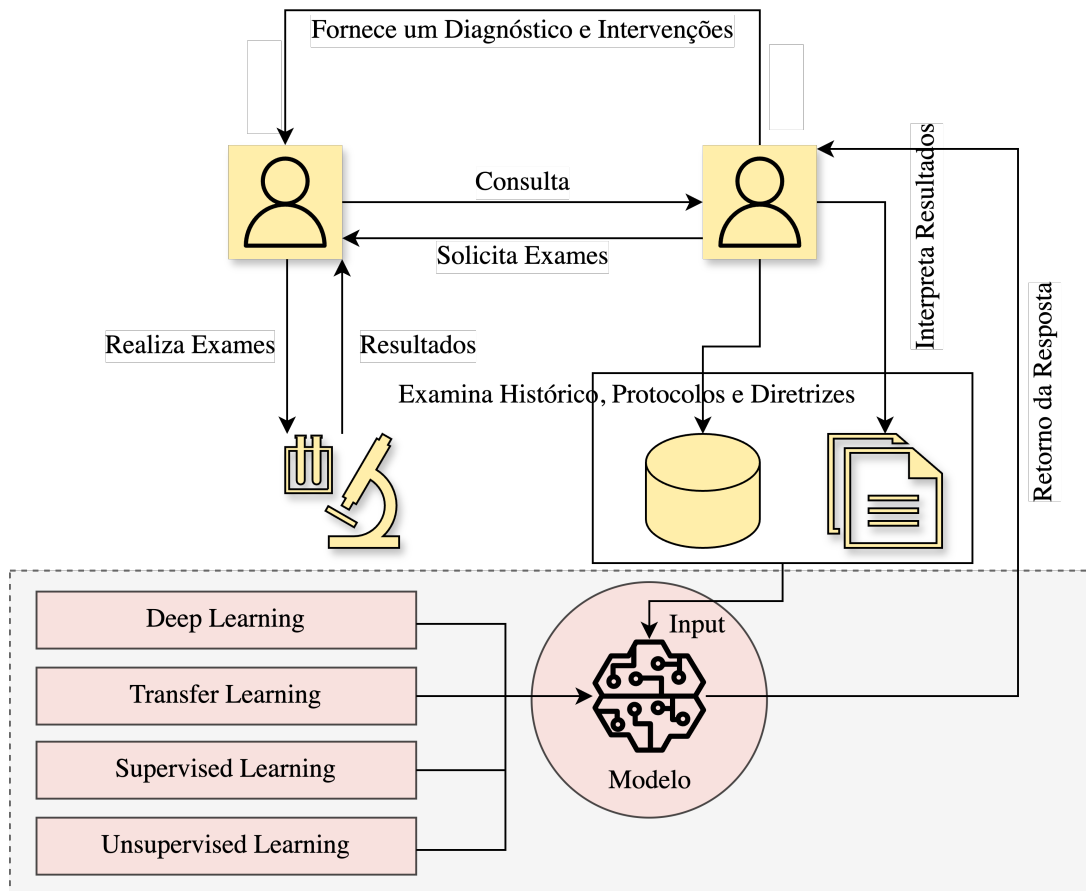


Figura 2.3: Atendimento ao paciente com uso de tecnologia.

cluindo o tratamento adequado de dados incompletos por meio de imputação, bem como a normalização dos atributos para evitar que diferentes escalas influenciem desproporcionalmente o desempenho dos modelos. Outros aspectos importantes incluem a calibragem e configuração dos hiperparâmetros dos modelos, como taxa de aprendizado, número de camadas em redes neurais ou critérios de parada. Esses processos são chamados de pré-processamento e demanda expertise, tempo e conhecimento dos classificadores.

## **TRABALHOS RELACIONADOS**

### **3.1 ABORDAGENS NO DIAGNÓSTICO**

Nos últimos anos, pesquisas têm explorado o potencial dos algoritmos de aprendizado de máquina no diagnóstico da doença falciforme, utilizando imagens de esfregaço de sangue como entrada. Os estudos têm se concentrado no uso de técnicas de visão computacional e algoritmos de classificação de rótulo único para analisar as características morfológicas das células sanguíneas presentes nas imagens. Diversas variáveis são consideradas nessa análise, incluindo forma, tamanho, textura e outras propriedades relevantes das células. Com base nessas informações, os algoritmos são treinados para identificar e distinguir células sanguíneas normais das células afetadas pela doença falciforme. A Tabela 3.1 apresenta uma visão geral dos algoritmos mais utilizados pelos pesquisadores nos últimos anos, destacando a abordagem adotada e métrica de acurácia para o desfecho.

No estudo conduzido por Javidi et al. (2018), foram utilizadas 300 amostras de células, incluindo pacientes saudáveis e indivíduos com doença falciforme. Todos os participantes eram adultos e não haviam recebido transfusões de sangue nos três meses anteriores à coleta dos dados. As características espaço-temporais foram medidas por meio do desvio padrão da média e do cubo formado pela sequência dos esfregaços de sangue ao longo do tempo para cada local na membrana celular. Além disso, foram calculadas características

Tabela 3.1: Principais algoritmos utilizados no diagnóstico da doença falciforme, tendo como entrada imagens de esfregaço de sangue.

<b>Autores</b>	<b>Algoritmos</b>	<b>Acurácia</b>	<b>Diagnóstico</b>
(JAVIDI et al., 2018)	Random Forest	93,33%	Doença falciforme
(CHY; RAHAMAN, 2019)	K-Nearest Neighbor, Support Vector Machine e Extreme Learning Machine	87,73%	Doença falciforme
(PETROVIĆ et al., 2020)	Gradient Boosting e Random Forest	93,98%	Doença falciforme
(HAAN et al., 2020)	Convolutional Neural Networks	98%	Doença falciforme
(TENGSHE et al., 2021)	Convolutional Neural Networks	94,57%	Doença falciforme
(SEN et al., 2021)	Convolutional Neural Networks	91%	Doença falciforme
(ABDULHAY; ALLOW; AL-JALOULY, 2021)	Convolutional Neural Networks	93,4%	Doença falciforme
(VICENT; SIMON; YONASI, 2022)	Support Vector Machine, Naïve Bayes e Logistic Regression	98,18%	Doença falciforme
(SONI; THAKKAR; SINGH, 2022)	Convolutional Neural Networks	95,5%	Doença falciforme versus Traço falciforme

morfológicas com base no comprimento do caminho óptico, incluindo o comprimento médio do caminho óptico, coeficiente de variação, volume óptico, área projetada, razão de PA e OV, assimetria e curtose. Após a seleção das características de entrada, os autores utilizaram o algoritmo de classificação *Random Forest* para prever se uma célula era normal ou apresentava doença falciforme. O modelo foi testado em três cenários diferentes. No primeiro cenário, em que apenas as características espaciais-temporais foram consideradas, o modelo obteve uma acurácia de 78,00%, com uma especificidade de 81,33% e sensibilidade de 74,67%. No segundo cenário, que levou em conta apenas as características morfológicas, o modelo apresentou uma acurácia de 92,67%, com especificidade de 96,00% e sensibilidade de 89,33%. Por fim, os autores treinaram o modelo utilizando tanto as características espaciais-temporais quanto as morfológicas, resultando em uma acurácia de 93,33%, com especificidade de 100% e sensibilidade de 86,67%.

Com uma abordagem diferente, a pesquisa realizada por Chy e Rahaman (2019), fez um estudo comparativo entre três algoritmos de classificação: *K-Nearest Neighbors*, *Support Vector Machine* e *Extreme Learning Machine*, para detectar a presença ou ausência de anemia falciforme. Foram utilizadas 80 imagens de esfregaço de sangue. O fluxo seguido pelos autores envolveu o processamento das imagens, incluindo a conversão das imagens para escala de cinza, filtragem de ruído e aprimoramento, segmentação usando a técnica de *fuzzy c* para separar os objetos do fundo. Em seguida, foram realizadas operações morfológicas e extração de características. Foram utilizados dois tipos de características: geométricas, que medem a circularidade das células, e estatísticas, como média, desvio padrão, variância, assimetria e curtose. Com base nas características extraídas, os autores treinaram os algoritmos de classificação e obtiveram os seguintes resultados: para o algoritmo KNN, precisão de 90%, sensibilidade de 75%, especificidade de 66,67%, acurácia de 73,33% e f-score de 81,81%; para o algoritmo SVM, precisão de 95,45%, sensibilidade de 83,33%, especificidade de 83,33%, acurácia de 83,33% e f-score de 88,89%; e para o algoritmo ELM, precisão de 95,45%, sensibilidade de 87,5%, especificidade de 83,33%, acurácia de 87,73% e f-score de 91,30%.

Corroborando com a ideia de ajustar os dados aos melhores parâmetros, Petrović et al.

(2020) seguiram tentando descobrir uma metodologia para selecionar o método de classificação e as características com melhor desempenho para suporte diagnóstico através de imagens de esfregaço de sangue periférico de hemácias. Os autores fizeram uso de imagens microscópicas obtidas no conjunto de dados público conhecido como erythrocytesIDB, as imagens consistem em amostras de pacientes com anemia falciforme classificadas do Hospital Geral em Santiago de Cuba. O trabalho seguiu três etapas principais em sistemas de processamento de imagem que são capazes de reconhecer diferentes classes de hemácias em imagens de esfregaço de sangue periférico: segmentação de imagem e detecção de células, extração e seleção de características e, finalmente, classificação. Para avaliar a eficácia dos diferentes classificadores, os pesquisadores utilizaram métricas como o *F-measure* e o *SDS-score* como indicadores da utilidade dos resultados para o diagnóstico de doença falciforme. Os dois classificadores mais promissores foram o GB, com *SDS-score* de 94,68%, *F-measure* de 94,67%, *Class Balance Accuracy* de 93,98% e Coeficiente de Correlação de Matthews de 88,72%, e o RF, com 94,44%, 94,42%, 93,66% e 88,19%, respectivamente.

Em um estudo conduzido por Haan et al. (2020), foi proposto um método de diagnóstico automatizado da doença falciforme usando imagens de esfregaço de sangue de 96 pacientes. Dentre esses pacientes, 32 foram diagnosticados com doença falciforme e 64 eram indivíduos saudáveis. As lâminas de sangue foram fotografadas utilizando um microscópio de smartphone, especificado no estudo, e foram obtidas no Hospital Ronald Reagan da UCLA. Os pacientes tanto com esfregaço de sangue normal quanto com DF incluíam uma mistura de gênero e uma ampla faixa etária. Os pesquisadores seguiram uma abordagem baseada em *deep learning* para processar as imagens, realizando etapas de aprimoramento, segmentação e classificação das células positivas para doença falciforme presentes nos esfregaços de sangue. A eficácia dos modelos foi avaliada com base na acurácia e na curva ROC (*Receiver Operating Characteristic*), obtendo-se uma acurácia de 98% e uma área de 0,998 na curva.

O estudo realizado por Tengshe et al. (2021) seguiu uma abordagem similar ao trabalho de Petrović et al. (2020), com o objetivo de classificar imagens de esfregaços de sangue

em três categorias: circular (normal), alongada (células falciformes) e outras. Os pesquisadores também utilizaram o conjunto de dados público erythrocytesIDB, que continha 629 imagens. Além disso, eles utilizaram dados sintéticos para aumentar a quantidade de dados disponíveis. O processo metodológico consistiu em adquirir as imagens e realizar o pré-processamento, que incluiu a aplicação de um filtro mediano que não apresentou resultados promissores. Em seguida, as imagens foram convertidas para escala de cinza e teve o equalizador de histograma aplicado. Posteriormente, os dados sintéticos foram incorporados para o treinamento do modelo de classificação, que utilizou uma rede neural convolucional composta por 5 camadas de convolução. A métrica de avaliação utilizada foi a acurácia de teste, e o estudo obteve um resultado de 94.57% de acurácia utilizando a CNN desenvolvida.

Ainda utilizando redes neurais profundas, Sen et al. (2021), realizaram testes de classificação em imagens de esfregaços de sangue utilizando o conjunto de dados do erythrocytesIDB, que consiste em três classes. Foram utilizadas 400 imagens para treinamento e 127 imagens para teste. Os autores empregaram diferentes classificadores, nomeadamente Resnet101, Resnet50, VGG 16, VGG19 e Inception V3, e obtiveram os seguintes resultados de acurácia durante o treinamento: Resnet101 com acurácia de 42,5%, Resnet50 com acurácia de 40,7%, VGG 16 com acurácia de 63,5%, VGG19 com acurácia de 66,5% e Inception V3 com acurácia de 91%.

No trabalho conduzido por Abdulhay, Allow e Al-Jalouly (2021), foi aplicada a aprendizagem de máquina para classificar imagens de microscópio em cinco categorias: hemácias normais, de pacientes com malária, DF, anemia megaloblástica ou talassemia. O conjunto de dados utilizado consistiu em 1815 imagens, extraídas do *kaggle* e coletadas pelos próprios autores, representando as cinco patologias. Os pesquisadores optaram pelo uso do classificador *Convolutional Neural Networks*, que foi implementado para processar as imagens sem a necessidade do protocolo padrão de teste de hemograma completo. Essa escolha foi justificada pelo benefício oferecido pela CNN de reduzir o número de parâmetros, especialmente porque eles são dependentes de aspectos espaciais nas imagens. Durante a fase de teste, a acurácia alcançada foi de 93,4%.



Em outro estudo, os pesquisadores propuseram o uso de um algoritmo de aprendizado de máquina para detectar a sobreposição de hemácias no diagnóstico da doença falciforme. Para isso, foram utilizadas 1.000 imagens de esfregaço de sangue obtidas da biblioteca digital do atlas hematológico. A técnica de pré-processamento de dados utilizada foi a limiarização, que teve como objetivo remover os componentes sanguíneos indesejados. Em seguida, foi aplicada a detecção de borda inteligente, que é uma técnica de segmentação de imagem baseada em aprendizado de máquina, para extrair biomarcadores utilizados na classificação dos resultados. Na fase de classificação, os autores utilizaram os classificadores *Support Vector Machine*, *Naïve Bayes* e *Logistic Regression* de aprendizado de máquina para determinar se um esfregaço sanguíneo era normal ou apresentava alguma alteração morfológica. Os resultados obtidos indicaram que o melhor algoritmo alcançou uma acurácia geral de 98,18%, sensibilidade de 98,29% e especificidade de 97,98% (VICENT; SIMON; YONASI, 2022).

Diferentemente dos estudos anteriores que se concentraram na classificação de células normais versus anormais, o estudo realizado por Soni, Thakkar e Singh (2022) adotou uma abordagem de aprendizado de transferência utilizando o modelo AlexNet, uma rede neural convolucional pré-treinada, para a classificação de casos de doença falciforme versus traços falciformes. Os autores utilizaram um conjunto de dados composto por 67 imagens de indivíduos com traço falciforme e 23 pacientes com doença falciforme. Para aumentar a quantidade de dados disponíveis para treinamento e validação, eles aplicaram técnicas como flip, rotação, redimensionamento e translação nas imagens. Como métricas de avaliação, os autores relataram uma taxa de acurácia de classificação de 95,5% utilizando um protocolo de divisão dos dados em 10 partes. Além disso, eles obtiveram uma sensibilidade de 94% e uma precisão de 95%.

### 3.2 ABORDAGENS NO MONITORAMENTO

As possibilidades de aplicação dos classificadores foram ampliadas quando se tratou de monitoramento, resultando em uma variedade de estudos que exploram o uso desses algoritmos em áreas como dosagem de medicação, classificação da dor e predição de

desfechos. A Tabela 3.2 resume os principais algoritmos encontrados nos últimos anos, abordando esses aspectos.

Na tentativa de utilizar dados obtidos pela técnica de fotopletismografia (PPG) durante o sono de pacientes com doença falciforme para prever um aumento na propensão a crises de dor, Ji et al. (2021) adotaram uma abordagem de aprendizado de máquina, utilizando um conjunto de dados derivado do estudo multicêntrico *Sleep and Asthma Cohort* (SAC), que envolve crianças e adolescentes. O conjunto de dados SAC é composto por 212 indivíduos, sendo 36 pertencentes à categoria positiva (“dor alta”) e 176 pertencentes à categoria negativa (“dor baixa”). Os autores utilizaram um modelo de empilhamento de dois níveis de aprendizado de máquina, que incorpora características detalhadas extraídas dos sinais PPG. No primeiro nível, foram empregados diferentes algoritmos de aprendizado de máquina para prever a categoria de dor de cada indivíduo com base nas características PPG e outras informações clínicas. No segundo nível, foi utilizado um metamodelo que utiliza as entradas do modelo de primeiro nível juntamente com as saídas do modelo para produzir a previsão final. Dentre os algoritmos utilizados no primeiro nível, foram empregados regressão logística múltipla, *Support Vector Machine* com *kernel* linear, SVM com *kernel* de função de base radial, *Random Forest*, árvores extremamente aleatórias (ExtraTrees), reforço adaptativo (AdaBoost), CNN unidimensional, *Multi-Layer Perceptron*, SVM com kernel polinomial, árvores de aumento de gradiente e árvores de aumento de gradiente extremo (XGBoost). O *Multi-Layer Perceptron* apresentou um desempenho substancialmente melhor em relação aos outros modelos base, sendo selecionado como o algoritmo representativo do “metamodelo” no segundo nível do modelo de empilhamento proposto. Como métricas de avaliação do modelo, foram utilizados o f1-score (0,4255), a precisão (0,2915), a sensibilidade (*recall*) (0,7917), a especificidade (0,6048) e a acurácia (0,6366).

Para lidar com a subjetividade da intensidade da dor em pacientes durante visitas hospitalares, Padhee et al. (2021) propuseram uma abordagem utilizando registros eletrônicos de saúde. O estudo utilizou dados coletados de 47 pacientes ao longo de cinco anos consecutivos no Hospital da Universidade Duke, totalizando 59.728 registros. Cada

Tabela 3.2: Principais algoritmos utilizados no monitoramento da doença falciforme.

<b>Autores</b>	<b>Algoritmos</b>	<b>Métrica</b>	<b>Desfecho</b>
(JI et al., 2021)	Multi-Layer Perceptron	Acurácia de 63,66%	Intensidade da dor
(PADHEE et al., 2021)	K-Nearest Neighbors, Support Vector Machine, Multinomial Logistic Regression, Decision Tree e Random Forest	Acurácia de 72,8%	Intensidade da dor
(PADHEE et al., 2022)	Random Forest, Regressão Lasso e Support Vector Machine	Acurácia de 82,3%	Intensidade da dor
(SINGH; THAKKAR, 2021)	Long short-term memory e Extreme Learning Machines	Acurácia de 78%	Dosagens de Hidroxiuréia
(PATEL et al., 2021)	Logistic Regression, Support Vector Machine e Random Forest	Estatística C de 0.77	Readmissão hospitalar
(PADRÃO et al., 2022)	Aprendizado de máquina não supervisionada de clusterização	Não se aplica	Admissões na UTI
(GÜNTÜRKÜN et al., 2021)	Não especificado	Acurácia de 80% para TFG	Declínio da função renal
(MOHAMMED et al., 2020)	Multi-Layer Perceptron, Support Vector Machine, Random Forest e Logistic Regression	Sensibilidade de 96% e Especificidade de 98%	Falência de órgão
(SACHDEV et al., 2021)	Random Forest	Estatística C corrigida por viés de 0,76	Mortalidade

registro continha medidas de seis sinais vitais: saturação capilar periférica de oxigênio (SpO<sub>2</sub>), pressão arterial sistólica (PA sistólica), pressão arterial diastólica (PA diastólica), frequência cardíaca (Pulso), frequência respiratória (Resp) e temperatura (Temp). Além dos sinais vitais, cada registro também continha o escore de dor autorrelatado pelo paciente, variando de 0 (sem dor) a 10 (dor intensa e insuportável). Os autores aplicaram cinco algoritmos de classificação supervisionada de aprendizado de máquina - *k-Nearest Neighbors*, *Support Vector Machine*, *Multinomial Logistic Regression*, *Decision Tree* e *Random Forest* - para prever os escores de dor dos pacientes com base nos sinais vitais. A métrica utilizada para avaliar os algoritmos foi a acurácia. O modelo de *Decision Tree* apresentou o melhor desempenho na previsão da dor em uma escala de gravidade de 0 a 10, com uma acurácia de 0,728 em nível interindividual (entre os pacientes) e 0,653 em nível intraindividual (dentro de cada paciente). A acurácia do modelo DT melhorou significativamente para 0,941 quando a escala de classificação foi reduzida para dois pontos (ou seja, sem dor/dor leve: 0-5, dor intensa: 6-10) em nível interindividual.

Na tentativa de melhorar a predição da intensidade da dor em pacientes durante visitas hospitalares, os mesmos autores Padhee et al. (2022) exploraram informações sobre os medicamentos administrados a pacientes com doença falciforme. O estudo utilizou uma base de dados composta por 126.519 registros de prontuários eletrônicos pessoais (EHR) de 496 pacientes com SCD, coletados ao longo de 5 anos consecutivos no Hospital da Universidade Duke. No entanto, apenas 33.000 registros foram incluídos devido a informações ausentes nos prontuários. Os sinais vitais foram registrados pela equipe médica a cada 4 horas durante a internação, enquanto os dados de medicação foram registrados conforme eram administrados aos pacientes. Isso incluiu informações como o tipo de medicamento, status da medicação (administrada, perdida, removida ou vencida) e dosagem total (mg/mL). A abordagem de aprendizado de máquina foi aplicada para imputar os dados ausentes nos prontuários médicos, utilizando redes neurais profundas de autoencoder. Em seguida, técnicas de aprendizado supervisionado, como *Random Forest*, regressão Lasso e *Support Vector Machine*, foram empregadas para prever a intensidade da dor com base nos dados disponíveis. A métrica de acurácia foi utilizada para avaliar o

desempenho do modelo. Os resultados mostraram a importância das informações sobre medicamentos, uma vez que o novo modelo alcançou uma acurácia de 82,3%, representando uma melhoria em relação ao modelo anteriormente utilizado pelos autores. Esses achados destacam a importância de uma coorte de dados mais ampla e o uso de técnicas de aprendizado de máquina representacional profunda para aprimorar a precisão da predição da intensidade da dor em pacientes durante visitas hospitalares.

Com o objetivo de aliviar os episódios de dor em pacientes com doença falciforme que estão em tratamento com hidroxureia (HU), Singh e Thakkar (2021) conduziram um estudo utilizando um conjunto de dados composto por 1128 amostras obtidas de um hospital localizado em Liverpool, Reino Unido. O objetivo principal do estudo foi comparar o desempenho de dois modelos de classificação para categorizar as dosagens de HU em faixas baixa, alta e moderada para pacientes com doença falciforme. Os autores avaliaram 12 parâmetros como entrada para os modelos de aprendizado de máquina. Esses parâmetros incluíam o peso do paciente em quilogramas, atividade da enzima aspartato aminotransferase, atividade da enzima alanina aminotransferase, atividade da enzima lactato desidrogenase, volume corpuscular médio (MCV), níveis de bilirrubina, níveis de hemoglobina, níveis de hemoglobina fetal, contagem de neutrófilos, contagem de reticulócitos (RETIC A e % RETIC) e contagem de plaquetas. No estudo, foram utilizados os modelos Long short-term memory e Extreme learning machines para realizar a classificação das dosagens de HU. Os resultados mostraram que o modelo LSTM alcançou uma acurácia geral de 78%, enquanto o modelo ELM obteve uma acurácia de 52%.

Em um estudo retrospectivo utilizando registros eletrônicos de saúde, Patel et al. (2021) investigaram a capacidade dos algoritmos de aprendizado de máquina superarem os sistemas padrão de pontuação de readmissão hospitalar, como os índices LACE e HOSPITAL. O estudo incluiu uma amostra de 446 pacientes adultos com doença falciforme e 3.299 hospitalizações não planejadas. Três algoritmos de aprendizado de máquina, *Logistic Regression*, *Support Vector Machine* e *Random Forest*, foram aplicados e avaliados pelos autores. As características consideradas incluíram dados laboratoriais, dados demográficos, número de consultas ambulatoriais anteriores à internação atual e número de

visitas ao Departamento de Emergência (PS) antes da hospitalização atual. Os dados foram obtidos a partir de registros de cinco hospitais pertencentes ao sistema hospitalar do Centro Médico da Universidade de Pittsburgh (UPMC). Os resultados obtidos no conjunto de dados mostraram que os algoritmos de aprendizado de máquina superaram os índices LACE (C-statistic 0,6, Intervalo de Confiança de 95% - IC 0,57–0,64) e HOSPITAL (C-statistic 0,69, IC 95% 0,66–0,72). O algoritmo de *Random Forest* apresentou uma C-statistic de 0,77 (IC 95% 0,73–0,79), enquanto a *Logistic Regression* obteve uma C-statistic de 0,77 (IC 95% 0,73–0,8), ambos demonstrando um desempenho superior.

O estudo realizado por Padrão et al. (2022) teve como objetivo analisar o perfil das admissões em unidades de terapia intensiva (UTI) de pacientes com doença falciforme. Para isso, os autores examinaram 125 registros de pacientes obtidos dos prontuários de todas as internações entre 1996 e 2020 na UTI do Hospital das Clínicas de São Paulo. Utilizando técnicas de aprendizado de máquina não supervisionado de clusterização, os autores agruparam os pacientes com base em características sem hipóteses prévias, com o objetivo de identificar grupos relacionados à gravidade, prognóstico ou nível de cuidado necessário. Esses agrupamentos são importantes para melhorar o atendimento ao paciente e auxiliar o sistema de saúde na alocação adequada de leitos de UTI para os casos mais graves. A partir do gráfico de silhueta, o número ótimo de clusters identificado foi três. O Cluster 1 foi o maior ( $n = 69$ ), apresentando a maior proporção do genótipo SS (87%). Esse cluster teve a menor prevalência de comorbidades como insuficiência cardíaca, doença renal crônica e vírus da hepatite C, mas uma alta proporção de pacientes com histórico de acidente vascular cerebral. Na admissão na UTI, esse cluster apresentou a maior prevalência de motivos relacionados à doença falciforme (72,5%) e disfunção hepática (76,8%). Uma baixa proporção de pacientes necessitou de ventilação mecânica invasiva (11,6%), e a mortalidade intra-hospitalar foi de 18,8%. O Cluster 2 ( $n = 25$ ) foi o mais grave e com pacientes mais velhos (idade média de 37 anos). A maioria era composta por homens (76%), com alta prevalência do genótipo SS (80%). Esse cluster apresentou a maior prevalência de hipertensão e doença renal crônica. Na admissão na UTI, esse cluster também apresentou os maiores escores de gravidade, além de alta prevalência de

disfunção renal e neurológica. Apenas 3 (12%) dos pacientes foram admitidos por motivos relacionados à doença falciforme. A maioria dos pacientes em ventilação mecânica estava nesse cluster, e a mortalidade hospitalar também foi alta (44%). Dos 11 óbitos ocorridos no Cluster 2, oito pacientes estavam em ventilação mecânica. Por fim, o Cluster 3 ( $n = 31$ ) consistia principalmente de mulheres (96,8%) com genótipo heterozigoto S/ $\beta$  (64,5%). Dentre esses pacientes, 19 (61,3%) tiveram internações relacionadas à doença falciforme. As comorbidades mais comuns nesse cluster foram insuficiência cardíaca (45,2%), acidente vascular cerebral (35,5%) e hepatite C crônica (41,9%). A mortalidade foi a mais baixa (12,9%). Como resultado, os autores concluíram que foram identificados com sucesso três clusters de doença falciforme capazes de distinguir claramente o grupo mais grave (Cluster 2), um grupo intermediário (Cluster 1) e um grupo menos grave (Cluster 3).

Explorou-se o potencial do uso de aprendizado de máquina para prever o declínio rápido da função renal em pacientes com doença falciforme (GÜNTÜRKÜN et al., 2021). A hipótese levantada foi de que os modelos poderiam prever o declínio rápido da função renal em genótipos graves de SCD com antecedência de 6 meses a 1 ano. A base de dados utilizada consistiu em um coorte interno com 236 pacientes e um coorte externo com 168 pacientes (HbSS). Os dados de entrada para os algoritmos incluíram informações demográficas, resultados de exames laboratoriais, complicações relacionadas à SCD, comorbidades e tratamentos. Em cada visita, a taxa de filtração glomerular (TFG) foi estimada utilizando a equação CKD-EPI (Colaboração de Epidemiologia de Doença Renal Crônica). Embora os autores não tenham especificado quais algoritmos foram utilizados no estudo, os resultados mostraram que o declínio estimado da  $TFG > 3 \text{ mL/min/1,73 m}^2$  em 6 meses foi previsto com sensibilidade de 82%, acurácia de 80% e AUC de 0,88 (intervalo de confiança [IC] de 95%: 0,79–0,97). Além disso, o declínio de  $eGFR > 5 \text{ mL/min/1,73 m}^2$  em 6 meses foi previsto com sensibilidade de 79%, precisão de 84% e AUC de 0,91 (IC de 95%: 0,85–0,97). Quanto ao declínio estimado da TFG em 12 meses, a sensibilidade foi de 53%, acurácia de 70% e AUC de 0,67 (IC de 95%: 0,53–0,81) com o limite  $> 3 \text{ mL/min/1,73 m}^2$ , e sensibilidade de 64%, precisão de 65% e AUC de 0,77 (IC

de 95%: 0,67–0,87) com o limite  $> 5 \text{ mL/min/1,73 m}^2$  para pacientes do coorte interno. Os autores concluíram que são necessários mais estudos para avaliar o uso de algoritmos de aprendizado de máquina na doença renal relacionada à doença falciforme.

O estudo conduzido por Mohammed et al. (2020) teve como objetivo principal a detecção precoce da falência de órgãos em pacientes adultos com doença falciforme. A pesquisa foi realizada em unidades de terapia intensiva (UTIs) de hospitais Methodist Le Bonheur Healthcare, localizados em Memphis, Tennessee, Estados Unidos, com a participação de 63 pacientes e a análise de 163 eventos. Durante o estudo, foram consideradas cinco características fisiológicas dos pacientes: frequência cardíaca, frequência respiratória, pressão arterial sistólica, pressão arterial diastólica e pressão arterial média. Esses dados foram utilizados como entrada para quatro algoritmos de aprendizado de máquina: *Multilayer perceptron*, *Support Vector Machine*, *Random Forest* e métodos de *Logistic Regression*. Para avaliação dos algoritmos, foram utilizadas as métricas de sensibilidade e especificidade. O MLP apresentou os melhores resultados, sendo capaz de prever a falência de órgãos com uma antecedência de até 6 horas em relação ao início da ocorrência. O MLP obteve médias de sensibilidade de 96% e especificidade de 98%.

Sachdev et al. (2021) propuseram um novo escore de risco que combina dados clínicos, laboratoriais e de imagem para prever a mortalidade em pacientes com doença falciforme. O estudo envolveu a análise de 600 pacientes adultos com essa patologia, e os dados foram obtidos a partir de pacientes matriculados no National Heart, Lung, and Blood Institute (NHLBI). As características utilizadas no estudo incluíram a velocidade regurgitante tricúspide, pressão atrial direita estimada, velocidade E mitral, espessura do septo ventricular esquerdo, índice de massa corporal, nitrogênio ureico no sangue, fosfatase alcalina, frequência cardíaca e idade dos pacientes. Para lidar com valores ausentes nas co-variáveis, os autores aplicaram *Random Survival Forest* (RSF) como método de imputação e para fornecer a classificação das variáveis com maior poder preditivo. Posteriormente, os autores utilizaram a regressão de riscos proporcionais de Cox regularizada para analisar os preditores de maior importância selecionados pelo RSF, com o objetivo de desenvolver o melhor modelo preditivo e escore de risco. Os resultados demonstra-



ram que o escore de risco prognóstico proposto teve um desempenho superior, com uma estatística C corrigida por viés de 0,76. Além disso, o modelo foi capaz de estratificar os pacientes em quatro grupos com taxas de mortalidade em 4 anos significativamente diferentes, sendo 3%, 11%, 35% e 75% respectivamente.

## **MATERIAL E MÉTODOS**

### **4.1 COLETA E DESCRIÇÃO DOS DADOS**

A base de dados utilizada neste estudo foi composta por informações provenientes de dois projetos de pesquisa: “Biomarcadores de lesão renal em indivíduos com doença falciforme” do Laboratório em Anemias da Faculdade de Farmácia da UFBA (Universidade Federal da Bahia) e “Proposição de score epidemiológico, clínico e laboratorial preditivo de gravidade dos principais desfechos clínicos da anemia falciforme” do Laboratório de Investigação em Genética e Hematologia Translacional do Instituto Gonçalo Moniz - FIOCRUZ (Fundação Oswaldo Cruz - Bahia). Esses projetos foram aprovados pelo Comitê de Ética em Pesquisa sob os registros 063/2011 (CEP COM-UFBA) e 29239220.8.0000.0040 (CEP FIOCRUZ-BA), respectivamente. A base de dados possui 609 registros de pacientes com doença falciforme, cada registro contém informações de 63 características distintas, após o devido tratamento e organização dos dados.

### **4.2 DEFINIÇÃO E CATEGORIZAÇÃO DAS VARIÁVEIS**

As variáveis dependentes representam os desfechos da doença falciforme que são de interesse principal nesta pesquisa. Esses desfechos incluíram vaso-oclusão, crises de dor,

hospitalizações e infecções. Cada desfecho foi considerado individualmente como variável dependente, buscando compreender sua ocorrência e os fatores associados a ele.

As variáveis independentes, por sua vez, representam os fatores e características que podem influenciar os desfechos da doença falciforme. Elas podem incluir informações demográficas (como idade e gênero), parâmetros laboratoriais (como concentração de hemoglobina total e hemoglobina S), características clínicas (como presença de outras condições de saúde), entre outros. A Tabela 4.1 mostra todas as variáveis independentes consideradas no estudo.

Tabela 4.1: Relatório de estatística descritiva da base de dados

Variável	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Idade	13,27	10,54	11,00	1,00	63,00
HM (106/mL)	3,29	0,92	3,11	1,50	6,31
HEMOG (g/dL)	9,48	1,90	9,40	4,60	15,50
HT (%)	27,56	5,77	27,10	14,00	47,00
VCM (fL)	85,82	10,25	85,50	50,40	117,50
HCM (pg)	29,65	4,02	29,50	16,50	41,40
CHCM (%)	34,48	1,50	34,48	22,30	40,10
LEUCO (/mL)	11280,56	4125,50	11000,00	11,50	28000,00
BAST (/mL)	11,10	50,89	0,00	0,00	640,00
SEG (/mL)	5579,67	2433,14	5579,67	599,00	21000,00
EOS (/mL)	693,35	649,27	490,00	0,00	4437,00
BAS (/mL)	85,76	79,77	85,76	0,00	476,00
LIT (/mL)	4065,75	1830,10	4065,75	2,86	12597,00
LIA (/mL)	59,53	117,42	0,00	0,00	834,00
MON (/mL)	778,45	403,89	720,00	63,60	3186,00
PLA (103/mL)	380,50	149,88	371,00	102,00	942,00
RET (%)	5,92	2,68	5,90	0,80	18,00
GLI (mg/dL)	77,14	17,50	76,00	50,00	400,00
COLTOT (mg/dL)	131,43	29,69	129,00	24,00	295,00

Variável	Média	Desvio Padrão	Mediana	Mínimo	Máximo
COLHDL (mg/dL)	34,96	9,87	34,00	7,00	78,00
COLLDL (mg/dL)	77,73	24,41	76,00	13,80	174,20
COLVLDL (mg/dL)	18,34	9,64	16,60	2,40	79,20
TRIG (mg/dL)	92,25	50,63	83,00	25,00	486,00
AURICO (mg/dL)	4,07	0,93	4,07	1,20	10,60
UREIA (mg/dL)	18,37	7,80	17,00	0,30	79,00
CREA (mg/dL)	0,56	0,73	0,50	0,10	17,00
FERRO (mg/dL)	88,64	45,79	80,00	8,30	403,00
AST (U/L)	44,67	19,83	44,00	8,00	213,00
ALT (U/L)	22,13	14,70	21,00	3,00	197,00
GGT (U/L)	29,39	18,62	29,39	5,00	309,00
BT (mg/dL)	1,93	1,23	1,90	0,17	7,70
BD (mg/dL)	0,48	0,28	0,48	0,00	2,90
BI (mg/dL)	1,44	1,12	1,30	0,04	7,10
FAL (U/L)	173,41	31,67	173,41	40,00	368,00
PT (g/dL)	7,78	0,71	7,78	5,00	10,60
ALBUM (g/dL)	4,38	0,39	4,38	1,40	7,40
GLOB (g/dL)	3,39	0,71	3,39	0,40	5,90
RELAG	1,41	0,79	1,41	0,50	18,50
LDH (U/L)	927,80	505,19	821,00	115,00	4394,00
FERRIT (ng/mL)	285,55	385,07	177,70	9,20	4537,00
ASLO (UI/mL)	177,68	258,01	177,68	25,00	4310,00
PCR (mg/L)	6,14	8,22	4,50	1,00	88,50
HAPTO (mg/dL)	8,40	13,27	5,83	5,83	215,00
A1ATS (mg/dL)	153,93	35,22	153,93	29,00	332,00
HBS (%)	71,06	17,21	75,60	24,70	95,40
HBA2 (%)	4,23	1,34	4,00	0,80	17,10
HBF (%)	7,29	6,32	5,90	0,20	39,50

Variável	Média	Desvio Padrão	Mediana	Mínimo	Máximo
HBA1 (%)	9,65	8,53	9,65	0,00	69,50
HBC (%)	31,54	12,34	31,54	0,00	45,80
HAPL	15,78	3,28	15,78	0,00	27,00
TAL	0,25	0,27	0,25	0,00	3,00
QCRISDOR	0,77	0,75	0,77	0,00	3,00
QVO	8,02	5,21	8,02	0,00	31,00

### 4.3 TREINAMENTO DO MODELO

O estudo abrangeu uma análise detalhada de um banco de dados com 609 registros de pacientes com a doença falciforme e 63 características dispostas nas colunas. Investigando os diversos desfechos relacionados à doença, como vaso-oclusão, crise de dor, hospitalização e infecção. Para garantir a qualidade dos dados e prepará-los para análise, todas as etapas importantes de pré-processamento foram aplicadas como indicado na Figura 4.1.

Inicialmente, características sem valor preditivo significativo foram excluídas, as variáveis que apresentaram mais de 25% de dados faltantes foram removidas, evitando distorções nos resultados. Após a exclusão e remoção de características e variáveis inadequadas, técnicas de imputação (mediana e moda) foram aplicadas para preencher os dados faltantes restantes. A normalização e padronização das variáveis foram realizadas, garantindo que estivessem em escalas comparáveis.

As variáveis mais relevantes e os algoritmos mais adequados para cada rótulo foram identificados utilizando o algoritmo de seleção de características *Recursive Feature Elimination* (RFE). Esse método seleciona recursivamente os atributos mais importantes, refinando o conjunto de treinamento do modelo e definindo as variáveis de entrada (ver Figura 4.2). Além disso, o *Grid Search* foi empregado para determinar os melhores parâmetros de configuração dos modelos. Os algoritmos selecionados serviram de base para a etapa de classificação multirrótulo, na qual foi adotada a estratégia de transformação do problema. Nessa etapa, utilizamos os classificadores *Binary Relevance*, *Classifier Chain*,

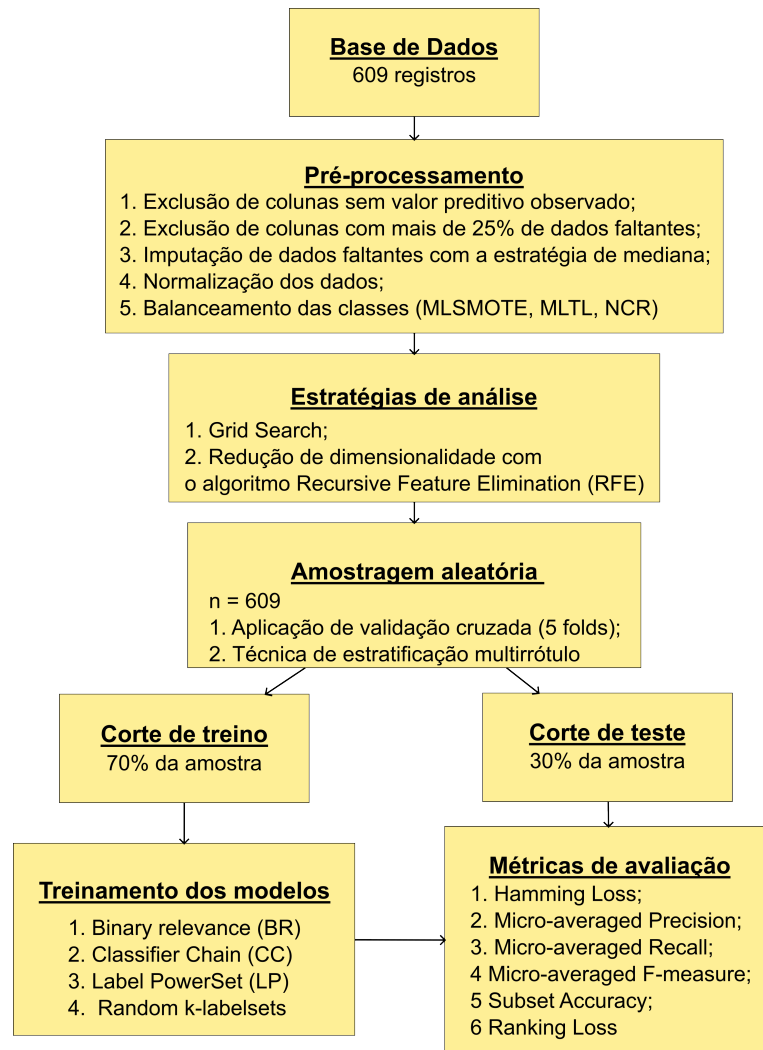


Figura 4.1: Pré-processamento dos dados.

*Label PowerSet* e *RAkELo* para explorar relações e associações entre os rótulos da doença falciforme.

Em seguida, dividimos os dados em 70% para treinamento e 30% para teste, garantindo uma distribuição estratificada que manteve as mesmas relações de rótulos em ambos os conjuntos. Aplicamos validação cruzada com 5 folds para ajustar os modelos, garantindo que todas as sessões de melhor desempenho fossem treinadas.

Validamos e avaliamos os resultados usando o conjunto de testes, que compreendeu 30% do total de dados, empregando diversas métricas para avaliar o desempenho dos modelos multirrótulo. Após a avaliação inicial das frequências de cada rótulo (ver Figura

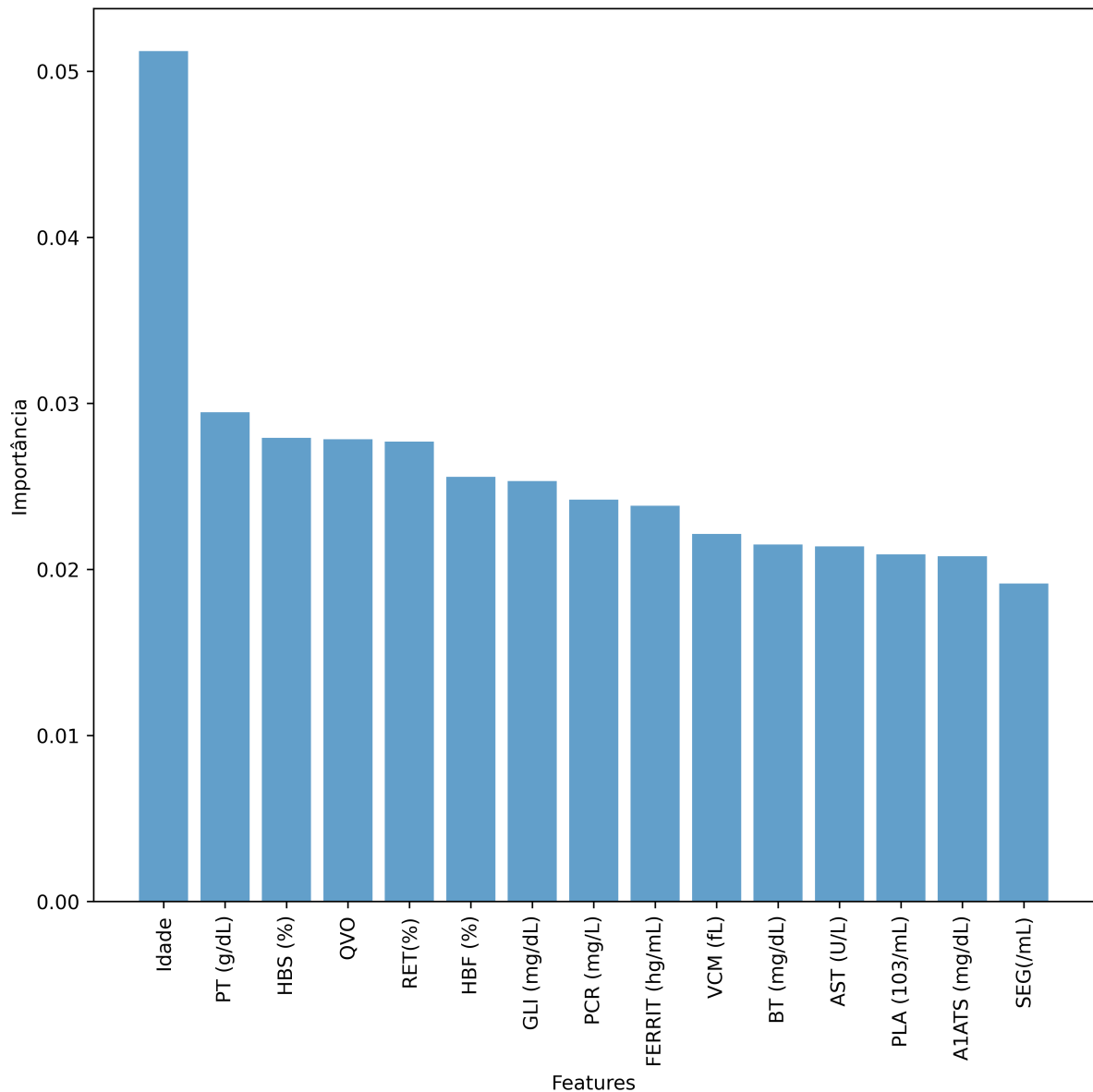


Figura 4.2: Features selecionadas na função Recursive Feature Eliminação e sua importância para o modelo.

4.3), aplicamos técnicas de balanceamento de rótulo aos dados de treinamento, treinamos novamente nossos modelos e balanceamos as ocorrências de rótulo usando métodos de sobre-amostragem MLSMOTE ou subamostragem MLTL (PEREIRA; COSTA; JR, 2020) para evitar sobre-ajuste das classes majoritárias. A principal decisão que precisa ser tomada nessa etapa é a de aumentar as classes minoritárias adicionando dados sintéticos ou reduzir as classes majoritárias reduzindo nosso conjunto de dados.

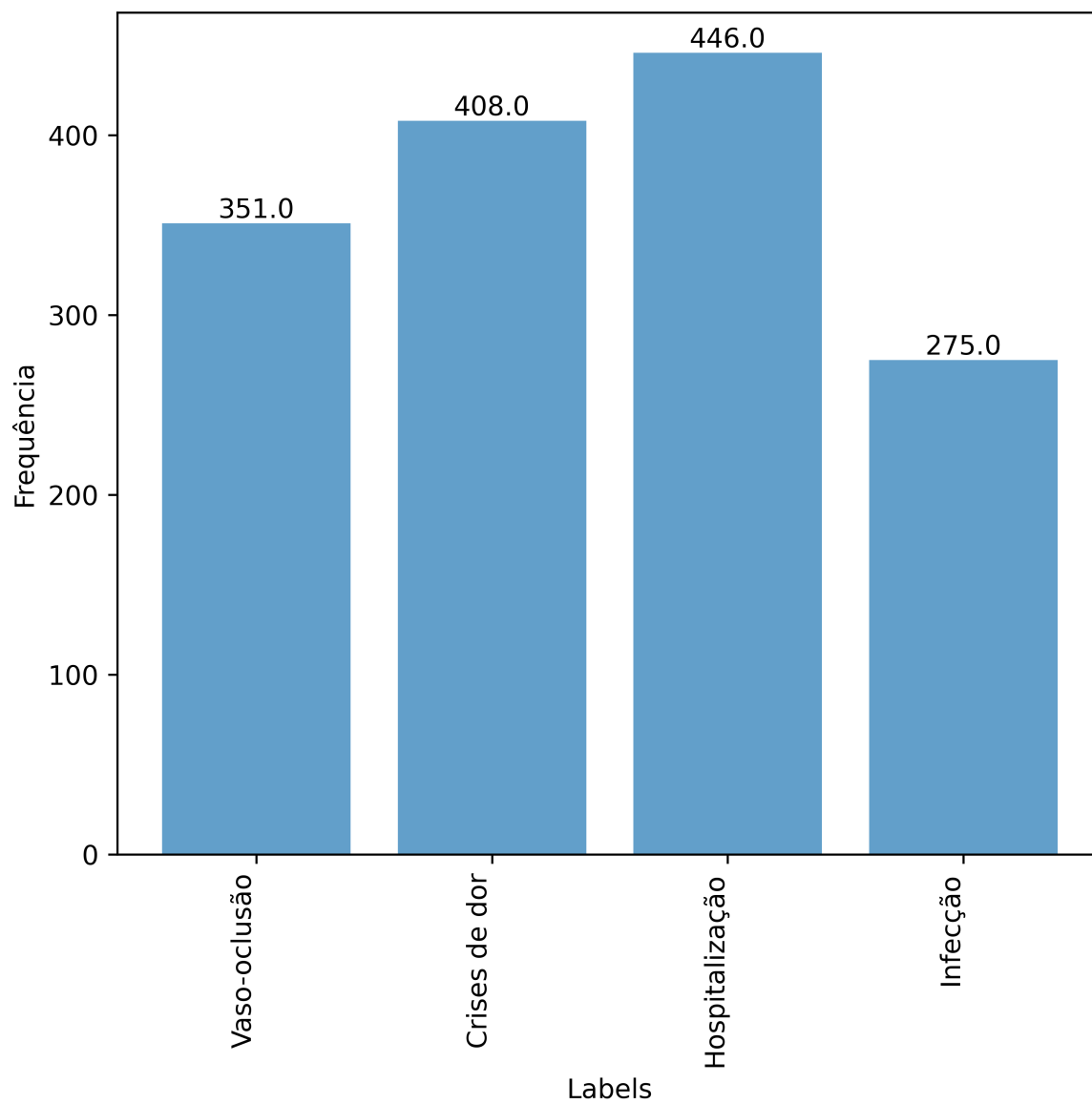


Figura 4.3: Frequência dos rótulos do conjunto de dados.



Após a aplicação dos classificadores multirrótulo aos dados dos pacientes com doença falciforme para treinamento, a validação e avaliação dos resultados foram realizadas no conjunto de dados de teste. O parâmetro `classifier`, essencial nesse processo, recebe um estimador (`BaseEstimator`), possibilitando o uso de modelos como `RandomForestClassifier`, `SVC`, `DT`, entre outros. Dessa forma, a estrutura do pipeline de classificação multirrótulo mantém a flexibilidade e integração com técnicas convencionais de aprendizado supervisionado.

Ao contrário dos modelos tradicionais de rótulo único, onde a maioria dos estudos se concentra em detalhes de acurácia, precisão, sensibilidade e especificidade, na classificação multirrótulo, as medidas de desempenho precisam ser diferentes, visto a especificidade do paradigma multirrótulo (GARCÍA-PEDRAJAS et al., 2024). Pensando nisso, de acordo com (PEREIRA et al., 2018) as métricas de avaliação dos modelos de rótulo único têm sido aplicadas arbitrariamente em experimentos de classificação multirrótulo sem uma análise objetiva que considera as correlações. Isso pode resultar em conclusões errôneas, pois os resultados dos experimentos podem favorecer um comportamento específico dependendo das métricas selecionadas. Além disso, a diversidade de subconjuntos de métricas utilizadas em diferentes estudos torna difícil comparar diretamente os resultados entre eles.

Com esta abordagem em mente, selecionamos um conjunto de métricas não correlacionadas para permitir uma avaliação completa dos modelos utilizados no estudo, nomeadamente *Hamming Loss*, *Micro-averaged Precision*, *Micro-averaged Recall*, *Micro-averaged F-measure*, *Subset Accuracy*, e *Ranking Loss*.

## RESULTADOS

Dividimos os experimentos em duas fases, para responder às perguntas de pesquisa. Na primeira fase buscamos determinar os melhores algoritmos de classificação para cada desfecho, tendo o Random Forest melhores métricas frente ao SVM, k-NN, e DT. Seguindo de uma quantidade de características suficientemente relevantes para a classificação dos rótulos, nesse primeiro momento utilizamos a configuração padrão do RFE. Por último procuramos identificar as variáveis mais importantes de forma individual, no estado da arte. Métricas como precisão, recall, f1-score e acurácia foram utilizadas para avaliar o desempenho dos classificadores nessa primeira fase dos experimentos.

A Figura 4.3 apresenta o gráfico com as ocorrências dos rótulos na base de dados, evidenciando o desbalanceamento das classes, e a ocorrência simultânea dos rótulos. Consideramos os quatro desfechos mais comuns: vaso-oclusão, crises de dor, hospitalização e infecções. No caso do desfecho de vaso-oclusão, o Random Forest com o hiper-parâmetro `max_depth=10` demonstrou o melhor desempenho entre os algoritmos avaliados (Random Forest, Support Vector Machine, K-Nearest Neighbors, Decision Tree).

Selecionamos 27 características relevantes para a classificação desse desfecho específico, através do algoritmo de *Recursive Feature Elimination* que faz uma busca gulosa das melhores variáveis. Essas características incluem variáveis como 'HM (106/mL)', 'HT (%)', 'VCM (fL)', 'CHCM (%)', 'SEG(/mL)', 'BAS (/mL)', 'RET(%)', 'GLI (mg/dL)',

'COLHDL (mg/dL)', 'COLLDL (mg/dL)', 'AURICO(mg/dL)', 'UREIA (mg/dL)', 'CREA (mg/dL)', 'FERRO (mg/dL)', 'ALT (U/L)', 'BD (mg/dL)', 'PT (g/dL)', 'RELAG', 'ASLO (UI/mL)', 'PCR (mg/L)', 'HAPTO (mg/dL)', 'GENERO', 'HBS (%)', 'HBA2 (%)', 'HBF (%)', 'HAPL' e 'QVO'.

*Legenda: Hemácias (milhões/uL) (HM), Hematócrito (%) (HT), Volume Corpuscular Médio (fL) (VCM), Concentração de Hemoglobina Corpuscular Média (g/dL) (CHCM), Segmentados Neutrófilos (mil/uL) (SEG), Basófilos (mil/uL) (BAS), Reticulócitos (%) (RET), Glicemia jejum (mg/dL) (GLI), Colesterol HDL (mg/dL) (COLHDL), Colesterol LDL (mg/dL) (COLLDL), Ácido úrico (mg/dL) (AURICO), Ureia (mg/dL) (UREIA), Creatinina (mg/dL) (CREA), Ferro sérico (mg/dL) (FERRO), Alanina aminotransferase (U/L) (ALT), Bilirrubina Direta (mg/dL) (BD), Proteínas Totais (g/dL) (PT), Relação Albumina/globulinas (g/dL) (RELAG), Antiestreptolisina O (UI/mL) (ASLO), Proteína C Reativa (mg/L) (PCR), Haptoglobina (mg/dL) (HAPTO), Gênero, Hemoglobina S (%) (HBS), Hemoglobina A2 (%) (HBA2), Hemoglobina Fetal (%) (HBF), Haplótipos ligados ao gene da globina beta (HAPL) e Quantidade de ocorrências de vaso-oclusão (QVO).*

A Figura 5.1 apresenta a matriz de confusão, e as métricas do estudo, como acurácia de 95%, precisão, recall e f1-score são observadas na Tabela 5.1. Essas medidas fornecem uma visão do desempenho do modelo para cada desfecho avaliado.

Tabela 5.1: Relatório de classificação do Random Forest para Vaso-oclusão.

	precision	recall	f1-score	sample
0.0	0.92	0.99	0.96	97
1.0	0.99	0.92	0.95	97
accuracy			0.95	194
macro avg	0.96	0.95	0.95	194
weighted avg	0.96	0.95	0.95	194

Na análise do rótulo crises de dor, o melhor classificador foi o Random Forest com hiper-parâmetro `n_estimators=150`, sendo selecionadas 27 características relevantes para classificação do desfecho, entre elas 'Idade', 'HM (106/mL)', 'LEUCO (/mL)', 'BAST

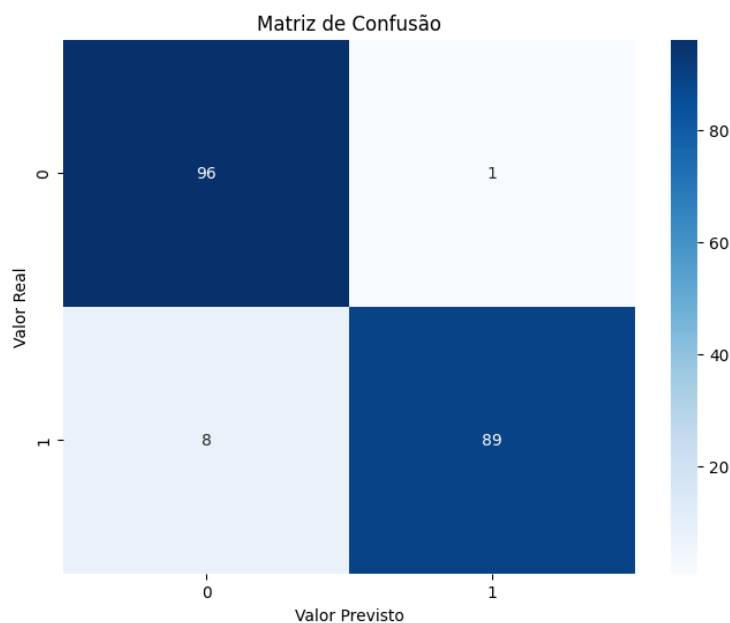


Figura 5.1: Matriz de confusão do rótulo Vaso-oclusão.

(/mL)', 'SEG(/mL)', 'BAS (/mL)', 'LIA (/mL)', 'PLA (103/mL)', 'RET(%)', 'COLVLDL (mg/dL)', 'TRIG (mg/dL)', 'UREIA (mg/dL)', 'GGT (U/L)', 'PT (g/dL)', 'ALBUM (g/dL)', 'RELAG', 'LDH (U/L)', 'FERRIT (hg/mL)', 'PCR (mg/L)', 'HAPTO (mg/dL)', 'HBS (%)', 'HBA2 (%)', 'HBC (%)', 'HAPL', 'TAL', 'QCRISDOR', 'QVO'. Os resultados são apresentados na matriz de confusão na Figura 5.2 e as métricas são exibidas na Tabela 5.2, com acurácia de 98%.

*Legenda: Idade, HM, Contagem Global Leucócitos (mil/uL) (LEUCO), Bastonetes Neutrófilos (mil/uL) (BAST), SEG, BAS, Linfócitos Atípicos (mil/uL) (LIA), Plaquetas (mil/uL) (PLA), RET, Colesterol VLDL (mg/dL) (COLVLDL), Triglicérides Colesterol HDL (mg/dL) (TRIG), UREIA, Gama Glutamil Transferase (U/L) (GGT), PT, Albumina (g/dL) (ALBUM), RELAG, Lactato Desidrogenase (U/L) (LDH), Ferritina (ng/mL) (FERRIT), PCR, HAPTO, HBS, HBA2, Hemoglobina C (%) (HBC), HAPL, Talassemia alfa (TAL), Quantidade de ocorrências de crises de dor nos últimos 6 meses (QCRISDOR) e QVO.*

Na análise do rótulo de hospitalizações, o melhor classificador foi o RF com hiperparâmetros `min_samples_split=5`, `n_estimators=150`, sendo selecionadas 27 caracterís-

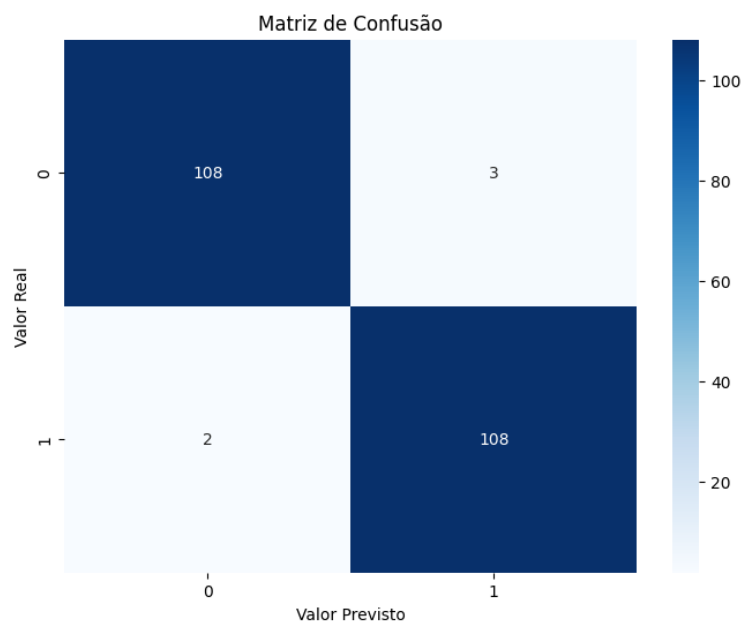


Figura 5.2: Matriz de confusão do rótulo Crises de Dor.

Tabela 5.2: Relatório de classificação do Random Forest para Crises de Dor.

	precision	recall	f1-score	sample
0.0	0.98	0.97	0.98	111
1.0	0.97	0.98	0.98	110
accuracy			0.98	221
macro avg	0.98	0.98	0.98	221
weighted avg	0.98	0.98	0.98	221

ticas relevantes para classificação do desfecho, entre elas 'Idade', 'HCM (rg)', 'CHCM (%)', 'LEUCO (/mL)', 'EOS(/mL)', 'BAS (/mL)', 'LIT (/mL)', 'MON (/mL)', 'PLA (103/mL)', 'RET(%)', 'COLHDL (mg/dL)', 'TRIG (mg/dL)', 'AST (U/L)', 'GGT (U/L)', 'BD (mg/dL)', 'ALBUM (g/dL)', 'GLOB (g/dL)', 'RELAG', 'ASLO (UI/mL)', 'GENERO', 'HBS (%)', 'HBA2 (%)', 'HBA1 (%)', 'HBC (%)', 'HAPL', 'QCRISDOR', 'QVO'. Os resultados são apresentados na matriz de confusão na Figura 5.3 e as métricas são exibidas na Tabela 5.3, com acurácia de 92%.

*Legenda: Idade, Hemoglobina Corpuscular Média (pg) (HCM), CHCM, LEUCO, Eosinófilos (mil/uL) (EOS), BAS, Linfócitos (mil/uL) (LIT), Monócitos (mil/uL) (MON), PLA, RET, COLHDL, TRIG, Aspartato aminotransferase (U/L) (AST), GGT, BD, ALBUM, Globulinas (g/dL) (GLOB),RELAG, ASLO, Gênero, HBS, HBA2, Hemoglobina A1 (%) (HBA1), HBC, HAPL, QCRISDOR e QVO.*

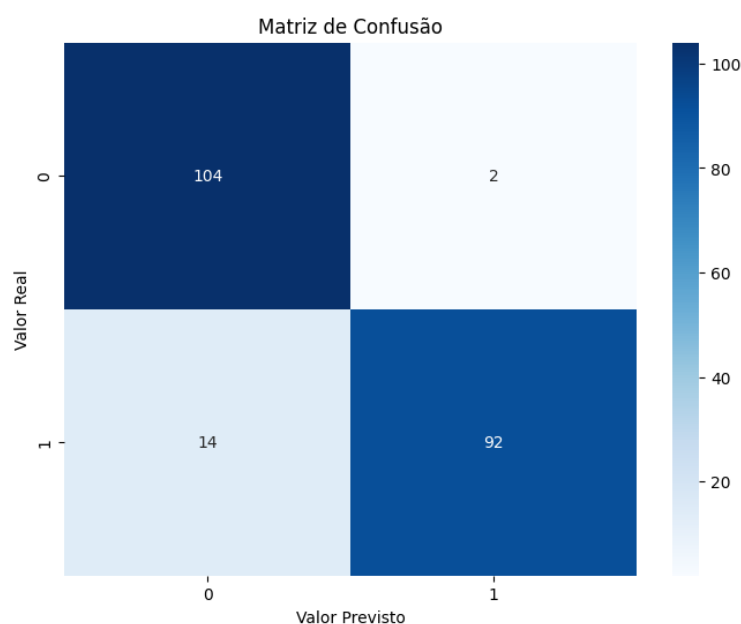


Figura 5.3: Matriz de confusão do rótulo Hospitalização.

Para o rótulo de infecções, os melhores resultados foram obtidos com o classificador RF e hiper-parâmetros  $\text{max\_depth}=10$ ,  $\text{min\_samples\_split}=5$ ,  $\text{n\_estimators}=50$ . Sendo selecionadas 27 características relevantes para classificação do desfecho, entre elas 'HM (106/mL)', 'BAST(/mL)', 'EOS(/mL)', 'BAS (/mL)', 'LIA (/mL)', 'MON (/mL)',

Tabela 5.3: Relatório de classificação do Random Forest para Hospitalização.

	precision	recall	f1-score	sample
0.0	0.88	0.98	0.93	106
1.0	0.98	0.87	0.92	106
accuracy			0.92	212
macro avg	0.93	0.92	0.92	212
weighted avg	0.93	0.92	0.92	212

'GLI (mg/dL)', 'COLHDL (mg/dL)', 'COLLDL (mg/dL)', 'COLVLDL (mg/dL)', 'AURICO(mg/dL)', 'UREIA (mg/dL)', 'FERRO (mg/dL)', 'GGT (U/L)', 'BD (mg/dL)', 'FAL (U/L)', 'PT (g/dL)', 'GLOB (g/dL)', 'RELAG', 'FERRIT (hg/mL)', 'ASLO (UI/mL)', 'HAPTO (mg/dL)', 'A1ATS (mg/dL)', 'HBA1 (%)', 'HAPL', 'TAL', 'QCRISDOR'.

*Legenda: HM, BAST, EOS, BAS, LIA, MON, GLI, COLHDL, COLLDL, COLVLDL, AURICO, UREIA, FERRO, GGT, BD, Fosfatase Alcalina (U/L) (FAL), PT, GLOB, RELAG, FERRIT, ASLO, HAPTO, Alfa 1 antitripsina (mg/dL) (A1ATS), HBA1, HAPL, TAL e QCRISDOR.*

Os resultados são apresentados na matriz de confusão na Figura 5.4 e as métricas são exibidas na Tabela 5.4, com acurácia de 62%.

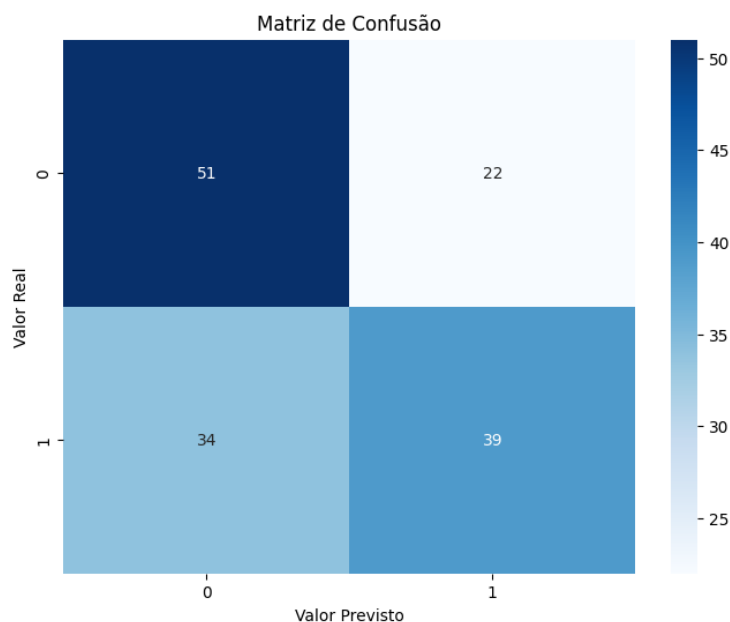


Figura 5.4: Matriz de confusão do rótulo Infecções.

Tabela 5.4: Relatório de classificação do Random Forest para Infecções.

	precision	recall	f1-score	support
0.0	0.60	0.70	0.65	73
1.0	0.64	0.53	0.58	73
accuracy			0.62	146
macro avg	0.62	0.62	0.61	146
weighted avg	0.62	0.62	0.61	146

Para os modelos de inteligência artificial testados no experimento, os atributos selecionados para o diagnóstico dos desfechos da doença falciforme são apresentados na Tabela 5.5. É importante observar que nesse primeiro momento utilizamos algoritmos tradicionais de rótulo único.

Na segunda fase dos experimentos, realizamos uma análise comparativa de diferentes algoritmos aplicados à classificação multirrótulo de desfechos em pacientes com doença falciforme. Os experimentos focaram nos classificadores mais conhecidos para transformação de problemas multirrótulo, permitindo a identificação do melhor algoritmo em diferentes cenários, desta vez, considerando que os desfechos podem ocorrer de forma simultânea.

A avaliação dos modelos incluiu métricas como *Hamming Loss*, *Micro-averaged Precision*, *Micro-averaged Recall*, *Micro-averaged F-measure*, *Subset Accuracy*, e *Ranking Loss*, conforme mostrado na tabela 5.6. Essas métricas fornecem uma compreensão do desempenho dos modelos com dados de pacientes com doença falciforme.



Tabela 5.5: Ocorrência dos Atributos selecionados para os desfechos de vaso-oclusão, crises de dor, hospitalização e infecções.

<b>Atributo</b>	<b>Ocorrências</b>
BAS (/mL)	4
RELAG	4
HAPL	4
HM (106/mL)	3
RET(%)	3
COLHDL (mg/dL)	3
UREIA (mg/dL)	3
BD (mg/dL)	3
PT (g/dL)	3
ASLO (UI/mL)	3
HAPTO (mg/dL)	3
HBS (%)	3
HBA2 (%)	3
QVO	3
GGT (U/L)	3
QCRISDOR	3
CHCM (%)	2
SEG(/mL)	2
GLI (mg/dL)	2
COLLDL (mg/dL)	2
AURICO(mg/dL)	2
FERRO (mg/dL)	2
PCR (mg/L)	2
GENERO	2
Idade	2
LEUCO (/mL)	2
LIA (/mL)	2
PLA (103/mL)	2
COLVLDL (mg/dL)	2
TRIG (mg/dL)	2
ALBUM (g/dL)	2
FERRIT (hg/mL)	2
HBC (%)	2
TAL	2
EOS(/mL)	2
MON (/mL)	2
GLOB (g/dL)	2
HBA1 (%)	2

Tabela 5.6: Desempenho dos algoritmos com diferentes métricas.

Algoritmo	Parâmetro	Hamming Loss	Micro-averaged precision	Micro-averaged recall	Micro-averaged f1-score	Label Ranking Loss	Subset accuracy
BR	RF	0.20	0.80	1.00	0.89	0.59	0.41
	DT	0.21	0.81	0.97	0.88	0.55	0.39
	SVM	0.20	0.80	1.00	0.89	0.59	0.41
CC	RF	0.21	0.82	0.95	0.88	0.52	0.38
	DT	0.21	0.80	0.99	0.88	0.58	0.40
	SVM	0.20	0.80	1.00	0.89	0.59	0.41
LP	RF	0.21	0.81	0.96	0.88	0.55	0.38
	DT	0.20	0.80	0.99	0.89	0.58	0.41
	SVM	0.20	0.80	1.00	0.89	0.59	0.41
RAkELo	RF	0.21	0.82	0.95	0.88	0.52	0.39
	DT	0.20	0.80	1.00	0.89	0.59	0.41
	SVM	0.21	0.80	0.99	0.88	0.57	0.40

A métrica de Hamming Loss, que mede a taxa de classificações incorretas, ficou em torno de 0,20 para a maioria dos modelos do estudo, o que indica que, em média, há uma taxa de erro de 20% por amostra. Os algoritmos BR (RF, SVM) e LP (DT, SVM) apresentam os menores valores, sugerindo que eles erram menos ao prever os rótulos. Essa métrica na classificação multirrótulo da doença falciforme reflete a taxa de erros de predição por rótulo. Um valor menor indica que o modelo comete menos erros, o que é relevante para condições onde a identificação de múltiplos desfechos impacta diretamente no tratamento e acompanhamento do paciente.

Considerando a métrica *micro-averaged f1-score* que é utilizada para avaliar o equilíbrio entre precisão e recall. Os algoritmos *Binary Relevance*, *Classifier Chains*, *Label Powerset* e *Random k-Labelsets* com SVM e RF apresentam resultados muito similares, com *f1-scores* próximos de 0,89. A *Micro-averaged Precision* e o *Micro-averaged Recall* também são altos nos experimentos, o que é um bom indicativo na classificação de múltiplos rótulos, em um ambiente clínico, uma alta precisão, indica que o modelo tem uma baixa taxa de falsos positivos. Isto evita intervenções desnecessárias, como tratamentos ou exames adicionais, que podem não ser requisitados se o modelo identifica erroneamente a presença de um desfecho, já a alta sensibilidade (*recall*) é utilizada para garantir que todos os desfechos relevantes sejam identificados. Na doença falciforme, onde múltiplas manifestações clínicas podem ocorrer simultaneamente, garantir que todas sejam detectadas é importante para um encaminhamento adequado da condição do paciente.

A *Label Ranking Loss* determina a ordenação correta dos rótulos, onde valores mais baixos indicam melhor desempenho. O algoritmo CC com RF e RAKELo com RF apresentam os menores valores de 0,52, sugerindo um desempenho superior nesses aspectos. Este ponto pode ser decisivo em situações onde a ordem das manifestações clínicas são importantes, no contexto da doença falciforme, onde alguns desfechos podem ser mais críticos que outros, ter um modelo que consiga ordenar corretamente a gravidade ou a prioridade dos desfechos ajuda na tomada de decisões clínicas, como priorizar intervenções para desfechos mais severos.

Outra métrica avaliada, a *Subset Accuracy*, considera a fração de amostras para as

quais todos os rótulos são corretamente previstos, ao contrário de outras métricas, esta é particularmente rigorosa e relevante em cenários onde o erro em qualquer desfecho pode ter implicações graves, nossos experimentos apontam que os modelos tiveram um desempenho similar, girando em torno de uma taxa de 0,41. Na gestão da doença falciforme, onde múltiplos desfechos podem coexistir e todos precisam ser tratados adequadamente, uma alta *Subset Accuracy* seria o cenário ideal de um modelo para fornecer predições mais completas, minimizando riscos ao paciente, porém mesmo para a inteligência artificial quando aumentamos o número de rótulos essa métrica é muito prejudicada.

A análise de correlação entre os rótulos foi realizada para investigar a relação entre os desfechos previstos, identificando possíveis dependências ou padrões comuns entre eles. Em nossos experimentos, observou-se que a maioria dos rótulos apresentaram correlações baixas, com coeficientes variando entre 0,14 e 0,29. E alta correlação 0,73 para vaso-oclusão e crises de dor (ver Figura 5.5).

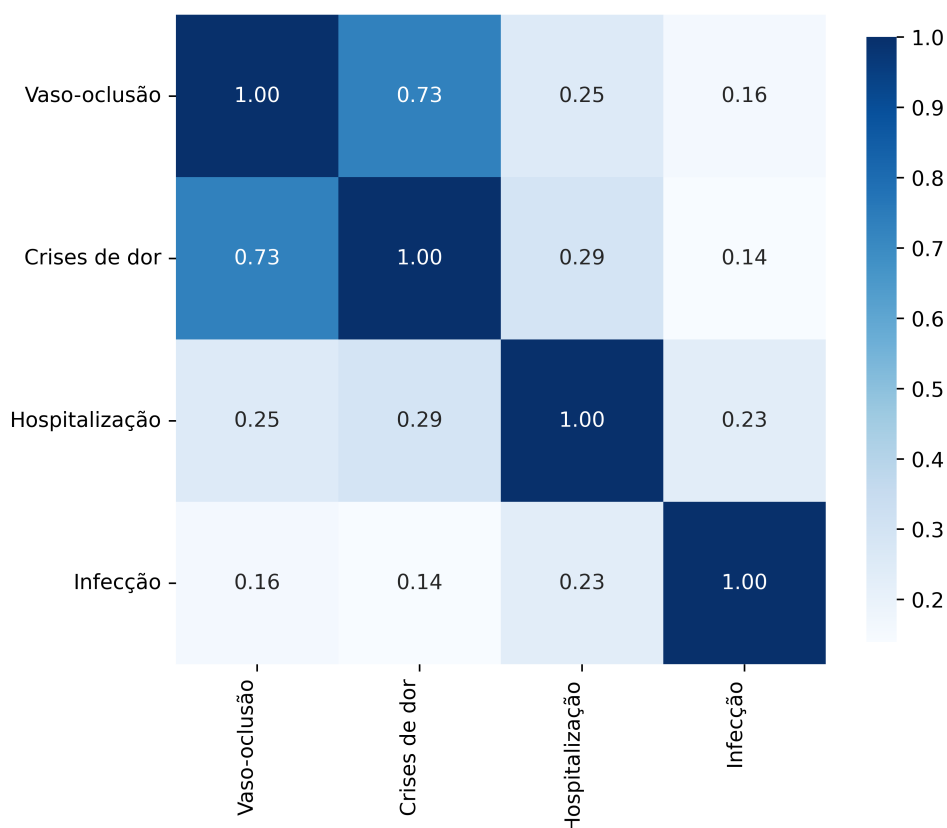


Figura 5.5: Correlações entre os desfechos.

## Capítulo

# 6

## DISCUSSÃO

Na análise de 609 registros de pacientes com doença falciforme obtidos de duas pesquisas científicas no Estado da Bahia conduzidas pelo Laboratório de Anemias da Faculdade de Farmácia da UFBA e pelo Laboratório de Investigação em Genética e Hematologia Translacional da FIOCRUZ, analisamos os classificadores multirrótulo no suporte ao diagnóstico de desfechos da doença falciforme.

Para responder a questão de pesquisa “**QP1. Quais são os parâmetros clínicos e laboratoriais mais relevantes para a previsão de múltiplos desfechos simultâneos na doença falciforme, e como esses fatores interagem para fornecer uma previsão?**”, fizemos dois testes, no primeiro, aplicamos algoritmos tradicionais de rótulo único para identificação dos melhores parâmetros e dos principais biomarcadores relacionados com a doença falciforme, foram obtidas 27 variáveis para servirem de entrada para os algoritmos, e a escolha desse *input* foi feita com uso do *Recursive Feature Elimination* que ficou responsável por indicar os preditores mais relevantes para cada um dos rótulos do estudo: vaso-oclusão, crises de dor, hospitalizações e infecções.

Os níveis HbF, talvez sejam o biomarcador mais importante da gravidade da doença falciforme, quando os níveis estão elevados os pacientes possuem sintomas mais brandos e ganham vantagem na qualidade e sobrevida. Níveis mais altos de HbF reduzem as

concentrações intracelulares de HbS e interrompem a polimerização da HbS desoxigenada. Uma ressalva se dá a grande variabilidade de sintomas da doença, ou seja, não está estabelecido na literatura, qualquer outros biomarcadores mais importantes que a concentração de hemoglobina, os níveis de HbF e as contagens de reticulócitos (REES; BROUSSE; BREWIN, 2022). Dessa forma, o algoritmo RFE utilizado no experimento, conseguiu identificar os atributos: RET e CHCM, o HbF acabou não sendo um preditor relevante para os modelos de classificação de rótulo único utilizados na primeira fase. Porém, quando observada a Figura 4.2 que relaciona as variáveis mais importantes nos algoritmos multirrótulo e responde a questão de pesquisa **Q1**. temos (VCM), (RET) e (HbF) entre os atributos mais importantes para os algoritmos Binary Relevance, Label Powerset, Classifier Chains e Random k-labelsets.

A ocorrência mais comum dos atributos selecionados para os modelos de rótulo único, podem ser consultados na Tabela 5.5, de forma geral os três atributos mais importantes para as predições foram basófilos (BAS) que geralmente é relacionado ao sistema imunológico, combatendo inflamações, a relação albumina/globulinas (RELAG) que está associada a infecções e Haplótipos ligados ao gene da globina beta (HAPL) que mostraram-se importantes na modulação dos níveis de hemoglobina fetal. Já quando utilizamos os algoritmos multirrótulo, as características mais importantes para os modelos incluem idade, proteínas totais, hemoglobina S, quantidade de vaso-oclusão nos últimos 6 meses e Proteína C-Reativa (PCR), corroborando com (REES; GIBSON, 2012) que indica biomarcadores de inflamação.

Para responder à questão de pesquisa QP2, que trata do pré-processamento dos modelos: **“Quais estratégias de pré-processamento de dados, como normalização, imputação de valores ausentes e redução de dimensionalidade, demonstram ter um impacto positivo nos modelos multirrótulo, contribuindo para a melhoria das métricas de desempenho?”**, observamos que a quantidade de rótulos é um fator crítico que impacta diretamente as métricas de desempenho. Quanto maior o número de rótulos, maior a complexidade algorítmica. Essa relação foi amplamente discutida por autores como (READ et al., 2009; TSOUMAKAS; VLAHAVAS, 2007; ZHANG;

ZHOU, 2013) e corroborada pelos resultados obtidos durante o treinamento dos modelos analisados neste trabalho.

A questão de pesquisa **“QP3. Os algoritmos de aprendizado de máquina multirrótulo são capazes de identificar padrões preditivos exclusivos e complexos que não seriam facilmente detectados por métodos tradicionais de classificação, como os algoritmos de rótulo único?”**, nossos resultados apontam que os algoritmos multirrótulo conseguem identificar as relações entre os desfechos simultâneos da doença falciforme. Os algoritmos baseados em SVM e RF são consistentes e apresentam um bom equilíbrio entre precisão e recall. A escolha do algoritmo deve ser guiada pelo objetivo clínico específico: se a ordenação correta dos desfechos for relevante, o modelo RAKELo com RF embora o *Micro-averaged f1-score* seja levemente inferior (0,88), o *Label Ranking Loss* de 0,52 sugere que este modelo pode ser preferível. Se o objetivo clínico não levar em conta essa ordenação, considerando o equilíbrio entre as métricas, os modelos baseados em SVM e RF, em todas as abordagens (BR, CC, LP, RAKELo), apresentaram uma combinação ideal de baixo *Hamming Loss* (0,20) e alto *Micro-averaged f1-score* (0,89).

As técnicas de sobre-amostragem ou sub-amostragem utilizadas não apresentaram melhorias significativas no nosso conjunto de dados, considerando o risco da inserção de dados sintéticos e o risco de excluir alguma condição importante nas classes de treinamento e validação dos classificadores, acreditamos que o uso dessas técnicas não vale o risco para um conjunto de dados reduzido.

Outro ponto que merece destaque é que selecionamos as técnicas de transformação em detrimento dos algoritmos adaptados para o problema de rótulos múltiplos, devido ao baixo desempenho dos classificadores adaptados ao nosso conjunto de dados. Acreditamos que esse cenário mereça mais experimentos para que seja possível determinar se os parâmetros foram configurados de forma correta, ou se os classificadores simplesmente não conseguiram superar a configuração dos rótulos no nosso conjunto de dados.

Ao analisar as correlações entre os rótulos, observamos que existe uma alta correlação entre vaso-occlusão e crises de dor, atingindo um valor de 0,73, o que indica uma forte

associação entre esses desfechos, o que é citado por outros autores (ROSSI et al., 2024; PADHEE et al., 2021; KATO et al., 2018). Por outro lado, identificamos uma baixa correlação entre os demais desfechos. Infecções e crises de dor apresentaram a correlação mais baixa do conjunto de dados, com um valor de 0,14, sugerindo que esses eventos ocorrem de forma independente. As hospitalizações demonstraram correlações com pouca variação em relação aos outros desfechos, com valores entre 0,23 e 0,29, evidenciando que todos os quadros são clinicamente relevantes para a necessidade de internação do paciente.

A tese apresenta algumas limitações relacionadas à generalização dos modelos, uma vez que os classificadores não foram validados com dados externos. Isso pode comprometer a capacidade dos algoritmos de identificar perfis biológicos e demográficos diferentes daqueles presentes no conjunto de treinamento, apesar do rigor científico adotado no desenvolvimento da pesquisa. Além disso, o tamanho da amostra é um fator relevante, pois acreditamos que a quantidade limitada de dados de treinamento pode influenciar diretamente o desempenho dos classificadores.



## CONCLUSÕES

Este estudo demonstra o potencial de algoritmos de inteligência artificial na classificação multirrótulo de desfechos clínicos na doença falciforme. Os resultados oferecem uma base robusta para pesquisas futuras que visem integrar esses modelos em sistemas de suporte à decisão clínica. Quando a ordem dos desfechos é relevante, o modelo RAKeLo com *Random Forest* se destaca, apresentando um *Label Ranking Loss* de 0,52. Por outro lado, quando a ordenação não é um fator clínico importante, os modelos baseados em SVM e *Random Forest*, em todas as abordagens (BR, CC, LP, RAKeLo), mostraram uma combinação ideal de baixo *Hamming Loss* (0,20) e alto *Micro-averaged f1-score* (0,89).

Embora os principais biomarcadores da doença falciforme já tenham sido amplamente documentados (REES; GIBSON, 2012; REES; BROUSSE; BREWIN, 2022), como a hemoglobina S (HbS), os níveis de hemoglobina fetal (HbF), e marcadores inflamatórios como a proteína C-reativa (PCR), ainda há espaço para aprofundar a compreensão de como esses biomarcadores se relacionam com os múltiplos desfechos clínicos da doença. Esses estudos sugerem que níveis elevados de HbF estão associados a um perfil clínico mais leve, enquanto a PCR tem sido frequentemente correlacionada com eventos vasculares.

Este parece ser o primeiro estudo a abordar a classificação multirrótulo dos desfechos clínicos na doença falciforme, uma abordagem diferente das pesquisas com os métodos

tradicionais de predição de rótulo único. A classificação multirrótulo permite que múltiplos desfechos sejam previstos simultaneamente, capturando a complexidade inerente da doença, na qual os pacientes frequentemente apresentam múltiplos eventos clínicos simultâneos, como crises vaso-oclusivas, crises de dor, hospitalizações e infecções. Ao adotar essa metodologia, conseguimos modelar as interações entre diferentes desfechos, oferecendo uma visão mais realista das ocorrências dos sintomas.

Trabalhos futuros devem focar na validação externa dos modelos preditivos utilizando conjuntos de dados maiores e mais diversos, representando diferentes perfis populacionais. Além disso, é importante explorar algoritmos adaptados para a classificação multirrótulo, a integração desses classificadores com biomarcadores emergentes pode aprimorar a estratificação dos pacientes e a previsão de desfechos. A inclusão de biomarcadores adicionais e dados clínicos longitudinais também permitirá o desenvolvimento de modelos mais precisos e adaptáveis, com potencial para melhorar intervenções personalizadas no manejo da doença falciforme.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABDULHAY, E. W.; ALLOW, A. G.; AL-JALOULY, M. E. Detection of sickle cell, megaloblastic anemia, thalassemia and malaria through convolutional neural network. In: IEEE. *2021 Global Congress on Electrical Engineering (GC-ElecEng)*. [S.l.], 2021. p. 21–25.

AGBOR, D. B. A. et al. Folic acid in the treatment of sickle cell disease: A systematic review. *Cureus*, Cureus Inc., v. 16, n. 4, 2024.

ARAÚJO, A. da S. et al. Sickle cell disease in brazil: Current management. *Hemoglobin*, Taylor & Francis, p. 1–13, 2024.

BARROS, G. d. S. et al. Real-world evidence of the burden of sickle cell disease: a 5-year longitudinal study at a brazilian reference center. *Hematology, Transfusion and Cell Therapy*, SciELO Brasil, v. 46, n. 2, p. 161–166, 2024.

BHASIN, N.; SARODE, R. Acute chest syndrome in sickle cell disease. *Transfusion Medicine Reviews*, Elsevier, p. 150755, 2023.

BOGATINOVSKI, J. et al. Comprehensive comparative study of multi-label classification methods. *Expert Systems with Applications*, Elsevier, v. 203, p. 117215, 2022.

CANÇADO, R. D. et al. Estimated mortality rates of individuals with sickle cell disease in brazil: real-world evidence. *Blood Advances*, American Society of Hematology Washington, DC, v. 7, n. 15, p. 3783–3792, 2023.

CHY, T. S.; RAHAMAN, M. A. A comparative analysis by knn, svm & elm classification to detect sickle cell anemia. In: IEEE. *2019 International conference on robotics, electrical and signal processing techniques (ICREST)*. [S.l.], 2019. p. 455–459.

CORDOVIL, K. et al. Social inequalities in the temporal trend of mortality from sickle cell disease in brazil, 1996-2019. *Cadernos de Saúde Pública*, SciELO Public Health, v. 39, p. e00256421, 2023.

DEBAUN, M. et al. American society of hematology 2020 guidelines for sickle cell disease: prevention, diagnosis, and treatment of cerebrovascular disease in children and adults. *Blood advances*, American Society of Hematology Washington, DC, v. 4, n. 8, p. 1554–1588, 2020.

GARCÍA-PEDRAJAS, N. E. et al. A thorough experimental comparison of multilabel methods for classification performance. *Pattern Recognition*, Elsevier, p. 110342, 2024.

- GOSWAMI, N. G. et al. Sickle cell classification using deep learning. In: IEEE. *2023 3rd International Conference on Intelligent Technologies (CONIT)*. [S.l.], 2023. p. 1–6.
- GOSWAMI, N. G. et al. Detection of sickle cell disease using deep neural networks and explainable artificial intelligence. *Journal of Intelligent Systems*, De Gruyter, v. 33, n. 1, p. 20230179, 2024.
- GÜNTÜRKÜN, F. et al. Using machine learning to predict rapid decline of kidney function in sickle cell anemia. *EJHaem*, Wiley Online Library, v. 2, n. 2, p. 257–260, 2021.
- HAAN, K. de et al. Automated screening of sickle cells using a smartphone-based microscope and deep learning. *NPJ digital medicine*, Nature Publishing Group UK London, v. 3, n. 1, p. 76, 2020.
- HOUWING, M. et al. Sickle cell disease: clinical presentation and management of a global health challenge. *Blood reviews*, Elsevier, v. 37, p. 100580, 2019.
- IDRIS, I. M.; BOTCHWEY, E. A.; HYACINTH, H. I. Sickle cell disease as an accelerated aging syndrome. *Experimental Biology and Medicine*, SAGE Publications Sage UK: London, England, v. 247, n. 4, p. 368–374, 2022.
- JAVIDI, B. et al. Sickle cell disease diagnosis based on spatio-temporal cell dynamics analysis using 3d printed shearing digital holographic microscopy. *Optics express*, Optica Publishing Group, v. 26, n. 10, p. 13614–13627, 2018.
- JENNIFER, S. S. et al. Sickle cell disease classification using deep learning. *Heliyon*, Elsevier, v. 9, n. 11, 2023.
- JI, Y. et al. Identifying elevated risk for future pain crises in sickle-cell disease using photoplethysmogram patterns measured during sleep: A machine learning approach. *Frontiers in digital health*, Frontiers Media SA, v. 3, p. 714741, 2021.
- KATO, G. J. et al. Sickle cell disease. *Nature reviews Disease primers*, Nature Publishing Group, v. 4, n. 1, p. 1–22, 2018.
- KAVANAGH, P. L.; FASIPE, T. A.; WUN, T. Sickle cell disease: a review. *Jama*, American Medical Association, v. 328, n. 1, p. 57–68, 2022.
- MOHAMMED, A. et al. Using machine learning to predict early onset acute organ failure in critically ill intensive care unit patients with sickle cell disease: retrospective study. *Journal of Medical Internet Research*, JMIR Publications Toronto, Canada, v. 22, n. 5, p. e14693, 2020.
- NEUMAYR, L. D.; HOPPE, C. C.; BROWN, C. Sickle cell disease: current treatment and emerging therapies. *Am J Manag Care*, v. 25, n. 18 Suppl, p. S335–43, 2019.

- PADHEE, S. et al. Pain intensity assessment in sickle cell disease patients using vital signs during hospital visits. In: SPRINGER. *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part II*. [S.l.], 2021. p. 77–85.
- PADHEE, S. et al. Improving pain assessment using vital signs and pain medication for patients with sickle cell disease: Retrospective study. *JMIR Formative Research*, JMIR Publications Toronto, Canada, v. 6, n. 6, p. e36998, 2022.
- PADRÃO, E. M. H. et al. Phenotypes of sickle cell intensive care admissions: an unsupervised machine learning approach in a single-center retrospective cohort. *Annals of Hematology*, Springer, v. 101, n. 9, p. 1951–1957, 2022.
- PATEL, A. et al. Machine-learning algorithms for predicting hospital re-admissions in sickle cell disease. *British Journal of Haematology*, Wiley Online Library, v. 192, n. 1, p. 158–170, 2021.
- PEREIRA, R. B. et al. Correlation analysis of performance measures for multi-label classification. *Information Processing & Management*, Elsevier, v. 54, n. 3, p. 359–369, 2018.
- PEREIRA, R. M.; COSTA, Y. M.; JR, C. N. S. Mtl: A multi-label approach for the torek link undersampling algorithm. *Neurocomputing*, Elsevier, v. 383, p. 95–105, 2020.
- PETROVIĆ, N. et al. Sickle-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images. *Computers in Biology and Medicine*, Elsevier, v. 126, p. 104027, 2020.
- PRASHANTHI, G.; SINGH, S. P. Identification of sickle cell anemia by employing hybrid optimization and recurrent neural network. In: IEEE. *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*. [S.l.], 2023. p. 853–858.
- PRIYA, K.; PETER, J. D. A federated approach for detecting the chest diseases using densenet for multi-label classification. *Complex & Intelligent Systems*, Springer, v. 8, n. 4, p. 3121–3129, 2022.
- RAGHURAMAN, A. et al. Role of gene therapy in sickle cell disease. *Disease-a-Month*, Elsevier, v. 70, n. 7, p. 101689, 2024.
- READ, J. et al. Classifier chains for multi-label classification. In: SPRINGER. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II 20*. [S.l.], 2009. p. 254–269.
- REES, D. C.; BROUSSE, V. A.; BREWIN, J. N. Determinants of severity in sickle cell disease. *Blood Reviews*, Elsevier, v. 56, p. 100983, 2022.

- REES, D. C.; GIBSON, J. S. Biomarkers in sickle cell disease. *British journal of haematology*, Wiley Online Library, v. 156, n. 4, p. 433–445, 2012.
- ROSSI, M. et al. Determinants of the haemoglobin level in patients with sickle cell disease living in sub-saharan africa: Major impact of the country of residence and independent effects of leucocyte and platelet counts and haemolysis. *British Journal of Haematology*, Wiley Online Library, 2024.
- SACHDEV, V. et al. A phenotypic risk score for predicting mortality in sickle cell disease. *British journal of haematology*, Wiley Online Library, v. 192, n. 5, p. 932–941, 2021.
- SANTO, A. H. Sickle cell disease related mortality in brazil, 2000-2018. *Hematology, Transfusion and Cell Therapy*, SciELO Brasil, v. 44, n. 2, p. 177–185, 2022.
- SEN, B. et al. Deep learning based diagnosis of sickle cell anemia in human rbc. In: IEEE. *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*. [S.l.], 2021. p. 526–529.
- SILVA-PINTO, A. C. et al. The neonatal screening program in brazil, focus on sickle cell disease (scd). *International Journal of Neonatal Screening*, MDPI, v. 5, n. 1, p. 11, 2019.
- SINGH, B. K.; THAKKAR, H. Hydroxyurea dosage classification for sickle cell disease patients. In: IEEE. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. [S.l.], 2021. p. 1289–1292.
- SONI, S.; THAKKAR, H.; SINGH, B. K. Transfer learning for sickle cell anemia and trait classification. In: IEEE. *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*. [S.l.], 2022. p. 1–5.
- TEBBI, C. K. Sickle cell disease, a review. *Hemato*, MDPI, v. 3, n. 2, p. 341–366, 2022.
- TENGSHI, R. et al. Sickle cell anemia detection using convolutional neural network. In: IEEE. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. [S.l.], 2021. p. 1–6.
- TSOUMAKAS, G.; KATAKIS, I. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, IGI Global, v. 3, n. 3, p. 1–13, 2007.
- TSOUMAKAS, G.; VLAHAVAS, I. Random k-labelsets: An ensemble method for multilabel classification. In: SPRINGER. *European conference on machine learning*. [S.l.], 2007. p. 406–417.
- VICENT, M.; SIMON, K.; YONASI, S. An algorithm to detect overlapping red blood cells for sickle cell disease diagnosis. *IET Image Processing*, Wiley Online Library, v. 16, n. 6, p. 1669–1677, 2022.
- WANG, Q. et al. Deep active learning for multi label text classification. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 28246, 2024.

ZHANG, M.-L.; ZHOU, Z.-H. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, IEEE, v. 26, n. 8, p. 1819–1837, 2013.

## **RESULTADOS EXPERIMENTAIS**

Este capítulo apresenta todos os resultados experimentais obtidos ao aplicar os algoritmos de aprendizado de máquina multirrótulo no conjunto de dados da tese. Como base de comparação, utilizamos três algoritmos tradicionais de aprendizado supervisionado como parâmetro, a saber: Random Forest (RF), Support Vector Machine (SVM) e Decision Tree (DT). A avaliação dos modelos foi realizada ao longo de 10 épocas, com o objetivo de garantir uma análise estatisticamente confiável e calcular os intervalos de confiança dos resultados. As tabelas apresentadas a seguir demonstram o desempenho dos algoritmos com base em métricas específicas de classificação multirrótulo, permitindo uma comparação quantitativa entre os métodos testados.



Tabela A1: Binary Relevance com parâmetros (RF, SVM, DT) executados em 10 épocas												
RF											Averaged	
Hamming loss	0,20	0,20	0,20	0,19	0,20	0,20	0,20	0,20	0,20	0,20	0,21	<b>0,20</b>
Micro-averaged precisic	0,80	0,80	0,80	0,81	0,80	0,80	0,80	0,80	0,80	0,80	0,79	<b>0,80</b>
Micro-averaged recall	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	<b>1,00</b>
Micro-averaged F1-scor	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	<b>0,89</b>
Subset accuracy	0,41	0,41	0,41	0,42	0,42	0,41	0,42	0,42	0,42	0,41	0,41	<b>0,41</b>
Label Ranking Loss	0,59	0,59	0,59	0,58	0,58	0,59	0,58	0,58	0,58	0,59	0,59	<b>0,59</b>
SVM												
Hamming loss	0,20	0,20	0,20	0,19	0,20	0,20	0,20	0,20	0,20	0,20	0,21	<b>0,20</b>
Micro-averaged precisic	0,80	0,80	0,80	0,81	0,80	0,80	0,80	0,80	0,80	0,80	0,79	<b>0,80</b>
Micro-averaged recall	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	<b>1,00</b>
Micro-averaged F1-scor	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	<b>0,89</b>
Subset accuracy	0,41	0,41	0,41	0,42	0,42	0,41	0,42	0,42	0,42	0,41	0,41	<b>0,41</b>
Label Ranking Loss	0,59	0,59	0,59	0,58	0,58	0,59	0,58	0,58	0,58	0,59	0,59	<b>0,59</b>
DT												
Hamming loss	0,20	0,21	0,20	0,22	0,21	0,23	0,21	0,19	0,21	0,21	0,21	<b>0,21</b>
Micro-averaged precisic	0,80	0,80	0,82	0,80	0,80	0,82	0,80	0,81	0,81	0,81	0,81	<b>0,81</b>
Micro-averaged recall	1,00	0,98	0,96	0,97	0,98	0,91	0,99	1,00	0,96	0,97	0,97	<b>0,97</b>
Micro-averaged F1-scor	0,89	0,88	0,88	0,88	0,88	0,87	0,89	0,89	0,88	0,88	0,88	<b>0,88</b>
Subset accuracy	0,42	0,38	0,38	0,36	0,38	0,34	0,41	0,43	0,39	0,38	0,38	<b>0,39</b>
Label Ranking Loss	0,58	0,57	0,51	0,56	0,56	0,48	0,59	0,56	0,52	0,54	0,54	<b>0,55</b>

Tabela A2: Classifier Chains com parâmetros (RF, SVM, DT) executados em 10 épocas											
RF											<b>Averaged</b>
Hamming loss	0,20	0,21	0,21	0,21	0,21	0,20	0,21	0,21	0,20	0,20	<b>0,21</b>
Micro-averaged precisic	0,82	0,82	0,82	0,82	0,83	0,83	0,81	0,81	0,83	0,82	<b>0,82</b>
Micro-averaged recall	0,96	0,95	0,94	0,94	0,93	0,94	0,96	0,95	0,95	0,96	<b>0,95</b>
Micro-averaged F1-scor	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,89	<b>0,88</b>
Subset accuracy	0,41	0,38	0,38	0,38	0,37	0,38	0,41	0,38	0,39	0,39	<b>0,38</b>
Label Ranking Loss	0,51	0,55	0,51	0,52	0,50	0,51	0,51	0,53	0,50	0,51	<b>0,52</b>
SVM											
Hamming loss	0,20	0,20	0,20	0,20	0,20	0,20	0,19	0,20	0,19	0,20	<b>0,20</b>
Micro-averaged precisic	0,80	0,80	0,80	0,80	0,80	0,80	0,81	0,80	0,81	0,80	<b>0,80</b>
Micro-averaged recall	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	<b>1,00</b>
Micro-averaged F1-scor	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	<b>0,89</b>
Subset accuracy	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,41	0,42	0,41	<b>0,41</b>
Label Ranking Loss	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,58	0,59	<b>0,59</b>
DT											
Hamming loss	0,22	0,20	0,21	0,19	0,21	0,20	0,20	0,21	0,21	0,21	<b>0,21</b>
Micro-averaged precisic	0,80	0,80	0,81	0,82	0,79	0,80	0,80	0,79	0,79	0,81	<b>0,80</b>
Micro-averaged recall	0,96	1,00	0,97	0,98	1,00	1,00	1,00	1,00	1,00	0,97	<b>0,99</b>
Micro-averaged F1-scor	0,88	0,89	0,88	0,89	0,88	0,89	0,89	0,88	0,88	0,88	<b>0,88</b>
Subset accuracy	0,37	0,41	0,38	0,42	0,41	0,41	0,41	0,40	0,41	0,37	<b>0,40</b>
Label Ranking Loss	0,56	0,59	0,56	0,55	0,59	0,59	0,58	0,59	0,59	0,54	<b>0,58</b>

Tabela A3: Label Powerset com parâmetros (RF, SVM, DT) executados em 10 épocas											
RF											<b>Averaged</b>
Hamming loss	0,21	0,22	0,21	0,22	0,20	0,22	0,21	0,18	0,20	0,22	<b>0,21</b>
Micro-averaged precisic	0,81	0,81	0,82	0,80	0,81	0,81	0,81	0,83	0,82	0,80	<b>0,81</b>
Micro-averaged recall	0,97	0,95	0,96	0,97	0,98	0,96	0,97	0,97	0,96	0,96	<b>0,96</b>
Micro-averaged F1-scor	0,88	0,87	0,88	0,87	0,89	0,87	0,88	0,89	0,88	0,87	<b>0,88</b>
Subset accuracy	0,39	0,38	0,38	0,36	0,40	0,37	0,39	0,42	0,37	0,37	<b>0,38</b>
Label Ranking Loss	0,56	0,55	0,53	0,56	0,55	0,56	0,54	0,51	0,54	0,57	<b>0,55</b>
SVM											
Hamming loss	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,20	0,19	<b>0,20</b>
Micro-averaged precisic	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,81	<b>0,80</b>
Micro-averaged recall	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	<b>1,00</b>
Micro-averaged F1-scor	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	0,89	<b>0,89</b>
Subset accuracy	0,42	0,41	0,41	0,41	0,41	0,41	0,42	0,41	0,41	0,41	<b>0,41</b>
Label Ranking Loss	0,58	0,59	0,59	0,59	0,59	0,59	0,58	0,59	0,59	0,59	<b>0,59</b>
DT											
Hamming loss	0,20	0,21	0,20	0,20	0,20	0,20	0,20	0,21	0,20	0,21	<b>0,20</b>
Micro-averaged precisic	0,80	0,80	0,80	0,80	0,81	0,80	0,80	0,80	0,80	0,79	<b>0,80</b>
Micro-averaged recall	1,00	0,97	1,00	1,00	0,98	1,00	1,00	0,99	1,00	1,00	<b>0,99</b>
Micro-averaged F1-scor	0,89	0,88	0,89	0,89	0,89	0,89	0,89	0,88	0,89	0,89	<b>0,89</b>
Subset accuracy	0,42	0,36	0,42	0,41	0,41	0,41	0,42	0,39	0,42	0,41	<b>0,41</b>
Label Ranking Loss	0,57	0,55	0,58	0,59	0,56	0,59	0,58	0,57	0,58	0,59	<b>0,58</b>

Tabela A4: Random k-labelsets com parâmetros (RF, SVM, DT) executados em 10 épocas											
RF											<b>Averaged</b>
Hamming loss	0,20	0,20	0,20	0,21	0,22	0,20	0,22	0,22	0,21	0,20	<b>0,21</b>
Micro-averaged precisic	0,82	0,82	0,83	0,81	0,80	0,83	0,80	0,82	0,82	0,82	<b>0,82</b>
Micro-averaged recall	0,95	0,96	0,95	0,96	0,97	0,95	0,97	0,94	0,94	0,95	<b>0,95</b>
Micro-averaged F1-scor	0,88	0,88	0,88	0,88	0,88	0,88	0,87	0,87	0,88	0,88	<b>0,88</b>
Subset accuracy	0,40	0,38	0,43	0,39	0,37	0,38	0,37	0,36	0,40	0,39	<b>0,39</b>
Label Ranking Loss	0,51	0,53	0,49	0,54	0,54	0,52	0,57	0,52	0,50	0,51	<b>0,52</b>
SVM											
Hamming loss	0,20	0,20	0,20	0,20	0,21	0,22	0,21	0,20	0,21	0,21	<b>0,21</b>
Micro-averaged precisic	0,80	0,80	0,81	0,81	0,81	0,79	0,80	0,80	0,79	0,80	<b>0,80</b>
Micro-averaged recall	0,99	0,99	0,98	0,98	0,97	0,98	0,99	0,99	0,99	0,99	<b>0,99</b>
Micro-averaged F1-scor	0,89	0,89	0,89	0,89	0,88	0,87	0,88	0,89	0,88	0,88	<b>0,88</b>
Subset accuracy	0,41	0,40	0,40	0,40	0,41	0,38	0,40	0,42	0,40	0,40	<b>0,40</b>
Label Ranking Loss	0,58	0,58	0,56	0,55	0,55	0,58	0,59	0,57	0,59	0,59	<b>0,57</b>
DT											
Hamming loss	0,20	0,19	0,20	0,21	0,20	0,20	0,20	0,21	0,20	0,20	<b>0,20</b>
Micro-averaged precisic	0,80	0,81	0,80	0,79	0,80	0,80	0,80	0,79	0,80	0,80	<b>0,80</b>
Micro-averaged recall	1,00	1,00	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	<b>1,00</b>
Micro-averaged F1-scor	0,89	0,89	0,89	0,88	0,89	0,89	0,89	0,88	0,89	0,89	<b>0,89</b>
Subset accuracy	0,42	0,42	0,41	0,41	0,42	0,41	0,42	0,41	0,41	0,41	<b>0,41</b>
Label Ranking Loss	0,58	0,58	0,59	0,59	0,58	0,59	0,58	0,59	0,59	0,59	<b>0,59</b>