UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS CENTRAL
PROGRAMA DE PÓS-GRADUAÇÃO EM DESIGN

Tainá Apoena Bueno de Oliveira

E-Triagem em Emergência: Modelo IA de Sistema para Suporte de Decisão Clínica

Florianópolis
2025

Tainá Apoena Bueno de Oliveira

E-Triagem em Emergência: Modelo IA de Sistema para Suporte de Decisão Clínica

Tese apresentada ao Programa de Pós-Graduação em Design da Universidade Federal de Santa Catarina para obtenção do título de Doutora em Design e Mídia.

Orientador(a): Prof. Milton Luiz Horn Vieira, Dr.
Coorientador(a): Profa. Sayonara de Fátima Faria Barbosa, Dra.

Florianópolis
2025

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Bueno de Oliveira, Taina Apoena
E-Triagem em Emergência : Modelo IA de Sistema para Suporte de Decisão Clínica / Taina Apoena Bueno de Oliveira ; orientador, Milton Luiz Horn Vieira, coorientadora, Sayonara de Fátima Faria Barbosa, 2025.
140 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, Programa de Pós Graduação em Design, Florianópolis, 2025.

Inclui referências.

1. Design. 2. Triagem eletrônica. 3. Interface. 4. Machine Learning. 5. Urgência e Emergência. I. Vieira, Milton Luiz Horn. II. Barbosa, Sayonara de Fátima Faria. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Design. IV. Título.

Tainá Apoena Bueno de Oliveira

E-Triagem em Emergência: Modelo IA de Sistema para Suporte de Decisão Clínica

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 7 de outubro de 2025 pela banca examinadora composta pelos seguintes membros:

Prof. Luiz Fernando Gonçalves de Figueiredo, Dr.
Universidade Federal de Santa Catarina - UFSC

Prof. Gabriel de Souza Prim, Dr.
Universidade Federal de Santa Catarina - UFSC

Prof. Milton José Cinelli, Dr.
Universidade do Estado de Santa Catarina - UDESC

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutora em Design.

Documento assinado digitalmente
Eugenio Andres Diaz Merino
Data: 03/11/2025 12:00:40-0300
CPF: ***.846.917-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Eugenio Andrés Díaz Merino, Dr.
Coordenação do Programa de Pós-Graduação

Documento assinado digitalmente
Milton Luiz Horn Vieira
Data: 03/11/2025 11:11:39-0300
CPF: ***.244.899-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Milton Luiz Horn Vieira, Dr
Orientador

Documento assinado digitalmente
SAYONARA DE FATIMA FARIA BARBOSA
Data: 03/11/2025 23:45:47-0300
CPF: ***.819.497-**
Verifique as assinaturas em https://v.ufsc.br

Profa. Sayonara de Fátima Faria Barbosa, Dra.
Coorientadora

Florianópolis, 2025.

Ao universo e
à minha família.

AGRADECIMENTOS

Agradeço ao universo pela vida.

Ao Prof. Dr. Milton Luiz Horn Vieira e. a Profa. Dra. Sayonara de Fátima Faria Barbosa pela orientação, pelos ensinamentos, pela confiança, pela paciência e, principalmente, por aceitar me orientar nessa pesquisa.

Aos membros da banca examinadora por terem aceitado o convite e colocado com excelência suas sugestões.

A todos os meus colegas de laboratório, principalmente ao Dr. Victor Nassar e ao Dr. Jonathan Nishida, a minha pesquisa não existiria sem antes vocês terem iniciado a pesquisa de vocês! E claro, por todo apoio, todas as risadas e jantares!

A todos os meus amigos por me apoiarem, me fazerem rir, me incentivarem, pelas conversas, pelas cervejas, por todo o companheirismo e amor.

Aos amigos Eduardo Zanella, Hérick Mendes e Cristina Link Rintzel, por todas as viagens, jantares, churrascos em entre inúmeras outras coisas, que fizeram meus dias mais leves, engraçados e divertidos.

Ao meu amor, Angela Santos pelo amor, paciência, ajuda, carinho, companheirismo, ensinamentos, inspiração e principalmente pela força em todos os momentos.

À minha mãe Carmen Bueno pela educação, pelo amor incondicional, pelo carinho, pela ajuda e por acreditar em mim e me apoiar em todas a mudanças.

À minha vó Joyce Bueno pelo amor, pelo carinho, pela torcida, pelo apoio e por todas as orações.

Ao meu padrinho Luiz Bueno e minha madrinha Luzia Bueno por toda ajuda durante minha jornada de vida e por serem apoio incondicional em todos os momentos.

Aos meus tios, tias, primos e primas pelo apoio e carinho até aqui.

Aos profissionais da saúde que se dispuseram a fazer parte da pesquisa, ensinando e compartilhando o seu conhecimento em prol da pesquisa.

Ao Eduardo Brandão por toda a ajuda na área de Machine Learning, sem ele essa tese não teria saído do jeito que saiu.

Aos governos pelas políticas de incentivo à educação, cultura e moradia que me proporcionaram toda a base necessária para chegar até aqui.

À UFSC e ao Programa de Pós-graduação em Design pela oportunidade.

À CAPES pela bolsa de doutorado e a Fapesc pelo financiamento do projeto.

Todas as grandes coisas são simples. E muitas podem ser expressas numa só palavra: liberdade; justiça; honra; dever; piedade; esperança (Winston Chirchill).

RESUMO

Os atendimentos feitos pela Urgência e Emergência (UE) dos hospitais devem apresentar, principalmente, agilidade de atendimento, sendo peça fundamental para isso a adoção de protocolos de estratificação de risco com o intuito de fornecer qualidade e acesso rápido ao atendimento. Por muitas vezes, no entanto, a equipe que faz a gestão da UE não consegue acompanhar os pacientes, mesmo utilizando os protocolos de estratificação, sem perceber a evolução da doença do paciente a tempo. Nesse sentido, a tecnologia se torna uma ferramenta útil para conectar profissionais da saúde e pacientes de forma a oferecer controle, autonomia e gerenciamento da UE. Diante disso, este estudo teve como objetivo desenvolver e avaliar uma interface para um sistema de e-Triagem, no intuito de integrar dados clínicos e multiparamétricos para apoiar a classificação de risco e a previsão de deterioração clínica. As etapas dessa pesquisa incluíram: (a) o desenvolvimento da interface fundamentada em princípios de *User Experience* (UX); (b) a avaliação da usabilidade do protótipo com profissionais de enfermagem; (c) a proposição de uma nova Classificação de Risco e Escores de Alertas para compor o sistema de e-Triagem; e (d) a aplicação e testagem de algoritmos de *Machine Learning* voltados à previsão da deterioração clínica e à geração de alertas de risco. Para isso, foram utilizados dados do MIMIC (*Medical Information Mart for Intensive Care*), um amplo banco de dados de livre acesso que reúne dados de saúde anonimizados de pacientes tratados no hospital *Beth Israel Deaconess Medical Center* (BIDMC) de Boston, Massachusetts, Estados Unidos da América. Os resultados demonstraram que a interface proposta apresentou elevada eficácia, com 100% de conclusão das tarefas propostas e tempos médios adequados ao fluxo de trabalho dos profissionais de saúde. A percepção dos usuários foi amplamente positiva, refletida por altos índices de usabilidade (*Net Promoter Score* de 75% e *System Usability Scale* de 79,7) e relatos qualitativos de facilidade de uso, intuitividade e impacto potencial no processo de triagem e no cuidado seguro ao paciente, embora ajustes visuais e de escopo sejam recomendados para reduzir sobrecarga cognitiva em telas iniciais. Quanto à classificação de risco, a nova classificação proposta (NEWS2-M) mostrou desempenho superior ao NEWS2 original na detecção precoce de deterioração clínica, reclassificando mais de 75% dos pacientes inicialmente considerados de baixo risco e ampliando a granularidade da avaliação. A comparação com o escore *Emergency Severity Index* (ESI) evidenciou alinhamento nos extremos de risco e redistribuição dos casos intermediários, contribuindo para maior precisão da estratificação clínica e redução de limitações de subtriagem. Por fim, diferentes algoritmos de *Machine Learning* foram testados para prever a deterioração clínica utilizando o conjunto de dados do MIMIC; as análises demonstraram a superioridade do modelo LightGBM em relação aos demais, destacando-se pelo melhor desempenho em métricas de discriminação, calibração e eficiência computacional, além de apresentar ótimo equilíbrio entre sensibilidade e taxa de falsos positivos, alinhado às demandas clínicas. Portanto, conclui-se que a interface de e-Triagem proposta, aliada à nova classificação de risco (NEWS2-M) e ao suporte de algoritmos de *Machine Learning*, mostrou-se uma solução promissora para otimizar o processo de triagem, ampliar a detecção precoce da deterioração clínica e apoiar decisões mais seguras e assertivas na gestão da Urgência e Emergência.

Palavras-chave: Triagem eletrônica; Interface; *Machine Learning*. Urgência e Emergência; Classificação de risco.

ABSTRACT

Emergency Department (ED) services must primarily ensure agility in patient care, for which the adoption of risk stratification protocols is essential to provide quality and timely access to treatment. Frequently, however, ED management teams are unable to adequately monitor patients, even when using these protocols, failing to detect disease progression in time. In this context, technology emerges as a useful tool to connect healthcare professionals and patients, enabling greater control, autonomy, and management of ED processes. This study aimed to develop and evaluate an interface for an e-Triage system designed to integrate clinical and multiparametric data in order to support risk classification and the prediction of clinical deterioration. The research included: (a) development of the interface grounded in User Experience (UX) principles; (b) usability assessment of the prototype with nursing professionals; (c) proposition of a new Risk Classification and Alert Scoring system to be incorporated into e-Triage; and (d) application and testing of Machine Learning algorithms for predicting clinical deterioration and generating risk alerts. Data from MIMIC (Medical Information Mart for Intensive Care), a large open-access database containing anonymized health records of patients treated at Beth Israel Deaconess Medical Center (BIDMC), Boston, USA, were used. The results showed that the proposed interface achieved high efficacy, with 100% task completion and average times consistent with healthcare workflows. User perception was highly positive, reflected in high usability indices (Net Promoter Score of 75% and System Usability Scale of 79.7), and qualitative reports highlighted ease of use, intuitiveness, and potential impact on triage and patient safety, although visual and scope adjustments were recommended to reduce cognitive overload on initial screens. Regarding risk classification, the new score (NEWS2-M) outperformed the original NEWS2 in early detection of clinical deterioration, reclassifying more than 75% of patients initially considered low risk and enhancing assessment granularity. Comparison with the Emergency Severity Index (ESI) score revealed consistency at the risk extremes and redistribution of intermediate cases, contributing to more precise clinical stratification and reducing under-triage limitations. Finally, Machine Learning analyses demonstrated the superiority of the LightGBM model over others, showing better discrimination, calibration, and computational efficiency, as well as a strong balance between sensitivity and false-positive rates, consistent with clinical requirements. In conclusion, the proposed e-Triage interface, combined with the new risk classification (NEWS2-M) and Machine Learning support, proved to be a promising solution to optimize triage, enhance early detection of clinical deterioration, and support safer, more accurate decision-making in Emergency Department.

Keywords: Electronic triage; Interface; Machine Learning. Emergency Department; Risk classification.

LISTA DE FIGURAS

Figura 1 – Categorias de prioridade do Protocolo Manchester 26
Figura 2 – *Emergency Severity Index* (ESI) Algoritmo de Triagem v5 27
Figura 3 – Pontuação do NEWS2 31
Figura 4 – Limites e gatilhos 32
Figura 5 – Metas de Usabilidade de Nielsen 41
Figura 6 – Validação Antecipada e Frequente 43
Figura 7 – Agentes e Fluxo de Atribuições 51
Figura 8 – Processo do Desenvolvimento Conceitual 52
Figura 9 – Mapa da Jornada do Enfermeiro(a) 53
Figura 10 – Módulo Enfermeiro 54
Figura 11 – Módulo Dashboard 55
Figura 12 – Módulo de Identificação do Paciente e Consulta em Andamento 56
Figura 13 – Módulo de Alerta Precoce e Gráfico de Estado 56
Figura 14 – Módulo de Exames, Editar Prontuário, Cartões de Informações Rápidas e Consultas Passadas 57
Figura 15 – Dashboard 58
Figura 16 – Lista de paciente 59
Figura 17 – Prontuário do Paciente 60
Figura 18 – Chamar Paciente 61
Figura 19 – Notificação de Atenção 62
Figura 20 – Atendimento do Paciente em Atenção 63
Figura 21 – Protótipo do Dashboard 64
Figura 22 – Protótipo do Registro de Paciente 65
Figura 23 – *Software* de triagem eletrônica (e-triagem) 67
Figura 24 – Valores calculados do *Net Promoter Score* (em A) e do *System Usability Scale* (em B) 89
Figura 25 – Parâmetros do NEWS2 Modificado 92
Figura 26 – Comparação de Modelos por PR-AUC 102
Figura 27 – Comparação de Modelos por AUROC 102
Figura 28 – Comparação de Modelos por *Brier Score* 103
Figura 29 – Comparação de Modelos por *F1-Score* 104
Figura 30 – Relação entre Performance (AUROC) x Tempo de Execução 105

Figura 31 – *Trade-off: Precision x Recall* com isolinhas de *F1-Score* 107
Figura 32 – Matriz de Confusão: a) XGB, b) CatBoost e c) LR 108
Figura 33 – Matriz de Confusão: SVM 109
Figura 34 – Matriz de Confusão: LGBM 110
Figura 35 – Mapa da Jornada do Atendente 135

LISTA DE TABELAS

Tabela 1 – Componentes e descrições das Tarefas 70
Tabela 2 – Tempo (em segundos) para a realização das tarefas pelos participantes. T1: Tarefa 1; T2: Tarefa 2; T3: Tarefa 3; T4: Tarefa 4..... 86
Tabela 3 – Número de toques para a realização das tarefas pelos participantes. T1: Tarefa 1; T2: Tarefa 2; T3: Tarefa 3; T4: Tarefa 4; TD: Número de toques desnecessários; TT: Número de toques totais (necessários e desnecessários)..... 87
Tabela 4 – Descrição das respostas clínica dos desencadeadores do NEW2-Modificado e sua relação com a classificação Emergency Severity Index (ESI)..... 93
Tabela 5 – Tabela de contingência: correlação NEWS2-M e NEWS2 Original, e risco clínico do NEWS2-M. N: Quantidade de pessoas (números absolutos); %: porcentagem. 95
Tabela 6 – Tabela de contingência: correlação NEWS2-M e ESI, e risco clínico do NEWS2-M. N: Quantidade de pessoas (números absolutos); %: porcentagem. 98
Tabela 7 – Resultado do *Decision Threshold* 112

LISTA DE ABREVIATURAS E SIGLAS

AB	Atenção Básica
AUC	*Area Under the Curve*
AUROC	*Area Under the Receiver Operating Characteristic Curve*
BIDMC	*Beth Israel Deaconess Medical Center*
CAAE	Certificado de Apresentação de Apreciação Ética
CDSS	*Clinical Decision Support System* (Sistema de Suporte à Decisão Clínica)
CFM	Conselho Federal de Medicina
CID-10	Classificação Internacional de Doenças
COFEN	Conselho Regional de Enfermagem
CR	Classificação de Risco
CTI	Centro de Tratamento Intensivo
EA	Escores de Alertas
ENA	Emergency Nurses Association
ESC	*European Society of Cardiology*
ESD	Estratégia de Saúde Digital
ESI	*Emergency Severity Index*
EWS	*Early Warning Score*
FiO_2	Fração Inspirada de Oxigênio
FN	*False Negatives* (Falsos Negativos)
FNR	*False Negative Rate* (Taxa de Falsos Negativos)
FP	*False Positives* (Falsos Positivos)
FPR	*False Positive Rates* (Taxa de Falsos Positivos)
GaussianNB	*Gaussian Naïve Bayes*
GBDT	*Gradient Boosting Decision*
GCS	*Glasgow Coma Scale*
IA	Inteligência Artificial
IDH	*Isolated Diastolic Hypertension* (Hipertensão Diastólica Isolada)
IOT	*Internet of Things*
k-NN	*k-Nearest Neighbors*
KPNC	*Kaiser Permanente Northern California*
L1 (LASSO)	*Logistic Regression 1 LASSO*

L2 (RIDGE)	*Logistic Regression 1 RIDGE*
LAI	Lei de Acesso à Informação
LGBM	*Light Gradient Boosting Machine*
LightGBM	*Light Gradient Boosting Machine*
LinearSVC	*Linear Support Vector Classification*
LOINC	*Logical Observation Identifiers, Names and Codes*
LR	*Logistic Regression*
MIMIC-IV	*Medical Information Mart for Intensive Care IV*
MIMIC-IV ED	*Medical Information Mart for Intensive Care IV - Emergency Department*
MIT	*Massachusetts Institute of Technology* (Instituto de Tecnologia de Massachusetts)
ML	*Machine Learning*
MS	Ministério da Saúde
MS	Ministério da Saúde
MTS	*Manchester Triage System* (Protocolo de Manchester)
MVP	Mínimo Produto Viável
NEWS	*National Early Warning Score*
NEWS2	*National Early Warning Score 2*
NEWS2-M	NEWS2 Modificado
NHS	*National Health Service*
NPS	*Net Promoter Score*
OBS	*Open Broadcaster Software*
OMS	Organização Mundial da Saúde
PA	Pressão Arterial
PAD	Pressão Arterial Diastólica
PAM&A	Plano de Ação, Monitoramento e Avaliação Estratégica de Saúde Digital
PAS	Pressão Arterial Sistólica
PEP	Prontuário Eletrônico do Paciente
PMT	Protocolo Manchester de Triagem
PNIIS	Política Nacional da Informação e Informática em Saúde
PR-AUC	*Area Under the Precision-Recall Curve*
PRD	*Product Requirements Document* (Documentos de requisitos do produto)
qSOFA	*Quick Sequential Organ Failure Assessment*

RF	*Random Forest*
RNDS	Rede Nacional de Dados em Saúde
ROC	Receiver Operating Characteristic
RUE	Rede de Atenção às Urgências e Emergências
RxNorm	*Standardized, Nomenclature for Clinical Drugs*
SAS	Secretaria de Atenção à Saúde
SIRS	*Systemic Inflammatory Response Syndrome*
SMOTE	*Synthetic Minority Over-sampling Technique*
SNIS	Sistema Nacional de Informação em Saúde
SNOMED CT	*Systematized Nomenclature of Medicine - Clinical Terms*
SOFA	*Sequential Organ Failure Assessment*
SpO2	Saturação
SUS	Sistema Único de Saúde
SUS	*System Usability Scale*
SVM	*Support Vector Machine*
SVS/MS	Secretaria de Vigilância em Saúde do Ministério da Saúde
TFP	*True False Positives* (Taxa de Falsos Positivos)
TIC	Tecnologias de Informação e Comunicação
TICS	Tecnologia da Informação e Comunicação em Saúde
TN	*True Negatives* (Verdadeiros Negativos)
TNR	*True Negative Rate* (Taxa de Verdadeiros Negativos)
TP	*True Positives* (Verdadeiros Positivos)
TPR	*True Positive Rates* (Taxa de Verdadeiros Positivos)
TPS	*Toyota Production System*
TruncatedSVD	*Truncated Singular Value Decomposition*
UBS	Unidades Básicas de Saúde
UE	Urgências e Emergências
UH	Unidades Hospitalares
UI	*User Interface*
UPA	Unidades de Pronto-Atendimento
UTI	Unidade de Terapia Intensiva
UX	*User Experience* (Experiência do Usuário)
XGB	*Extreme Gradient Boosting*
XGBoost	*Extreme Gradient Boosting*

SUMÁRIO

1 INTRODUÇÃO 15
1.1 PROBLEMA DE PESQUISA 19
1.2 OBJETIVOS 19
1.2.1 Objetivo geral 19
1.2.2 Objetivos específicos 20
1.3 JUSTIFICATIVA 20
2 REVISÃO BIBLIOGRÁFICA 23
2.1 POLÍTICAS REGULAMENTADORAS DA SAÚDE BRASILEIRA 23
2.2 INFORMÁTICA EM SAÚDE 25
2.2.1 Urgência/Emergência e Protocolos de Triagem 25
2.2.2 Escores de Alerta Precoce 29
2.2.3 Dados do registro de atendimento clínico 33
2.3 INTELIGÊNCIA ARTIFICIAL NA SAÚDE 34
2.3.1 *Machine Learning* como ferramenta no suporte à decisão clínica 35
2.3.2 Base de dados MIMIC-IV 37
2.4 USABILIDADE E EXPERIÊNCIA DO USUÁRIO 39
2.4.1 Metas de usabilidade 39
2.4.2 *Lean* UX 41
2.4.3 PRD – Documentos de requisitos do produto 43
2.4.4 Wireframes, dashboards e protótipos 44
2.5 PULSEIRA MULTIPARAMÉTRICA 45
3 PROCEDIMENTOS METODOLÓGICOS 47
3.1 CARACTERIZAÇÃO DA PESQUISA 47
3.1.1 Primeira Etapa: Revisão da Literatura 47
3.1.2 Segunda Etapa: Investigação Qualitativa do Contexto Clínico 47
3.1.3 Terceira Etapa: Desenvolvimento do Modelo de Classificação e Monitoramento 47
3.1.4 Quarta Etapa: Desenvolvimento do Protótipo Funcional 48
3.1.5 Quinta Etapa: Avaliação de Usabilidade 48
3.1.6 Sexta Etapa: Análise e Validação dos Modelos 49
3.2 DESENVOLVIMENTO DA INTERFACE 49
3.2.1 Participantes 49

3.2.2 Questionário Pré-teste 49
3.2.3 Agentes e suas atribuições no sistema de triagem 50
3.2.4 Desenvolvimento conceitual da interface do sistema 51
3.2.5 Enfermeiro(a): mapa da jornada 52
3.2.6 Módulos 53
3.2.7 Wireframes 57
3.2.8 Protótipo de baixa fidelidade 63
3.3 TESTES DE USABILIDADE 66
3.3.1 Materiais 66
3.3.2 Sistema de triagem eletrônica 66
3.3.3 Local, layout e papel do pesquisador 67
3.3.4 Testes de usabilidade aplicados 68
3.3.5 Questões da pesquisa 68
3.3.6 Tarefas 69
3.3.7 Lista de tarefas 69
3.3.8 Cenário das tarefas 70
3.3.9 Questionário pós-teste 71
3.3.10 Dados a serem coletados e medidas de avaliação 73
3.4 MODIFICAÇÃO DO ESCORE DE ALERTA PRECOCE – NEWS2-M 73
3.5 MACHINE LEARNING 76
3.5.1 Delineamento do Estudo e Definição do Problema 76
3.5.2 Condução dos experimentos 77
3.5.3 Pré-processamento e Engenharia de Atributos 78
3.5.4 Estratégia de Modelagem 79
3.5.5 Seleção e Justificativa dos Algoritmos 81
3.5.6 Avaliação e Calibração dos Modelos 82
4 RESULTADOS E DISCUSSÃO 85
4.1 TESTES DE USABILIDADE COM PROFISSIONAIS DA SAÚDE 85
4.2 SISTEMA NEWS MODIFICADO 92
4.2.1 Correlação NEWS2-M + NEWS2 Original 93
4.2.2 Correlação do NEWS2-M com o ESI 96
4.3 MACHINE LEARNING 101
4.3.1 Análise Comparativa de Desempenho dos Modelos 101
4.3.2 Análise de Métricas de Calibração e Erro 103

4.3.3 Trade-off entre Performance e Custo Computacional 105
4.3.4 Análise entre Precision e Recall 106
4.3.5 Análise Inferida da Matriz de Confusão e Implicações dos Tipos de Erro 107
4.3.6 Seleção de Decision Thresholds para modelos preditivos 110
4.3.7 Análise de Performance dos Modelos 111
5 CONCLUSÃO 114
REFERÊNCIAS 117
APÊNDICE A – QUESTIONÁRIO PRÉ-TESTE 134
APÊNDICE B – MAPA DA JORNADA DO ATENDENTE 135
ANEXO A – CÓDIGO MACHINE LEARNING 136

1 INTRODUÇÃO

Atualmente, a conectividade e a tecnologia desempenham um papel contundente na vida cotidiana. Na saúde, a tecnologia atua conectando médicos e pacientes de forma a oferecer com transparência o controle, a autonomia e o gerenciamento de dados, no intuito de produzir resultados melhores no ecossistema da saúde (Weinstock, 2016). Iniciativas para a informatização na área da saúde buscam oferecer benefícios como: “acesso rápido às informações de saúde e intervenções realizadas, melhoria na efetividade do cuidado e possível redução de custos com otimização dos recursos além de aprimorar e automatizar o processo de envio de informações” às bases de dados do governo (Brasil; Ministério da Saúde, 2022).

Em 2021, por meio da Política Nacional da Informação e Informática em Saúde (PNIIS), resolução nº 659 (atualização da Portaria nº 589 de 2015), que foi uma das estratégias iniciais do governo para modernização do Sistema Nacional de Informação em Saúde (SNIS) tinha como essência a utilização de modernas Tecnologias de Informação e Comunicação (TIC) com foco na eficiência e efetividade para aumentar e ampliar a qualidade de acesso à saúde. Entre as melhorias pretendidas, inclui-se a qualificação das equipes, do atendimento e do fluxo de informações, para que pudesse apoiar o Sistema de Suporte à Decisão Clínica em Saúde (CDSS – em inglês *Clinical Decision Support System*) (Brasil; Ministério da Saúde, 2015, 2021).

Em 2020, começa a Estratégia de Saúde Digital do Departamento de Informática do SUS (DataSUS) da Secretaria Executiva do Ministério da Saúde (MS) que cria o Plano de Ação, Monitoramento e Avaliação Estratégica de Saúde Digital (PAM&A) para o Brasil utilizando o termo Saúde Digital; este termo compreende “uso de recursos de Tecnologias de Informação e Comunicação (TICs) para produzir e disponibilizar informações confiáveis, sobre o estado de saúde para quem precisa, no momento que precisa” utilizando-se dos avanços tecnológicos na área de Internet das Coisas (IoT – em inglês *Internet of Things*) e Inteligência Artificial (IA) (Brasil; Ministério da Saúde, 2020a). O PAM&A segue as diretrizes do *National eHealth Strategy Toolkit* desenvolvido pela Organização Mundial da Saúde (OMS) com o intuito de compartilhar com os países o emprego e o uso das TICs na área da saúde (*World Health Organization and International Telecommunication Union,* 2012).

Como parte da Estratégia de Saúde Digital (ESD) foi criada a Rede Nacional de Dados em Saúde (RNDS), uma plataforma informacional de integração de dados em saúde para ser utilizada em inovação, pesquisa e desenvolvimento do Brasil e o Informatiza APS, um Programa de Apoio à Informatização e Qualificação dos Dados da Atenção Primária à Saúde, pela portaria nº 2.983, de 11 de novembro de 2019 (Brasil; Ministério da Saúde, 2019). Estas regulamentações do Ministério da Saúde (MS), portanto, contemplam diretrizes que tratam sobre a adoção de tecnologia na área da saúde, além de esforços para categorizar e otimizar o fluxo de atendimento, por meio do uso de estratégias que compreendem nesse processo, a identificação, a priorização e a integração de programas, projetos e ações de saúde em toda a extensão do Sistema Único de Saúde (SUS) — sendo eles, Atenção Básica (AB), Unidades de Pronto-Atendimento (UPA) e Unidades Hospitalares (UH) e Unidades Básicas de Saúde (UBS).

Contudo, grandes desafios deverão ser vencidos para que essas tecnologias sejam implementadas de forma correta e segura. De um lado, tem-se toda a estruturação e integração tecnológica que deverá ser realizada em um país em desenvolvimento e com uma grande área de extensão territorial; de outro lado, há o desafio em termos do próprio uso de IoT e IA pelo profissional da saúde, barreira no desenvolvimento de hardware e de software que perpassa campos multidisciplinares da ciência. Nesse contexto, o Design aliado à IA pode ser utilizado para contribuir na área da saúde, tanto para os profissionais quanto para os pacientes que se utilizam de Urgências e Emergências (UE) públicas. Como ferramenta para auxiliar nessa questão, tem-se o Design Centrado na Experiência do Usuário (UX – em inglês *User Experience*), que busca “atender às necessidades exatas do cliente sem confusão ou incômodo” no processo de criação de um produto (Don Norman; Jakob Nielsen, 2022).

O Design UX é, portanto, um processo iterativo de entendimento dos usuários em seu contexto a partir do qual se dá a criação do produto. Logo, seu papel pode ser utilizado como ferramenta no amparo ao CDSS, fornecendo aos profissionais da saúde e pacientes o conhecimento e as informações específicas, filtradas de forma inteligente ou apresentadas em momentos apropriados, para melhorar a saúde e os cuidados de saúde. O CDSS também “engloba ferramentas que aprimoram tomadas de decisão no fluxo de trabalho clínico, que incluem alertas e lembretes computadorizados para profissionais de saúde e pacientes” além de “diretrizes clínicas; pedidos específicos de condições; relatórios e resumos de dados de

pacientes; modelos de documentação; suporte de diagnóstico e informações de referência contextualmente relevantes" (USA; *HealthIT*, 2022).

A partir disso, pode-se utilizar a IA para que um sistema aprenda com um conjunto de dados e use uma variedade de algoritmos para melhorar, descrever e prever resultados (Judith Hurwitz; Daniel Kirsch, 2018). Neste caso, busca-se obter um maior entendimento da situação clínica do paciente, a partir de dados da saúde fornecidos e interações do usuário com a interface pode ser relacionado com o entendimento da situação clínica do paciente, criando modelos de Aprendizagem de Máquina (ML - em inglês *Machine Learning*) que poderão ser apresentados de forma estruturada pela Interface do Usuário (UI – em inglês *User Interface*) parte integrantes do design UX.

A utilização de tecnologias baseadas em dados, algoritmos e IA é especialmente importante para a saúde no contexto das Urgências e Emergências (UEs). Nos casos atendidos pelas UEs, a agilidade no atendimento é fundamental e a adoção de protocolos de estratificação de risco impacta na qualidade do acesso ao serviço. Logo, esses protocolos buscam oferecer atendimento prioritário com escuta qualificada, avaliação de necessidades de saúde e análise de vulnerabilidade que, como resultado, permitem atendimento prioritário a doenças agudas potencialmente fatais (Brasil; Ministério da Saúde, 2012, 2013). Como exemplo desses protocolos, tem-se o Índice de Gravidade de Emergência (ou ESI, sigla de *Emergency Severity Index*), um dos principais algoritmos de triagem de emergências utilizados no mundo; e o Protocolo Manchester de Triagem (ou MTS, sigla de *Manchester Triage System)*, principal protocolo utilizado nas emergências públicas do Brasil – ambos os protocolos classificam pacientes em cinco níveis de urgência e prioridade de atendimento (Ausserhofer et al., 2021).Contudo, protocolos como estes dependem muito do julgamento do profissional da triagem, e, portanto, estão sujeitos a uma grande variabilidade de decisões (Christ et al., 2010; Costa et al., 2020). Além disso, esses protocolos de triagem são muitas vezes falhos em diferenciar pacientes gravemente doentes e em detectar piora no estado clínico dos que estão em espera e já classificados (Hitchcock et al., 2014; Yurkova; Wolf, 2011). Por outro lado, protocolos como o *National Early Warning Score 2* (NEWS2) podem auxiliar a preencher essa lacuna na classificação. O NEWS2 é um sistema de alerta precoce de rastreamento da deterioração de pacientes, que vem sendo usado cada vez mais internacionalmente (Smith et al., 2013; Welch; Dean; Hartin, 2022). Aliado a isso, a

integração do NEWS2 com a automatização de sinais vitais e com a inteligência artificial demonstra aumentar ainda mais a precisão preditiva e a eficiência operacional em prever eventos críticos em pacientes de risco (Huete-Garcia; Rodriguez-Lopez, 2024).

Como parte das iniciativas para a informatização da saúde, o grupo de pesquisa DesignLab da UFSC projetou uma pulseira multiparamétrica para aplicação em ambientes hospitalares das UEs. O dispositivo realiza a aferição contínua de sinais vitais — pressão arterial, frequência cardíaca, temperatura corporal e oximetria de pulso — e transmite os dados em tempo real para uma interface dedicada. Este produto tem como finalidade aprimorar o monitoramento constante do paciente, otimizando a classificação de risco e agilizando a resposta clínica com base em seu estado momentâneo (Nishida, 2023).

Diante do exposto, a presente pesquisa buscou, inicialmente, desenvolver uma interface de prontuário eletrônico que apresentasse os dados oriundos da pulseira multiparamétrica desenvolvida por Nishida (2023) de forma visual, com a finalidade de fazer a triagem da UE de forma automatizada (e-Triagem ou Triagem Eletrônica). Após isso, a usabilidade de um protótipo dessa interface foi testada com profissionais da área da saúde. Paralelamente, esta pesquisa também propôs novos parâmetros de Classificação de Risco (CR) e Escores de Alertas (EA), onde desenvolveu-se um sistema de e-Triagem potencialmente útil para ser acoplado aos dados da pulseira multiparamétrica, com a proposta da reclassificação automática dos pacientes. Para isso, foi analisado um conjunto de dados de sinais vitais proveniente de pacientes classificados com o *Emergency Severity Index* (ESI), onde esses dados foram utilizados para testar uma nova CR com a adaptação do sistema de alerta precoce NEWS2. A partir da definição dos parâmetros de CR e EA, criou-se a classificação de risco NEWS2-M (ou NEWS2-modificado), e o conjunto de dados oriundos do ESI foram reclassificados utilizando essa nova classificação. Por fim, este projeto também avaliou o uso da Inteligência Artificial, no campo da ciência conhecido como *Machine Learning*, para prever e detectar, a partir da deterioração dos sinais vitais, o risco de piora no estado clínico de pacientes.

Desse modo, esta pesquisa busca preencher a lacuna deixada pelos sistemas de triagem tradicionais, que são estáticos e subjetivos, ao propor e avaliar um ecossistema tecnológico que integra monitoramento contínuo, uma nova classificação

de risco e inteligência artificial para uma detecção preditiva e dinâmica da deterioração clínica.

Cabe ressaltar que a presente pesquisa se insere no contexto de um projeto mais amplo desenvolvido no âmbito do grupo de pesquisa DesignLab da UFSC. A contribuição original desta tese concentra-se especificamente na concepção da interface de usuário (UI/UX) para o prontuário eletrônico, na proposição e validação do escore de alerta precoce modificado (NEWS2-M) e na exploração do componente de Inteligência Artificial para suporte à decisão clínica. Desta forma, a Pulseira Multiparamétrica desenvolvida por Nishida (2023) é aqui tratada como um dispositivo de entrada de dados permitindo que o foco da investigação recaia sobre a interpretação e apresentação da informação para otimizar o processo de triagem na Urgência e Emergência.

1.1 PROBLEMA DE PESQUISA

Esta pesquisa se propôs a responder o seguinte problema: Como desenvolver um sistema de triagem eletrônica automatizada, fundamentado na experiência do usuário (UX), para apoiar o Sistema de Suporte à Decisão Clínica (CDSS), utilizando Inteligência Artificial (IA) na área de *Machine Learning* (ML) a fim de criar correlações entre alterações de sinais vitais do paciente e possibilidades de tratamento na Urgência e Emergência.

1.2 OBJETIVOS

1.2.1 Objetivo geral

Propor e validar uma estratégia de e-Triagem para o monitoramento e predição da deterioração clínica em tempo real. O estudo propõe criar uma estratégia que integre a aquisição de dados clínicos e multiparamétricos, advindos de um dispositivo vestível (pulseira), com modelos preditivos baseados no escore NEWS2-M e algoritmos de *Machine Learning* para estratificação de risco, e uma interface de visualização dedicada (e-Triagem) centrada no usuário (UX). O propósito é desenvolver um ecossistema integrado onde a IA identifica padrões de risco complexos, a interface os traduz em informações clinicamente acionáveis e de baixo atrito cognitivo e, desta forma, avaliar o impacto desta integração na acurácia da estratificação de risco, na redução da carga cognitiva dos profissionais e na otimização do tempo de resposta clínica.

1.2.2 Objetivos específicos

a) Desenvolver a interface fundamentada na experiência do usuário (UX) para um sistema de e-Triagem a ser alimentado com dados de uma pulseira multiparamétrica;
b) Testar a usabilidade do protótipo da interface com enfermeiros;
c) Propor parâmetros de Classificação de Risco e Escores de Alertas, e, a partir deles, criar um sistema de e-Triagem potencialmente útil para ser acoplado aos dados da pulseira multiparamétrica;
d) Aplicar algoritmos de *Machine Learning* para prever a deterioração do estado clínico do paciente e alertar para risco de piora.

1.3 JUSTIFICATIVA

A crescente demanda por serviços de Urgência e Emergência (UE) é um fenômeno mundial, observado não só em países em desenvolvimento, mas também em nações de primeiro mundo, como Estados Unidos, Canadá, Dinamarca, Itália, França, entre outros (Morley et al., 2018; Pines et al., 2011). No Brasil, como consequência da alta procura pelo serviço, há frequentemente uma superlotação das emergências, o que torna a assistência a esse serviço de difícil prática pelos profissionais, interferindo também na segurança do paciente (Silva et al., 2016).

Neste cenário, os protocolos de Classificação de Risco (CR) surgiram como ferramenta para o acolhimento hospitalar, avaliação inicial dos pacientes e determinação da necessidade de um atendimento mais urgente. Entre os protocolos, o Protocolo Manchester de Triagem (PMT) é utilizado em diversos países e é o principal protocolo utilizado nas emergências públicas do Brasil, o qual distribui os níveis de gravidade em cinco cores, que tria os pacientes distribuindo a gravidade dos quadros nas cinco cores e estabelece um limite de tempo para que recebam o atendimento (Christ et al., 2010; Kevin Mackway-Jones; Janet Marsden; Jill Windle, 2014). Além deste, outros protocolos de triagem utilizados mundialmente são o *Australian Triage Scale* (Escala de Triagem da Australásia), *Canadian Emergency Department Triage and Acuity Scale* (Escala Canadense de Triagem e Acuidade) e o *Emergency Severity Index* (Índice de Gravidade de Emergência); todos esses protocolos, assim como o Manchester, também utilizam uma classificação em 5 níveis de prioridade (Ausserhofer et al., 2021).

No Brasil, segundo dados da Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS), após avaliação do perfil epidemiológico e demográfico brasileiro, evidencia-se que os principais problemas de saúde dos usuários na área de UE estão relacionados a alta morbimortalidade de doenças do aparelho circulatório, como o Infarto Agudo do Miocárdio e Acidente Vascular Cerebral (Brasil; Ministério da Saúde, 2013). Nos casos atendidos pela UE, a agilidade no atendimento é fundamental e a adoção de protocolos de estratificação de risco impactam na qualidade do acesso ao serviço. Logo, esses protocolos, buscam oferecer atendimento prioritário com escuta qualificada, avaliação de necessidades de saúde e análise de vulnerabilidade que, como resultado, permite atendimento prioritário a doenças agudas potencialmente fatais (Brasil; Ministério da Saúde, 2012, 2013). De fato, há associação entre a classificação na triagem e os desfechos clínicos, sendo a mortalidade associada aos atendimentos urgentes e a alta, aos não urgentes (Oliveira et al., 2016).

Apesar da importância destes protocolos, os processos manuais de triagem, como neste caso do PMT, são baseados no julgamento e na experiência humana. Dessa forma, com a prática clínica manual de triagem, podem surgir problemas de inexatidão, devido a erros de classificação, treinamento inconsistente do pessoal responsável, no âmbito da experiência e competência do pessoal, subjetividade e a tendência à subtriagem em UEs saturadas (Hitchcock et al., 2014; Sánchez-Salmerón et al., 2022; Zachariasse et al., 2018). Devido a isso, modelos de pontuações de alerta precoce (os chamados *Early Warning Score*) têm sido testados em diversos hospitais pelo mundo, inclusive no Brasil, utilizado no Distrito Federal (GDF; SES, 2024), para auxiliar na detecção de pacientes críticos ou potencialmente críticos precocemente, no intuito de permitir cuidados precoces e diminuir as taxas de mortalidade hospitalar (Huete-Garcia; Rodriguez-Lopez, 2024; Welch; Dean; Hartin, 2022).

Aliado aos sistemas de alerta precoce, outra maneira de minimizar o erro humano em procedimentos convencionais de triagem para urgências e emergências pode ser com sistema de inteligência artificial, utilizando o *Machine Learning* (ML) (Stewart; Sprivulis; Dwivedi, 2018). A utilização de ML, com algoritmos incorporados em sua programação, pode ajudar a identificar, por exemplo, padrões na entrada de dados – a grande vantagem clínica é que eles podem ser projetados para fazer previsões de registros de saúde, analisando uma ampla gama de condições médicas, ajudando no diagnóstico ou condições crônicas e níveis de gravidade das doenças (Moulik; Kotter; Fishman, 2020). Além disso, se alimentado em tempo real com novos

dados do paciente durante a triagem, o sistema poderia também ajudar na tomada de decisões e na própria classificação da triagem. Para além, com a nova Política Nacional de Informação e Informática em Saúde (PNIIS) de 2021, e a nova Estratégia de Saúde Digital (ESD) do PAM&A, há o endosso da Organização Mundial de Saúde (OMS) para a utilização de tecnologia para identificar, priorizar e integrar, de forma coordenada a resolução dos problemas no sistema de saúde utilizando Internet das Coisas e Inteligência Artificial. Além destes, investimentos e modernizações em redes de atenção à saúde justificam-se principalmente no que se refere aos serviços das UEs, que passa a ser o primeiro local de busca do usuário doente nos casos em que a Unidade Básica de Saúde (UBS) não atende.

A relevância social desta pesquisa está em sua abordagem a um desafio crítico da saúde: a detecção tardia da deterioração de pacientes em unidades de Urgência e Emergência. Este problema é frequentemente intensificado por um conjunto de fatores que inclui a superlotação, a subjetividade inerente aos protocolos de triagem tradicionais, a baixa granularidade dos dados coletados e aos amplos intervalos entre os acompanhamentos de triagem. O ineditismo do trabalho está na combinação de soluções para enfrentar essas lacunas de forma integrada. O diferencial da pesquisa é a proposição de uma estratégia de e-Triagem para o monitoramento e predição da deterioração clínica em tempo real. Esta abordagem visa melhorar o rastreamento do paciente com exatidão e agilidade fornecendo uma base de dados objetiva e contínua. A não trivialidade da solução é demonstrada pela complexidade técnica de sua arquitetura e pelo rigor metodológico exigido para validar um sistema que integra hardware, software e modelos preditivos em um ambiente de alta pressão.

A presente tese foi desenvolvida dentro do programa de pós-graduação em Design da Universidade Federal de Santa Catarina/UFSC, que possui as linhas de pesquisa em Gestão e Mídia, que é definida por pesquisas que envolvam interatividade, interação, usabilidade, informação e comunicação, e ênfase em tecnologia. Esta pesquisa está contida junto às pesquisas de gestão com características tecnológicas que envolvem prototipagem, processos e serviços para a área da saúde com foco na experiência do usuário.

2 REVISÃO BIBLIOGRÁFICA

No presente capítulo apresenta-se a contextualização teórica da pesquisa que está dividida em: políticas regulamentadoras de digitalização da saúde no âmbito nacional, protocolos de triagem em urgência e emergência, aplicação da informática na saúde e o uso de inteligência artificial na área saúde.

2.1 POLÍTICAS REGULAMENTADORAS DA SAÚDE BRASILEIRA

Em 2011, com o regulamento nº1.600 de 7 de julho, começa a reformulação da Política Nacional de Atenção às Urgências apresentada pelo Ministério da Saúde (MS) em 2003, que conduz uma nova estratégia de gestão e expansão da qualidade do Sistema Único de Saúde (SUS). Assim, a Secretaria de Atenção à Saúde (SAS) do MS implanta a Rede de Atenção às Urgências e Emergências (RUE) no SUS de prontidão às situações clínicas envolvidas na saúde e ao atual contexto de superlotação dos prontos-socorros (Brasil; Ministério da Saúde; Conasems, 2013).

Seguindo o Plano Nacional do setor público de saúde no contexto de Urgência e Emergência (UE), em 2016 foi institucionalizada a Política Nacional de Informação e Informática em Saúde (PNIIS) e atualizada em 2021 para o uso de tecnologia na área saúde que sejam norteadas por ações em Tecnologia da Informação e Comunicação em Saúde (TICS) em todo o sistema de saúde Brasileiro ao mesmo tempo a adesão às diretrizes da Lei de Acesso à Informação (LAI) e da Política de Governo Eletrônico (e-GOV). A tecnologia da informação pode compor decisões no atendimento às estratégias de decisão da gestão pública e no desenvolvimento técnico e científico da saúde em sincronia com as demais políticas sociais e econômicas do país (Brasil; Ministério da Saúde, 2016).

> PNIIS, de caráter orientador, apresenta princípios e diretrizes norteadores de uma organização institucional, tais como: a melhoria da qualidade e do acesso ao sistema de saúde brasileiro; a transparência e segurança da informação em saúde; o acesso à informação de saúde pessoal como um direito do cidadão; o suporte da informação para tomada de decisão por parte do gestor e profissional de saúde; e, por fim, o desenvolvimento institucional do SUS e de todo o sistema de saúde brasileiro, com ganhos de eficiência na redução do número de sistemas de informação em saúde existentes ou sua simplificação, gestão e formação de pessoas, aquisição de insumos, monitoramento e avaliação das ações, logística, pagamento e transferência de recursos e outros processos-meio (Brasil; Ministério da Saúde, 2016, p.7).

Conceitos como “e-Saúde” foram replicados a partir de propostas de políticas globais de informação em saúde como base para um Sistema Nacional de Informação

em Saúde (SNIS) no Brasil. Em consequência, a sistematização da informação em saúde com o objetivo de promover o uso inovador, criativo e transformador da tecnologia para melhorar os processos de trabalho em saúde beneficiando usuários e profissionais de saúde (Brasil; Ministério da Saúde, 2016).

Para o Conselho Federal de Medicina (CFM) na Resolução nº 2.077 de 2014 em seu artigo 2º e parágrafo único é "obrigatória" a implantação do acolhimento com Classificação de Risco (CR) para atendimento dos pacientes nos serviços hospitalares de Urgência e Emergência. O tempo de acesso do paciente à CR deve ser imediato, sendo necessário dimensionar o número de classificadores para atingir este objetivo. No Anexo I da mesma resolução o CFM discorre de forma particular sobre o acolhimento e CR, sendo:

> É obrigatória a implantação nos ambientes dos Serviços Hospitalares de Urgência e Emergência de um sistema de classificação de pacientes de acordo com a gravidade do agravo à saúde que apresentam, e que deve ser realizado por profissionais médicos ou enfermeiros capacitados. O paciente classificado por enfermeiro não pode ser liberado ou encaminhado a outro local sem ser consultado por médico. A classificação deve ser feita obrigatoriamente em local que assegure a privacidade e o sigilo do paciente, podendo este ter duas ou mais salas de classificação para os momentos de maior fluxo de atendimento, resguardadas as condições de privacidade. Ao chegar ao Serviço Hospitalar de Urgência e Emergência, o acesso dos pacientes ao Setor de Classificação de Risco deve ser imediato. Assim, o tempo de espera para ser classificado deverá tender a zero, com os tempos de espera diferenciais para acesso ao médico emergencista não ultrapassando, na categoria de menor urgência, 120 minutos. O médico poderá, no contato com o paciente, rever a classificação para fins de prioridade maior ou menor nas etapas subsequentes de atendimento. A classificação das necessidades urgentes do paciente define seu fluxo em todo o processo de atenção hospitalar e todos os setores devem estar cientes dessa definição. O bom uso da classificação afeta o fluxo e a qualidade assistencial dos pacientes do Serviço Hospitalar de Urgência e Emergência no conjunto da instituição hospitalar (Brasil; Conselho Federal de Medicina - CFM, 2014).

Portanto, o acolhimento com CR caracteriza-se “como uma das intervenções potencialmente decisivas na reorganização das portas de urgência e na implementação da produção de saúde em rede [...]” (Brasil; Ministério da Saúde, 2009, p.28).

No Brasil, os casos de acolhimento em urgência e emergência utilizado pelo SUS em Unidades Hospitalares (UH) e Unidade de Pronto Atendimento (UPA) — atendimento de complexidade intermediária entre Atenção Primária (AP) e Unidades

Hospitalares — aplica protocolos baseados no Protocolo Manchester de Triagem (PMT) (Brasil; Ministério da Saúde, 2009).

Segundo Goldzweig e Shekelle (2009) em casos de uso de Tecnologia da Informação e Comunicação (TIC) no setor bancário e no setor aeronáutico provaram ser eficazes na redução do erro humano, automatizando processos ineficientes e trabalhosos, impactando substancialmente na qualidade e segurança. O uso da TIC na Saúde tem como intuito melhorar a eficiência, a qualidade e a segurança da prestação de cuidados médicos reduzindo o erro médico (Goldzweig et al., 2009).

2.2 INFORMÁTICA EM SAÚDE

Oficialmente conhecida como estudo da informação, no mundo da saúde, a informática em saúde é o gerenciamento da informação em saúde. Usando registros médicos eletrônicos, dispositivos que coletam informações de saúde eletronicamente e outros padrões de informação eletrônica, os enfermeiros de informática de saúde são responsáveis por gerenciar, interpretar e comunicar os dados que entram e saem das instalações de saúde, todos com um propósito principal: melhorar a qualidade da assistência ao paciente (Scott Rupp, 2016).

2.2.1 Urgência/Emergência e Protocolos de Triagem

De acordo com o Conselho Federal de Medicina (CFM) em sua Resolução 1451/1995, entende-se por urgência a "ocorrência imprevista de agravo à saúde com ou sem risco potencial de vida, cujo portador necessita de assistência médica imediata" e emergência é a "constatação médica de condições de agravo à saúde que impliquem em risco iminente de vida ou sofrimento intenso, exigindo tratamento médico imediato" (Brasil; Conselho Federal de Medicina - CFM, 1995).

No pronto-socorro, as ferramentas de triagem visam priorizar pacientes mais graves quando a capacidade do setor é excedida (Farrohknia et al., 2011). No Brasil, essas ferramentas são essenciais uma vez que a superlotação hospitalar é rotina nas urgências e emergências (Da Silva Júnior et al., 2020).

Nas últimas duas décadas, foram desenvolvidas algumas escalas de triagem que visam priorizar os cuidados de acordo com a gravidade do paciente. Os principais sistemas de triagem internacionais são o *Australian Triage Scale* (Escala de Triagem da Australásia), *Canadian Emergency Department Triage and Acuity Scale* (Escala Canadense de Triagem e Acuidade), o *Emergency Severity Index* (Índice de

Gravidade de Emergência) e o *Manchester Triage System* (Protocolo Manchester de Triagem PMT) (Ausserhofer et al., 2021).

Entre essas metodologias de classificação de risco, o Protocolo Manchester de Triagem (PMT) é largamente o mais utilizado, tanto no Brasil quanto no mundo. Essa metodologia é baseada em categorias de sinais e sintomas, com 52 fluxogramas que representam sintomas predefinidos. Ao ser atendido, o paciente relata seus sintomas e o enfermeiro avalia os sinais, escolhendo o fluxograma que melhor se encaixe no conjunto sintoma/sinal do paciente. Cada fluxograma é composto por seis discriminadores-chave, como perigo de vida, ou dor intensa, por exemplo, a fim de distinguir entre categorias de urgência (Christ et al., 2010; Kevin Mackway-Jones; Janet Marsden; Jill Windle, 2014). Além disso, certos parâmetros vitais, como saturação de oxigênio e temperatura, são integrados dentro os fluxogramas do Protocolo Manchester. O paciente, pós análise no fluxograma pelo enfermeiro, pode ser alocado nas seguintes categorias, que compreendem prioridade, cor e tempo de espera até o atendimento médico (Figura 1):

Figura 1 – Categorias de prioridade do Protocolo Manchester

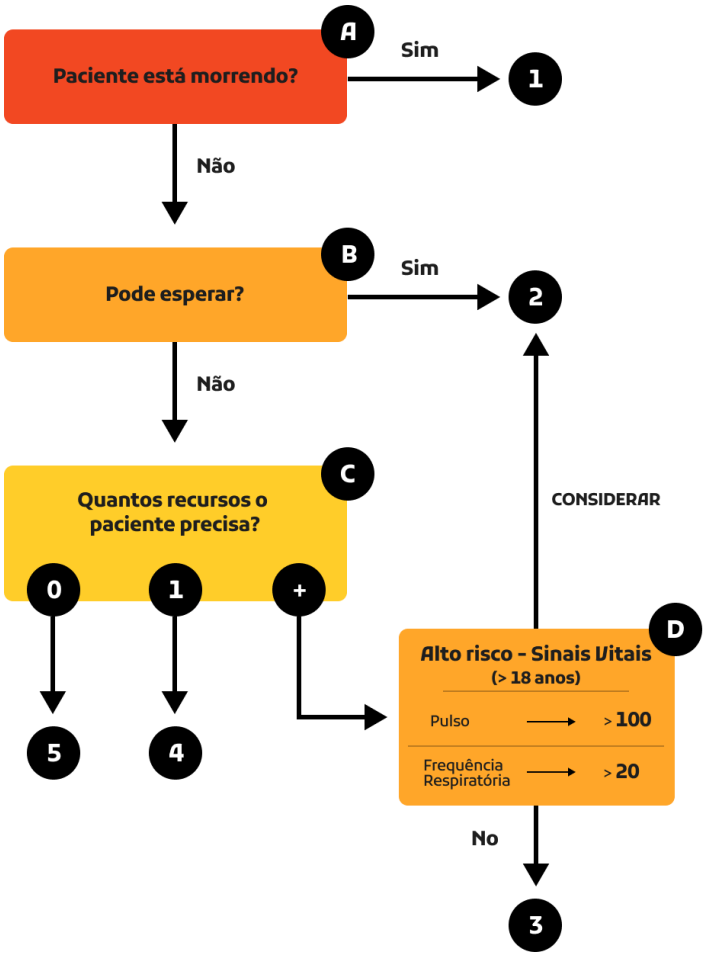
Fonte: O autor.

a) Emergência (cor vermelho – atendimento imediato);
b) Muito urgente (cor laranja – atendimento em até 10 minutos);
c) Urgente (cor amarelo – atendimento em até 60 minutos);
d) Pouco urgente (cor verde - atendimento em até 120 minutos);
e) Não urgente (cor azul - atendimento em até 240 minutos).

Outro protocolo de triagem, amplamente utilizado principalmente nos Estados Unidos da América, é o Índice de Gravidade de Emergência (ESI - *Emergency Severity Index*). No ESI, os pacientes são avaliados e organizados de acordo com a gravidade de seu quadro clínico, considerando a quantidade de recursos que provavelmente precisarão para o atendimento (Figura 2). Nessa classificação, os indivíduos com

condições menos complexas tendem a demandar menos recursos, sendo enquadrados na categoria 5. Já os casos mais críticos geralmente requerem quatro ou mais recursos, sendo classificados como prioridade máxima (1). Pacientes com essa classificação exigem atendimento imediato, enquanto aqueles designados como ESI 2 ou 3 devem ser atendidos em até 15 minutos. Já os classificados como ESI 4 ou 5 podem aguardar até 30 minutos para receber assistência médica (*Emergency Severity Index Handbook*, 2023).

Figura 2 – *Emergency Severity Index* (ESI) Algoritmo de Triagem v5



Fonte: *Emergency Severity Index Handbook*, 2023.

Considerando que protocolos de triagem como estes descritos acima estão presentes em diferentes prontos-socorros mundiais, alguns trabalhos vêm se preocupando em demonstrar o quão eficaz e confiável são esses métodos, inclusive apontando seus pontos mais vulneráveis. Uma pesquisa de 2017 feita em três departamentos de emergência europeus, mostra que a validade do PMT em classificar

os riscos varia entre moderada e boa, mas apresenta um desempenho baixo em classificar corretamente crianças e idosos (Zachariasse et al., 2017). De fato, embora o PMT tenha alguns parâmetros fisiológicos integrados, esse método não requer medição de sinais importantes como frequência cardíaca e frequência respiratória, por exemplo. Sinais desse tipo, no entanto, podem ser bons preditores de doenças graves, sobretudo em crianças e idosos (Lambert et al., 2017; Zachariasse et al., 2021).

Outras questões que ocorrem na classificação de risco baseada em protocolos de triagem são a subtriagem e a supertriagem. Na supertriagem, o paciente recebe uma atribuição de categoria maior do que urgência real; na subtriagem, ocorre o inverso, o paciente recebe uma categoria inferior à de sua verdadeira urgência. A subtriagem contribui para atrasos em intervenções sensíveis ao tempo, morbidade e mortalidade (Hitchcock et al., 2014). De maneira similar, a supertriagem também pode ter efeitos prejudiciais, e pode resultar no desvio de tempo e de recursos à pacientes menos urgentes e afetar indiretamente os mais graves (Hinson et al., 2018). Pesquisas demonstram, por exemplo, que o protocolo ESI de triagem induz a uma grande variação nas atribuições de triagem pelos profissionais, o que pode reduzir a confiabilidade entre os avaliadores (Christ et al., 2010; Martin et al., 2014; Mistry et al., 2018). Além disso, quando se utiliza a classificação ESI, há um considerável número de pacientes com diferentes doenças que são alocados no grupo 3 da classificação (vide Figura 2), grupo considerado crítico pois o ESI se demonstra falho ao identificar as diferenças e os riscos atrelados a cada paciente dentro desse grupo (Arya et al., 2013; Levin et al., 2018).

Além do erro na triagem inicial, outro importante ponto igualmente impactante na classificação de risco é a reavaliação do paciente. O Ministério da Saúde, no documento Acolhimento com Classificação de Risco, explica que a classificação de risco é um processo dinâmico, e que é necessário que, periodicamente, os pacientes que já receberam uma classificação de risco e que estão na espera sejam reavaliados. Isso deve ser feito porque um paciente classificado inicialmente como não sendo de risco, pode piorar clinicamente durante a espera, e, devido a isso, vir a demandar cuidados emergenciais (Brasil; Ministério da Saúde, 2009).

Nesse contexto da reavaliação, uma pesquisa de Costa e Colaboradores (2015), realizada em dois serviços hospitalares de emergência do estado do Paraná, demonstrou que o item reavaliação dos casos em espera obteve a menor pontuação

entre todos do sistema de acolhimento com classificação de risco, evidenciando o pior desempenho neste procedimento entre todos os necessários. Inoue e colaboradores (2015), em pesquisa com 151 profissionais de enfermagem, a reavaliação dos casos de espera também foi um item o qual a pontuação demonstrou baixa desempenho. De maneira similar, ainda, Hermida e colaboradores (2017), avaliaram o acolhimento com classificação de risco de uma UPA, do estado de Santa Catarina, na perspectiva de médicos e enfermeiras, e os autores demonstram que um dos itens com a menor pontuação avaliada também foi a reavaliação dos casos em espera. Em conjunto, esses trabalhos indicam a precariedade do sistema de reavaliação dos pacientes nas urgências e emergências brasileiras (Costa et al., 2015; Hermida et al., 2018; Inoue et al., 2015).

Segundo o Conselho Regional de Enfermagem (COFEN) a classificação de risco e priorização de assistência é função privativa do enfermeiro(a). No entanto, o enfermeiro deverá ser capacitado para o uso do protocolo adotado pela instituição (Brasil; Conselho Federal de Enfermagem - COFEN, 2021). Os estudos supracitados, bem como o próprio Ministério da Saúde, sugerem que, uma vez realizada a classificação de risco, o paciente deve ser reavaliado periodicamente pelo enfermeiro antes do atendimento médico. No entanto, a reavaliação constante do paciente depende da disponibilidade de recursos humanos, que geralmente é limitada nos prontos-socorros e usualmente está sob grande pressão. Dessa forma, novas metodologias para manutenção da avaliação do paciente em espera seriam bem-vindas, o que poderia assegurar a identificação de modificações nos sinais/sintomas dos pacientes, sem demandar mais trabalho dos profissionais responsáveis pela triagem.

2.2.2 Escores de Alerta Precoce

A evolução do Sistema de Suporte a Decisão Clínica (CDSS) se dá pela introdução de novas tecnologias na área da saúde. Estas novas tecnologias auxiliam no aperfeiçoamento da prestação de serviços de saúde por meio do fornecimento de informações de saúde precisas e direcionadas (*National Institute for Health and Care Research* - NHIR, 2020).

A deterioração clínica pode ser entendida como a piora repentina da condição fisiológica do paciente podendo aumentar o risco de morte. Em seus estudos comparativos houve a relação da deterioração fisiológica do paciente ser precedida

da alteração de sinais vitais, aumento da taxa de respiração, e se demorada a detecção desses fatores antecedentes, a disfunção do órgão alvo. Ainda foi entendido que doenças pré-existentes são fatores a serem considerados pois influencia em eventos adversos precedidos por alterações nos sinais vitais (Jones et al., 2013).

A identificação de pacientes em deterioração é uma tarefa complexa, e uma forma de obter estes dados de forma a reconhecer antecipadamente eventos adversos são o uso de modelos de alerta precoce. Além destes modelos, é preciso o uso de dispositivos vestíveis que obtém dados em tempo real dos sinais vitais do paciente. Desta forma é possível o desenvolvimento de critérios objetivos afim de identificar a deterioração clínica do paciente em tempo real que ajudam a equipe médica no CDSS. Alguns modelos de alerta precoces conhecidos como o SOFA e o qSOFA (*Quick Sequential Organ Failure Assessment*) para a detecção da deterioração do paciente por Sepse, que é a presença de disfunção orgânica ameaçadora à vida seguida por uma resposta desregulada do organismo à infecção, internado em Centro de Tratamento Intensivo (CTI) e a SIRS (*Systemic Inflammatory Response Syndrome*) também para o alerta de Sepse (Silva et al., 2021).

O GCS (*Glasgow Coma Scale*) publicado em 1974 e é utilizado para descrever objetivamente a extensão do comprometimento da consciência em pacientes traumatizados. Esta escala avalia os pacientes de acordo com três aspectos da capacidade de resposta: abertura dos olhos, respostas motoras e verbais (Jain; Margetis; Iverson, 2025).

O NEWS2 (*National Early Warning Score*) é um sistema de medições fisiológicas com o objetivo de identificar pacientes com doenças agudas, incluindo sepse. O protocolo foi criado para o sistema de saúde britânico, o NHS (*National Health Service*) em 2012. Previamente conhecido como EWS e posteriormente NEWS, o NEWS2 foi implementado em 2017 (National Institute for Health and Care Excellence - NICE, 2020). A diferença entre o EWS e o NEWS foi a adição do parâmetro saturação (SpO2) e do NEWS para o NEWS2 foi a melhora na segurança para pacientes com insuficiência respiratória hipercapnia (Bilben; Grandal; Søvik, 2016; Hodgson et al., 2018). O NEWS2 é, portanto, uma abordagem pragmática com ênfase na padronização no uso dos sinais vitais em um sistema de pontuação que mede seis parâmetros fisiológicos:

a) Saturação (SpO2);

b) Pressão Arterial (Sistólica);

c) Pulsação;

d) Temperatura;

e) Taxa de Respiração;

f) Nível de consciência ou confusão mental.

Uma pontuação de 0 até 3 é atribuída a cada parâmetro conforme (Figura 3) e quanto mais alta mais longe do intervalo normal.

Figura 3 – Pontuação do NEWS2

Parâmetros Fisiológicos	Pontuação						
	3	2	1	0	1	2	3
Frequência Respiratória	≤ 8		9 - 11	12 - 20		21 - 24	≥ 25
SpO_2 Escala 1 (%)	≤ 91	92 - 93	94 - 95	≥ 96	93 - 94 com oxigênio	95 - 96 com oxigênio	≥ 97 com oxigênio
SpO_2 Escala 2 (%)	≤ 83	84 - 85	86 - 87	≥ 88-92 ar ambiente			≥ 97 com oxigênio
Ar ambiente ou oxigênio		Oxigênio		Ar ambiente			
Pressão Arterial Sistólica (mmHg)	≤ 90	91 - 100	101 - 110	111 - 219			≥ 220
Pulso	≤ 40		41 - 50	51 - 90	91 - 110	111 - 130	≥ 131
Nível de consciência				Alerta			CI/PU
Temperatura	≤ 35.0		35.1 - 36	36.1 - 38	38.1 - 39	≥ 39.1	

Fonte: *Royal College of Physicians* (2022). Adaptado pelo autor.

As respostas clínicas apropriadas devem ser dadas conforme os níveis limites (Figura 4).

Figura 4 – Limites e gatilhos

Pontuação NEWS	Frequência de Monitoramento	Resposta Clínica
0	Mínimo 12 horas	• Monitoramento rotineiro;
Total 1-4	Mínimo 4-6 horas	• Enfermeiro(a) deverá avaliar o paciente; • E decidirá se aumentará a frequência de monitoramento e/ou escalonamento de cuidado é necessário;
3 em parâmetro individual	Mínimo 1 hora	• Enfermeiro(a) informará a equipe médica que cuida do paciente, que revisará e decidirá se a escalação de cuidados é necessária;
Total 5 ou mais Limite para resposta de urgência	Mínimo 1 hora	• Enfermeiro(a) informará imediatamente a equipe médica que cuida do paciente; • Avaliação de urgência por uma equipe médica com competências essenciais no cuidado de pacientes com doenças agudas; • Fornecerá atendimento clínico em um ambiente com instalações de monitoramento;
Total 7 ou mais Limite para resposta de emergência	Monitoramento contínuo dos Sinais Vitais	• Enfermeiro(a) informará imediatamente a equipe médica que cuida do paciente; • Avaliação de emergência por uma equipe médica com competências em cuidados intensivos, incluindo profissionais com habilidades avançadas de gerenciamento de vias aéreas; • Considerará a transferência de cuidados para uma unidade de cuidados clínicos de nível 2 ou 3, ou seja, unidade de alta dependência ou UTI; • Atendimento clínico em ambiente com facilidades de monitoramento;

Fonte: *Royal College of Physicians* (2022). Adaptado pelo autor.

Apesar do protocolo de escore de alerta precoce NEWS2 considerar seis métricas fisiológicas, ele não leva em conta alguns outros parâmetros importantes na avaliação do paciente, como a pressão arterial diastólica e o nível de dor. A análise da pressão arterial diastólica, por exemplo, tem recebido bastante atenção nos últimos anos pela comunidade médica, principalmente devido à chamada Hipertensão Diastólica Isolada (IDH na sigla em inglês *Isolated Diastolic Hypertension*). Estudos têm definido a IDH como um paciente que possui a pressão arterial sistólica < 140 mmHg e a pressão arterial diastólica ≥ 90 mmHg (Kaneko et al., 2020; Xie et al., 2021; Yano et al., 2022). Essa definição também é apoiada pela *European Society of Cardiology* (ESC) (Mancia et al., 2023). Investigações sugerem que a IDH pode estar associada a eventos adversos principalmente em pacientes jovens, incluindo problemas cardiovasculares, insuficiência cardíaca, fibrilação atrial e doença renal crônica (Huang et al., 2022; Kaneko et al., 2020; Li et al., 2014). Por outro lado, a

hipotensão arterial diastólica também vem sendo relatada como um importante alerta emergencial. Em um amplo estudo de 2011, foi verificado que adultos mais velhos (65 anos ou mais) que apresentaram pressão arterial diastólica < 60 mmHg com uma pressão sistólica normal (> 100 mmHg), tiveram maior risco de desenvolver insuficiência cardíaca de início recente (Guichard et al., 2011). Ainda, (Li et al., 2021), em ensaios clínicos com 7.515 pacientes adultos, verificaram que aqueles com uma pressão arterial sistólica tratada (menor que 130 mm Hg) e pressão arterial diastólica menor que 60 mmHg, foram associados com um alto risco cardiovascular (Li et al., 2021). Além da importância da pressão arterial diastólica, o nível de dor, também ausente no NEWS2, é bastante relevante na avaliação do paciente porque ajuda a determinar o tipo apropriado de analgesia a ser administrado e a urgência do alívio da dor necessário, especialmente para dores muito intensas (Romero-Hall, 2015).

Uma revisão recente discute que o escore NEWS2 apresenta também outras limitações para além da ausência de algumas métricas fisiológicas, como descrito anteriormente. A pesquisa encontrou que o principal gargalo na aplicação desse escore de alerta é que a medição dos parâmetros vitais precisa ser feita manualmente por profissionais de saúde treinados, o que demanda tempo e está sujeito a erros de cálculo (Holland; Kellett, 2023). Nesse sentido, o monitoramento automático contínuo das tendências dos sinais vitais, juntamente com algoritmos de *Machine Learning*, poderia promover previsões mais precisas e menos desgaste dos profissionais de saúde (Muralitharan et al., 2021).

2.2.3 Dados do registro de atendimento clínico

Com o progresso da obtenção e armazenamento de dados por meio dos sistemas de registro é possível utilizar algoritmos personalizados para prever resultados para os pacientes. Cada vez se tem usado os registros eletrônicos para registrar sinais vitais e demais informações dos pacientes. Portanto, com a utilização de algoritmos personalizados e dados mais robustos poderão ser usados para calcular pontuações de alerta precoce e prever resultados para os pacientes (Gerry et al., 2020).

A partir da Resolução nº 33 de 22 de março de 2018 que regulariza o Registro de Atendimento Clínico estabelece um modelo de informação que garanta o conjunto mínimo de informações clínicas de forma padronizadas - padrões de

interoperabilidade e informação em saúde no âmbito do Sistema Único de Saúde (SUS) (Brasil; Ministério da Saúde, 2018a).

A Resolução nº 1.638/02 do CFM determina itens considerados obrigatórios que devem conter no prontuário do paciente, os quais destacam-se:

a) **Identificação do paciente:** nome completo, data de nascimento (dia, mês e ano com quatro dígitos), sexo, nome da mãe, naturalidade (indicando o município e o estado de nascimento), endereço completo (nome da via pública, número, complemento, bairro/distrito, município, estado e CEP);
b) Anamnese, exame físico, exames complementares solicitados e seus respectivos resultados, hipóteses diagnósticas, diagnóstico definitivo e tratamento efetuado;
c) Evolução diária do paciente, com data e hora, discriminação de todos os procedimentos aos quais o mesmo foi submetido e identificação dos profissionais que os realizaram, assinados eletronicamente quando elaborados e/ou armazenados em meio eletrônico;
d) Nos casos emergenciais, nos quais seja impossível a colheita de história clínica do paciente, deverá constar relato médico completo de todos os procedimentos realizados e que tenham possibilitado o diagnóstico e/ou a remoção para outra unidade (Conselho Federal de Medicina - CFM, 2002).

2.3 INTELIGÊNCIA ARTIFICIAL NA SAÚDE

A saúde possui dados como nunca antes relacionados a várias áreas, como paciente, suprimentos, faturamento entre outros. Dados na área da saúde podem ser um grande ativo, desde que simplificados e analisados de maneira adequada, retornando aos pacientes melhores resultados (Andy De, 2017).

Os dados hoje podem ser transmitidos em tempo real de variadas fontes e é preciso que ele esteja fundamentalmente interpretável para que sejam tomadas as melhores decisões (Panesar, 2021). O desafio de lidar com os dados obtidos na área da saúde é o maior obstáculo, já que de forma geral os dados não são obtidos de forma estruturada que possam ser acionados a fim de apresentar apenas os elementos relevantes para uma decisão médica (Jay Anders, 2019).

De fato, tornar os dados em informações acionáveis é o maior desafio para o uso de IA) no meio da saúde, principalmente quando se trata de dados que ajudem no CDSS. Algumas formas de dados estruturados e acionáveis na área da saúde

estão organizadas como a Classificação Internacional de Doenças e Problemas Relacionados à Saúde (CID-10), SNOMED *Clinical Terms, Standardized, Nomenclature for Clinical Drugs* (RxNorm) e *Logical Observation Identifiers, Names and Codes* (LOINC).

A inteligência artificial pode ser definida como a "simulação de comportamento inteligente em agentes (computadores) de uma maneira que nós, como humanos, consideraríamos ser inteligente ou humano", e em seus conceitos básicos incluem "conhecimento, raciocínio, solução de problemas, percepção, aprendizagem, planejamento e capacidade de manipulação e movimentação" (Panesar, 2021, p.4).

Em particular para este projeto, a IA pode compreender técnicas de: a) raciocínio automatizado, planejamento de prova, resolução de restrições e raciocínio baseado em casos; b) técnicas de aprendizagem de máquina, mineração de dados e descobertas de conhecimento científico (Panesar, 2021).

De acordo com Artasanchez e Joshi (2020, p.2) uma definição otimista de IA seria “a área da ciência da computação que estuda como as máquinas podem realizar tarefas que normalmente exigiria um agente senciente”. E de forma mais cética poderia ser "a área da ciência da computação que estuda como as máquinas podem imitar de perto a inteligência humana”. Definir o que pode ser inteligível e como pode ser processado tem sido as questões que permeiam a área de IA, visto que em ambas as definições há contradições do que realmente pode ser considerado inteligência artificial (Artasanchez; Joshi, 2020).

2.3.1 *Machine Learning* como ferramenta no suporte à decisão clínica

A medicina está receptiva ao uso de dados e IA na saúde, mesmo demorando tanto, em comparação com outros setores. O uso de ML na saúde configura um novo cenário estimulante e inovador em que os dados estão se revelando empoderadores que pode ser democratizada e personalizada.

Este movimento vem com instituições acadêmicas e startups desenvolvendo tecnologias e protótipos em saúde para medir e quantificar mais aspectos de nossa vida. A previsão é que mais modelos de ML estarão disponíveis para uso e serão constantemente atualizados com novos dados otimizados para análises preditivas (Panesar, 2021).

Podemos definir o ML como um mecanismo pelo qual um sistema de IA pode ser desenvolvido, por meio da criação de algoritmos que se modificam em resposta a

padrões, e fazem inferências quando aplicados a novos dados (Obermeyer; Emanuel, 2016). Na área da saúde, muitos trabalhos recentes têm explorado o ML para desenvolver sistemas de suporte à decisão clínica, e essa ferramenta têm se mostrado útil em vários campos da medicina, como radiologia, neurocirurgia e dermatologia. É importante salientar que objetivo desses sistemas não é substituir os médicos, os pacientes ou as organizações de saúde, mas fornecer conhecimento relevante e suporte na sua tomada de decisão (Fernandes et al., 2020a).

Alguns estudos têm utilizado o ML para prever resultados ou injurias específicos em pacientes, como o risco de mortalidade e parada cardiorrespiratória (Coslovsky et al., 2015; Fernandes et al., 2020b; Teubner et al., 2015); admissão nas Unidades de Terapia Intensiva (UTI) (Raita et al., 2019); e mortalidade por sepse intra-hospitalar (Scott; Colborn, 2016; Taylor et al., 2016). Na área de radiologia, algoritmos já foram desenvolvidos e testados para realizar varreduras em tomografias computadorizadas e detectar hemorragias, hidrocefalias ou lesões cerebrais traumáticas, apresentando bons resultados (Prevedello et al., 2017; Yuh et al., 2008). Recentemente, estudos têm se dedicado a utilizar ML para predizer o risco de desenvolvimento de condições críticas em pacientes com a COVID-19, ou até mesmo auxiliar no diagnóstico da doença (De Moraes Batista et al., 2020; Fernandes et al., 2021).

Outras pesquisas com ML têm mostrado resultados promissores para melhorar os atuais métodos de triagem e classificação de risco em emergências. Em 2013, Azeez e colaboradores utilizaram um modelo de triagem primário com métodos baseados em ML. No caso deste estudo, quando o paciente chegava ao balcão de triagem, o sistema capturava seus sinais vitais e suas queixas principais, além da aparência geral do paciente, estes dados eram gerenciados e analisados em um servidor de dados e, prontamente, após análise, o sistema gerava o status de emergência do paciente. Outro estudo, de Levin e colaboradores (2018), desenvolveu a chamada Triagem Eletrônica (*e-triage)*, e aplicaram em um departamento de emergência dos Estados Unidos. Esse sistema baseado em ML prevê a probabilidade de resultados agudos, permitindo a melhora na estratificação de pacientes mais graves logo na chegada à emergência. Como resultado da aplicação, a triagem eletrônica foi mais precisa em classificar pacientes do que o método tradicional utilizado no país, o *Emergency Severity Index* (ESI) (Azeez et al., 2013; Levin et al., 2018).

Em decorrência da pandemia da COVID-19, alguns estudos trouxeram a pesquisa e implementação de aplicativos e sensores para a detecção de sintomas fisiopatológicos relacionados à essa doença ou monitorar pacientes já infectados, alguns inclusive utilizam ML. Um grupo de pesquisa dos Estados Unidos utilizou um *smartwatch* Fitbit para monitorar a frequência respiratória, a frequência cardíaca e a variabilidade da frequência cardíaca em pacientes já confirmados com a doença (testados positivos) e desenvolver um sistema para prever a necessidade de internação (Natarajan; Su; Heneghan, 2020). Fyntanidou et al. (2021) desenvolveram e implementaram um dispositivo de pulso para extrair continuamente dados de frequência cardíaca, níveis de saturação de oxigênio no sangue, estimativa de temperatura corporal e classificação respiratória de pacientes que chegavam a um departamento de emergência; o dispositivo foi testado no departamento de emergência de um hospital da Grécia. Com os dados gerados em tempo real, com um modelo de algoritmo baseado em ML, e com um aplicativo móvel disponibilizado para a equipe de triagem, o grupo conseguiu aprimorar a detecção de deterioração da saúde dos pacientes e permitiu destacar os casos mais críticos de Covid-19, culminando numa triagem mais eficiente (Fyntanidou et al., 2020).

Para além da COVID-19, o monitoramento constante de pacientes gerais dos departamentos de emergência também poderia ser aplicado, onde mortes e eventos adversos graves também podem ocorrer se uma resposta imediata não for aplicada (Cardona-Morrell et al., 2016). Nesse sentido, dispositivos móveis, vestíveis e ML podem ser explorados não só para triagem inicial de pacientes, mas também para monitorar o constantemente o paciente durante o tempo de espera, a fim de: a) gerar dados contínuos de sinais vitais importantes do paciente em espera, b) cruzar esses dados com dados integrados pré-existentes de forma inteligente, utilizando ML e c) reclassificar os pacientes em tempo real, de acordo com os dados gerados e análise do algoritmo. Dessa forma, questões como triagem falha e reavaliação precária poderiam ser evitados, aumentando a eficiência do pronto-socorro.

2.3.2 Base de dados MIMIC-IV

O MIMIC (*Medical Information Mart for Intensive Care*) é uma ampla base de dados de livre acesso que reúne dados de saúde anonimizados de pacientes tratados no hospital *Beth Israel Deaconess Medical Center* (BIDMC) em Boston, Massachusetts, EUA. Este banco é disponibilizado pelo do trabalho de pesquisadores

do Laboratório de Fisiologia Computacional do Instituto de Tecnologia de Massachusetts (MIT) (Johnson et al., 2021).

O MIMIC-IV contém dados de mais de 65 mil pacientes que foram internados na UTI e mais de 200 mil que foram admitidos na urgência/emergência. Essa base inclui dados demográficos, sinais vitais, resultados laboratoriais, medicamentos, notas clínicas e outros parâmetros relevantes para pesquisa em saúde de forma modular.

Apesar de ser um banco de dados livre, é preciso concluir o treinamento e obter a certificação do *CITI Data or Specimens Only Research,* para ter acesso a determinados ao *PhysioNet,* onde esses dados estão localizados. O treinamento envolve aspectos importantes sobre a pesquisa com dados de participantes humanos, como:

a) Definindo Pesquisa com Seres Humanos
b) Privacidade e Confidencialidade
c) Avaliando Riscos
d) Pesquisa com Crianças
e) Pesquisa Internacional
f) História e Princípios Éticos
g) Regulamentos e Processos
h) Metodologias SBR em Pesquisa Biomédica
i) Pesquisa Genética
j) Pesquisa baseada em registros
k) Populações em Pesquisa que Requerem Considerações e/ou Proteções Adicionais
l) HIPAA e Pesquisa com Seres Humanos
m) Conflitos de interesse em pesquisas envolvendo seres humanos

O MIMIC tem sido utilizado para pesquisas na área de *Machine Learning* aplicado à medicina para previsão de desfechos clínicos e desenvolvimento de algoritmos que suportem o CDSS. O banco possui várias versões ao longo dos anos, bem como subdivisões. A versão utilizada para esta pesquisa é o MIMIC-IV ED, divisão com dados do departamento de emergência. O banco é de 2021, datando sua última atualização em outubro de 2024.

É importante ressaltar que há uma limitação no uso direto da base de dados do MIMIC devido à diferença nos protocolos de triagem adotados no Brasil e nos

Estados Unidos. O MIMIC utiliza o Protocolo de triagem *Emergency Severity Index* (ESI), enquanto o padrão brasileiro é o Protocolo de Manchester. Por essa razão, os modelos desenvolvidos utilizando os dados do MIMIC devem ser cuidadosamente validados e adaptados com dados locais antes de uma possível implementação clínica.

2.4 USABILIDADE E EXPERIÊNCIA DO USUÁRIO

A eficiência do uso de um sistema pelo usuário depende de sua usabilidade, que significa facilidade de uso ou uso “amigável”, isso ocorre quando o sistema é projetado tendo em vista as necessidades do usuário e utilizando métricas de satisfação e eficiência (Jan Dul; Bernard Weerdmeester, 2012).

Segundo Norman (2013) o design centrado no usuário coloca as necessidades, capacidades e comportamento do usuário em primeiro lugar. Para Norman, o bom design começa com a compreensão do usuário por meio da psicologia e a compreensão do contexto tecnológico (Norman, 2013).

O Design centrado no usuário é ter a compreensão das pessoas e das suas necessidades dentro do projeto. E uma das formas de compreender as necessidades está na observação da tarefa, pois por muitas vezes o usuário desconhece suas verdadeiras necessidades. Quando se interage com o produto, é preciso descobrir o que ele faz, como funciona e quais operações são possíveis, que é a fase de descoberta (Norman, 2013).

2.4.1 Metas de usabilidade

Quando se projeta um sistema, a importância está em saber quais serão as capacidades e limitações do usuário do sistema, a fim de determinar as características do sistema. Para os autores, as características mais importantes são: faixa etária, nacionalidade, escolaridade, pessoas com necessidades especiais, experiências anteriores com tarefas e sistemas semelhantes, frequência da execução da tarefa entre outros (Jan Dul; Bernard Weerdmeester, 2012).

Na legislação brasileira a usabilidade é a “medida na qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso” (Brasil; ABNT, 1998)

A abordagem centrada no uso se traduz em sistemas intuitivos, fáceis de aprender e de usar. Tais sistemas causarão menos fadiga proporcionando conforto

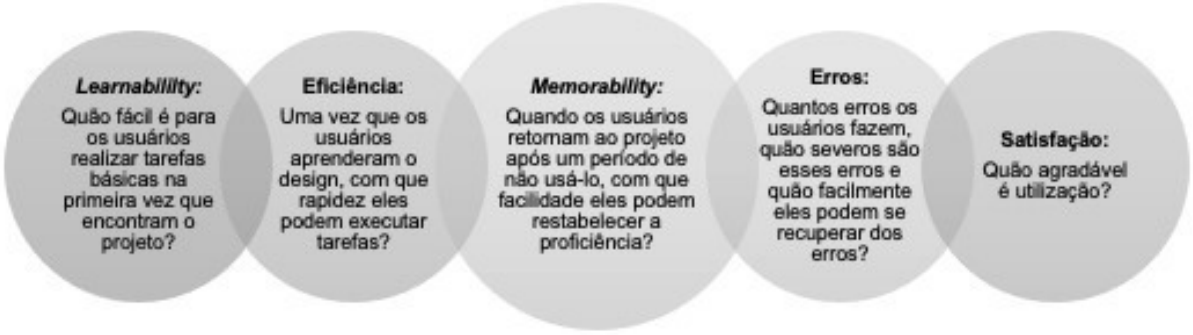
ao usuário e a garantia de qualidade no resultado (Walter Cybis; Adriana Holtz Betiol; Richard Faust, 2015).

Jakob Nielsen listou dez princípios gerais de para o design de interação conhecidas por "Heurísticas de Nielsen" porque são regras gerais e não diretrizes de usabilidade específicas (Jakob Nielsen, 1994).

1. **Visibilidade do status do sistema:** o sistema deve sempre manter os usuários informados sobre o que está acontecendo, por meio de feedback adequado e dentro de um prazo razoável;
2. **Correspondência entre o sistema e o mundo real:** o sistema deve falar o idioma dos usuários, com palavras, frases e conceitos familiares, em vez de termos orientados ao sistema;
3. **Liberdade e controle do usuário:** os usuários precisam ter um suporte em desfazer e refazer ações;
4. **Consistência e padrões:** os usuários não devem ter que se perguntar se diferentes palavras, situações ou ações significam a mesma coisa;
5. **Prevenção de erros:** melhor do que boas mensagens de erro é um design cuidadoso que impede que um problema ocorra em primeiro lugar;
6. **Reconhecimento ao invés de memorização:** minimize a carga de memória do usuário, tornando visíveis objetos, ações e opções. O usuário não deve se lembrar de informações de uma parte do diálogo para outra;
7. **Flexibilidade e eficiência do uso:** o sistema tem que servir tanto para usuários inexperientes quanto para experientes;
8. **Design estético e minimalista:** os diálogos não devem conter informações irrelevantes ou raramente necessárias;
9. **Ajude os usuários a reconhecerem, diagnosticarem e se recuperarem dos erros:** as mensagens de erro devem ser expressas em linguagem simples (sem códigos), indicar com precisão o problema e sugerir construtivamente uma solução;
10. **Ajuda e documentação:** mesmo que seja o melhor sistema e possa ser usado sem documentação, pode ser necessário que o usuário precise de ajuda, essas informações devem ser fáceis de pesquisar e focadas.

E posteriormente, Nielsen elencou cinco atributos de usabilidade que avalia o uso de interfaces e a concepção delas. São eles: *learnabiliy*, eficiência, *memorability*, erros e satisfação (Figura 5) (Jakob Nielsen, 2012a).

Figura 5 – Metas de Usabilidade de Nielsen



Fonte: Adaptado pelo autor.

2.4.2 *Lean* UX

O conceito foi elaborado Gothelf e Seiden (2017) com base no conceito *Toyota Production System* (TPS), posteriormente conhecida como *Lean Manufacturing* desenvolvida pelo fundador da Toyota, o Toyoda Sakichi, seu filho Toyoda Kiichiro e seu engenheiro Taiichi Ohno (Gothelf; Seiden, 2021). Diferente de outras ferramentas, o *Lean* é uma filosofia que almeja eliminar completamente os resíduos em busca de métodos mais eficientes onde se produz apenas o que é necessário para o próximo fluxo em um sistema de fluxo contínuo (Sakichi Toyoda et al., 2020).

No contexto do *Lean UX*, são utilizadas diversas ações para se pensar a gestão de processos de forma enxuta, reduzir o desperdício e priorizar o valor. Posteriormente, este sistema originou o *Lean Thinking*, que tem a mesma premissa e é amplamente utilizado em variados segmentos, onde se chegou ao *Lean UX*. O *Lean* pode ser conhecido também como Pensamento Enxuto, que é a tradução oficial para o idioma português.

O design de software tem como pai o design industrial que tem como objetivo o projeto de materiais físicos. O processo de criação e desenvolvimentos de produtos físicos é extremamente caro e por isso seguia um processo de criação engessado sem margem para melhorias contínuas (Gothelf; Seiden, 2021).

Entretanto, hoje, com a produção de softwares temos uma nova realidade que permite que seja utilizado a melhoria contínua, não necessitando processos longos

para desenvolvimento do projeto. A vantagem do processo ágil é a integração e implantação contínua, o que reduz drasticamente o tempo do ciclo. Empresas gigantes da tecnologia utilizam de processos ágeis de melhoramento contínuo todos os dias atualizando seus sistemas, e o que se ganha com isso são ciclos de lançamento curtos e feedbacks mais ágeis. Assim se tem a capacidade de aprender, rápido e continuamente se os produtos estão atendendo as necessidades do cliente, além de elevar as expectativas dos clientes em temos de qualidade e tempo de resposta.

O *Lean UX*, como o nome diz, é o processo de pensamento enxuto no desenvolvimento voltado para a experiência do usuário. *Lean UX* aplica os quatro valores centrais do desenvolvimento ágil ao projeto de um produto:

a) **Indivíduos e interações** acima de processos e ferramentas;

b) **Software funcionando** acima de uma documentação;

c) **Colaboração do cliente** acima da negociação de contrato;

d) **Responder às mudanças** ao invés de seguir um plano;

e) ***Build-Measure-Learn*** para minimizar o risco do projeto e ter times que aprendam rápido para construir o Mínimo Produto Viável (MVP), o para iniciar o processo de aprendizagem o mais cedo possível.

O MVP é um método do *Lean Startup*, metodologia cunhada por Ries (2017) para desenvolvimento ágil de *Startups*. O autor defende a criação de protótipos rápidos projetados para testar suposições de mercado e usar o feedback do cliente para evoluí-lo, reduzindo o desperdício e aumentando a frequência do contato com clientes reais que testam o produto, assim a equipe garante corrigir suposições incorretas mais cedo (Ries, 2017).

Lean UX utiliza dessa filosofia para a prática de design de produto. Para cada projeto uma hipótese, e o objetivo é de validar a solução da maneira mais eficiente possível usando o *feedback* do cliente. Assim, entende-se o um sistema com o essencial de recursos para executar uma tarefa (a hipótese a se testar) como sendo um MVP. Além disso, entende-se que o MVP não necessariamente precisa ser o produto final, mas sim uma aproximação da experiência que o usuário terá com o projeto completo.

Se coleta os dados fornecidos por esse MVP, desenvolve-se novas e ideias e testa novamente, nesse ciclo de melhoria contínua. Por definição, o *Lean UX* é:

a) *Lean UX* é uma abordagem de design que traz a verdadeira natureza de um produto para leve mais rápido, de forma colaborativa, multifuncional e centrada no usuário.

b) Os métodos *Lean UX* constroem uma compreensão compartilhada do usuário, suas necessidades, nossas soluções propostas e nossa definição de sucesso.

c) *Lean UX* prioriza o aprendizado contínuo para construir evidências para as decisões da equipe e garantir a entrega cada vez melhor de produtos, serviços e valor.

Segundo Patton e Economy (2014) o processo de se manter focado nos usuários e em suas experiências como objetivo de se ter e compartilhar entendimento acerca do produto é o diferencial de um método voltado para a usabilidade (Patton; Economy, 2014), e com o Mapa de Jornada do Serviço se tem a estrutura da maneira como o usuário se move pelo produto, esse processo tem o objetivo de ser capaz de visualizar o fluxo de maneira que se possa entender o comportamento que é mais importante para o usuário dentro do sistema (Gothelf; Seiden, 2021).

2.4.3 PRD – Documentos de requisitos do produto

No desenvolvimento de um MVP a validação antecipada (Figura 6) é uma ferramenta para reduzir incertezas e garantir o sucesso do produto. Este processo deve ser permanente, em todas as etapas do processo de desenvolvimento do produto (Klein, 2013).

Figura 6 – Validação Antecipada e Frequente

Fonte: Laura Klein, 2013.

O Documento de Requisitos do Produto, mais conhecido por sua sigla em inglês (PRD - *Product Requirements Document*) é uma descrição detalhada dos requisitos de um produto, isto inclui os recursos e funcionalidades. Os PRDs devem seguir a premissa de serem claros e sucintos, de forma a descrever como o produto

será realmente construído (Ulrich; Eppinger, 2016). Os componentes chaves de um PRD incluem:

a) **Objetivos do Produto:** declaração clara do problema;

b) **Pesquisa de clientes:** pesquisas dos possíveis clientes para este produto que podem influenciar no processo de desenvolvimento;

c) **Versionamento:** o que será entregue;

d) **Componentes:** listagem de todos os recursos e funcionalidades que precisa ser construído do produto;

e) **Requisitos do Sistema:** detalhamento sobre os sistemas;

f) **Fluxo do Usuário e Design:** como será a aparência do produto, bem como seu processo de interação com o usuário;

g) **Métricas:** como será a forma de medir o sucesso dos recursos e;

h) **Trabalhos Futuros:** informações sobre como o produto deverá evoluir.

Desta forma é possível organizar de forma descritiva e visual como este produto deverá ser concebido e quais as funcionalidades deverão estar contidas nele.

2.4.4 Wireframes, dashboards e protótipos

Um dos processos de desenvolvimento de um produto é a criação de *wireframes* (Levy; Dullea, 2021). Os *wireframe*s são comumente utilizados para mostrar ideias de layout no contexto visual, é um guia gráfico da representação de um design sem elementos visuais ou elementos da marcam (Deacon, 2020; Laubheimer, 2016).

Normalmente é utilizado para definir itens que estarão em tela de modo a construir estratégias para organizar as informações com o objetivo de compreender como o usuário interagem com o do produto ou serviço (Deacon, 2020). Para Deacon (2020), *wireframes* são como a planta de uma construção, os pontos mais importantes são:

a) Fornecer aos designers de interface uma base para começar a criar telas;

b) Serve como ponto de referência para o desenvolvimento;

c) Ajuda na comunicação da experiência com usuários em potencial sem elementos de design gráfico;

d) Ajuda a explorar ideias facilmente;

e) Ele fornece um meio para prototipagem e teste de ideias.

Wireframes são como a espinha dorsal de um produto, eles ajudam a evitar erros e custos altos no desenvolvimento, visto que se pode ter um panorama acerca do produto ou serviço antes mesmo de uma etapa de desenvolvimento, assim, pode-se também, evitar funcionalidades desnecessárias (Levy; Dullea, 2021; Paul Boag, 2020).

Dashboards são coleções de dados essenciais apresentados de forma visual e em uma única página que transmite informações rápidas sobre as quais os usuários podem tomar decisões rapidamente. O objetivo do dashboard é expor informações que podem ser consumidas rapidamente com o mínimo de interação ou processo cognitivo (Laubheimer, 2017). Segundo o autor, existem dois tipos de dashboards, os Operacionais que transmite informações críticas rapidamente aos usuários e os Analíticos que transmitem informações rápidas usadas para análise e tomada de decisões.

2.5 PULSEIRA MULTIPARAMÉTRICA

A utilização de tecnologias no contexto hospitalar tem sido apontada como um recurso estratégico para otimizar o monitoramento contínuo de pacientes e a resposta clínica. No caso específico das unidades de urgência e emergência, a integração entre dispositivos multiparamétricos e sistemas informatizados tem o potencial de possibilitar a coleta e transmissão de dados em tempo real, contribuindo para decisões mais rápidas e precisas. Nesse sentido, como parte das iniciativas para a informatização da saúde, o grupo de pesquisa DesignLab da UFSC projetou uma pulseira multiparamétrica para aplicação em ambientes hospitalares das UEs. O dispositivo realiza a aferição contínua de sinais vitais — pressão arterial, frequência cardíaca, temperatura corporal e oximetria de pulso — e transmite os dados em tempo real para uma interface dedicada. Este produto tem como finalidade aprimorar o monitoramento constante do paciente, otimizando a classificação de risco e agilizando a resposta clínica com base em seu estado momentâneo (Nishida, 2023).

Além da aferição contínua, estudos desenvolvidos no âmbito do DesignLab da UFSC têm enfatizado a necessidade de interfaces que facilitem a interpretação das informações coletadas e apoiem protocolos clínicos baseados em evidências. A concepção da interface voltada ao prontuário eletrônico, proposta nesta tese, tem

como diferencial a incorporação do escore de alerta precoce modificado (NEWS2-M), cujo objetivo é aprimorar a estratificação de risco e oferecer subsídios para uma resposta assistencial mais eficaz. Dessa forma, a pesquisa avança ao associar parâmetros clínicos validados à experiência do usuário (UX), assegurando que a tecnologia atenda às demandas práticas do ambiente hospitalar.

Embora a pulseira multiparamétrica desenvolvida por Nishida (2023) seja tratada aqui como um dispositivo de entrada de dados, o foco da presente tese recai sobre a interpretação e a apresentação dessas informações por meio de uma interface eficiente e responsiva. Com isso, espera-se não apenas agilizar a classificação de risco, mas também contribuir para a evolução dos sistemas de triagem eletrônica nas unidades de urgência e emergência, em consonância com os objetivos do grupo de pesquisa DesignLab da UFSC (Nishida, 2023).

3 PROCEDIMENTOS METODOLÓGICOS

3.1 CARACTERIZAÇÃO DA PESQUISA

A presente pesquisa foi estruturada em seis etapas metodológicas sequenciais, concebidas para investigar e propor soluções, por meio do *software* de triagem eletrônica (e-triagem), para os desafios gerenciais na triagem de pacientes em Unidades de Emergência (UE).

3.1.1 Primeira Etapa: Revisão da Literatura

A fase inicial da pesquisa compreendeu uma revisão sistemática da literatura, cujo objetivo foi estabelecer o estado da arte sobre a triagem em serviços de emergência, a utilização de escores de alerta precoce e a aplicação de sistemas de apoio à decisão clínica. Este levantamento teórico-conceitual foi fundamental para a precisa delimitação do problema, a formulação dos objetivos e o embasamento da hipótese central do estudo.

3.1.2 Segunda Etapa: Investigação Qualitativa do Contexto Clínico

Nesta etapa, foi conduzido um estudo qualitativo por meio de entrevistas semiestruturadas com profissionais de saúde (médicos, enfermeiros e técnicos de enfermagem) com experiência prévia ou atual em unidades de urgência e emergência. O objetivo foi mapear o fluxo de trabalho (*workflow*) da triagem, identificar os pontos críticos de decisão e diagnosticar as percepções sobre os sistemas de *software* atuais. A investigação focou-se em desafios relacionados à usabilidade, armazenamento de dados, correlação e visualização de dados.

3.1.3 Terceira Etapa: Desenvolvimento do Modelo de Classificação e Monitoramento

Esta etapa centrou-se no desenvolvimento da solução proposta. Procedeu-se à adaptação do escore de alerta precoce NEWS2 (*National Early Warning Score 2*), resultando em uma versão modificada capaz de integrar tanto os parâmetros fisiológicos obtidos em tempo real por uma pulseira multiparamétrica quanto os dados extraídos do Prontuário Eletrônico do Paciente (PEP). O propósito desta fase foi duplo: viabilizar não apenas uma triagem eletrônica automatizada, mas também instituir o monitoramento contínuo do paciente durante o período de espera. Este

monitoramento visa: a) gerar dados contínuos de sinais vitais; b) analisar inteligentemente estes dados em conjunto com informações pré-existentes do prontuário, utilizando técnicas de *Machine Learning* (ML) com o objetivo de predizer a deterioração do estado clínico do paciente; e c) reclassificar dinamicamente o risco do paciente em tempo real. Dessa forma, busca-se mitigar falhas de triagem e a precariedade na reavaliação, aumentando a eficiência e a segurança do serviço de emergência. Foram utilizados os dados do MIMIC-IV ED (*Medical Information Mart for Intensive Care IV - Emergency Department*) nos testes dos algoritmos de ML.

3.1.4 Quarta Etapa: Desenvolvimento do Protótipo Funcional

Com base nos requisitos levantados na segunda etapa e nos modelos da terceira, procedeu-se ao desenvolvimento de um protótipo funcional de alta fidelidade do *software* de triagem eletrônica (e-triagem). Utilizando uma metodologia de desenvolvimento ágil, a interface do sistema foi projetada para receber e exibir, em tempo real, os dados do paciente. O objetivo central da interface é fornecer alertas visuais, baseados em padrões e anomalias identificados pela Inteligência Artificial (IA), que sinalizem a deterioração do quadro clínico e apoiem a equipe de saúde na reclassificação do paciente com base no escore NEWS2 modificado, visando conferir maior assertividade e celeridade ao processo de atendimento. O protótipo foi populado com dados fictícios, criados especificamente para esta pesquisa, não sendo necessária a obtenção de consentimento para o uso dessas informações nesta fase.

3.1.5 Quinta Etapa: Avaliação de Usabilidade

A quinta etapa consistiu na avaliação de usabilidade do protótipo com seus usuários-alvo (médicos, enfermeiros e técnicos de enfermagem). O estudo buscou observar a interação dos profissionais com a interface enriquecida de informações e analisar a influência desses dados na qualidade e na tomada de decisão clínica. Para a coleta de dados, empregou-se uma abordagem mista, utilizando métricas de desempenho (e.g., tempo na tarefa, taxa de erros) na manipulação do sistema e métricas de preferência, aferidas por meio da aplicação de questionários padronizados.

3.1.6 Sexta Etapa: Análise e Validação dos Modelos

Finalmente, a sexta etapa foi dedicada à análise quantitativa e à validação dos artefatos desenvolvidos. Foram executados os experimentos computacionais para aferir o desempenho preditivo dos algoritmos de *Machine Learning.* Adicionalmente, realizou-se uma análise estatística comparativa da acurácia do escore NEWS2 modificado frente aos padrões de referência estabelecidos, o NEWS2 original e o ESI (*Emergency Severity Index*). Os achados desta etapa constituem a base para a seção de resultados e discussão desta tese.

3.2 DESENVOLVIMENTO DA INTERFACE

Nessa etapa, a proposta era entender quais eram atribuições de cada pessoa dentro da UE. Para assim criar-se o fluxo do sistema.

3.2.1 Participantes

Para esta pesquisa, foram utilizados 8 (oito) participantes, todos voluntários, qualificados, ou seja, enfermeiros(as), e/ou médicos(as), e/ou técnicos(as) de enfermagem, com idades entre 18 e 65 anos. Para a pesquisa com os participantes, o projeto foi submetido ao Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina (CEPSH/UFSC), como prevê a Resolução 196 (BRASIL, 1996, art. 3º) — para avaliar e garantir o espaço de cada um na pesquisa, buscando a premissa do equilíbrio e evitando práticas questionáveis — e aprovado na data de 29 de abril de 2025 com o número CAAE: 86552925.9.0000.0121. Todos os participantes preencheram o Termo de Consentimento Livre e Esclarecido (TCLE) aprovado pelo comitê e baseado nos princípios do Conselho Nacional de Saúde (CNS), por meio da Resolução nº 466/2012.

3.2.2 Questionário Pré-teste

De forma não estruturada foi aplicado um questionário pré-teste aos profissionais que trabalham ou trabalharam em UE com a finalidade de entender como é a rotina de trabalho dentro do plantão. Normalmente são plantões de 12h com um fluxo de trabalho intenso. A partir deste questionário obteve-se os agentes do sistema,

as atribuições de cada agente e a rotina diária dentro da UE. O questionário pré-teste está detalhado no Apêndice A.

3.2.3 Agentes e suas atribuições no sistema de triagem

A primeira fase de desenvolvimento do processo de e-Triagem *Vital Signs* consiste na definição dos tipos de elementos e agentes que serão partes integrantes na Urgência e Emergência - UE do Hospital, bem como suas atribuições e jornadas dentro do plantão. Para cada um dos agentes, foram analisadas suas atribuições, a fim de poder definir sua função dentro do processo de e-Triagem. Os agentes e elementos participantes que serão considerados nesta pesquisa, são:

a) Pessoas:

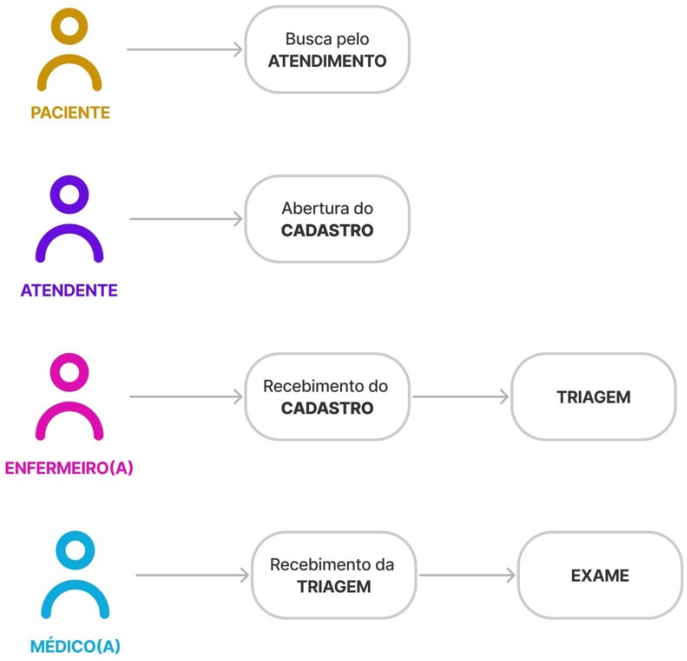
- **Paciente:** a ser monitorado pelo sistema;
- **Atendente:** agente de inserção de Dados de ID;
- **Enfermeiro(a):** agente de inserção de dados de triagem básica e classificação de risco;
- **Médico(a):** agente do exame clínico;

b) e-Triagem dentro do Sistema *Vital Signs*:

- **Dados de Identificação:** informações principais sobre a identificação do paciente;
- **Dados de Triagem Básica:** informações principais dos sistemas vitais;
- **Inserção da Pulseira:** coletor de informações dos principais sistemas vitais em tempo real;
- **Dados de Prontuário:** informações do paciente constantes no sistema;

Com base nesta definição, têm-se os agentes e suas atribuições que formarão a base de dados do sistema. A partir desta primeira caracterização buscou-se organizar o fluxo de atendimento da UE e as atribuições para cada agente. Por meio da entrevista com profissionais da área pode-se classificar o que é atribuição de cada agente segue a ordem descrita na Figura 7.

Figura 7 – Agentes e Fluxo de Atribuições



Fonte: O autor.

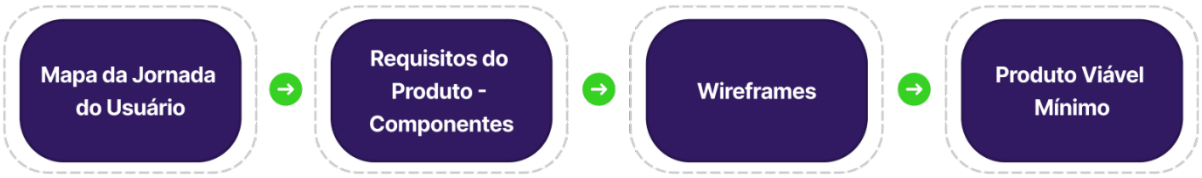
O paciente é o agente que inicia todo o processo de atendimento, bem como a utilização do sistema de e-Triagem. O paciente busca o atendimento na UE que é atendido inicialmente pelo atendente. A função do atendente é a abertura de cadastro de atendimento, ele pode iniciar um cadastro novo ou complementar um cadastro antigo para posteriormente enviá-lo a lista de triagem. A lista de triagem é de responsabilidade do enfermeiro, ele deve receber esta lista e fazer o atendimento conforme o recebimento. Até então não há uma triagem em relação ao estado do paciente. O enfermeiro, então, recebe o paciente e faz a e-Triagem dele. Neste momento gera uma lista ordenada conforme a relação do estado de saúde que o paciente apresenta. Conforme a gravidade de seu problema, o paciente deve ser atendido pelos médicos que recebem esta lista ordenada e conduz a parte de hipótese diagnóstica, exame físico e conduta.

3.2.4 Desenvolvimento conceitual da interface do sistema

Para o desenvolvimento conceitual da interface e construção do protótipo funcional de triagem eletrônica, utilizou-se a metodologia ágil proposta por Gothelf e Seiden (2021). Nesta etapa, considerou-se a definição dos agentes que irão compor o sistema de e-triagem eletrônica, os fluxos de tarefa, as informações a serem

coletadas, os meios humanos e tecnológicos para o desenvolvimento do sistema (Gothelf; Seiden, 2021). Desta forma, a fim de organizar esta etapa de modo objetivo, foi dividida em quatro fases: a) A definição da história dos usuários e o mapa de jornada do serviço - os agentes do sistema - (*User Story e Service Journey Map*); b) A coletânea do Documento de Requisitos do Produto em forma de Componentes - PRD; c) Wireframe; e d) Desenvolvimento da Interface para um Produto Viável Mínimo - MVP (Figura 8).

Figura 8 – Processo do Desenvolvimento Conceitual



Fonte: O autor.

Os itens obrigatórios da Resolução nº 33/2018 e da Resolução CFM nº 1.638/02 servirão de orientação para o desenvolvimento da interface do e-Triagem dentro do Sistema *Vital Signs*. Ressalta-se que pelo proposto da pesquisa ser um MVP do software, ele deverá conter as informações pertinentes ao processo de triagem e ao processo de suporte à decisão clínica. Desse modo, é o próprio processo de MVP irá determinar os requisitos de funcionalidades necessários para o objetivo da pesquisa (Brasil; Conselho Federal de Medicina - CFM, 1995; Brasil; Ministério da Saúde, 2018).

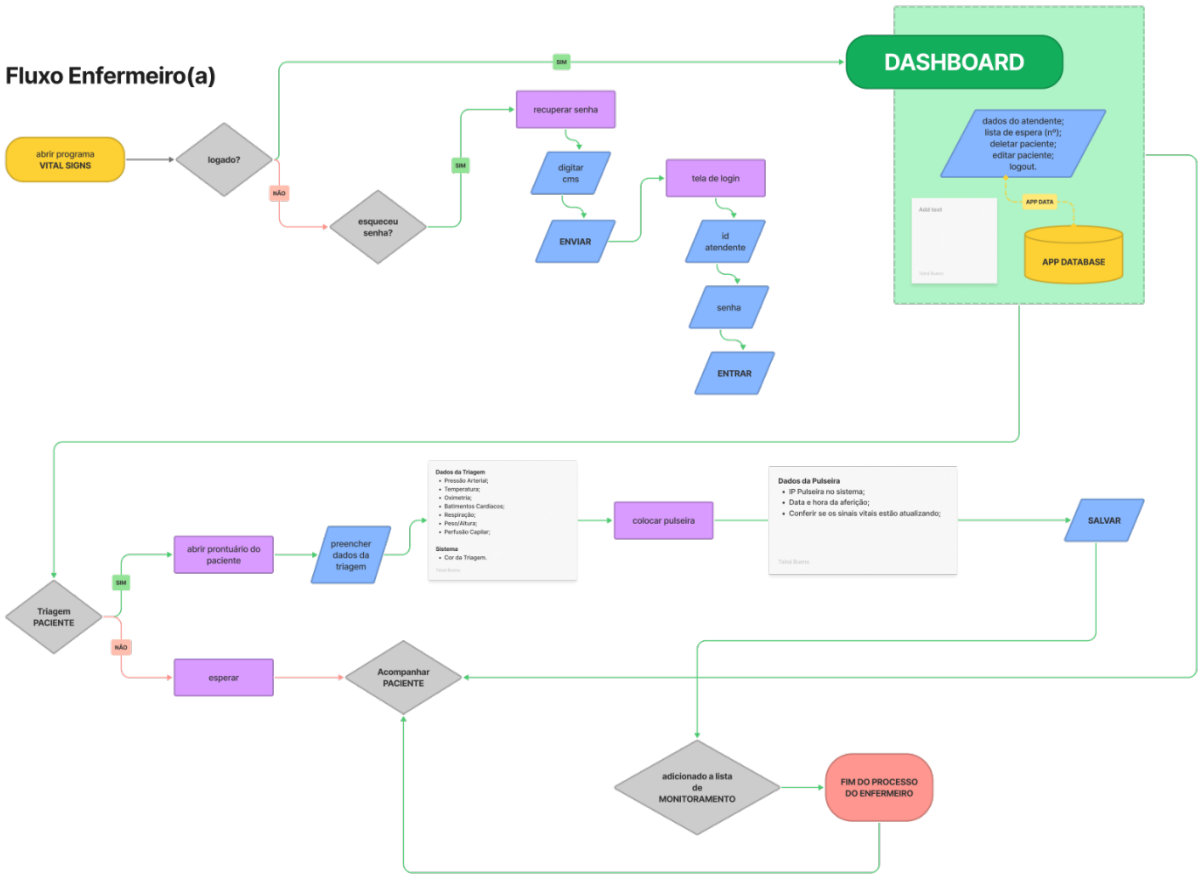
3.2.5 Enfermeiro(a): mapa da jornada

Partiremos do mapa do Enfermeiro, pois o fluxo do atendente não é objeto desse estudo. No entanto para que se tenha uma visão geral do sistema, o mapa do atendente está em Apêndice B. Segundo o Protocolo Manchester de Triagem (PMT) a principal função do enfermeiro de triagem é entender o que o paciente precisa da forma mais objetiva. O enfermeiro de triagem é o primeiro contato clínico do paciente (Kevin Mackway-Jones; Janet Marsden; Jill Windle, 2014). Ele deve coletar o máximo de informações clínicas do paciente, dentro de suas atribuições, para que se possa ter um entendimento do que o paciente está sentindo para que o próximo agente do sistema (médico) consiga elaborar a hipótese diagnóstica de forma assertiva.

Entendendo a função do enfermeiro, criou-se um mapa de sua jornada com as atribuições mínimas viáveis para o MVP (Figura 9).

O fluxo do enfermeiro consiste em forma resumida, abrir o sistema, fazer toda a parte de autenticação, chegar ao dashboard. No dashboard o enfermeiro tem uma visualização de como está a UE, pacientes aguardando a triagem, pacientes sendo monitorados após triados, alertas de possíveis deteriorações clínicas.

Figura 9 – Mapa da Jornada do Enfermeiro(a)



Fonte: O autor.

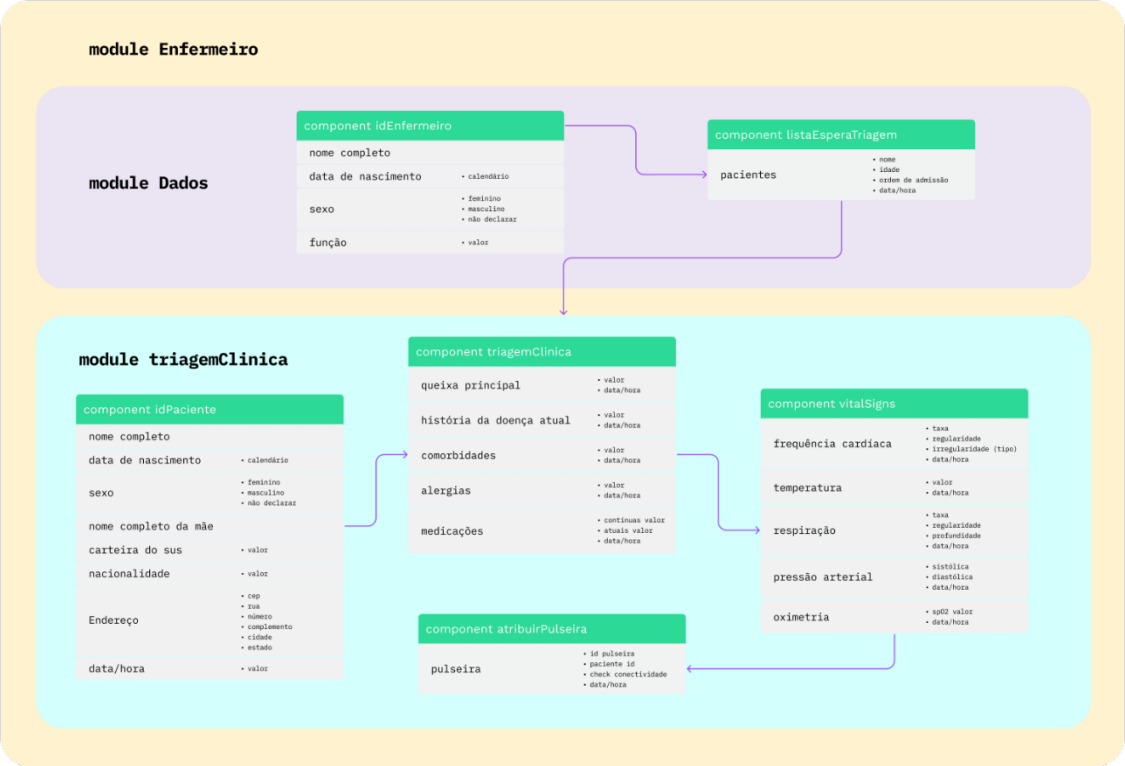
3.2.6 Módulos

Para que o aplicativo de e-Triagem funcione conforme o objetivo desta pesquisa, foram listados módulos e componentes essenciais para a caracterização do MVP de acordo com cada atribuição dos agentes.

O módulo do Enfermeiro (Figura 10) compreende dois módulos: a) módulo Dados e b) módulo Triagem Clínica. O módulo Dados contém o *component IDEnfermeiro* onde têm as informações necessárias de identificação do Enfermeiro

que estará responsável pela triagem Clínica. O segundo componente do módulo Dados é o *component ListaEsperaTriagem* onde estão os pacientes aguardando a triagem clínica. A ordem de atendimento deste componente é a de chegada na UE, a partir desta lista é determinada a ordem da triagem clínica.

Figura 10 – Módulo Enfermeiro



Fonte: O autor.

Quando o paciente é encaminhado para a triagem clínica, inicia-se o módulo Triagem Clínica, esta é a parte mais extensa da pesquisa por conter todas as informações básicas para que o sistema de e-Triagem *Vital Signs* funcione. É neste módulo que se coleta os dados do paciente que servirão como guia o CDSS e para o ML. Os componentes deste módulo são:

a) **component IDPaciente:** este componente é o mesmo componente que está no módulo Atendente e contêm todas as informações de identificação do paciente coletadas pelo atendente;

b) **component TriagemClínica:** este componente contém informações coletadas pelo enfermeiro de triagem do paciente que se subdividem nos dados:

- **Queixa Principal:** consiste na queixa que o paciente apresenta ao enfermeiro na anamnese inicial;
- **História da Doença Atual:** consiste na história da doença que foi apresentada pelo paciente na anamnese;
- **Antecedentes médicos pessoal e familiar:** consiste nas doenças pré-existentes do paciente;
- **Alergias:** consiste nas alergias que o paciente manifesta;
- **Medicações:** consiste nas medicações que o paciente está em uso no momento da anamnese (medicação contínua; c) component VitalSigns: o que será entregue;

c) **component AtribuirPulseira:** listagem de todos os recursos e funcionalidades que precisa ser construído do produto;

Após o término da coleta dos dados do módulo Enfermeiro, estes dados são compreendidos em um dashboard de monitoramento em tempo real de responsabilidade do Enfermeiro que está na função de triagem clínica. O *dashboard* da *Vital Signs* (Figura 11) contêm as principais informações dos pacientes monitorados para que se acompanhe a possíveis evoluções da doença.

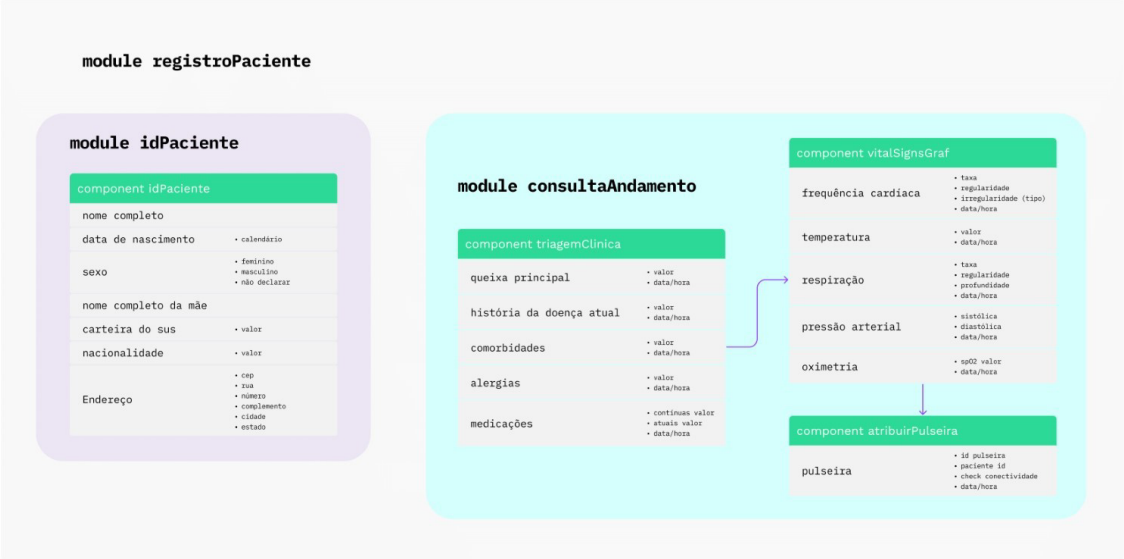
Figura 11 – Módulo Dashboard

Fonte: O autor.

O módulo *registroClinicoPaciente* (Figura 12) compreende todas as informações do paciente já coletadas previamente em outras consultas, bem como a

atualização em tempo real da consulta atual do paciente. Nesta primeira parte do módulo *registroClinicoPaciente* temos os módulos internos *idPaciente* e *consultaAndamento*.

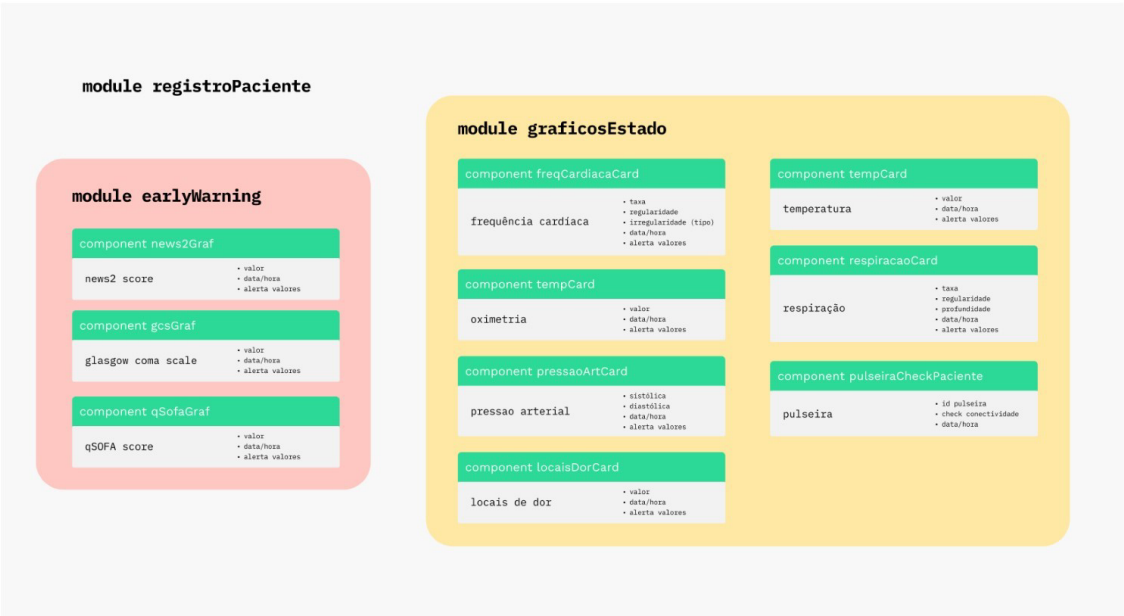
Figura 12 – Módulo de Identificação do Paciente e Consulta em Andamento



Fonte: O autor.

Na segunda parte do módulo *registroClinicoPaciente* (Figura 13) temos os módulos internos *earlyWarning* e *graficosEstado*.

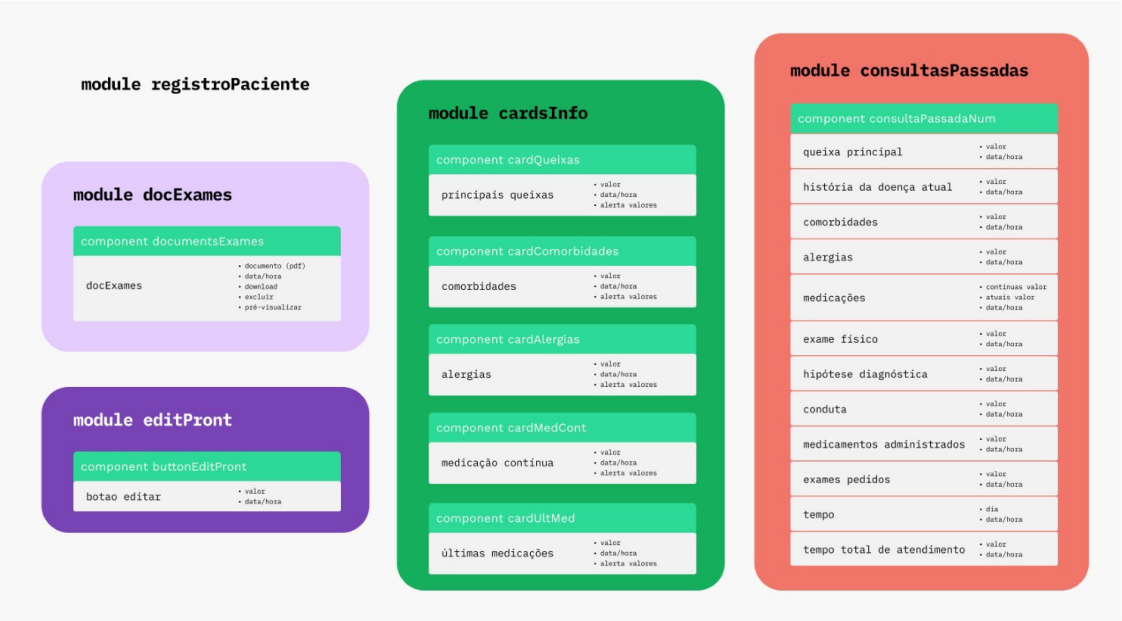
Figura 13 – Módulo de Alerta Precoce e Gráfico de Estado



Fonte: O autor.

E na terceira parte do módulo *registroClinicoPaciente* temos os módulos internos *docExames*, *editPront*, *cardsInfo* e *consultasPassadas* (Figura 14).

Figura 14 – Módulo de Exames, Editar Prontuário, Cartões de Informações Rápidas e Consultas Passadas



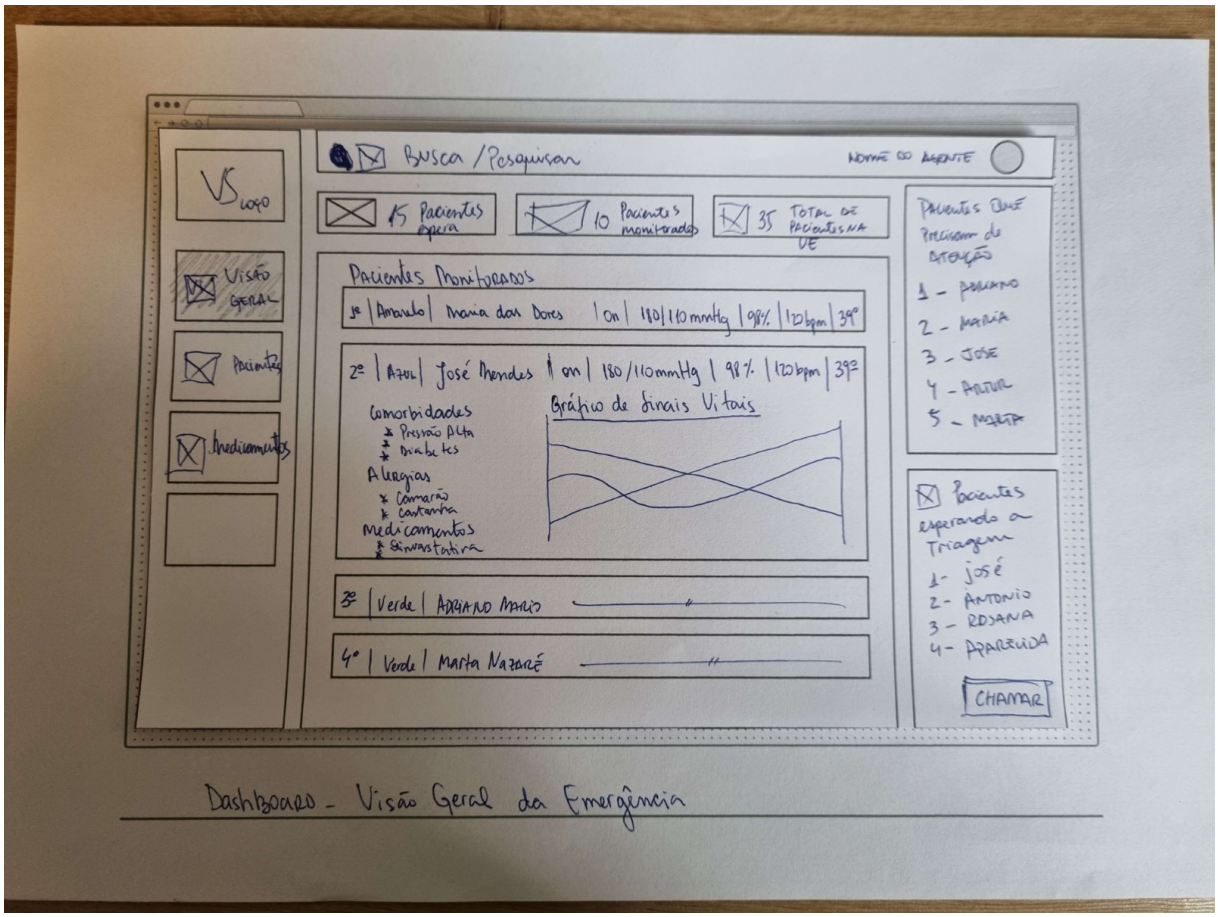
Fonte: O autor.

Após catalogar todos os módulos de coleta de dados determinantes para a pesquisa que serão mostrados de forma visual no sistema, passou-se para a fase de desenvolvimento dos *wireframes*.

3.2.7 Wireframes

Após o término da coleta dos dados essenciais do projeto, bem como sua caracterização e organização, desenvolveu-se os wireframes do *software Vital Signs*. Na Figura 15 temos o *wireframe* do *dashboard* simulando com suas funções principais.

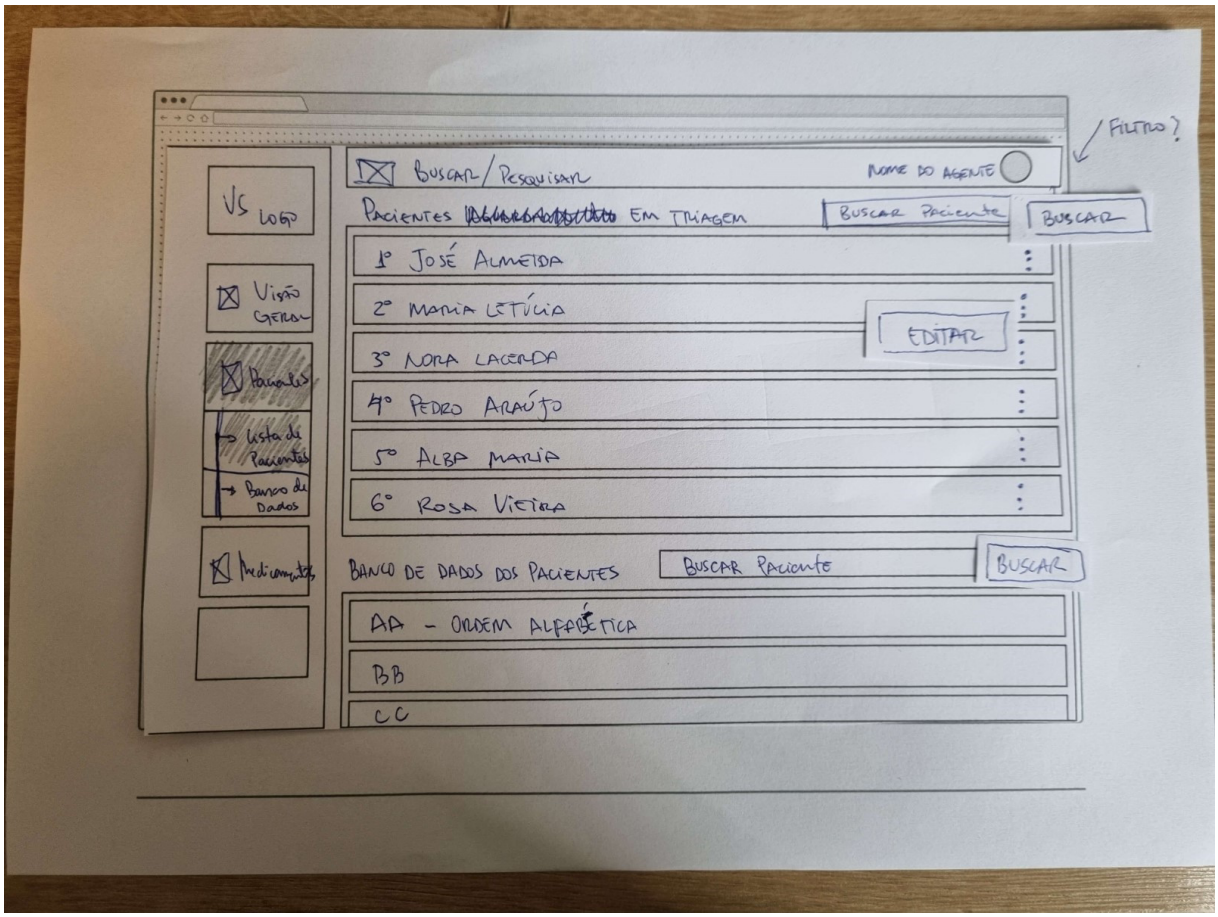
Figura 15 – Dashboard



Fonte: O autor.

Na Figura 16 temos o wireframe da lista dos pacientes, a lista que aguarda triagem e o banco de dados com todos os pacientes cadastrados no sistema.

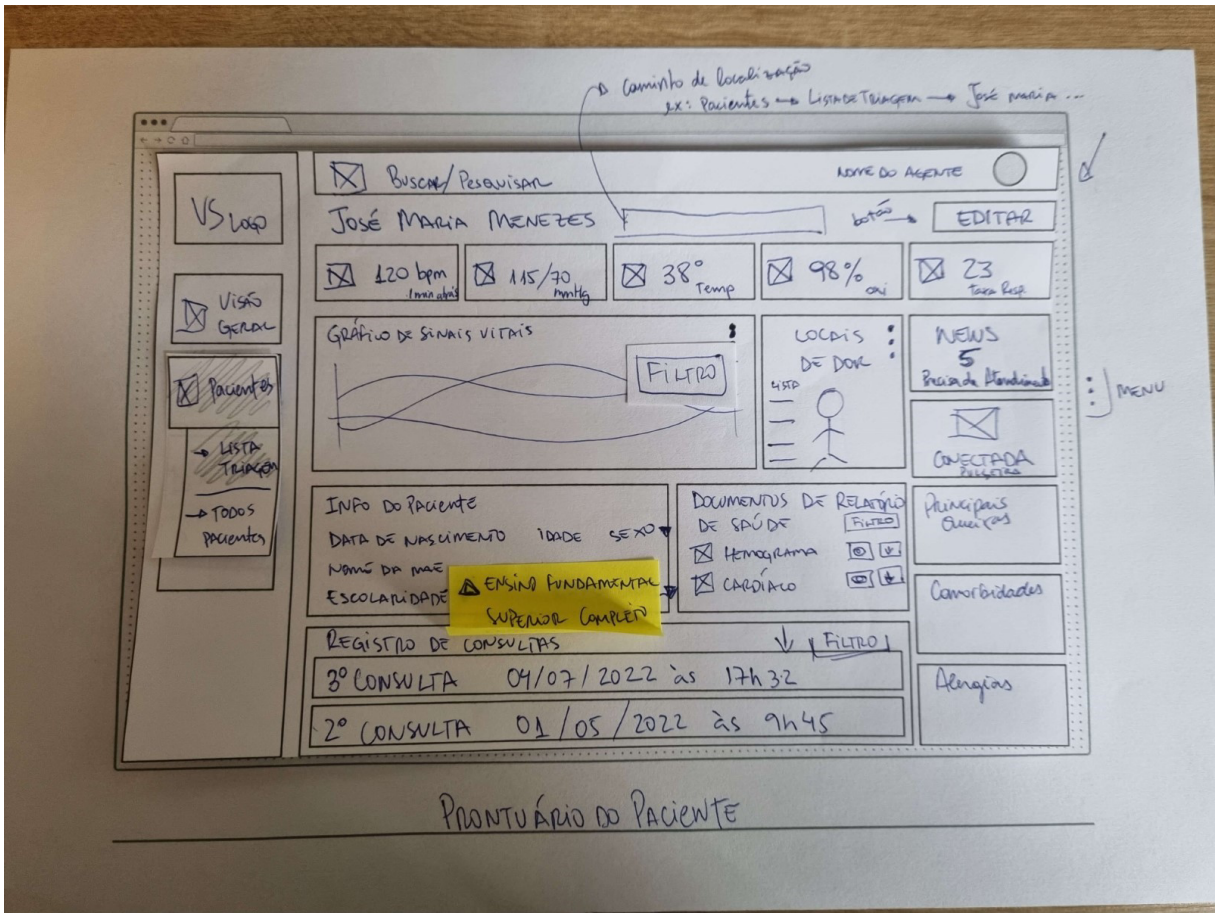
Figura 16 – Lista de paciente



Fonte: O autor.

Na Figura 17 temos o *wireframe* do Prontuário do Paciente com todas as informações cadastradas previamente e a parte de consulta que é o cadastro da consulta que está ocorrendo.

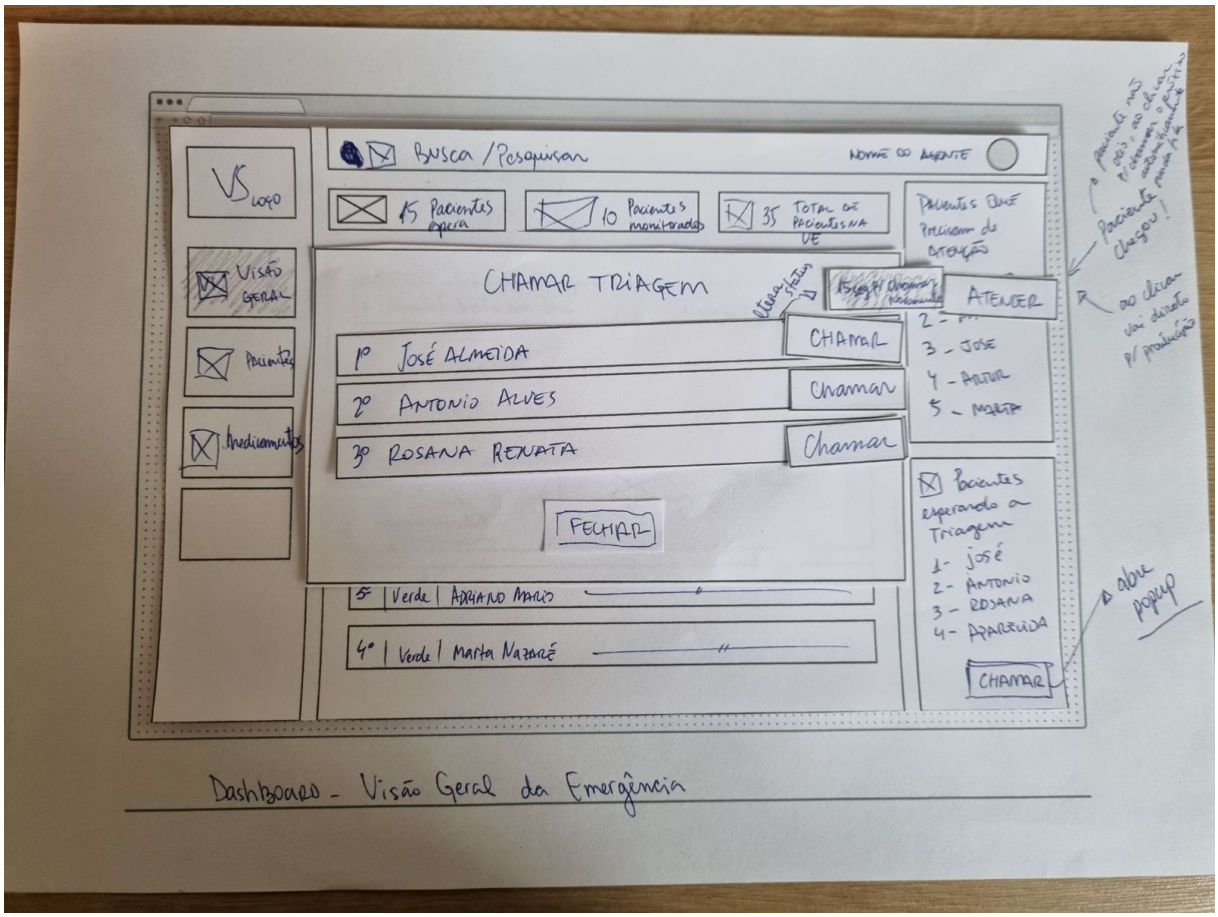
Figura 17 – Prontuário do Paciente



Fonte: O autor.

Na Figura 18 temos a simulação de chamar o paciente que está esperando para a triagem com o enfermeiro. Abre-se um pop-up em tela com os pacientes que já foram previamente cadastrados pelo atendente e estão aguardando a triagem. O enfermeiro tem a opção de chamar todos, mas o sistema organiza conforme a entrada na UE, assim que o enfermeiro chama o paciente, o sistema cria uma contagem de 15 segundos como espera para o paciente se direcionar até a sala de atendimento, caso o paciente não chegue nesse tempo, o sistema abre para chamá-lo novamente ou chamar o próximo. Caso o paciente já esteja esperando o atendimento, é só clicar no nome do paciente que abrirá o prontuário para atendimento.

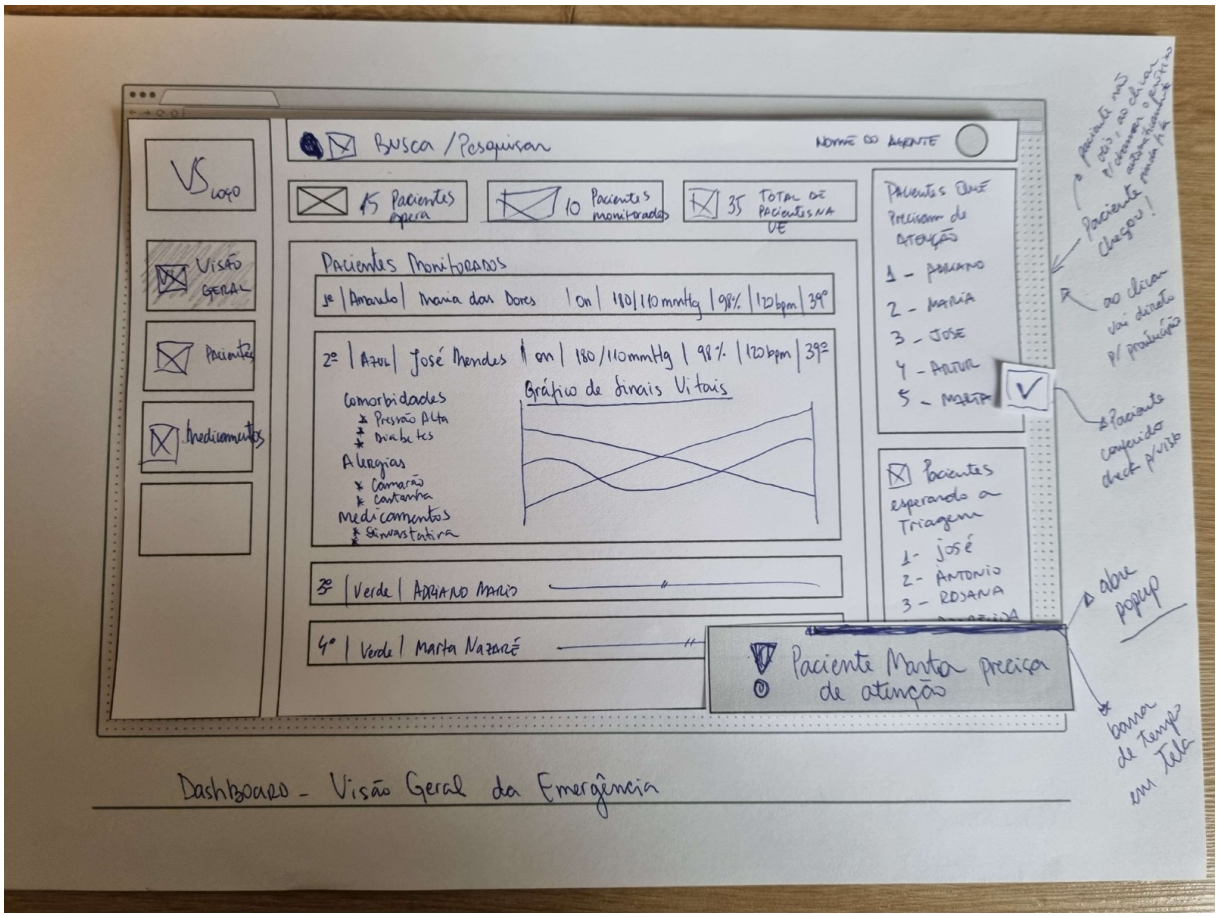
Figura 18 – Chamar Paciente



Fonte: O autor.

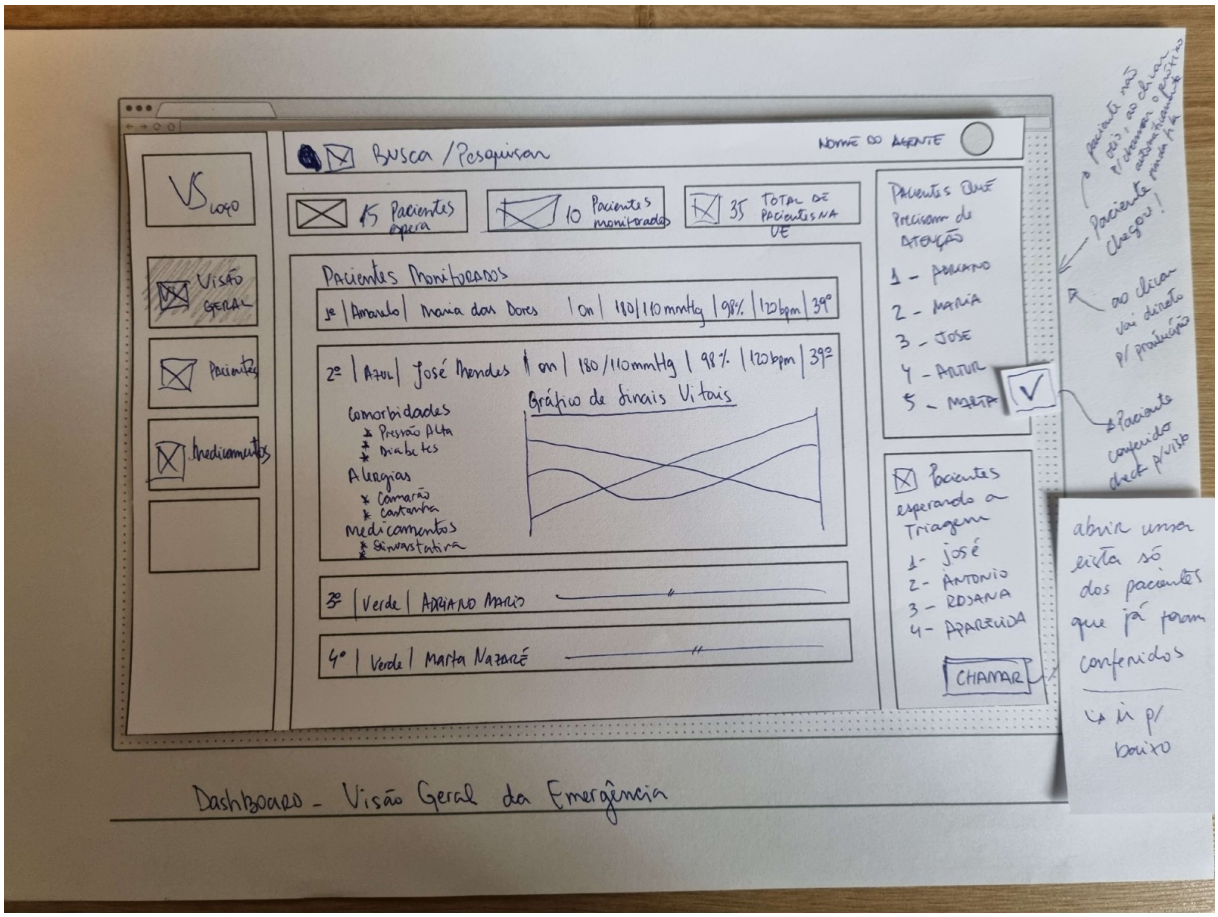
E na Figura 19 temos a notificação de que o paciente precisa de atendimento. O enfermeiro deve ir até o paciente, observar sua mudança de estado, e ao voltar ao sistema, deve informar que já foi feita a verificação, marcando o *check* ao lado do nome do paciente, como na Figura 20. Logo abaixo, o sistema criará uma lista de que o paciente que já foi monitorado, mas se houver alguma modificação no seu estado de saúde, o sistema gera um novo alerta ao enfermeiro.

Figura 19 – Notificação de Atenção



Fonte: O autor.

Figura 20 – Atendimento do Paciente em Atenção



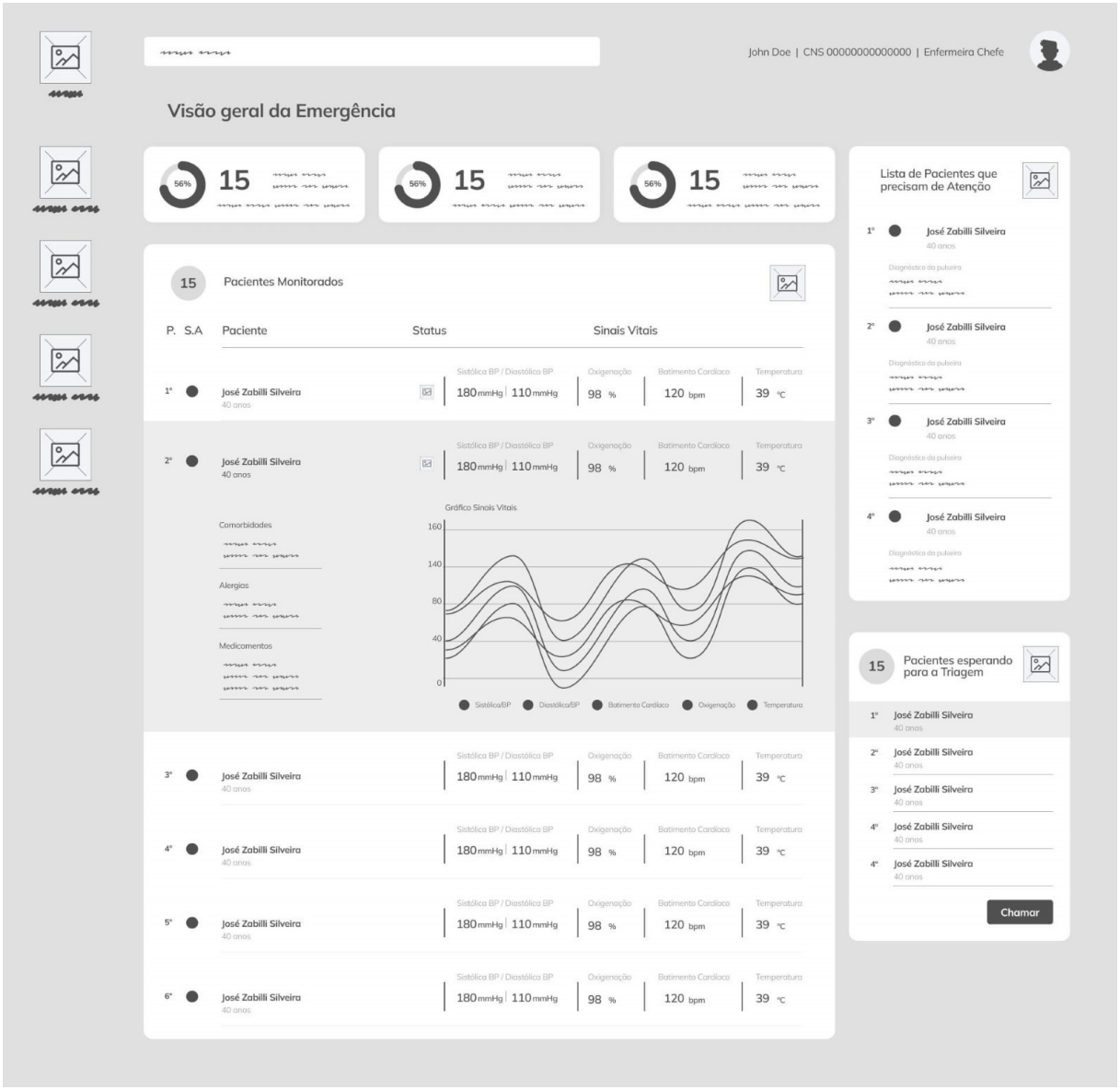
Fonte: O autor.

Os *wireframes* foram criados considerando as heurísticas de Nielsen (1994) focado nos itens sobre consistência dos dados, visibilidade do sistema, prevenção e recuperação de erro, eficiência no uso e minimalismo. A etapa que se seguiu foi o desenvolvimento de um protótipo de baixa fidelidade.

3.2.8 Protótipo de baixa fidelidade

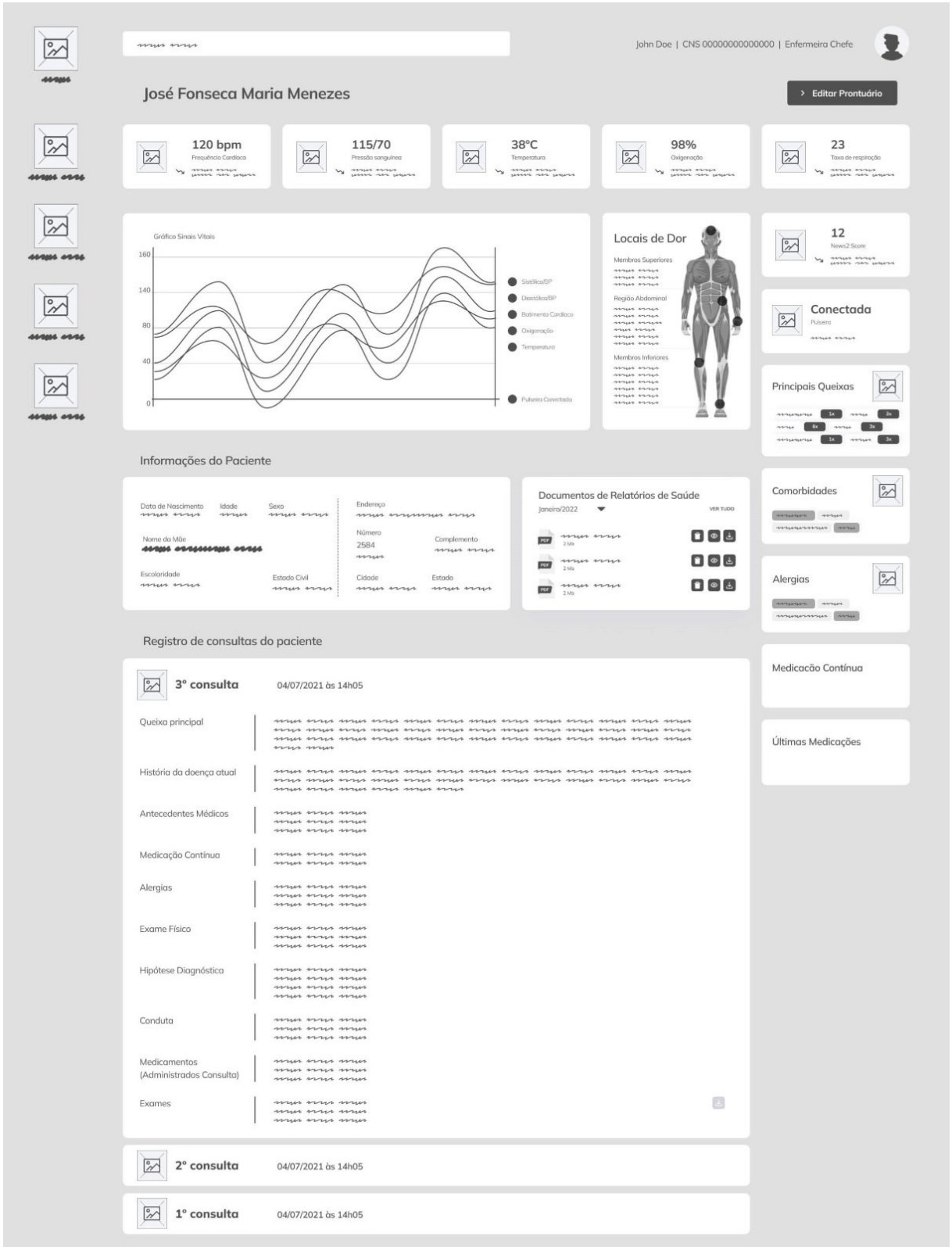
A partir dos wireframes criou-se o protótipo em baixa fidelidade.

Figura 21 – Protótipo do Dashboard



Fonte: O autor.

Figura 22 – Protótipo do Registro de Paciente



Fonte: O autor.

O próximo passo da pesquisa foi a validação do protótipo funcional por meio dos testes de usabilidade com os usuários, simulando ações do dia a dia na urgência

e emergência a fim de iniciar a programação e o desenvolvimento do MVP do *software* de e-Triagem *Vital Signs*.

3.3 TESTES DE USABILIDADE

Como proposta da presente pesquisa foi observar a interação da equipe que trabalha em urgência e emergência com o *software* de triagem eletrônica (e-triagem), serão apresentadas abaixo as etapas do processo metodológico que foi aplicado para validar as hipóteses levantadas na pesquisa.

Após os testes de usabilidade, foi aplicado um questionário de satisfação geral, com perguntas baseadas em dois testes, o NPS (*Net Promoter Score*) e o SUS (*System Usability Scale*), a fim de coletar informações gerais sobre o sistema proposto e os testes realizados. Também foram feitas anotações dos dados de preferências que consiste em dados mais subjetivos que medem os sentimentos ou opiniões de um participante acerca da utilização do sistema.

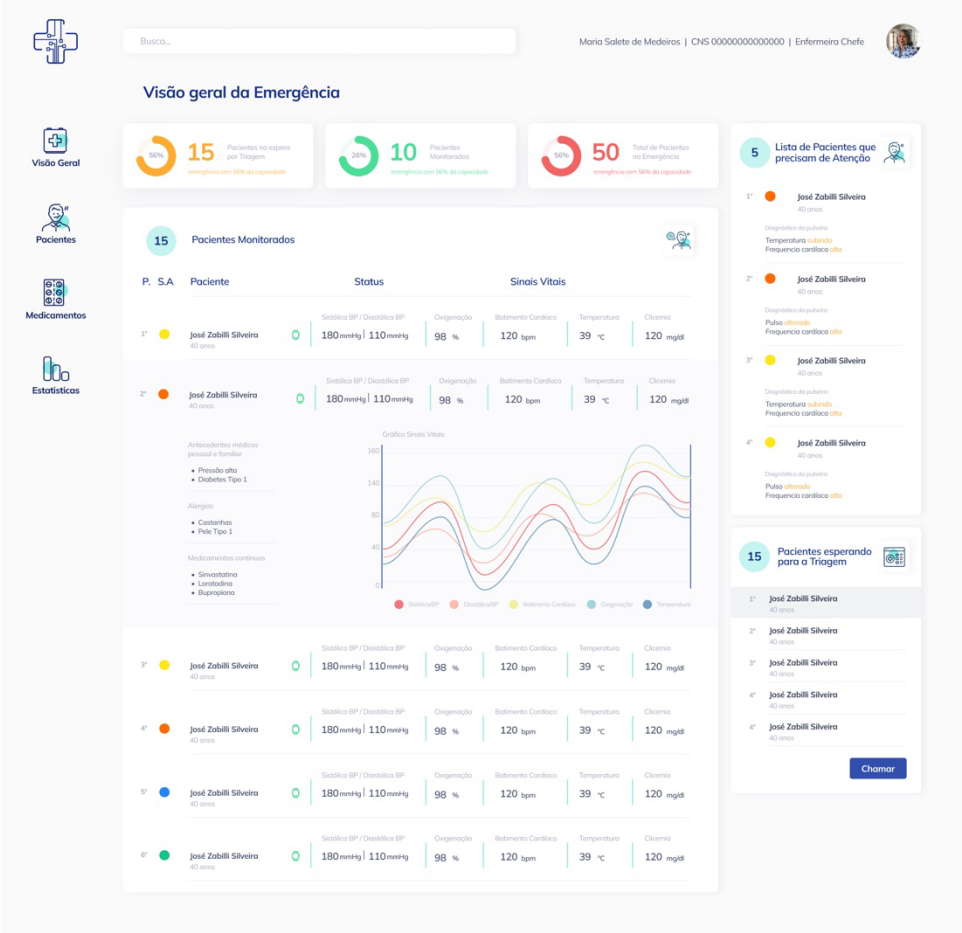
3.3.1 Materiais

Para as medições do experimento, foi utilizado notebook com sistema de gravação de tela, o *Open Broadcaster Software* (OBS), para a gravação do teste de usabilidade do *software* de triagem eletrônica (e-Triagem) desenvolvido pelo pesquisador.

3.3.2 Sistema de triagem eletrônica

Com o objetivo de otimizar e tornar mais precisa a triagem de pacientes em situações de urgência e emergência, foi desenvolvido um *software* que permite à equipe de enfermagem registrar os pacientes em todas as etapas do processo. Além disso, o sistema possibilita o monitoramento contínuo e em tempo real dos pacientes após a triagem.

Figura 23 – *Software* de triagem eletrônica (e-triagem)



Fonte: O autor.

3.3.3 Local, layout e papel do pesquisador

A condução dos testes de usabilidade foi realizada em um ambiente controlado, simulando a execução de tarefas em um cenário de uso realista. Este procedimento seguiu as diretrizes metodológicas propostas por Rubin e Chisnell (2008) para avaliações de usabilidade (Rubin; Chisnell, 2008). A abordagem adotada fundamentou-se na ergonomia da interface, a qual, segundo Moraes e Mont'Alvão (2010), prioriza os aspectos físicos da interação humano-computador com o objetivo de otimizar a manipulação dos controles e facilitar o uso do sistema (Anamaria de Moraes; Claudia Mont'Alvão, 2010).

Durante as sessões de teste, o pesquisador atuou como facilitador, sendo responsável por conduzir o participante por todo o processo. As atribuições do pesquisador incluíram: (a) uma breve explicação sobre os objetivos da pesquisa; (b) a apresentação das tarefas a serem executadas e o acompanhamento de sua

realização; e (c) o registro sistemático do desempenho do participante para a subsequente análise dos resultados.

3.3.4 Testes de usabilidade aplicados

A metodologia para a coleta de dados envolveu a aplicação de testes de usabilidade para aferir métricas de desempenho e de preferência relativas ao software de triagem eletrônica (e-triagem) proposto. Foram desenvolvidas tarefas com graus de dificuldade distintos, direcionadas ao público-alvo da pesquisa, composto por enfermeiros, médicos e técnicos de enfermagem. A coleta de dados de desempenho focou-se em indicadores quantitativos, como a taxa de erros e o tempo para a realização da tarefa, enquanto os dados de preferência capturaram as percepções e opiniões qualitativas dos participantes. A análise dos dados obtidos por meio destes procedimentos permitiu a avaliação sistemática da solução proposta.

3.3.5 Questões da pesquisa

As questões que este estudo se propôs a responder foram as seguintes:

I. O profissional da saúde consegue identificar todos os campos de entrada?

II. A ordem de inserção dos dados no software reflete o fluxo de trabalho da triagem tradicional?

III. O profissional de saúde encontra dificuldades na edição dos dados inseridos?

IV. O profissional de saúde consegue identificar a saída dos parâmetros da triagem automática no software?

V. O profissional de saúde consegue monitorar o paciente e observar as mudanças de estado de forma automática?

VI. O profissional da saúde encontra dificuldade na realização de alguma tarefa no software proposto?

VII. Com que facilidade o profissional da saúde consegue realizar a tarefa?

VIII. Com que facilidade profissional da saúde desfaz uma tarefa incorreta?

IX. Qual é a porcentagem/quantidade de erro na realização da tarefa?

X. O software dá um feedback adequado ao usuário quando executa a tarefa?

XI. Como o profissional da saúde se sente em relação a execução da tarefa?

XII. Quanto tempo, em média, o profissional da saúde leva para conseguir obter uma tarefa completa?

XIII. Quais são as dúvidas que mais aparecem em relação ao uso do software?

3.3.6 Tarefas

Foram criadas 4 tarefas relacionadas ao uso do *software* de triagem eletrônica (e-triagem) executadas por cada participante, a fim de obter dados para responder a hipótese da pesquisa. Os testes foram feitos em um notebook com o protótipo funcional do sistema rodando no Figma e duraram de 15 a 30 minutos.

Partindo do pressuposto a dificuldade de utilização de um sistema novo, que propõe uma triagem eletrônica em tempo real, o presente estudo tem o objetivo de testar a usabilidade do software em tarefas cotidianas da UE. Em termos específicos, o estudo busca:

I. Identificar possíveis dificuldades na inserção de dados;

II. Testar a eficácia geral do controle de pacientes inseridos do sistema;

III. Identificar possíveis problemas de controle dos pacientes inseridos no sistema;

IV. Identificar possíveis problemas ou soluções em relação ao software.

3.3.7 Lista de tarefas

A lista de tarefas simula a realização das tarefas mais comuns no preenchimento do prontuário eletrônico do paciente em uma UE. No teste de usabilidade o participante realizou as tarefas (Tabela 1) em um protótipo funcional simulando os comandos no software, inserindo dados a pacientes fictícios.

- **Tarefa 1 (considerada fácil):** observar se houve mudança no estado de algum paciente monitorado no sistema;
- **Tarefa 2 (considerada fácil)** observar se há paciente novo aguardando triagem e chamar paciente para atendimento;
- **Tarefa 3 (consideradas medianas):** preencher queixa principal e história da doença;

- **Tarefa 4 (consideradas difíceis):** editar queixa principal e história da doença.

Tabela 1 – Componentes e descrições das Tarefas

Tarefa	Descrição
Tarefa 1	Observar se houve mudança + se (sim) abrir prontuário
Tarefa 2	Observar se entrou um novo paciente para triagem + se (sim) chamar para triagem
Tarefa 3	Selecionar queixa principal + escrever sobre a queixa + preencher queixas secundárias se existentes + salvar
Tarefa 4	Voltar ao prontuário e editar queixa principal + editar a queixa + salvar + voltar ao prontuário
Critério de tarefa bem-sucedida	Encontrar os botões corretos e conseguir finalizar a tarefa.

Fonte: O autor

3.3.8 Cenário das tarefas

Tarefa #1

Observar se houve mudança no estado de algum paciente monitorado no sistema: Você está monitorando os pacientes dentro do sistema. No sistema, os sinais vitais do paciente são aferidos em tempo real. Conforme o sistema encontra mudanças dos dados do paciente que necessitam de atenção, ele emite um alerta. Nesse momento, o sistema emitiu um alerta. Então:

- Acesse o prontuário do paciente;
- Ao avaliar os dados, você percebeu que o paciente precisa de atenção, chame-o para uma nova avaliação.

Tarefa #2

Observar se há paciente novo aguardando triagem e (C) **Chamar paciente para atendimento:** Você está monitorando os pacientes dentro do sistema. No sistema, você tem acesso em tempo real dos pacientes que deram entrada no UE e foram inseridos em ordem de chegada para serem chamados para a triagem. Ao clicar em chamar, o sistema seguirá a ordem de chegada. Ao clicar no paciente, o sistema chamará o paciente escolhido. Então:

- Observe a lista de pacientes esperando a triagem;
- Chame um paciente para triagem;

Tarefa #3

Preencher queixa principal e história da doença: Agora que você chamou para a triagem e o paciente está na sua frente, você precisa preencher o prontuário da consulta na emergência. Você já tem um prontuário prévio desse paciente no sistema e é preciso acrescentar uma nova consulta.

- Vá na área de consultas;
- Crie uma consulta;
- Preencha a queixa principal e a história da doença;
- Salve a consulta;
- Observe se está correto;
- Volte ao prontuário.

Tarefa #4

Editar queixa principal e história da doença: Você precisa editar a consulta para acrescentar os sinais vitais aferidos por você. Você já tem a consulta iniciada desse paciente no sistema e é necessário acrescentar essas novas informações.

- Vá na área de consultas;
- Abra a consulta em andamento;
- Preencha os sinais vitais aferidos por você;
- Salve a consulta;
- Volte para o prontuário.

3.3.9 Questionário pós-teste

O questionário pós-teste serviu para coletar as medidas de preferência do usuário após os testes. Esse questionário foi composto por 10 perguntas para calcular o SUS (*System Usability Scale*), em que cada uma das questões foi apresentada em uma escala *Likert* de 5 pontos, onde 1 significa "Discordo totalmente" e 5 significa "Concordo totalmente" (Brooke, 1995). Nesse questionário, 5 perguntas são afirmações positivas e 5 perguntas são afirmações negativas. As 10 afirmações do questionário foram as seguintes:

- Eu gostaria de usar esse sistema com frequência;

- Eu achei o sistema desnecessariamente complexo;
- Eu achei o sistema fácil de usar;
- Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema;
- Eu acho que as várias funções do sistema estão muito bem integradas;
- Eu acho que o sistema apresenta muita inconsistência;
- Eu imagino que as pessoas aprenderão como usar esse sistema rapidamente;
- Eu achei o sistema atrapalhado de usar;
- Eu me senti confiante ao usar o sistema;
- Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.

Para calcular a pontuação SUS, as respostas dos participantes foram avaliadas, e o cálculo da pontuação foi realizado da seguinte forma: Para as questões ímpares (1, 3, 5, 7 e 9), subtraiu-se 1 da resposta. Para as questões pares (2, 4, 6, 8 e 10), subtraiu-se a resposta do participante do número 5. Em seguida, somaram-se os novos valores de cada item e multiplicou-se a soma total por 2,5 para obter a pontuação SUS (que varia de 0 a 100). A pontuação final foi a média da pontuação de todos os participantes.

Além das 10 questões aplicadas para o cálculo do SUS, o questionário pós-teste também contou com uma pergunta para cálculo do *Net Promoter Score* (NPS). A pergunta foi: *em uma escala de 0 a 10 (onde 0 é não recomendar e 10 recomendar), qual a probabilidade de recomendar o software?* De acordo com as respostas obtidas, os participantes foram classificados em três categorias: promotores (9-10), neutros (7-8) e detratores (0-6). O valor de NPS foi calculado subtraindo a porcentagem de participantes que que foram classificados como detratores da porcentagem de participantes que foram classificados como promotores (Reichheld, 2003).

Por fim, uma pergunta complementar também foi realizada: *Considerando a sua experiência no teste realizado hoje, qual a probabilidade de você adotar o sistema no seu dia a dia?" Em uma escala de 0 a 10, onde 0 não é "de jeito nenhum", 10 é extremamente provável.* Ainda, os participantes tiveram um espaço final no questionário onde ficaram livres para dar suas sugestões e realizar comentários.

3.3.10 Dados a serem coletados e medidas de avaliação

Os dados coletados servirão para responder as questões de pesquisa propostas. De acordo com a literatura, serão avaliados dados de desempenho (sem assistência por parte do pesquisador) e dados de preferência. Sendo eles:

I. Dados de desempenho

- Número e percentagem de tarefas concluídas com sucesso;
- Número de toques além dos especificados para cada tarefa;
- Número de comandos totais (com e sem toques desnecessários) médio (entre todos os participantes) realizados para cumprir todas as tarefas;
- Tempo que o participante leva para completar a tarefa;

II. Dados de preferência:

- Protocolo *Think-Aloud* (pense em voz alta);
- Facilidade/Dificuldade geral de uso;
- Facilidade/Dificuldade geral de aprendizagem;
- Utilidade do produto;
- Expectativas relacionadas ao uso;
- Comentários negativos em relação às modificações propostas;
- Sugestões para melhorar as modificações propostas;
- Identificação dos comentários (negativos e positivos) mais frequentes.

3.4 MODIFICAÇÃO DO ESCORE DE ALERTA PRECOCE – NEWS2-M

Para que o sistema funcionasse de forma automatizada, apenas aplicar os protocolos de Classificação de Risco não seria suficiente, uma vez que a proposta da pesquisa está centrada em utilizar dados em tempo real coletados pela pulseira multiparamétrica e dados prévios do prontuário do paciente. Os protocolos de risco atualmente utilizados como o PMT e ESI utilizam-se de um fluxograma sobre o estado do paciente assim que chega à urgência/emergência, em que, neles, são utilizados alguns dados categóricos para classificação, mas há outros dados que são levados em consideração, como a quantidade de procedimentos que o paciente irá precisar, entre outros. Dessa forma, o protocolo acaba dependendo do conhecimento técnico do profissional da saúde para fazer a classificação.

Considerando que a proposta central desta pesquisa é o monitoramento contínuo e precoce da deterioração do paciente a partir de dados automatizados provenientes da pulseira multiparamétrica, a seleção de um protocolo de avaliação de gravidade era uma etapa fundamental. Optou-se pelo *National Early Warning Score 2* (NEWS2), um escore de alerta de ampla utilização e que abrange seis parâmetros fisiológicos vitais: SpO_2, Pressão Arterial Sistólica, Pulso, Temperatura, Frequência Respiratória e Nível de Consciência (vide Figura 3, Página 31).

Apesar do protocolo de risco de deterioração NEWS2 considerar um número maior de sinais vitais que o *Emergency Severity Index* (ESI), ele não leva em conta alguns parâmetros importantes como a pressão arterial diastólica do paciente e o nível de dor. A análise da pressão arterial diastólica, por exemplo, tem recebido bastante atenção nos últimos anos pela comunidade médica, principalmente devido à chamada Hipertensão Diastólica Isolada (IDH na sigla em inglês *Isolated Diastolic Hypertension*). Estudos têm definido que pacientes com IDH são aqueles que apresentam pressão arterial sistólica inferior a 140 mmHg e diastólica igual ou superior a 90 mmHg (Kaneko et al., 2020; Xie et al., 2021; Yano et al., 2022). Essa definição também é apoiada pela *European Society of Cardiology* (ESC) (Williams et al., 2018). Investigações sugerem que a IDH pode estar associada a eventos adversos principalmente em pacientes jovens, incluindo problemas cardiovasculares, insuficiência cardíaca, fibrilação atrial e doença renal crônica (Huang et al., 2022; Kaneko et al., 2020; Li et al., 2014). Portanto, para aumentar a sensibilidade na detecção da Hipertensão Diastólica Isolada (IDH), adaptamos o escore de risco NEWS2. A modificação consiste em atribuir uma pontuação +2 ao algoritmo para pacientes que apresentam pressão arterial sistólica inferior 140 mmHg e diastólica igual ou superior a 90 mmHg.

Além de acrescentar o dado referente à Hipertensão Diastólica Isolada, também foi feita outra adaptação do NEWS2 relacionada a pressão arterial diastólica. Neste caso, acrescentamos um peso de +3 pontos para uma pressão arterial diastólica inferior a 60 mmHg em pacientes idosos (65 anos ou mais) com uma pressão sistólica normal (superior a 100 mmHg). Em um amplo estudo de 2011, foi verificado que adultos mais velhos que se enquadram nessas condições têm maior risco de desenvolver insuficiência cardíaca de início recente (Guichard et al., 2011). Além disso, Li et al. (2021), em ensaios clínicos com 7.515 pacientes adultos, verificaram que aqueles com uma pressão arterial sistólica tratada (inferior a 130

mmHg) e pressão arterial diastólica inferior a 60 mmHg, foram associados com um alto risco cardiovascular (Li et al., 2021).

Outra modificação fundamental no protocolo foi a inclusão do nível de dor, um parâmetro crítico no *Emergency Severity Index* (ESI), mas ausente no NEWS2. Propôs-se, adicionalmente, a incorporação do nível de dor ao escore, uma variável não contemplada pelo NEWS2. A fundamentação para tal ajuste reside no protocolo ESI, em cujo algoritmo a dor severa (≥ 7 em 10) constitui o Passo B, um critério de triagem de alta relevância, superado em importância apenas pela necessidade de intervenções imediatas para manutenção da vida (Passo A) (Emergency Nurses Association - ENA, 2023).

A implementação no NEWS2 modificado buscou espelhar essa urgência. No escore NEWS2 original, os desvios fisiológicos considerados extremos recebem uma pontuação de +3, sinalizando a necessidade de uma resposta clínica emergencial (NICE, 2020). Dessa forma, foi estabelecido que um relato de dor severa (≥ 7) também resultaria na adição de +3 pontos, equiparando seu peso ao dos indicadores de maior gravidade do sistema.

Adicionalmente, realizou-se outro ajuste metodológico fundamental na estratificação da Pressão Arterial Sistólica (PAS) do escore NEWS2. No protocolo original, um intervalo excessivamente amplo, de 111 a 219 mmHg, é classificado como normal, não recebendo qualquer pontuação (0 pontos). Essa faixa, contudo, diverge significativamente dos padrões clínicos estabelecidos para o diagnóstico de hipertensão. Segundo as diretrizes de 2023 da Sociedade Europeia de Hipertensão (ESH), que norteiam a prática clínica, a hipertensão estágio 1 é definida por níveis de PAS entre 140-159 mmHg ou de Pressão Arterial Diastólica (PAD) entre 90-99 mmHg (Mancia et al., 2023). Isso evidencia que a faixa de normalidade do NEWS2 é tão ampla que falha em identificar e estratificar o risco de pacientes com hipertensão estabelecida, diminuindo a sensibilidade do escore nesta população.

Seguindo os protocolos e referências descritas acima, foi desenvolvido, então, o novo NEWS2, chamado de NEWS2 modificado ou NEWS2-M, que está apresentado na seção Resultados e Discussão.

Por fim, utilizando a nova classificação de risco proposta (NEWS2-M), aplicaram-se as pontuações dessa nova classificação a dados de uma coorte retrospectiva composta por 390.039 pacientes, cujas informações foram extraídas do banco de dados MIMIC-ED (v2.2) (*Medical Information Mart for Intensive Care* –

Emergency Department). Estes dados, originalmente classificados pelo escore ESI, foram reclassificados com a nova classificação proposta (NEWS2-M), bem como com o NEWS2 original, utilizando o software Excel (Microsoft Corporation). A relação entre as classificações obtidas pelos escores foi analisada por meio de tabelas de contingência, permitindo a avaliação da frequência de distribuição dos pacientes em cada combinação de categorias e a identificação de padrões de concordância ou divergência entre os métodos.

3.5 MACHINE LEARNING

Esta seção detalha a metodologia empregada no teste de algoritmos de *Machine Learning* para predição de deterioração clínica em pacientes no setor de emergência. A metodologia foi desenhada em etapas sequenciais e interdependentes, partindo da fundamentação clínica e definição do problema, passando pela preparação dos dados, e culminando no treinamento, avaliação e calibração de um conjunto de algoritmos preditivos.

3.5.1 Delineamento do Estudo e Definição do Problema

O ponto de partida para esta pesquisa foi a necessidade de um sistema automatizado para monitoramento contínuo e precoce do risco de deterioração do paciente. O objetivo do modelo de *Machine Learning* é prever a necessidade de internação hospitalar com um *proxy* para a deterioração clínica.

É um desafio para um sistema computacional medir diretamente a 'deterioração clínica', pois ela envolve uma combinação de sinais vitais, resultados de exames, relatos de pacientes e a avaliação de um médico (Chua et al., 2023). Por isso, em vez de tentar prever esse conceito abstrato, um modelo pode focar em prever um evento concreto que serve como um forte indicativo dessa piora: a necessidade de internação hospitalar que pode ser registrada como "sim" ou "não" (Agulnik et al., 2017, 2018; Chua et al., 2023; Rakers et al., 2024).

O uso de dados binários como *proxy* pode ser útil para monitorar o estado de pacientes em ambientes clínicos, onde a necessidade de intervenções rápidas é crítica (Agulnik et al., 2017; Rakers et al., 2024).

A variável-alvo foi definida como binária: pacientes com desfecho ADMITTED (admitido para internação) foram classificados como classe positiva (1), enquanto aqueles com desfecho HOME (alta para casa) foram classificados como classe

negativa (0) que estão presentes na base de dados. Outros desfechos foram excluídos da análise para garantir um problema de classificação binário.

Como fonte de dados, o estudo utilizou o banco de dados MIMIC-IV-ED (v2.2), um repositório público e anonimizado de prontuários eletrônicos do *Beth Israel Deaconess Medical Center,* já mencionado anteriormente nesse estudo (item 2.3.2). Foram utilizados três arquivos principais: edstays.csv (informações demográficas e de estadia), triage.csv (sinais vitais e queixa principal na triagem) e vitalsign.csv (séries temporais de sinais vitais durante a estadia na emergência).

3.5.2 Condução dos experimentos

Os experimentos foram conduzidos e uma máquina rodando o MacOS Sonoma apenas com CPU e também em uma máquina rodando Windows com GPU NVIDIA A4000 (para os modelos *Gradient Boosting*) com o ambiente configurado utilizando Python 3.13.5. As principais bibliotecas e suas versões incluem:

- **Manipulação e visualização de dados:** Pandas (2.3.1), NumPy (2.2.6), Matplotlib (3.10.3) e Seaborn (0.13.2)
- **Aprendizado de Máquina:** Scikit-learn (1.6.1), Imbalanced-learn (0.13.0), Shap (0.48.0), Flaml (2.3.5), H2o (3.46.0.7), Nltk (3.9.1), Spacy (3.8.7) e Medspacy (1.3.1)
- **Algoritmos de *Gradient Boosting* e *Boosting*:** LightGBM (4.6.0), XGBoost (3.0.2), CatBoost (1.2.8), Gradient Boosting Decision Tree (GBDT) e Adaboost.
- **Algoritmos:** *Logistic Regression* (LR), L1 (LASSO), L2 (RIDGE) Random Forest (RF), Extra Trees, SVM, LinearSVC (Support Vector Machine), Ridge (L2), k-Nearest Neighbors (k-NN) e Gaussian Naïve Bayes (GaussianNB).

Para garantir a reprodutibilidade, que é capacidade de obter os mesmos dados toda vez que o mesmo código for executado, é preciso controlar a aleatoriedade e para isso utiliza-se uma semente aleatória (em inglês *random seed* ou *random state*). Esse método cria como se fosse um “ponto de partida”. No caso desse estudo, a semente aleatória global (*random_state*) foi fixada em 42. Dessa forma a divisão dos dados de treino, validação e teste será sempre a mesma, ou seja, os pacientes estarão sempre nos mesmos conjuntos em todas as execuções.

Por curiosidade, a utilização do 42 não tem a ver com algum número específico da matemática e sim um número derivado do livro o Guia do Mochileiro das

Galáxias, sendo o número 42 a resposta para a "Questão Fundamental da Vida, do Universo e de Tudo Mais". No caso da pesquisa, poderia ser qualquer número que o resultado garantiria reprodutibilidade, mas por uma questão de convenção nerd, foi utilizado o 42.

3.5.3 Pré-processamento e Engenharia de Atributos

Esta etapa é fundamental para o *Machine Learning*, nela é onde se transformam os dados brutos de uma base de dados em um formato estruturado e informativo para que possam ser usados pelos modelos de forma a garantir o melhor resultado possível.

Inicialmente foi feita a limpeza e padronização dos dados fisiológicos, iniciando com a temperatura, que estava originalmente em Fahrenheit na base original, para Celsius a medida padrão nacional. Para diminuir o desbalanceamento dos dados, foram aplicados cortes que removeram os *outliers* clinicamente implausíveis nos sinais vitais como frequência cardíaca 20 ou 250 bpm e saturação fora desses limites foram convertidos para nulos. Para a variável *pain* (dor) do banco de dados, foi padronizada na escala numérica inteira de 0 a 10.

Os dados brutos da tabela de sinais vitais (*vitalsign.csv*) são caracterizados por múltiplas medições por paciente. Com o objetivo de modelar a dinâmica da condição clínica, foram agregados os seguintes atributos para cada tipo de sinal vital (temperatura, frequência cardíaca etc.) em relação ao identificador de internação (*stay_id* - *si*). Como essas medições não são iguais (em quantidade) para todos os pacientes, foram utilizadas a primeira e a última ($si_{_primeiro}$, $si_{_última}$) medição registrada e o mínimo e o máximo ($si_{_min}$, $si_{_max}$) de cada dado fisiológico registrado durante a estadia do paciente utilizando como referência o estudo feito (Johnson; Mark, 2017) em suas coletas de dados. Além desses valores, também foi calculado um Δ (delta) que é a diferença entre a última e a primeira medição ($\Delta_{_prim}$, $\Delta_{_última}$) e o slope que é a taxa de variação de delta dividido pela duração da medição em horas ($\Delta \div D_{_horas}$). Isso foi pensado por que os registros não são regulares em relação ao tempo para coleta (ex: a cada 30 minuto), dessa forma se teria uma média aproximada entre todos.

Após as etapas de limpeza e padronização os *dataframes* processados (*triage_clean.csv*) e as *features* acima agregadas em (*vitals_agg.csv*) foram salvos como artefatos intermediários para auditoria e reuso.

A criação do *dataset* final (*master_dataset.csv*) foi a partir da fusão das tabelas *triage_clean.csv, vitals_agg.csv* e *edstays.csv* (contendo a variável-alvo – *stay_id*), resultando em um conjunto de dados com 399.642 linhas e 49 colunas de atributos antes da modelagem.

3.5.4 Estratégia de Modelagem

Para garantir a robustez do modelo e evitar o vazamento de dados (*data leakage*), a divisão do *dataset* foi realizada no âmbito do de paciente, uma prática fundamental em estudos clínicos com dados de múltiplas visitas. A estratégia, conhecida como "*Patient Wise Split"*, utilizou-se da proporção de 70% para treino, 15% para validação e 15% para teste, baseando a divisão no identificador único de cada paciente (*subject_id*). Essa abordagem garante que todas as visitas e observações de um mesmo indivíduo pertençam exclusivamente a um único conjunto de dados (treino, validação ou teste).

Para garantir a consistência do pré-processamento, foi implementado um fluxo de trabalho automatizado (*pipeline)* de preparação de dados utilizando a classe *ColumnTransformer* da biblioteca *Scikit-learn*. Esta abordagem permite a aplicação automatizada de sequências de transformação distintas para subconjuntos de variáveis heterogêneas, como as numéricas e categóricas presentes no dataset (scikit-learn.org).

Para as variáveis numéricas, o *pipeline* consistiu em um processo de duas etapas. Primeiramente, a imputação de dados faltantes foi realizada utilizando a mediana como estratégia, escolhida por sua robustez a valores atípicos (*outliers*). Subsequentemente, as variáveis foram submetidas à padronização por meio do *StandardScaler*. Esta etapa é metodologicamente crucial, pois transforma os dados para que possuam média zero e desvio padrão unitário. A padronização assegura que todas as variáveis contribuam de forma equitativa para o treinamento de algoritmos sensíveis à escala, como a Regressão Logística e as Máquinas de Vetores de Suporte (SVMs), evitando que *features* com maiores magnitudes dominem indevidamente o processo de aprendizagem. A adoção desta técnica é uma prática padrão em *pipelines* de Machine Learning na área da saúde (scikit-learn.org).

Para as variáveis categóricas, um *pipeline* análogo foi aplicado. A imputação dos valores ausentes foi feita com o valor mais frequente da respectiva coluna. Em seguida, as variáveis foram transformadas utilizando o *OneHotEncoder*. Esta técnica

é a abordagem padrão para variáveis nominais em modelos preditivos clínicos, pois converte cada categoria em um formato numérico binário (0 ou 1) que o algoritmo pode utilizar (scikit-learn.org). A aplicação do *OneHotEncoder* resultou em uma matriz de alta dimensionalidade, que foi armazenada de forma eficiente como uma matriz esparsa (.npz) para otimizar o uso de memória computacional.

O conjunto de dados utilizado neste estudo apresentou um desbalanceamento de classes moderado, com uma prevalência de desfechos positivos (internações) de aproximadamente 39,54%. Para mitigar o potencial viés do classificador em favor da classe majoritária, foram implementadas estratégias específicas de manejo do desbalanceamento utilizando o SMOTE (*Synthetic Minority Over-sampling Technique*) e o *scale_pos_weight* em modelos *Gradiente Boosting* (imbalanced-learn.org; xgboost.readthedocs.io)

A técnica SMOTE tem sido amplamente empregada para mitigar os problemas relacionados ao desbalanceamento de classes em modelos de base, como a regressão logística e as máquinas de vetores de suporte. O SMOTE gera novas instâncias sintéticas da classe minoritária por interpolação entre uma amostra e seus vizinhos próximos (Chawla et al., 2002). Estudos subsequentes reforçam que a interpolação realizada pelo SMOTE amplifica a área onde as decisões são tomadas, melhorando a capacidade de generalização dos classificadores e prevenindo o viés introduzido pela replicação direta (Delgado; Núñez-González, 2022). Além disso, pesquisas que avaliam a aplicação de técnicas de *oversampling* em contextos envolvendo regressão logística e máquinas de vetores de suporte evidenciam a eficácia do SMOTE na melhoria do desempenho dos modelos ao corrigir o desbalanceamento das classes (Wibowo; Masruriyah; Rahmawati, 2025).

Em contrapartida, para os modelos de *Gradient Boosting Decision Tree* (*XGBoost*, *LightGBM*, *CatBoost*), optou-se por uma abordagem de ponderação de custo, utilizando parâmetros internos como *scale_pos_weight*. Esta técnica ajusta a função de perda para penalizar mais severamente os erros de classificação na classe minoritária, forçando o algoritmo a dar maior importância a essas instâncias (Chen; Guestrin, 2016). A escolha por esta última abordagem justifica-se por sua maior eficiência, pois não há aumento no volume de dados de treinamento, tornando-a preferível para algoritmos que já exigem alto poder computacional.

3.5.5 Seleção e Justificativa dos Algoritmos

Para a tarefa de predição, foi avaliado um espectro diversificado de algoritmos de aprendizado de máquina, com o objetivo de estabelecer um *benchmark* robusto e identificar a arquitetura de modelo com maior poder preditivo. A seleção dos algoritmos foi guiada por serem utilizados na literatura e apresentarem sucesso em problemas de classificação com dados tabulares, especialmente na área da saúde.

Para estabelecer uma linha de base de desempenho, foram empregados modelos lineares *Logistic Regression* (LR), em sua forma padrão e com a regularização L1 (Lasso) e L2 (Ridge) e o *Ridge Classifier* como uma alternativa de classificação linear regularizada. Foram empregados por sua reconhecida simplicidade, solidez, interpretabilidade e eficiência computacional.

A *Decision Tree* foi incorporada ao estudo com um duplo propósito: primeiramente, para contextualizar o ganho de performance obtido pelos modelos de *ensemble* que a utilizam como bloco de construção fundamental; e, em segundo lugar, para avaliar o desempenho de um único classificador não linear.

Para explorar a capacidade de modelagem de interações complexas e não lineares, foram selecionados os algoritmos de *ensemble Random Forest* e *Extra Trees*, estes modelos agregam predições de múltiplos estimadores, o que lhes confere robustez na captura de relações não lineares e mitiga o risco de *overfitting* (Breiman, 2001).

Foram implementados, também, modelos baseados em *gradiente boosting*, *LightGBM*, *XGBoost*, *CatBoost*, *AdaBoost* e *Gradient Boosting Decision Tree* (GBDT) por construírem árvores de decisão de forma sequencial, em que cada nova árvore é treinada para corrigir os erros residuais da anterior, especialmente utilizados em modelagem para dados tabulares. Esta abordagem iterativa resulta, consistentemente, em um desempenho preditivo superior (Chen; Guestrin, 2016).

Para robustez em espaços de alta dimensionalidade, foram utilizadas as *Support Vector Machines* (*SVM*), no caso, a *LinearSVC* por uma condição frequentemente encontrada após a etapa de codificação de variáveis categóricas. O objetivo deste classificador é encontrar o hiperplano ótimo que maximiza a margem de separação entre as classes (Scikit Learn, 2025).

Por fim, os algoritmos *k-Nearest Neighbors* (k-NN) e *Gaussian Naive Bayes* (GNB), no caso do k-NN para assegurar uma análise criteriosa e mitigar os efeitos da "maldição da dimensionalidade", foi usado um pré-processamento específico,

transformandos os dados de entrada por meio da técnica de *Truncated Singular Value Decomposition* (TruncatedSVD) que realiza a redução da dimensionalidade linear (Scikit Learn, 2025).

3.5.6 Avaliação e Calibração dos Modelos

A avaliação de desempenho dos modelos foi conduzida por meio de um conjunto de métricas consolidadas na literatura de *Machine Learning*, visando a uma análise robusta pela capacidade preditiva de cada um. As métricas utilizadas foram *Area Under the Receiver Operating Characteristic Curve* (AUROC), *Area Under the Precision-Recall Curve* (PR-AUC), *F1-Score* (Combinação do *Precision* e *Recall*), *Recall* (Sensibilidade), *Precision* (Precisão do valor preditivo positivo) e o *Brier Score* (Acurácia da previsão probabilística).

Quando se trata de avaliar o desempenho de classificadores binários, duas métricas são normalmente utilizadas o PR-AUC e o AUROC.

A curva PR-AUC é uma representação gráfica da relação entre Sensibilidade x Precisão, ou seja, a Sensibilidade mede a eficácia do modelo na detecção de casos positivos — Taxa de Verdadeiros Positivos (TPR - *True Positive Rates*) —, enquanto a precisão mede a eficácia do modelo na prevenção de falsos positivos — Taxa de Falsos Positivos (FPR - *False Positive Rates*). A PR-AUC é geralmente utilizada para avaliar conjuntos de dados com acentuando desbalanceamento de classes, dessa forma, as curvas do *Precision-Recall* (PR) fornecem uma imagem mais informativa sobre o desempenho do algoritmo (Davis; Goadrich, 2006).

A curva AUROC é uma representação gráfica da relação Sensibilidade x Especificidade, ou seja, a Taxa de Verdadeiros Positivos (TPR) e Taxa de Falsos Positivos (FPR) e é uma métrica padrão bastante utilizada na tomada de decisões médicas e serve para avaliar a capacidade de discriminação de um modelo, ou seja, uma função para taxa de falsos positivos (Fawcett, 2006). Embora seja uma métrica importante em *Machine Learning*, em conjuntos desbalanceados a AUROC não é tão eficiente e isso se dá por ser afetada pela Taxa de Verdadeiros Negativos (TNR - *True Negative Rate*) não afetando o desequilíbrio de classe.

Portanto, a AUROC mede a capacidade do modelo de diferenciar casos positivos e negativos enquanto a PR-AUC mede a capacidade do modelo de prever corretamente os casos positivos.

Enquanto as métricas de curva avaliam o desempenho em todos os limiares, as métricas *Precision* (Precisão), o *Recall* (Sensibilidade) e *F1-Score* avaliam o modelo em um ponto de decisão específico.

No contexto de triagem para deterioração clínica, um alto *Recall* (Sensibilidade) é crucial para minimizar Falsos Negativos (FN), que é a não identificação de um paciente em risco, ou seja, mede a proporção de todas as instâncias positivas reais que foram corretamente identificadas pelo modelo.

A *Precision* (Precisão) é importante para evitar sobrecarga do sistema com falsos alarmes, sendo a proporção de instâncias classificadas como positivas que são de fato positivas.

O *F1-Score* é a média harmônica entre a *Precision* e *Recall*, fornece um valor único que equilibra ambas as métricas, sendo uma métrica valiosa para comparar o desempenho prático dos modelos, e especialmente útil quando existe um custo assimétrico entre falsos positivos (FP) e falsos negativos (FN) ou em presença de desbalanceamento de classes (Batista; Prati; Monard, 2004; Saito; Rehmsmeier, 2015).

O *Brier Score* é conhecido na literatura por ser eficaz em avaliar a qualidade de previsões probabilísticas de eventos binários, para (Siegert, 2014) *Brier Score* é uma medida bem estabelecida para essa finalidade e se caracteriza por diferentes atributos de previsão, como confiabilidade, resolução e incerteza. Portanto, o *Brier Score* mede a acurácia e a calibração das probabilidades de um modelo. O resultado que se espera é um valor entre 0 e 1, sendo 0 previsões corretas e 1 indica previsões erradas e essa acurácia das probabilidades previstas é essencial para a confiança na predição (Steyerberg et al., 2010).

Modelos como LR e de *Gradient Boosting* geram "probabilidades brutas" e nem sempre refletem a verdadeira probabilidade de um evento ocorrer. Um modelo pode, por exemplo, prever um risco de 80% para um conjunto de pacientes, quando, na realidade, apenas 60% desse grupo específico desenvolverá a condição. Isso acontece porque os modelos são treinados para otimizar métricas de classificação e essas probabilidades nem sempre correspondem à probabilidade real do evento (Guo et al., 2017).

Para garantir que a predição corresponda a um risco real, e nesse contexto a predição é essencial para a decisão clínica, é importante que os modelos finais sejam calibrados usando calibração Isotônica e de Platt (S*igmoidal*) e para isso, utilizou o

menor *Brier Score* no conjunto de validação (Platt, 1999; Steyerberg, 2019; Zadrozny; Elkan, 2002). Dessa forma, para encontrar um limiar de probabilidade que otimizasse o equilíbrio entre sensibilidade e especificidade, de forma a maximizar o *Recall* e manter a Taxa de Falsos Positivos (FPR) em um nível clínico.

4 RESULTADOS E DISCUSSÃO

4.1 TESTES DE USABILIDADE COM PROFISSIONAIS DA SAÚDE

Usabilidade é definida como um atributo de qualidade que avalia a facilidade de uso de uma interface. Essa avaliação inclui componentes como capacidade de aprendizagem, eficiência, eficácia, memorização, erros e satisfação (Nielsen, 2012).

Na presente pesquisa, o foco foi a avaliação da usabilidade da interface desenvolvida (e apresentada na seção 3.3) utilizando quatro desses componentes fundamentais: a eficiência, a eficácia, os erros e a satisfação dos usuários ao interagir com a interface proposta. Esses aspectos permitiram uma análise objetiva e subjetiva da experiência dos usuários, e forneceram subsídios importantes para identificar os pontos fortes e as oportunidades de melhoria.

Participaram da avaliação 8 (oito) enfermeiras com diferentes níveis de experiência profissional e familiaridade com tecnologias digitais aplicadas à saúde. A etapa de testes de usabilidade foi conduzida com o objetivo de avaliar a experiência desses usuários ao interagir com a interface do software proposto. Os testes foram realizados em ambiente controlado, seguindo um protocolo estruturado que incluiu 4 (quatro) tarefas específicas a serem executadas (de forma simulada) no sistema (testes de desempenho), conforme descrito em material e métodos. Durante a execução, foram observados 1) o tempo para realização de cada tarefa (eficiência); 2) se a tarefa foi executada em sua totalidade (eficácia); e, ainda, 3) o número de toques totais (com e sem toques desnecessários, para avaliar os erros cometidos) para cumprir cada uma das tarefas. Todas as tarefas também foram realizadas por um especialista na interface, para fins comparativos.

Conforme pode ser visto na Tabela 2, para todas as tarefas testadas, os tempos de execução foram, para a maioria dos casos, menores que um minuto. Portanto, qualquer diferença em segundos pode representar grandes mudanças nesta avaliação de usabilidade. De fato, os dados demonstram variações consideráveis nos tempos de execução de cada tarefa entre os participantes e também em comparação com um tempo de referência (medido com a execução dessas tarefas pelo usuário especialista). Isso foi especialmente visto para as Tarefas 1 e 3. Para a Tarefa 1, essa variação pode ser explicada pelo fato de a realização dela ser atrelada ao primeiro contato do usuário com a interface. Contudo, vale ressaltar que, mesmo sendo o primeiro contato do usuário com a interface, a Tarefa 1 foi realizada com sucesso em menos de meio minuto por 5 dos 8 participantes avaliados (com uma média de 37.88

segundos, sendo o valor de referência 20 segundos). Já a Tarefa 3, considerada a mais complexa, teve média de tempo de 56.25 segundos, que foi muito próxima ao valor de referência (55 segundos), destacando a adequação da interface para operações mais detalhadas. Vale salientar, ainda, que a Tarefa 2, mais simples, apresentou tempos também próximos (média 11.25 segundos) ao valor de referência (8 segundos); enquanto a Tarefa 4 (última tarefa), embora de complexidade intermediária, também apresentou valores próximos ao de referência (média 29.38 versus 24 do valor de referência), indicando um rápido aprendizado da interface pelos profissionais de saúde ao longo das 4 tarefas.

Tabela 2 – Tempo (em segundos) para a realização das tarefas pelos participantes. T1: Tarefa 1; T2: Tarefa 2; T3: Tarefa 3; T4: Tarefa 4.

Participante	**T1 Tempo (s)**	**T2 Tempo (s)**	**T3 Tempo (s)**	**T4 Tempo (s)**	**%C***
Enfermeira 1	73	12	85	20	100%
Enfermeira 2	20	18	57	30	100%
Enfermeira 3	86	31	71	36	100%
Enfermeira 4	29	4	67	33	100%
Enfermeira 5	35	6	62	40	100%
Enfermeira 6	10	6	48	14	100%
Enfermeira 7	20	5	26	32	100%
Enfermeira 8	30	8	34	30	100%
Média	37,88	11,25	56,25	29,38	
Desvio padrão	27,03	9,21	19,54	8,47	
Valor de Referência**	20	8	55	24	100%

Nota: *%C: porcentagem de tarefas concluídas com sucesso.
**Valor de referência: valores da realização das tarefas por um usuário especialista.
Fonte: O autor.

Apesar da variabilidade individual no tempo de realização das tarefas, a totalidade de conclusão de todas as tarefas (%C = 100% para as 4 tarefas, Tabela 2) evidencia a eficácia das operações, o que demonstra que os elementos visuais e operacionais do sistema são compreensíveis mesmo para usuários com diferentes perfis e que nunca tiveram contato com a interface.

Ao encontro dos resultados quantitativos aqui apresentados, (Wunsch; Da Costa; Righi, 2017) também avaliaram o tempo para a realização de tarefas pré-

estabelecidas, bem como a conclusão de tais tarefas, em testes com um aplicativo chamado UbiTriagem. No trabalho dos autores, o aplicativo foi desenvolvido no intuito de dar suporte ao processo de triagem em departamentos de emergência, e os testes de usabilidade foram executados por enfermeiros. De maneira similar aos achados da presente pesquisa, no trabalho dos autores todos os enfermeiros conseguiram realizar todas as tarefas propostas (100% de eficácia), porém, com uma média de tempo para a realização das tarefas maior, de maneira geral, do que o tempo de referência (mensurado por meio da execução da tarefa por um especialista no software).

O número de toques desnecessários para a realização de cada uma das tarefas, bem como o número de toques totais (que incluem os necessários e os desnecessários), também foram avaliados (Tabela 3). Neste caso, a Tarefa 2 se destacou positivamente, sendo executada com exatidão por todos os participantes, sem toques desnecessários, o que indica um fluxo claro e eficiente. A Tarefa 3 também apresentou ótimo desempenho, com apenas um usuário excedendo o número de toques de referência, sugerindo que sua estrutura está bem compreendida.

Por outro lado, a Tarefa 1 apresentou maior variação entre os usuários, com um desvio de até 250% em relação ao valor ideal, o que pode indicar confusão no início do uso ou pouca clareza no caminho da tarefa. Já a Tarefa 4 foi corretamente executada pela maioria, mas teve um participante com o dobro de toques necessários, possivelmente por dificuldade na navegação ou repetição de ações. De maneira geral, apesar dos desvios pontuais, a interface se mostrou funcional e compreensível.

Tabela 3 – Número de toques para a realização das tarefas pelos participantes. T1: Tarefa 1; T2: Tarefa 2; T3: Tarefa 3; T4: Tarefa 4; TD: Número de toques desnecessários; TT: Número de toques totais (necessários e desnecessários).

Participante	**TD-T1**	**TT-T1**	**TD-T2**	**TT-T2**	**TD-T3**	**TT-T3**	**TD-T4**	**TT-T4**
Enfermeira 1	5	7	0	1	5	5	0	4
Enfermeira 2	0	2	0	1	0	5	0	4
Enfermeira 3	1	3	0	1	1	5	0	4
Enfermeira 4	0	2	0	1	0	5	0	4
Enfermeira 5	0	2	0	1	0	5	4	8
Enfermeira 6	0	2	0	1	0	5	0	4
Enfermeira 7	1	3	0	1	1	6	1	5
Enfermeira 8	1	3	0	1	0	5	0	4

Valor de Referência**	-	2	-	1	-	5	-	4

Nota: *Valor de referência: valores da realização das tarefas por um usuário especialista.

Fonte: O autor.

A avaliação qualitativa da usabilidade baseou-se na aplicação de um questionário (Apêndice A), após a realização dos testes de desempenho. Por meio do questionário, foi possível calcular o *Net Promoter Score* (NPS) e a *System Usability Scale* (SUS), conforme detalhado em material e métodos. Além disso, os participantes também responderam a questões auxiliares e deixaram comentários e sugestões de melhoria.

Os resultados obtidos pela da aplicação do questionário revelam uma percepção bastante positiva dos participantes em relação ao sistema proposto. A partir da questão base para cálculo do *Net Promoter Score* (NPS), foi possível calcular o valor do NPS (Figura 24, A), que foi de 75%, valor que se enquadra na faixa de excelente (Reichheld, 2003). Além disso, o score *System Usability Scale* (SUS) obtido foi de 79,7 (Figura 24, B). Assim como o valor obtido de NPS, o valor de SUS encontrado indica uma usabilidade excelente (Bangor; Kortum; Miller, 2009). Ainda, a média da resposta à pergunta complementar — “Considerando a sua experiência no teste realizado hoje, qual a probabilidade de você adotar o sistema no seu dia a dia?” — foi de 8,625, com uma mediana de 9, em uma escala de 0 a 10, onde 0 não é “de jeito nenhum”, 10 é extremamente provável. Esses números indicam uma alta intenção de adoção do sistema pelos profissionais, sugerindo que a interface atende, em grande parte, às expectativas de usabilidade e aplicabilidade na prática clínica.

Figura 24 – Valores calculados do *Net Promoter Score* (em A) e do *System Usability Scale* (em B)

Fonte: O autor.

Similar aos resultados aqui encontrados, Rocha et al. (2021) — avaliando a usabilidade de um sistema de monitoramento das internações em um pronto-socorro brasileiro, com 17 especialistas — obtiveram um score SUS de 86,5, o que também se enquadra em excelente. Em outro estudo, Padrini-Andrade et al. (2019) avaliaram a usabilidade de um sistema de informação em saúde neonatal por profissionais da Rede Brasileira de Pesquisas Neonatais. Os pesquisadores encontraram um score SUS de 73,3, considerado como bom. A escala SUS, embora não tenha sido originalmente desenvolvida para sistemas de saúde, é a mais frequentemente utilizada para avaliar a satisfação dos usuários com sistemas de apoio à decisão clínica e sistemas de informação em saúde (Ghorayeb et al., 2023). Em relação aos valores de SUS, uma pontuação SUS > 68 é considerada acima da média na medida geral de satisfação com o sistema e nas subescalas de usabilidade e capacidade de aprendizagem (Yi et al., 2023). Já o NPS serve como uma métrica para medir a

fidelidade e a satisfação dos usuários, avaliando a disposição dos usuários em recomendar o serviço ou produto a terceiros. Altos valores de NPS, como o encontrado pela presente pesquisa, indicam que a experiência e a satisfação dos usuários atingiram valores bastante satisfatórios. (Reichheld, 2003).

As respostas à questão “você tem sugestões?” também corroboram essa visão positiva da interface proposta aqui. Muitos participantes destacaram que o sistema é intuitivo, de fácil aprendizagem e bem integrado, aspectos essenciais para a adoção em contextos em que o tempo é um fator crítico. Além disso, houve reconhecimento do impacto positivo que o sistema pode ter na eficiência da triagem, no acompanhamento do paciente e na redução de erros, o que contribui para um cuidado mais seguro e resolutivo. Por outro lado, as sugestões oferecidas pelos participantes também trazem insights importantes. Um dos pontos levantados foi o excesso de informações na tela inicial, o que pode causar uma certa sobrecarga cognitiva, dificultando a rápida identificação de sinais de alarme — especialmente durante os primeiros contatos com a interface. Isso evidencia a necessidade de aperfeiçoamentos no design visual, como o destaque de alertas e elementos críticos, favorecendo uma navegação mais eficiente em situações de alta pressão. Por fim, outro aspecto mencionado foi a aplicabilidade do sistema em diferentes níveis de atenção. Um participante observou que a ferramenta se justifica principalmente em cenários de emergência, sendo relativamente dispensável na Atenção Primária à Saúde. Essa distinção é relevante para guiar futuras decisões sobre a adaptação ou o foco do sistema conforme o contexto de uso.

Em síntese, os dados de desempenho e satisfação apontam para uma interface bem aceita, com alto potencial de uso e aprovação entre os profissionais de saúde que realizaram os testes. No entanto, a pesquisa de satisfação indica também indicam interface pode ser aprimorada em termos de interface visual e delimitação mais clara do seu escopo de atuação. Esses ajustes podem tornar a ferramenta ainda mais eficaz e aumentar a perspectiva de sua futura adoção em conjunto com um sistema aplicado.

Recentemente, Çetin & Cebeci (2024) conduziram uma pesquisa que avaliou as experiências de 14 enfermeiros com um sistema computacional de suporte à decisão de triagem baseado em um departamento de emergência da Turquia. No trabalho em questão, os enfermeiros avaliaram o software de triagem positivamente e enfatizaram que ele facilitou o processo de tomada de decisão na triagem. Contudo,

uma das recomendações do trabalho foi conduzir uma pesquisa de usabilidade do sistema, o que foi primeiramente realizado na presente pesquisa. Por outro lado, considerando os resultados positivos obtidos na avaliação de usabilidade da interface proposta no presente estudo — que abordou aspectos como eficiência, eficácia, ocorrência de toques desnecessários e satisfação dos usuários —, trabalhos futuros poderiam se concentrar no desenvolvimento de um software funcional com base nessa interface. A implementação prática do sistema permitiria verificar, em contexto real de uso, se o funcionamento do sistema contribui efetivamente para apoiar os enfermeiros no processo de tomada de decisão durante a triagem, assim como foi realizado no trabalho de Çetin & Cebeci (2024). Além disso, essa etapa seria essencial para avaliar a integração da ferramenta na rotina de trabalho e sua aceitação pelos profissionais, fornecendo dados complementares àqueles obtidos na presente fase experimental.

4.2 SISTEMA NEWS MODIFICADO

Conforme detalhado nos procedimentos metodológicos (seção 3.4), esta pesquisa também se propôs a desenvolver uma versão modificada do score de alerta NEWS2, que foi então chamada de NEWS2 modificado (ou NEWS2-M), considerando que a proposta central desta pesquisa é o monitoramento contínuo e precoce da deterioração do paciente a partir de dados automatizados provenientes de uma pulseira multiparamétrica. Essa nova versão do NEWS2 (NEWS2-M), apresentada na Figura 25, mantém os parâmetros vitais básicos do NEWS2, mas adiciona a avaliação da Pressão Arterial Diastólica (PAD) e nível de dor. Além disso, refina as faixas de pontuação da Pressão Arterial Sistólica (PAS) para uma avaliação mais detalhada da hipertensão e especifica a "Confusão" como um indicador de alteração do nível de consciência.

Figura 25 – Parâmetros do NEWS2 Modificado

NEWS2 MODIFICADO

	3	2	1	0	1	2	3
Frequência Respiratória	≤ 8		9 - 11	11 - 20		21 - 24	≥ 25
SpO2 Escala 1 (%)[1]	≤ 91	92 - 93	94 - 95				
Pressão Arterial Sistólica (mmHg)	≤ 90	91 - 100	101 - 110	111 - 139[2]		140 - 179[2]	≥ 180[2]
Pulso	≤ 40		41 - 50	51 - 90	91 - 110	111 - 130	≥ 131
Temperatura	≤ 35.0		35.1 - 36	36.1 - 38	38.1 - 39	≥ 39.1	
Pressão Arterial Diastólica (mmHg)	≤ 60[4] (> 65 anos)		0	80		≥ 90[2] (sbp < 140)	
Nível de dor							≥ 7[3]
Nível de consciência				Alerta			Confuso

1. SpO2 Escala 2 (%) foi desconsiderada nesse estudo por falta de informação na base de dados.
2. European Society of Hypertension (ESH).
3. Severe Pain (ESI).
4. Estudos.

Fonte: O autor.

O NEWS2 original possui uma tabela de pontuação para classificar o risco clínico. Para que ela pudesse ser utilizada no NEWS2 modificado, devido ao

acréscimo de parâmetros, foram feitos primeiramente ajustes que categorizassem o ESI a essas respostas clínicas e posteriormente a reclassificação do risco clínico conforme a pontuação obtida no NEWS2.

A estratificação de risco proposta neste estudo adapta o escore NEWS2 para criar quatro níveis de ação clínica (Tabela 4). Primeiramente, a pontuação 0 foi isolada para classificar pacientes fisiologicamente estáveis, mas com outras queixas que requerem tratamento, sendo seu risco de piora considerado mínimo. A faixa de escore de 1 a 4 foi definida como **Risco Médio**, indicando a necessidade de vigilância clínica contínua. Um patamar superior, o **Risco Alto**, foi atribuído a escores de 5-6 ou a um único parâmetro com valor 3; este nível demanda uma resposta de avaliação, embora não configure, a princípio, um risco de vida iminente. Por fim, o **Risco Urgente** foi estabelecido para qualquer escore agregado igual ou superior a 7, representando uma emergência médica que exige intervenção da equipe devido ao provável risco de vida iminente.

Tabela 4 – Descrição das respostas clínica dos desencadeadores do NEW2-Modificado e sua relação com a classificação Emergency Severity Index (ESI).

Pontuação NEWS2-M	RISCO CLÍNICO	RESPOSTA	ESI
Pontuação de 0	Baixo	12 horas de monitorização	5 e 4
Pontuação AGREGADA de 1 – 4	Médio	4h de monitorização de rotina	3
Pontuação AGREGADA de 5 – 6 ou 3 individual	Alto	Limite chave para resposta urgente	2
Pontuação AGREGADA de 7 ou mais	Alto	Resposta de urgência ou emergência	1

Fonte: O autor.

Após a criação da ferramenta modificada, foi feita a verificação com base nos dados dos MIMIC para testar sua eficácia e propensão a ser de fato mais completa e sensível, no intuito de capturar um espectro maior de sinais de alerta para orientar intervenções clínicas com maior rapidez e eficácia, conforme será detalhado abaixo.

4.2.1 Correlação NEWS2-M + NEWS2 Original

A intuito primordial do desenvolvimento do NEWS2-M nesta pesquisa, em relação ao NEWS2 original, é aumentar a sensibilidade da classificação na detecção de casos mais graves. Para verificar se houve diferença entre uma classificação

baseada no NEWS2 e uma nova classificação baseada no NEWS2-M, realizaram-se testes de classificação de pacientes reais baseado em dados oriundos do MIMIC (*Medical Information Mart for Intensive Care*), conforme detalhado nos procedimentos metodológicos (seção 2.3.2). Esse banco de dados inclui dados demográficos, sinais vitais, resultados laboratoriais, medicamentos e notas clínicas de milhares de pacientes estratificados pelo sistema ESI.

Para investigar a associação entre o escore de alerta precoce modificado NEWS2-M, proposto neste estudo, e o NEWS-2 original, foi conduzida uma análise retrospectiva de uma coorte de 390.039 pacientes a partir dos dados extraídos do banco de dados MIMIC-ED (*Medical Information Mart for Intensive Care - Emergency Department)*. A relação entre as classificações categóricas geradas por ambos os escores foi avaliada por meio de uma análise de tabela de contingência, conforme descrito nos procedimentos metodológicos. Este método foi selecionado para determinar a frequência de distribuição dos pacientes em cada combinação de categorias dos dois sistemas de triagem, permitindo a identificação de padrões de associação entre eles.

A partir da Tabela 5 é possível verificar que, enquanto o NEWS2 original classifica massivamente os casos desse banco de dados em “Baixo”, a nova modificação proposta (NEWS2-M) consegue categorizar melhor os casos apresentando uma melhoria fundamental na segurança do paciente. Dos 350.607 pacientes classificados como "Baixo" pelo NEWS2, o NEWS2-M reclassificou a grande maioria para categorias de maior risco, sendo:

- 173.684 (44.5% do total de pacientes) foram elevados para "Médio";
- 77.295 (19.8% do total) foram elevados para "Alto";
- 42.506 (10.9% do total) foram elevados diretamente para "Urgente".

A análise da performance do escore NEWS2 revela uma limitação intrínseca em sua capacidade de estratificação, evidenciada pela aglutinação de 89,9% da população de pacientes em uma única categoria, a de risco "Baixo". Essa baixa granularidade resulta em uma capacidade discriminatória inadequada, que poderia comprometer a identificação de indivíduos com sinais sutis de deterioração fisiológica e, consequentemente, mascarar a deterioração clínica em sua fase incipiente.

Em contrapartida, o modelo modificado, aqui denominado NEWS2-M, exibe uma sensibilidade marcadamente superior. A implementação deste modelo permitiu a

reestratificação de um contingente substancial de mais de 293 mil pacientes (correspondendo a 75,2% da coorte originalmente classificada como de baixo risco), elevando seu status para um nível de alerta que demanda vigilância clínica ativa (Tabela 5).

Tabela 5 – Tabela de contingência: correlação NEWS2-M e NEWS2 Original, e risco clínico do NEWS2-M. N: Quantidade de pessoas (números absolutos); %: porcentagem.

	NEWS2-M							
NEWS2	**Baixo**		**Médio**		**Alto**		**Urgente**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Baixo	57.122	14,65%	173.684	44,53%	77.295	19,82%	42.506	10,90%
Médio	0	0%	9.160	2,35%	9.562	2,45%	8.929	2,29%
Alto	0	0%	0	0%	1.096	0,28%	8.364	2,14%
Urgente	0	0%	0	0%	0	0%	2.321	0,60%
Total	**350.607**	**14,65%**	**27.651**	**46,88%**	**9.460**	**22,55%**	**2.321**	**15,93%**

Fonte: O autor.

A principal implicação da maior acurácia apresentada pelo NEWS2-M reside na antecipação da janela para a intervenção terapêutica. Ao identificar com maior eficácia os pacientes que necessitam de atenção imediata, o NEWS2-M se estabelece como uma ferramenta com potencial significativo para otimizar os desfechos clínicos, e poderia contribuir para a redução da morbimortalidade associada ao reconhecimento tardio da instabilidade fisiológica. Vale ressaltar, ainda, que uma propriedade fundamental do escore NEWS2-M, conforme elucidado pela análise de dados, é sua relação de não-inferioridade em relação ao NEWS2 original. Isto é, para qualquer dado paciente, o NEWS2-M invariavelmente preserva ou escalona a categoria de risco atribuída pelo escore padrão, sem nunca diminuir o nível de risco. Essa característica de design assegura que o NEWS2-M funcione como um mecanismo de vigilância adicional ao mitigar o risco de uma complacência clínica induzida por uma reclassificação de risco inferior, o que pode garantir que a vigilância para pacientes já identificados como vulneráveis pela métrica estabelecida seja, no mínimo, mantida.

A literatura traz o NEWS2 como excelente na sensibilidade e especificidade para a predição de mortalidade precoce em pacientes em ambientes pré-hospitalares e em departamentos de emergência (Milne-Ives et al., 2024). No entanto, esse escore apresentou baixo desempenho na predição de mortalidade intra-hospitalar e em 30

dias (Wei et al., 2023). Além disso, um recente estudo encontrou que, embora o NEWS2 tenha apresentado desempenho moderado a excelente na previsão de desfechos clínicos, ele não foi eficaz na previsão de desfechos críticos em subgrupos de pacientes com sepse e com COVID-19, limitando sua utilidade como ferramenta de alerta precoce em diversos cenários clínicos (Mhd Yunin; Tan, 2025). Ainda, o trabalho de Goodacre et al., 2025 avaliou que, de prontuários médicos onde 45 pacientes necessitaram de cirurgia de emergência, antibióticos para fraturas expostas, infusão de insulina ou manipulação de lesões com risco de morte; grande parte deles (37 pacientes) foram classificados com um NEWS2 ≤ 4, o que sugere que o paciente não está gravemente doente e que é improvável que seu quadro se deteriore rapidamente.

Em outro estudo recente, feito a partir de 12.867 referências, foi encontrado que 11 versões modificadas do NEWS2 demonstraram maior precisão preditiva para pelo menos um desfecho, em comparação com o NEWS2 tradicional. Modificações que incorporaram variáveis demográficas, dados de tendência e ajustes na ponderação dos componentes da pontuação foram consideradas particularmente úteis para aumentar a precisão preditiva do NEWS2. O estudo encontrou que três modificações importantes no NEWS2 — incorporando idade, tratamento diferenciado dos dados de FiO_2 e análise de tendências — tiveram o potencial de melhorar a precisão preditiva sem aumentar a sobrecarga do clínico (Riccalton et al., 2025). Esses dados corroboram com os resultados apresentados pela presente pesquisa, onde incorporações como a pressão arterial diastólica, nível de dor, “confusão” como indicador de alteração em nível de consciência, além do refinamento das faixas de pontuação da pressão arterial sistólica, resultaram em um modelo muito mais sensível e proativo na detecção da deterioração clínica.

4.2.2 Correlação do NEWS2-M com o ESI

Uma vez desenvolvido o escore NEWS2-M, tornou-se necessário avaliar sua correspondência com métodos de triagem já consolidados na prática clínica. Optou-se por utilizar o Índice de Severidade de Emergência (ESI) como referência para investigar a associação entre o escore de alerta precoce modificado NEWS2-M, proposto neste estudo, e um sistema de triagem já estabelecido. Além disso, é importante ressaltar que os dados utilizados para testar a nova classificação foram originalmente classificados pelo ESI, já que é o sistema de escore utilizado no

departamento de emergência do Beth Israel Deaconess Medical Center (de onde os dados foram originados).

Para investigar a associação entre o escore de alerta precoce modificado NEWS2-M, proposto neste estudo, e o Índice de Severidade de Emergência (ESI), foi conduzida uma análise retrospectiva de uma coorte de 390.039 pacientes a partir dos dados extraídos do banco de dados MIMIC-IV-ED, assim como realizado na comparação entre o NEWS2-M e o NEWS2 original (na seção anterior).

Conforme pode ser visualizado na Tabela 6, a estratificação dos 390.039 pacientes admitidos no departamento de emergência, com base no escore NEWS2-M, revelou a seguinte distribuição: a categoria de risco “Médio” foi a mais prevalente, abrangendo ~47% da amostra. Seguiram-se as categorias de risco “Alto” em ~23%, “Urgente” em ~16% e “Baixo" em ~15%.

A tabela 6 também demonstra a relação entre a nova classificação dos pacientes (distribuída conforme o novo escore NEWS2-M) e a classificação convencional destes mesmos pacientes pelo ESI. Pode-se notar que o maior volume de pacientes está nas células que cruzam ESI 3 com NEWS2-M “Médio”, totalizando 105.410 pessoas (27% do total). Isso significa que uma parcela expressiva dos pacientes classificados como moderadamente urgentes pelo ESI (nível 3) foi identificada pelo NEWS2-M como apresentando risco clínico médio, o que sugere uma convergência entre os dois sistemas quanto à priorização desses casos intermediários. Pode-se notar também que o segundo maior volume de pacientes está no cruzamento do ESI 2 com NEWS2-M “Médio” com 58.846 pessoas (15.1% do total) e ESI 3 com NEWS2-M “Alto” com 51.638 pacientes, totalizando 13.2% do total de pacientes do coorte. Isso indica que há uma forte concordância entre os sistemas em identificar um grande grupo de pacientes de risco moderado a alto. Além disso, o escore NEWS2-M classificou 62.120 pacientes como "Urgente", e a maior parte desses pacientes urgentes vieram do ESI 3 (27.732) e ESI 2 (25.969). Isso indica um alinhamento entre os dois sistemas, visto que dos 13.967 pacientes classificados como ESI 1 (os mais graves pela triagem ESI), o NEWS2-M classificou 6.511 (cerca de 47%) como "Urgente". Isso demonstra um bom alinhamento nessa parcela dos pacientes, pois o escore fisiológico confirma a alta gravidade percebida na triagem inicial para quase metade desses pacientes.

Tabela 6 – Tabela de contingência: correlação NEWS2-M e ESI, e risco clínico do NEWS2-M. N: Quantidade de pessoas (números absolutos); %: porcentagem.

ESI	NEWS2-M							
	Baixo		Médio		Alto		Urgente	
	N	%	N	%	N	%	N	%
0	24	0,01%	68	0,02%	36	0,01%	28	0,01%
1	967	0,25%	3.872	0,99%	2.617	0,67%	6.511	1,67%
2	17.741	4,55%	58.846	15,09%	27.876	7,15%	25.969	6,66%
3	32.113	8,23%	105.410	27,03%	51.638	13,24%	27.732	7,11%
4	5.942	1,52%	14.132	3,62%	5.648	1,45%	1.837	0,47%
5	335	0,09%	516	0,13%	138	0,04%	43	0,01%
Total	**57.122**	**14,65%**	**182.844**	**46,88%**	**87.953**	**22,55%**	**62.120**	**15,93%**

Fonte: O autor.

Por outro lado, um ponto de discordância entre os escores é que 1.837 pacientes foram classificados como ESI 4 (pouco urgente), foram identificados como “Urgente” no NEWS2-M (Tabela 6). Isso pode ser um sinal de que a queixa principal pode parecer simples (levando a um ESI 4) mas os sinais vitais já se apresentam gravemente alterados, indicando um risco de deterioração rápida que poderia passar despercebido no ESI. De fato, Grossmann et al. (2012) relataram que uma subtriagem ocorreu em 117 casos de um total de 519 pacientes com 65 anos ou mais, sendo os principais motivos situações de alto risco e falha na interpretação adequada dos sinais vitais. Em contrapartida, pode-se notar também o que poderia indicar uma potencial subtriagem pelo NEWS2-M, haja vista que há 17.741 pacientes (4.5% do total) no ESI 2 (que, pela definição do ESI, necessitam de recursos e atenção rápida) que foram classificados como NEWS2-M Baixo. Embora essa se mostre a maior divergência entre os dois sistemas de triagem ao se observar a queixa principal (que não é levada em consideração no NEWS2-M, apenas os sinais vitais), ao analisar os dados a fundo, observou-se que desse número, apenas 2.453 pacientes relataram dor no peito (dados não mostrados), e a maioria deles a dores relacionadas a trauma, não necessariamente cardíaca. Isso indica o que a essência de classificação do ESI que é trabalhar com a contagem de procedimentos (raio-x, ressonância etc.) classifica melhor pacientes que precisam desse tipo de intervenção. Clinicamente, isso poderia indicar situações em que o estado fisiológico do paciente está estável no momento, mas poderia se deteriorar, no entanto observando os dados clínicos é possível observar que não é esse o caso na maioria desses pacientes. Nesses casos, pesquisas futuras poderiam testar uma classificadora (um peso) para pacientes que

apresentarem dor no peito, colocando uma pontuação no NEWS2-M para esses casos.

Alguns trabalhos vêm demonstrando que, em urgências e emergências que usam o escore ESI, mais da metade das visitas são triadas para o nível 3, o que gera muitos pacientes indiferenciados e limita a estratificação adequada da gravidade clínica nesse nível (Dugas et al., 2016; Hinson et al., 2019; Levin et al., 2018). Nesse sentido, o modelo de triagem proposto na presente pesquisa (NEWS2-M) demonstrou uma redistribuição dos pacientes triados como nível 3 pelo ESI: embora a maioria deles tenha sido mantida com nível médio de urgência (NEWS-2-M médio, 105.410 pacientes); 32.113 foram realocados para o nível baixo, 51.638 para o nível alto e 27732 para o nível urgente. Dessa forma, nota-se que a adoção de sistemas complementares ao ESI, como o NEWS2-M, pode contribuir para uma estratificação mais precisa do risco clínico dos pacientes classificados como ESI 3. Similar a estes resultados, Hilson et al. (2018) desenvolveram um modelo de triagem eletrônica (utilizando inteligência artificial) que realocou mais de 10 % (14.326) dos pacientes originalmente classificados como ESI 3, identificando-os como de maior risco.

Sax et al. (2023) avaliaram a taxa de erros da triagem pelo ESI, em um amplo estudo de coorte com 5.315.176 atendimentos de adultos em prontos-socorros de 21 hospitais que utilizam o sistema de saúde Kaiser Permanente Northern California (KPNC). Os dados daquele estudo foram coletados entre janeiro de 2016 e dezembro de 2020, e os resultados apontaram que a triagem incorreta ocorreu em um terço dos atendimentos, sendo 3,3% de subtriagem (casos graves não identificados) e 28,9% de supertriagem (priorização desnecessária). Além disso, a sensibilidade do ESI para identificar um paciente grave foi de apenas 66% (Sax et al., 2023). Além deste estudo, o trabalho de Malinovska et al. (2019) demonstrou que, ao adicionar ao ESI um parâmetro de “primeira impressão do médico” (neste caso, cada médico designado apontou um valor de 0 a 10 para a pergunta: “quão doente este paciente parece?”), houve um aumento significativo no poder preditivo da mortalidade de pacientes idosos em 30 dias. Esses achados reforçam que, assim como proposto no presente estudo, a incorporação de escores fisiológicos complementares à triagem convencional pode aprimorar a acurácia na identificação precoce de pacientes em risco (Malinovska et al., 2019).

Em conjunto, os resultados aqui apresentados demonstram, portanto, o NEWS2-M desenvolvido como uma ferramenta em potencial para auxiliar na

estratificação de risco, sobretudo por é utilizar os sinais vitais como parâmetros de classificação e reclassificação de risco. Como visto, o NEWS-2-M se mostra robusto e bem alinhado aos resultados de urgência obtidos a um dos maiores classificadores de risco, o ESI: o NEWS2-M identificou com sucesso um grande grupo de pacientes "Urgentes" (16%), mostrando boa sensibilidade para os casos mais graves classificados pelo ESI e apresenta boa concordância com o ESI tanto nos extremos de baixo risco quanto nos de alto risco. Por outro lado, as divergências entre os scores, como os casos de ESI “Baixo” e NEWS2-M “Urgente”, poderiam funcionar como um importante sinal de segurança, onde alertariam a equipe para pacientes com risco fisiológico elevado que poderiam ser subestimados pela triagem inicial utilizando o ESI. Outra vantagem da utilização do NEWS2-M é a de que, se acoplado à pulseira multiparamétrica e um software correspondente, poderia possibilitar a reclassificação dos pacientes de uma forma automatizada, podendo tê-la em tempo real sem precisar de um profissional da saúde para ficar fazendo a medição de forma manual para identificar possíveis mudanças no estado clínico do paciente.

4.3 MACHINE LEARNING

Os dados utilizados para treinar o modelo são provenientes do MIMIC (*Medical Information Mart for Intensive Care*) que é um amplo banco de dados de livre acesso que reúne dados de saúde anonimizados de pacientes tratados no hospital *Beth Israel Deaconess Medical Center* (BIDMC) de Boston, Massachusetts, EUA, conforme já comentado anteriormente. Este banco é disponibilizado por meio do trabalho de pesquisadores do Laboratório de Fisiologia Computacional do Instituto de Tecnologia de Massachusetts (MIT). A versão que se utiliza para esta pesquisa, é o MIMIC-IV de 2021.

Este capítulo apresenta os resultados da avaliação comparativa de 16 modelos de algoritmos (incluindo *ensembles*) de aprendizagem de máquina com a intenção de responder a uma das questões de pesquisa que é a predição de piora dos pacientes que aguardam dentro das UEs. O objetivo desta análise quantitativa é identificar o modelo com a *performance* preditiva mais robusta, considerando métricas de avaliação que medem não apenas a capacidade de discriminação, mas também a calibração das previsões e o balanço entre diferentes tipos de erros de classificação. Adicionalmente, foi medido o tempo de treino, para avaliar o custo computacional e a viabilidade para potencial implementação prática dos modelos.

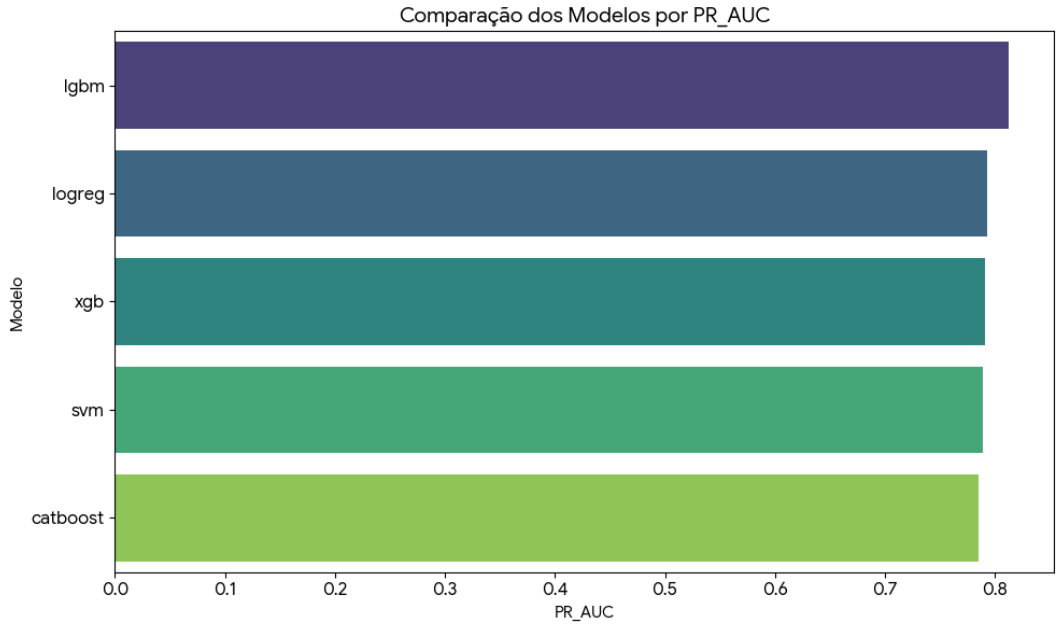
Os modelos foram avaliados com base nas seguintes métricas: *Area Under the Receiver Operating Characteristic Curve* (AUROC), *Area Under the Precision-Recall Curve* (PR-AUC), *F1-Score*, *Recall*, *Precision* e *Brier Score,* conforme descrito nos procedimentos metodológicos.

4.3.1 Análise Comparativa de Desempenho dos Modelos

A presente análise detalha os resultados de um *leaderboard* de modelos de *Machine Learning*, com o objetivo de extrair dados sobre o desempenho e a eficiência de diferentes algoritmos em uma tarefa de classificação binária. Foram treinados 16 modelos, mas aqui na discussão dos resultados, serão apresentados o top 5 melhores colocados. O restante dos resultados estará nos anexos dessa tese.

A análise do *leaderboard* revela uma clara hierarquia de desempenho entre as famílias de modelos avaliadas. Os modelos foram ordenados pela métrica *Precision-Recall Area Under the Curve* (PR-AUC) por ser mais informativa em cenários com desequilíbrio de classe, muito comum em conjuntos de dados biomédicos (Saito; Rehmsmeier, 2015)

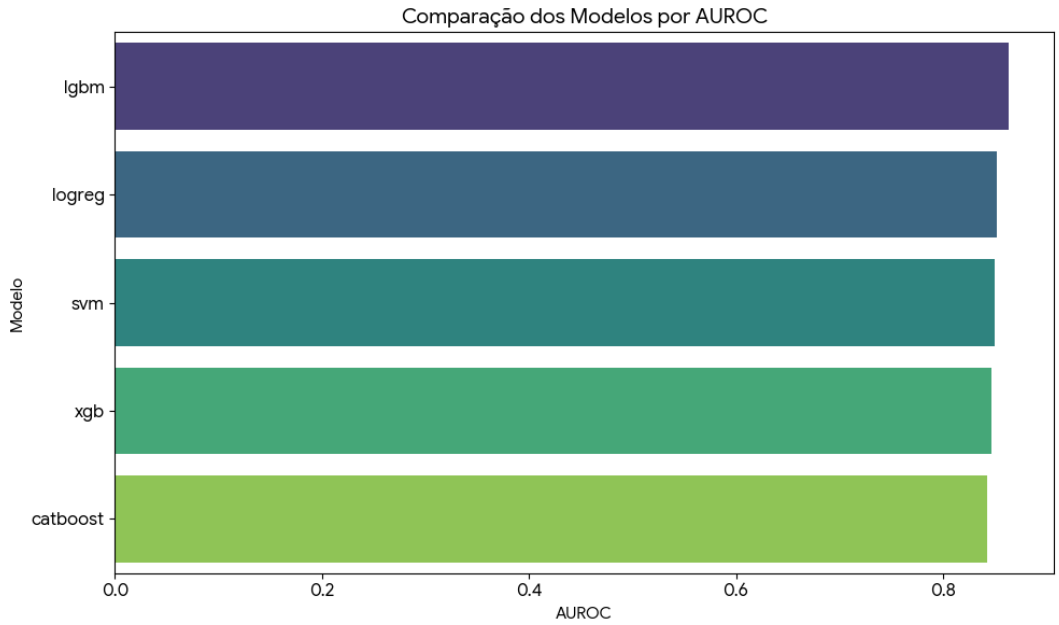
Figura 26 – Comparação de Modelos por PR-AUC



Fonte: O autor.

O modelo que melhor performou foi o LGBM alcançando um PR-AUC de 0.813, AUROC, que é a capacidade de um modelo em distinguir corretamente entre as classes (ex: doente x saudável), foi de 0.864 e um *F1-Score* de 0.729. Os modelos baseados em *Gradient Boosting*, como o *XGBoost* e o *CatBoost*, também figuram entre os de melhor desempenho, embora com performance ligeiramente inferior ao LGBM (Figura 26). A eficácia de algoritmos baseados em *Gradient Boosting* em problemas de classificação, é extensivamente documentada para a área biomédica (Dissanyake; Johar, 2023). A sua capacidade de modelar interações complexas e não-lineares nos dados, mantendo a eficiência computacional, posiciona-os como uma escolha robusta, o que é consistente com os achados em estudos sobre prognósticos e diagnósticos assistidos por computador.

Figura 27 – Comparação de Modelos por AUROC



Fonte: O autor.

O segundo modelo que melhor performou foi o LR com um PR-AUC de 0.793, AUROC de 0.852 e um F1-Score de 0.716 (Figuras 26 e 27). Os modelos baseados em LR apresentou um desempenho competitivo, com um PR-AUC em torno de 0.793.

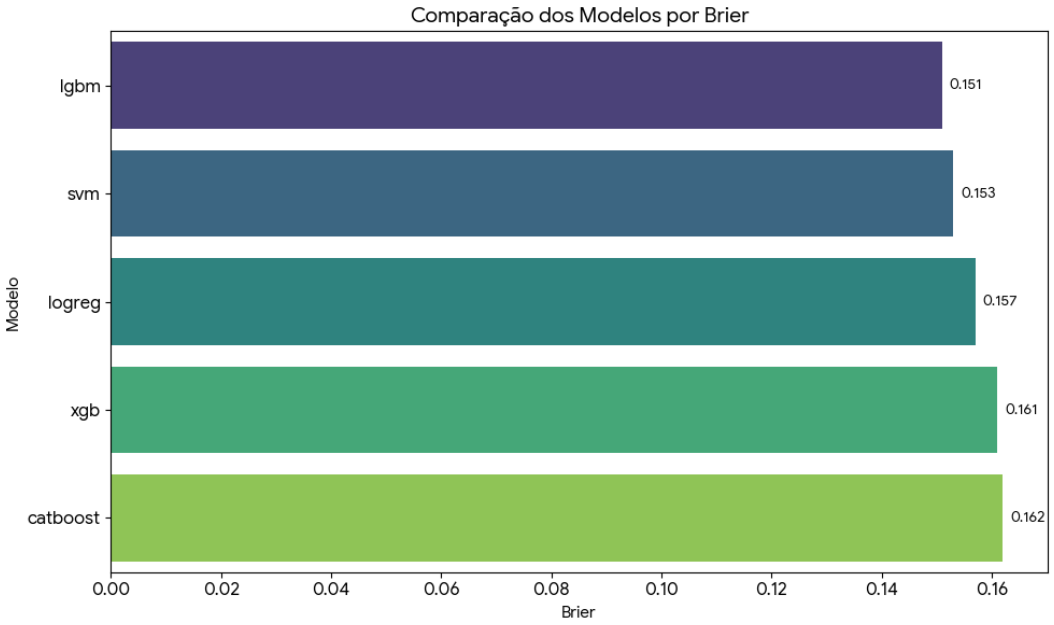
O SVM (*Support Vector Machine*) demonstrou a maior *Precision* entre os modelos de ponta (0.746), porém com um *Recall* consideravelmente menor (0.644), o que indica uma tendência mais conservadora na classificação da classe positiva.

4.3.2 Análise de Métricas de Calibração e Erro

O *Brier Score*, que mede a acurácia das predições probabilísticas, fornece probabilidades que podem ser confiavelmente interpretadas como o risco real de um paciente. Para os modelos testados, ele se mostrou consistentemente baixo entre os melhores modelos. Conforme a Figura 28, que ilustra a comparação entre os modelos, o LGBM (0.151) e o SVM (0.153) foram os modelos mais bem calibrados. Foram seguidos do LR com o valor de 0.157 indicando

A boa calibração é crucial para a tomada de decisão clínica, onde um médico pode usar a probabilidade de risco para decidir sobre a necessidade de exames adicionais ou intervenções terapêuticas.

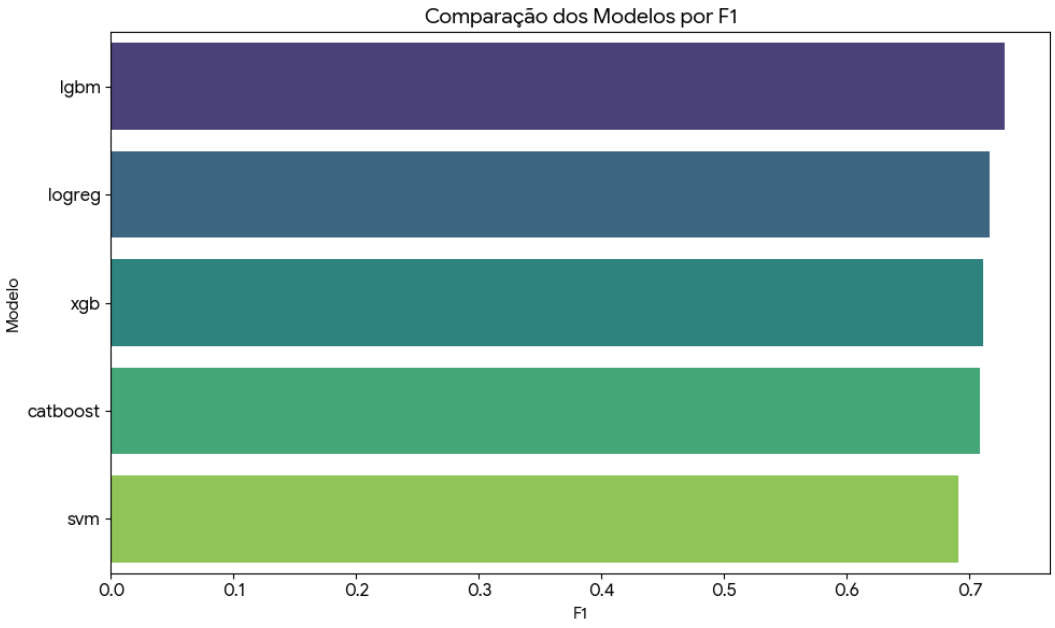
Figura 28 – Comparação de Modelos por *Brier Score*



Fonte: O autor.

O *F1-Score*, que é a média harmônica entre *Precision* e *Recall*, é maximizado nos modelos SVM com o valor de 0.691, o CatBoost com o valor de 0.708 seguido pelo XGB 0.71. O LGBM, embora com maior valor que os primeiros, performou com valores próximos (0.729) que pode ser visto na Figura 29, todos esses modelos reforçando a superioridade em equilibrar a identificação de verdadeiros positivos e a minimização de falsos positivos.

Figura 29 – Comparação de Modelos por *F1-Score*.

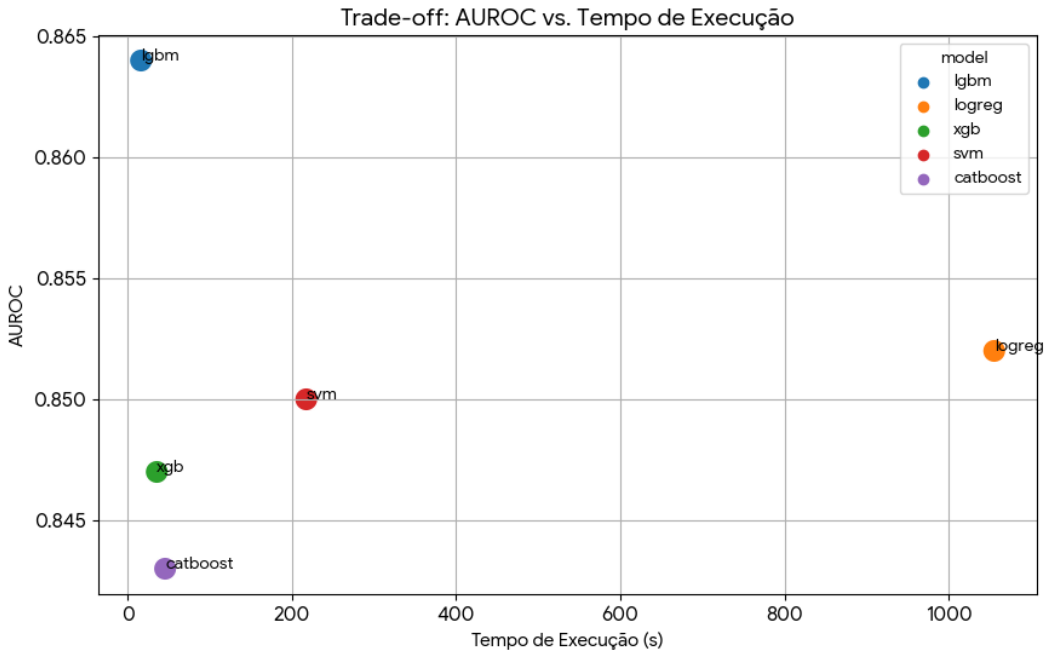


Fonte: O autor.

4.3.3 Trade-off entre Performance e Custo Computacional

Uma dimensão importante quando se trata de Machine Learning é o trade-off entre a acurácia preditiva e a eficiência computacional, ou seja, melhor modelo que performa entre resultado e tempo. A viabilidade de um modelo em ambientes clínicos ou de produção não depende apenas da sua acurácia, mas também da sua eficiência, visto que a importância é de que o modelo consiga trazer resultados quase que em tempo real do estado clínico do paciente.

Figura 30 – Relação entre Performance (AUROC) x Tempo de Execução



Fonte: O autor.

Na figura de dispersão (Figura 30) que ilustra a relação entre o AUROC e o tempo de execução (em segundos) é possível observar a competitividade dos modelos de *Gradient Boosting*. O LGBM (16.4s) ao lado do *XGBoost* (35.6s) e do *CatBoost* (51.5s) se comportam como um trio de modelos de alta *performance* e grande eficiência computacional. O LGBM ainda se mantém em vantagem, por oferecer melhor desempenho também no PR-AUC (vide Figura 26) e com o menor tempo de execução entre os três. No entanto, o *XGBoost*, performa com PR-AUC de 0.791 (vide Figura 26) e se torna uma alternativa altamente competitiva e muito mais

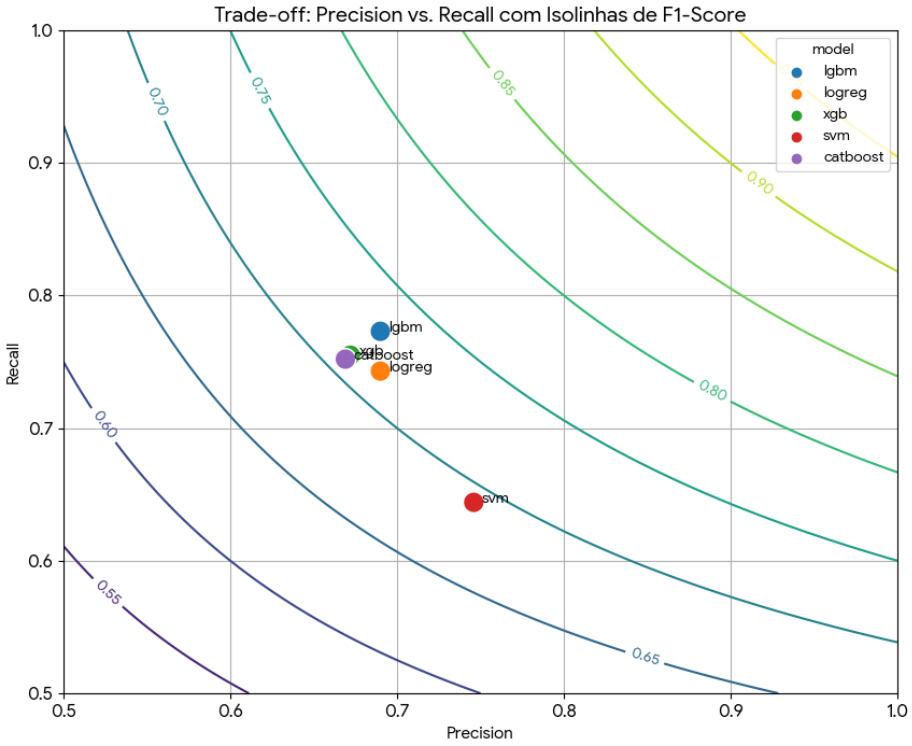
rápida que os modelos de LR e SVM, notavelmente mais lentos. Logo, o LGBM é o que possui melhor equilíbrio entre performance preditiva e velocidade de treinamento/inferência. Logo, além da acurácia preditiva, fatores como eficiência computacional, facilidade de implementação e interpretabilidade dos modelos são essenciais para a adoção bem-sucedida de algoritmos de aprendizado de máquina na prática clínica (Gama et al., 2014).

Apesar de contraintuitivo, pois os modelos lineares são esperados como sendo computacionalmente mais eficientes que modelos baseados em árvores, o tempo elevado pode ser devido a um grande volume de dados, que de fato a base de dados do MIMIC apresenta. Embora, em relação ao *trade-off,* o modelo LR não se saia tão bem (1055.3s) ainda possui vantagem sobre o modelo *CatBoost* em relação ao PR-AUC, AUROC e *F1-Score* não podendo ser descartado como uma possível solução também. Com base na análise dos resultados do *leaderboard*, conclui-se que o modelo LGBM representa a solução mais promissora para o problema em questão, pois apresenta o melhor equilíbrio entre alta performance e eficiência computacional seguido pelos demais modelos de G*radient Boosting,* SVM e LR.

4.3.4 Análise entre Precision e Recall

Para muitas aplicações na área da saúde, a natureza dos erros de classificação não é simétrica. Por exemplo, num cenário de diagnóstico de câncer, um falso negativo (não detectar a doença quando ela existe: baixo *Recall*) é frequentemente mais grave do que um falso positivo (diagnosticar a doença quando ela não existe: baixo *Precision*). A Figura 31 mapeia os modelos no espaço entre *Precision-Recall*.

Figura 31 – *Trade-off: Precision x Recall* com isolinhas de *F1-Score*



Fonte: O autor.

O *trade-off*, juntamente com as isolinhas de *F1-Score*, permite uma análise mais granular. O LGBM novamente demonstra um excelente equilíbrio, alcançando o maior *F1-Score* (0.729) por possuir tanto um alto *Recall* (0.773) quanto uma *Precision* competitiva (0.690). É interessante notar o comportamento do SVM, que se destaca pela maior *Precision* (0.746), mas ao custo do menor *Recall* (0.644). Esta característica poderia torná-lo útil num cenário onde a confirmação de um positivo precisa ser muito confiável, minimizando falsos alarmes. Por outro lado, o XGB e o *CatBoost* mostraram um perfil mais orientado ao *Recall* (0.756 e 0.752, respetivamente), o que pode ser preferível em fases de triagem, onde o objetivo principal é não perder nenhum caso potencialmente positivo. A escolha entre estes modelos dependeria, portanto, do objetivo clínico específico, um dilema bem explorado na literatura sobre modelos de prognóstico (Steyerberg et al., 2010).

4.3.5 Análise Inferida da Matriz de Confusão e Implicações dos Tipos de Erro

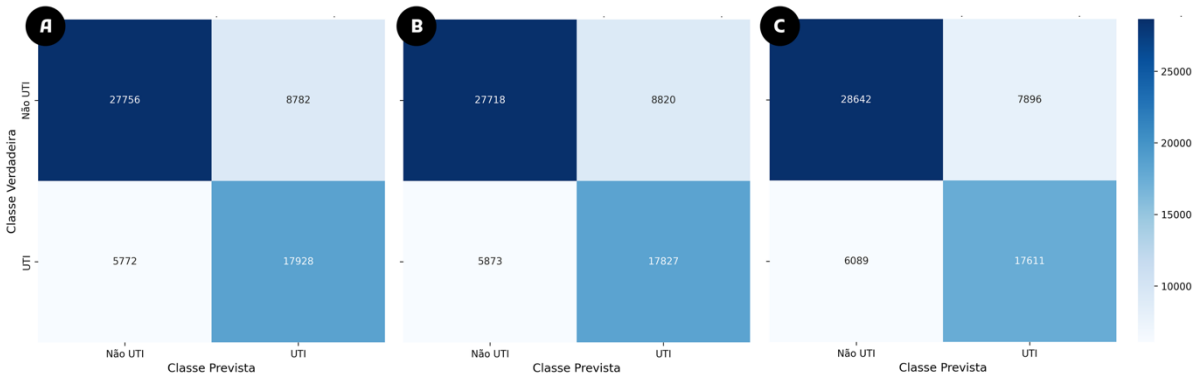
Para uma avaliação mais completa dos modelos de classificação, principalmente os que permeiam a área da saúde, é preciso fazer uma análise que vá

além das métricas de performance agregadas. É importante compreender a natureza dos erros cometidos pelos modelos, e para isso utiliza-se da matriz de confusão, que é uma ferramenta diagnóstica primária que decompõe as previsões em quatro categorias: Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), Falsos Positivos (FP) e Falsos Negativos (FN). Essa previsão foi feita a partir da inferência de comportamento dos modelos a partir das métricas Precision e Recall já calculadas.

O Recall é definido como *TP/(TP+FN)* – Sensibilidade – e quantifica a capacidade do modelo de identificar todos os casos positivos reais (baixo número de FN é igual a alto *Recall*). No contexto de um diagnóstico médico, um FN representa um paciente doente incorretamente classificado como saudável, um erro frequentemente considerado o mais crítico. A *Precision* é definida como *TP/(TP+FP)* – Valor Preditivo Positivo – e quantifica a proporção de acertos nas previsões positivas (baixo número de FP resulta em alta *Precision*). Um FP, nesse contexto, representa um paciente saudável que foi incorretamente classificado como doente.

Os modelos que apresentaram baixa Taxa de Falsos Negativos (FNR - Alto Recall) foram os XGB que apresentou *Recall* de 0.756, o *CatBoost* de 0.752 e o LR de 0.743 e demonstraram a maior capacidade de detectar os casos positivos indicando que são melhores em minimizar FN. Este resultado, do ponto de vista clínico, são modelos mais adequados para tarefas de triagem em que o objetivo é minimizar a probabilidade de não se detectar um indivíduo com a condição de interesse, garantindo que o maior número possível de casos positivos fosse encaminhado para investigação (Figura 33).

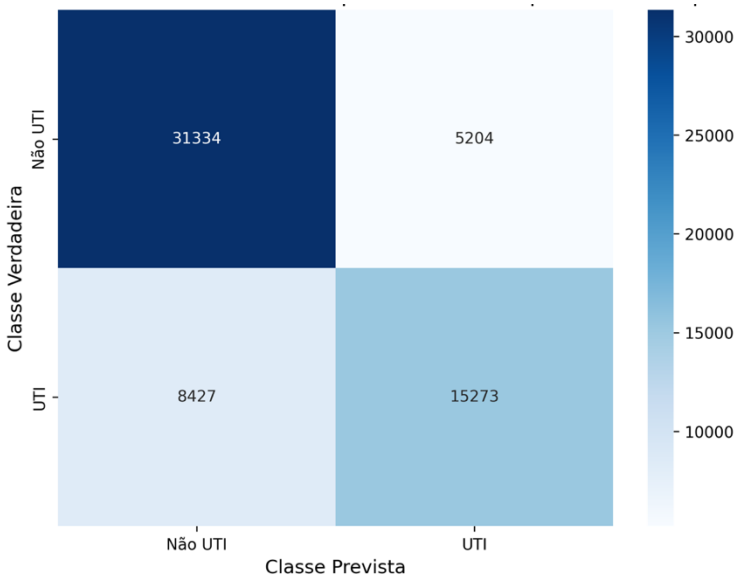
Figura 32 – Matriz de Confusão: a) XGB, b) CatBoost e c) LR



Fonte: O autor.

O modelo que apresentou baixa Taxa de Falsos Positivos (TFP - alta *Precision*) foi o SVM e se destacou pela maior *Precision* de 0.746. Logo, sua matriz de confusão apresenta uma contagem de FP relativamente baixa e indica que a proporção de alertas positivos corretos gerados por este modelo é superior à dos demais tornando-o mais confiável (Figura 34). Nesse contexto, o SVM um forte candidato para uma função de confirmação diagnóstica ou como uma "segunda opinião" algorítmica, sendo o modelo mais conservador. O que resulta em um modelo de alta precisão conferindo maior confiança ao clínico para proceder com intervenções mais invasivas ou dispendiosas.

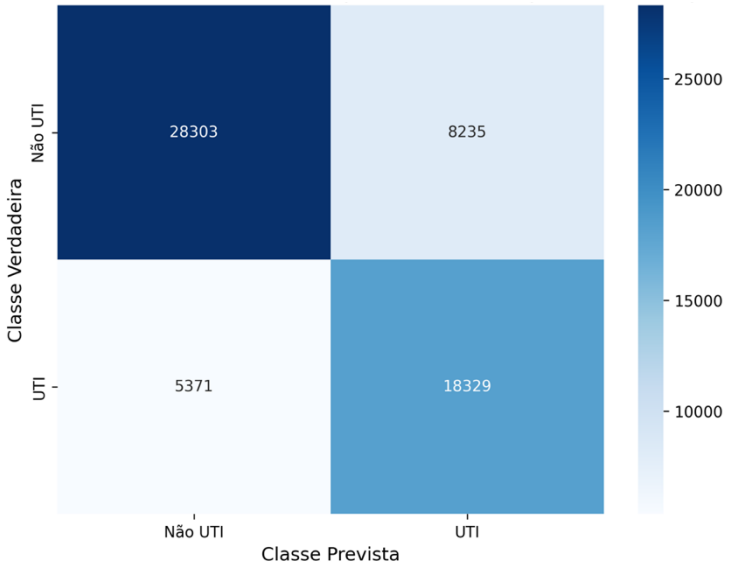
Figura 33 – Matriz de Confusão: SVM



Fonte: O autor.

O modelo que se apresentou mais equilibrado foi o LGBM, apresentando o *Recall* maior que os demais (0.773), ou seja, indicando que é o melhor em minimizar FN. Logo, sua matriz de confusão se apresenta como a mais "equilibrada" do conjunto, conseguindo um elevado número de TP (alto *Recall*) sem inflacionar excessivamente o número de FP (boa *Precision*) e sua métrica *F1-Score* representa o melhor compromisso entre a minimização dos dois tipos de erro. O que resulta em um modelo ótimo para um sistema de apoio à decisão (Figura 35).

Figura 34 – Matriz de Confusão: LGBM



Fonte: O autor.

4.3.6 Seleção de Decision Thresholds para modelos preditivos

Após os resultados obtidos no teste de *Machine Learning*, foram selecionados os três algoritmos que melhor performaram para fazer a seleção de *thresholds* clínicos. Isso foi necessário, pois, um modelo de classificação, tal como os desenvolvidos neste estudo, não produz, em sua essência, uma decisão binária (como "internar" x "não internar"). Em vez disso, sua saída fundamental é uma probabilidade, um valor contínuo no intervalo [0,1] que representa a confiança do modelo de que uma determinada instância pertence à classe positiva.

Para converter essa probabilidade em uma decisão binária é necessário a aplicação de um *Decision Threshold* (Limiar de Decisão). Por convenção, esse *threshold* é definido como 0.5, onde se assume que os custos associados a erros d FP e FN são iguais, no entanto em contexto médico essa premissa pode levar a conclusões perigosas. Um exemplo prático é que o custo de um FN, que é liberar um paciente que necessita de cuidados, é imensamente superior ao custo de um FP, que é admitir um paciente para observação que não precisaria. Portanto, selecionar criteriosamente o *Decision Threshold* é um passo importante para se obter um desempenho estatístico de um modelo que possa ser aplicado em uma clínica real.

Nesse contexto, então, buscou-se encontrar um *threshold* que satisfizesse um conjunto de restrições operacionais e clínicas pré-definidas. Os modelos que foram

utilizados foram: o LGBM, o XGB e o LR na tarefa de predição de admissão em Unidade de Terapia Intensiva (UTI).

O protocolo utilizado para a geração de candidatos foi a função AUROC complementada pela PR-AUC utilizada para calcular a Taxa de Verdadeiros Positivos (TPR, ou Sensibilidade/*Recall*) e a Taxa de Falsos Positivos (FPR). A definição do critério clínico foi que o *Recall* fosse de no mínimo 85% (*TPRgeq*0.85 - restrição de segurança), o objetivo é garantir que o modelo identifique corretamente pelo menos 85% de todos os pacientes que necessitam de internação, priorizando a minimização dos FN. Já a FPR deve ser de no máximo 25% (*FPRleq*0.25 - restrição de eficiência e recursos), o objetivo é garantir que ao atingir a sensibilidade desejada o sistema não sobrecarregue a UTI com um número excessivo de alarmes falsos.

O código foi desenhado para assegurar robustez e reprodutibilidade, empregando um treinamento exaustivo com 4000 estimadores, buscando, entre todos os *thresholds* possíveis, aquele que melhor satisfaz ambas a condições. Se múltiplos *thresholds* atenderem aos critérios, é selecionado o primeiro encontrado (que geralmente corresponde ao *threshold* mais alto que ainda cumpre a meta de sensibilidade, ajudando a controlar o FPR). Caso nenhum *threshold* consiga satisfazer as duas restrições, uma política de *fallback* é acionada, este cálculo encontra o ponto na curva ROC que maximiza a distância vertical da linha de chance, conhecido como Índice J de Youden ($J = Sensibilidade + Especificidade - 1$) ou seja, matematicamente equivalente a maximizar a diferença TPR – FPR ($D = argmax\ (TPR - FPR)$). É uma escolha para encontrar um equilíbrio geral entre sensibilidade e especificidade quando os critérios estritos não podem ser atendidos.

4.3.7 Análise de Performance dos Modelos

Os resultados de performance no conjunto de teste demonstraram uma clara superioridade dos modelos baseados em *Gradient Boosting* em relação ao modelo de Regressão Logística. O LGBM obteve o melhor desempenho geral, com o AUROC de 0.864 e o PR-AUC de 0.816, seguido pelo XGB com AUROC de 0.859 e PR-AUC de 0.810 e por fim o LR com AUROC de 0.848 e PR-AUC de 0.790.

Estes achados são consistentes com a literatura, que frequentemente reporta a dominância de algoritmos de *Gradient Boosting* em dados tabulares complexos, como os encontrados em prontuários eletrônicos. Uma AUROC de 0.864 é considerada de excelente a notável poder discriminatório em contextos de predição

clínica (Hosmer; Lemeshow; Sturdivant, 2013). O desempenho marginalmente superior do LGBM sobre o XGBoost, tanto em AUROC quanto em PR-AUC, alinha-se a estudos comparativos que, embora dependentes do *dataset*, por vezes apontam uma ligeira vantagem para o LGBM devido à sua estratégia de crescimento de árvores *leaf-wise* (Ke et al., 2017).

Em relação ao custo computacional, o LR levou 48 segundos, o LGBM levou 2 minutos e 43 segundos e o XGB levou 5 minutos e 45. A eficiência da LR é esperada, dada sua menor complexidade algorítmica. O resultado mais relevante, contudo, é a comparação entre os modelos de *boosting*. O LightGBM demonstrou ser aproximadamente ≈ 52% mais rápido que o XGBoost neste hardware específico (GPU NVIDIA) e com estes hiperparâmetros. Esta vantagem de eficiência é um dos principais argumentos para a adoção do LGBM e é atribuída ao seu algoritmo baseado em histogramas e à amostragem exclusiva de gradientes (*Gradient-based One-Side Sampling*), que otimizam o processo de busca pelos melhores pontos de divisão (Ke et al., 2017).

Em relação ao *threshold*, os modelos apresentaram os seguintes resultados: LGBM com *Recall* de 0.788 (78.8%), FPR de 0.239 (23.9%) e *threshold* de 0.369, o XGB com *Recall* de 0.803 (80.3%), FPR de 0.263 (26.3%) e *threshold* de 0.341 e o LR com *Recall* de 0.781 (78.1%), FPR de 0.255 (25.5%) e *threshold* de 0.358 (Tabela 7).

Tabela 7 – Resultado do *Decision Threshold*

Modelo	***Threshold***	***Recall***	**FPR**
LGBM	0.369	0.788 – 78.8%	0.239 – 23.9%
LR	0.358	0.781 – 78.1%	0.255 – 25.5%
XGB	0.341	0.803 – 80.3%	0.263 – 26.3%

Fonte: O autor.

Nenhum dos modelos conseguiu atingir o alvo de 85% de *Recall* sob a restrição de 25% de FPR. O modelo LGBM foi o único que respeitou estritamente a restrição de FPR (0.239), mas ao custo de um *Recall* de 78.8%. Os outros dois modelos excederam ligeiramente o FPR, indicando que o algoritmo de seleção

recorreu ao seu critério de *fallback* que também não foi suficiente para atingir o *Recall* desejado.

Dessa forma, o modelo LGBM consegue identificar quase 8 em cada 10 pacientes que necessitarão de cuidados intensivos, enquanto mantém a taxa de encaminhamentos desnecessários (alarmes falsos) abaixo de um quarto dos pacientes saudáveis sendo do ponto de vista da gestão de recursos, o *trade-off* mais favorável, alinhando alta sensibilidade com o menor FPR entre os competidores. Nos outros casos, o modelo XGB somente seria justificado se a prioridade máxima fosse a minimização absoluta do FN e o LR embora obtenha resultados aceitáveis, não supera o balanço oferecido pelo LGBM e a sensibilidade do XGB.

5 CONCLUSÃO

A partir dos resultados encontrados no presente estudo, pode-se concluir que:

1. A interface desenvolvida demonstrou elevada eficácia, com 100% de conclusão das tarefas e tempos médios próximos aos valores de referência, indicando boa adequação ao fluxo de trabalho dos profissionais de saúde;

2. A percepção dos participantes dos testes diante da interface proposta foi amplamente positiva, refletida pelos altos valores de NPS (75%) e SUS (79,7), além dos relatos qualitativos de facilidade de uso, intuitividade e potencial de impacto positivo no processo de triagem e no cuidado seguro ao paciente;

3. Apesar dos resultados favoráveis em relação à interface, ajustes visuais e de escopo são necessários, sobretudo para reduzir a sobrecarga cognitiva em telas iniciais;

4. A nova classificação de risco desenvolvida (NEWS2-M) demonstrou capacidade superior ao NEWS2 original na detecção precoce de pacientes em deterioração clínica, reclassificando mais de 75% dos casos inicialmente considerados de baixo risco e ampliando a granularidade da avaliação, o que pode resultar em um aumento da segurança do paciente;

5. A comparação com o ESI mostrou que o NEWS2-M mantém alinhamento nos extremos de risco (casos graves e estáveis), ao mesmo tempo em que redistribui pacientes classificados como intermediários, o que tem o potencial de reduzir as limitações de subtriagem do ESI e fortalecer a precisão da estratificação clínica.

6. O NEWS2-M, por utilizar apenas sinais vitais, apresenta viabilidade para integração com dispositivos automatizados, como a pulseira multiparamétrica, o que poderia tornar possível o monitoramento contínuo e a reclassificação em tempo real dos pacientes;

7. A análise dos resultados do *Machine Learning* demonstra a superioridade do modelo LightGBM frente aos outros modelos do comparativo. Este modelo não só alcançou a maior capacidade de discriminação (PR-AUC e AUROC) e o melhor balanço entre *Precision* e *Recall* (*F1-Score*), como também provou ser o modelo mais bem calibrado (menor *Brier Score*) e, de longe, o mais eficiente em termos computacionais;

8. Em relação ao *thresrold*, o LightGBM demonstrou um desempenho ótimo representando um ponto de equilíbrio, alinhado com as necessidades de um sistema

de saúde, enquanto garante a segurança dos pacientes e evita a sobrecarga de suas unidades de terapia intensiva. Sua sensibilidade próxima a 80% com a menor taxa de falsos positivos entre os modelos de alta performance valida sua superioridade não apenas em métricas de ranking (como AUROC), mas também em termos de utilidade clínica aplicada;

9. Apesar da superioridade do LGBM, a análise revelou que modelos como o SVM podem ser considerados se a prioridade for maximizar a *Precision*, enquanto a a LR, conhecida pela sua interpretabilidade, continua a ser uma *baseline* forte e relevante, embora com desafios de escalabilidade no contexto deste *dataset*.

Embora os resultados se mostrem promissores, é fundamental reconhecer as limitações em relação a este estudo. O primeiro ponto a ser observado é a validação do escore NEWS2-M e dos modelos de *Machine Learning*, pois foram realizados com base nos dados americana, o MIMIC-IV. Como falado anteriormente, essa base utiliza o protocolo ESI que impõe restrições à generalização direta dos achados para o contexto brasileiro. O segundo ponto é a avaliação da interface e a performance dos algoritmos que ocorreram em um ambiente controlado e offline. Portanto, para se verificar a eficácia e usabilidade da solução em um cenário clínico real, com as pressões e a variabilidade do atendimento de emergência, ainda precisam ser comprovadas.

As conclusões deste estudo estabelecem uma base sólida para futuras investigações, essenciais para utilização na prática clínica. Como perspectiva futura, seria relevante avaliar tanto a interface e a nova classificação de risco NEWS2-M desenvolvidas, quanto o algoritmo que melhor performou, o LightGBM, em cenários reais de atendimento, alimentando o sistema com dados contínuos provenientes da pulseira multiparamétrica um estudo piloto para integrar e validar o sistema completo em um ambiente real de atendimento. A aplicação em pacientes reais permitiria verificar não apenas a acurácia na detecção precoce de deterioração clínica, mas também a usabilidade e a efetividade da interface em condições práticas de triagem. Essa etapa representaria um avanço fundamental para validar o potencial do sistema em apoiar decisões clínicas em tempo real e aumentar a segurança do cuidado em ambientes de urgência e emergência. Sugere-se também a condução de novos ensaios de usabilidade com profissionais brasileiros para avaliar a ferramenta em seu fluxo de trabalho real. Futuramente, um estudo clínico prospectivo e randomizado

poderia comparar o desempenho da e-Triagem proposta com o protocolo padrão, a fim de mensurar seu impacto em desfechos clínicos, como tempo para tratamento e taxas de mortalidade.

REFERÊNCIAS

AGULNIK, Asya *et al.* Improved outcomes after successful implementation of a pediatric early warning system (PEWS) in a resource-limited pediatric oncology hospital. **Cancer**, v. 123, n. 15, p. 2965–2974, 25 ago. 2017.

AGULNIK, Asya *et al.* Pediatric Early Warning Systems aid in triage to intermediate versus intensive care for pediatric oncology patients in resource-limited hospitals. **Pediatric Blood & Cancer**, v. 65, n. 8, 10 ago. 2018.

ANAMARIA DE MORAES; CLAUDIA MONT'ALVÃO. **Ergonomia. Conceitos e Aplicações**. Rio de Janeiro: 2AB, 2010.

ANDY DE. **4 Ways Data is Improving Healthcare Operations**.

ARTASANCHEZ, Alberto.; JOSHI, Prateek. **Artificial intelligence with Python : your complete guide to building intelligent apps using Python 3.x and TensorFlow 2**. *[S.l.]*: Packt Publishing, 2020.

ARYA, Rajiv *et al.* Decreasing Length of Stay in the Emergency Department With a Split Emergency Severity Index 3 Patient Flow Model. **Academic Emergency Medicine**, v. 20, n. 11, p. 1171–1179, nov. 2013.

AUSSERHOFER, Dietmar *et al.* Errors in nurse-led triage: An observational study. **International Journal of Nursing Studies**, v. 113, p. 103788, jan. 2021.

AZEEZ, Dhifaf *et al.* Comparison of adaptive neuro-fuzzy inference system and artificial neutral networks model to categorize patients in the emergency department. **SpringerPlus**, v. 2, n. 1, p. 1–10, 2013.

BANGOR, Aaron; KORTUM, Philip; MILLER, James. **Determining What Individual SUS Scores Mean: Adding an Adjective Rating ScaleJournal of Usability Studies**. *[S.l.: S.n.]*.

BATISTA, Gustavo E. A. P. A.; PRATI, Ronaldo C.; MONARD, Maria Carolina. A study of the behavior of several methods for balancing machine learning training data. **ACM SIGKDD Explorations Newsletter**, v. 6, n. 1, p. 20–29, jun. 2004.

BILBEN, Bente; GRANDAL, Linda; SØVIK, Signe. National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient – a prospective observational study. **Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine**, v. 24, n. 1, p. 80, 2 dez. 2016.

BIŞKIN ÇETIN, Songül; CEBECI, Fatma. Nurses' experiences of using a computer-based triage decision support system in the emergency department. **Nursing in Critical Care**, v. 29, n. 5, p. 1078–1085, 5 set. 2024.

BRASIL; ABNT. **NBR 9241-11/1998**.

BRASIL; CONSELHO FEDERAL DE ENFERMAGEM - COFEN. **Resolução COFEN Nº 661 de 2021**.

BRASIL; CONSELHO FEDERAL DE MEDICINA - CFM. **Resolução CFM nº 1451/1995**.

BRASIL; CONSELHO FEDERAL DE MEDICINA - CFM. **Resolução CFM nº 2.077/2014**.

BRASIL; MINISTÉRIO DA SAÚDE. **Acolhimento e Classificação de Risco nos Serviços de Urgência**. Brasília: https://bvsms.saude.gov.br/bvs/publicacoes/acolhimento_classificaao_risco_servico_urgencia.pdf, 2009.

BRASIL; MINISTÉRIO DA SAÚDE. **Acolhimento à Demanda Espontânea: queixas mais comuns na Atenção Básica**. 28. ed. Brasília: *[S.n.]*. v. 2

BRASIL; MINISTÉRIO DA SAÚDE. **Manual instrutivo da Rede de Atenção às Urgências e Emergências no Sistema Único de Saúde (SUS)**. *[S.l.: S.n.]*.

BRASIL; MINISTÉRIO DA SAÚDE. **Portaria Nº 589, de 20 de Maio de 2015**.

BRASIL; MINISTÉRIO DA SAÚDE. **Política Nacional de Informação e Informática em Saúde**. Brasília: https://bvsms.saude.gov.br/bvs/publicacoes/politica_nacional_infor_informatica_saude_2016.pdf, 2016.

BRASIL; MINISTÉRIO DA SAÚDE. **Resolução Nº 33, de 22 de Março de 2018**.

BRASIL; MINISTÉRIO DA SAÚDE. **Resolução Nº 33, de 22 de Março de 2018**.

BRASIL; MINISTÉRIO DA SAÚDE. **Portaria Nº 2.983, de 13 de Novembro de 2019**.

BRASIL; MINISTÉRIO DA SAÚDE. **Estratégia de Saúde Digital para o Brasil 2020-2028 [recurso eletrônico]** . Brasília: Ministério da Saúde, 2020.

BRASIL; MINISTÉRIO DA SAÚDE. **Prontuário eletrônico**.

BRASIL; MINISTÉRIO DA SAÚDE. **Resolução Nº 659, de 26 de Julho de 2021**.

BRASIL; MINISTÉRIO DA SAÚDE; CONASEMS. **Manual instrutivo da Rede de Atenção às Urgências e Emergências no Sistema Único de Saúde (SUS)**. *[S.l.: S.n.]*.

BREIMAN, Leo. **Random Forests**. *[S.l.: S.n.]*.

BROOKE, John. **SUS - A quick and dirty usability scale**. *[S.l.: S.n.]*.

CARDONA-MORRELL, M. *et al.* Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. **International Journal of Clinical Practice**, v. 70, n. 10, p. 806–824, out. 2016.

CHAWLA, N. V. *et al.* SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 1 jun. 2002.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost. *In*: New York, NY, USA: ACM, 13 ago. 2016.

CHRIST, Michael *et al.* Modern Triage in the Emergency Department. **Deutsches Ärzteblatt international**, 17 dez. 2010.

CHUA, Wei Ling *et al.* Automated rapid response system activation—Impact on nurses' attitudes and perceptions towards recognising and responding to clinical deterioration: Mixed-methods study. **Journal of Clinical Nursing**, v. 32, n. 17–18, p. 6322–6338, 23 set. 2023.

CONSELHO FEDERAL DE MEDICINA - CFM. **Resolução Nº 1.638, de 9 de Agosto de 2002**.

COSLOVSKY, Michael *et al.* A clinical prediction model to identify patients at high risk of death in the emergency department. **Intensive Care Medicine**, v. 41, n. 6, p. 1029–1036, 20 jun. 2015.

COSTA, Jaqueline Pereira da *et al.* The accuracy of the Manchester Triage System in an emergency service. **Revista gaucha de enfermagem**, v. 41, p. e20190327, 2020.

COSTA, Maria Antonia Ramos *et al.* Admittance of Risk-Classified Cases: Assessment of Hospital Emergency Services. **Escola Anna Nery - Revista de Enfermagem**, v. 19, n. 3, 2015.

DA SILVA JÚNIOR, Sergio Vital *et al.* Superlotação dos serviços de urgência e emergência hospitalar. **Enfermagem Brasil**, v. 19, n. 1, p. 49–57, 22 mar. 2020.

DAVIS, Jesse; GOADRICH, Mark. The relationship between precision-recall and ROC curves. **ACM International Conference Proceeding Series**, v. 148, p. 233–240, 2006.

DE MORAES BATISTA, André Filipe *et al.* **COVID-19 diagnosis prediction in emergency care patients: a machine learning approach**. , 7 abr. 2020.

DEACON, Pamala B. .. **UX and UI design strategy : a step-by-step guide on UX and UI design**. *[S.l.]*: [publisher not identified], 2020.

DELGADO, Rosario; NÚÑEZ-GONZÁLEZ, J. David. Bayesian network-based over-sampling method (BOSME) with application to indirect cost-sensitive learning. **Scientific Reports**, v. 12, n. 1, p. 8724, 24 maio 2022.

DISSANYAKE, Kaushalya; JOHAR, Md Gapar Md. Two-level boosting classifiers ensemble based on feature selection for heart disease prediction. **Indonesian Journal of Electrical Engineering and Computer Science**, v. 32, n. 1, p. 381, 1 out. 2023.

DON NORMAN; JAKOB NIELSEN. **The Definition of User Experience (UX)**.

DUGAS, Andrea Freyer *et al.* An Electronic Emergency Triage System to Improve Patient Distribution by Critical Outcomes. **The Journal of Emergency Medicine**, v. 50, n. 6, p. 910–918, jun. 2016.

EMERGENCY NURSES ASSOCIATION - ENA. **Emergency Severity Index Handbook**. 5th Edition ed. *[S.l.: S.n.]*.

FARROHKNIA, Nasim *et al.* Emergency Department Triage Scales and Their Components: A Systematic Review of the Scientific Evidence. **Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine**, v. 19, n. 1, p. 42, 2011.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 1 jun. 2006.

FERNANDES, Fernando Timoteo *et al.* A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil. **Scientific Reports**, v. 11, n. 1, p. 3343, 8 fev. 2021.

FERNANDES, Marta *et al.* Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. **PLoS ONE**, v. 15, n. 4, 2020a.

FERNANDES, Marta *et al.* **Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review**. **Artificial Intelligence in Medicine**Elsevier B.V., , 1 jan. 2020b.

FYNTANIDOU, Barbara *et al.* IoT-based smart triage of Covid-19 suspicious cases in the Emergency Department. *In*: IEEE, dez. 2020.

GAMA, João *et al.* A survey on concept drift adaptation. **ACM Computing Surveys**, v. 46, n. 4, p. 1–37, abr. 2014.

GDF; SES. **Segurança do Paciente: Prevenção da Deterioração Clínica em Pacientes Adultos em Serviço Hospitalar**. Distrito Federal: *[S.n.]*.

GERRY, Stephen *et al.* Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. **BMJ**, p. m1501, 20 maio 2020.

GHORAYEB, Abir *et al.* Design and validation of a new Healthcare Systems Usability Scale (HSUS) for clinical decision support systems: a mixed-methods approach. **BMJ Open**, v. 13, n. 1, p. e065323, 30 jan. 2023.

GOLDZWEIG, Caroline Lubick *et al.* Costs And Benefits Of Health Information Technology: New Trends From The Literature. **Health Affairs**, v. 28, n. Supplement 2, p. w282–w293, jan. 2009.

GOODACRE, Steve *et al.* Accuracy of the National Early Warning Score version 2 (NEWS2) in predicting need for time-critical treatment: retrospective observational cohort study. **Emergency Medicine Journal**, v. 42, n. 6, p. 396–402, jun. 2025.

GUICHARD, Jason L. *et al.* Isolated Diastolic Hypotension and Incident Heart Failure in Older Adults. **Hypertension**, v. 58, n. 5, p. 895–901, nov. 2011.

GUO, Chuan *et al.* **On Calibration of Modern Neural Networks**. *[S.l.: S.n.]*.

HERMIDA, Patrícia Madalena Vieira *et al.* User embracement with risk classification in an Emergency Care Unit: An evaluative study. **Revista da Escola de Enfermagem**, v. 52, 2018.

HINSON, Jeremiah S. *et al.* Accuracy of emergency department triage using the Emergency Severity Index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis. **International Journal of Emergency Medicine**, v. 11, n. 1, p. 3, 15 dez. 2018.

HINSON, Jeremiah S. *et al.* Triage Performance in Emergency Medicine: A Systematic Review. **Annals of Emergency Medicine**, v. 74, n. 1, p. 140–152, jul. 2019.

HITCHCOCK, Maree *et al.* Triage: an investigation of the process and potential vulnerabilities. **Journal of Advanced Nursing**, v. 70, n. 7, p. 1532–1541, 23 jul. 2014.

HODGSON, Luke E. *et al.* NEWS 2 – too little evidence to implement? **Clinical Medicine**, v. 18, n. 5, p. 371–373, out. 2018.

HOLLAND, Mark; KELLETT, John. The United Kingdom's National Early Warning Score: should everyone use it? A narrative review. **Internal and Emergency Medicine**, v. 18, n. 2, p. 573–583, 5 mar. 2023.

HOSMER, David W.; LEMESHOW, Stanley; STURDIVANT, Rodney X. **Applied Logistic Regression**. *[S.l.]*: Wiley, 2013.

HUANG, Mingyan *et al.* Isolated Diastolic Hypertension and Risk of Cardiovascular Events: A Systematic Review and Meta-Analysis of Cohort Studies With 489,814 Participants. **Frontiers in Cardiovascular Medicine**, v. 8, 5 jan. 2022.

HUETE-GARCIA, Alejandro; RODRIGUEZ-LOPEZ, Sara. Development and implementation of an artificial intelligence–enhanced care model to improve patient safety in hospital wards in Spain. **Acute and Critical Care**, v. 39, n. 4, p. 488–498, 30 nov. 2024.

INOUE, Kelly Cristina *et al.* Avaliação da qualidade da Classificação de Risco nos Serviços de Emergência. **Acta Paulista de Enfermagem**, v. 28, n. 5, p. 420–425, ago. 2015.

JAIN, Shobhit; MARGETIS, Konstantinos; IVERSON, Lindsay M. **Glasgow Coma Scale**. *[S.l.: S.n.]*.

JAKOB NIELSEN. **10 Usability Heuristics for User Interface Design**.

JAKOB NIELSEN. **Usability 101: Introduction to Usability**.

JAKOB NIELSEN. **Usability 101: Introduction to Usability**.

JAN DUL; BERNARD WEERDMEESTER. **Ergonomia na Prática**. 3ª ed. São Paulo: Blucher, 2012.

JAY ANDERS. **Here's what physicians really need – and it's not more data**.

JEFF GOTHELF; JOSH SEIDEN. **Lean UX: Creating Great Products with Agile Teams**. Sebastopol: O'Reilly Media, 2021.

JOHN C. PLATT. **Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods**. Redmond: *[S.n.]*.

JOHNSON, A. *et al.* MIMIC-IV (version 1.0). **PhysioNet**, 2021.

JOHNSON, Alistair E. W.; MARK, Roger G. **Real-time mortality prediction in the Intensive Care Unit**. *[S.l.: S.n.]*.

JONATHAN NISHIDA. **O design como ferramenta de desenvolvimento no processo interativo para aferição e transmissão de dados dos sinais vitais em ambiente hospitalar**. Florianópolis: Universidade Federal de Santa Catarina, 2023.

JONES, Daryl *et al.* Defining clinical deterioration. **Resuscitation**, v. 84, n. 8, p. 1029–1034, ago. 2013.

JUDITH HURWITZ; DANIEL KIRSCH. **Machine learning for dummies**. *[S.l.]*: John Wiley & Sons, Inc, 2018. v. 75

KANEKO, Hidehiro *et al.* Association of Isolated Diastolic Hypertension Based on the Cutoff Value in the 2017 American College of Cardiology/American Heart Association Blood Pressure Guidelines With Subsequent Cardiovascular Events in the General Population. **Journal of the American Heart Association**, v. 9, n. 19, 6 out. 2020.

KE, Guolin *et al.* **LightGBM: A Highly Efficient Gradient Boosting Decision Tree**. *[S.l.: S.n.]*. Disponível em: <https://github.com/Microsoft/LightGBM.>.

KEVIN MACKWAY-JONES; JANET MARSDEN; JILL WINDLE. **Emergency Triage**. *[S.l.]*: John Wiley & Sons, 2014.

KLEIN, Laura. **UX for Lean Startups**. *[S.l.]*: O'Reilly Media, Inc., 2013.

LAMBERT, Veronica *et al.* Paediatric early warning systems for detecting and responding to clinical deterioration in children: a systematic review. **BMJ Open**, v. 7, p. 14497, 2017.

LEVIN, Scott *et al.* Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. **Annals of Emergency Medicine**, v. 71, n. 5, p. 565-574.e2, maio 2018.

LEVY, Jaime.; DULLEA, Kate. **UX strategy : product strategy techniques for devising innovative digital solutions**. *[S.l.]*: O'Reilly Media, Inc., 2021.

LI, Jingen *et al.* Evaluation of Optimal Diastolic Blood Pressure Range Among Adults With Treated Systolic Blood Pressure Less Than 130 mm Hg. **JAMA Network Open**, v. 4, n. 2, p. e2037554, 17 fev. 2021.

LI, Yan *et al.* Cardiovascular Risks Associated With Diastolic Blood Pressure and Isolated Diastolic Hypertension. **Current Hypertension Reports**, v. 16, n. 11, p. 489, 3 nov. 2014.

MALINOVSKA, Alexandra *et al.* Modification of the emergency severity index improves mortality prediction in older patients. **Western Journal of Emergency Medicine**, v. 20, n. 4, p. 633–640, 2019.

MANCIA, Giuseppe *et al.* **ESH Guidelines for the management of arterial hypertension. The Task Force for the management of arterial hypertension of the European Society of Hypertension.www.jhypertension.com**. *[S.l.: S.n.]*. Disponível em: <http://journals.lww.com/jhypertension>.

MARTIN, Andrew *et al.* An Examination of ESI Triage Scoring Accuracy in Relationship to ED Nursing Attitudes and Experience. **Journal of Emergency Nursing**, v. 40, n. 5, p. 461–468, set. 2014.

MHD YUNIN, Nor Safiahani; TAN, Toh Leong. National Early Warning Score 2 (NEWS2) as a prognostic tool for adult patients in emergency department: A retrospective observational study. **PLOS One**, v. 20, n. 6, p. e0326058, 16 jun. 2025.

MILNE-IVES, Madison *et al.* Modifications to the National Early Warning Score: a scoping review protocol. **BMJ Open**, v. 14, n. 10, p. e089061, out. 2024.

MISTRY, Binoy *et al.* Accuracy and Reliability of Emergency Department Triage Using the Emergency Severity Index: An International Multicenter Assessment. **Annals of Emergency Medicine**, v. 71, n. 5, p. 581- 587.e3, maio 2018.

MORLEY, Claire *et al.* Emergency department crowding: A systematic review of causes, consequences and solutions. **PLOS ONE**, v. 13, n. 8, p. e0203316, 30 ago. 2018.

MOULIK, Supratik K.; KOTTER, Nina; FISHMAN, Elliot K. **Applications of artificial intelligence in the emergency department**. **Emergency Radiology**Springer, , 1 ago. 2020.

MURALITHARAN, Sankavi *et al.* Machine Learning–Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. **Journal of Medical Internet Research**, v. 23, n. 2, p. e25187, 4 fev. 2021.

NATARAJAN, Aravind; SU, Hao-Wei; HENEGHAN, Conor. Assessment of physiological signs associated with COVID-19 measured using wearable devices. **npj Digital Medicine**, v. 3, n. 1, p. 156, 30 nov. 2020.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE - NICE. **National Early Warning Score systems that alert to deteriorating adult patients in hospital**.

NATIONAL INSTITUTE FOR HEALTH AND CARE RESEARCH - NHIR. **Clinical decision support tools**.

NORMAN, Donald A. .. **The Design of Everyday Things : Revised and Expanded Edition**. *[S.l.]*: Basic Books, 2013.

OBERMEYER, Ziad; EMANUEL, Ezekiel J. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. **New England Journal of Medicine**, v. 375, n. 13, p. 1216–1219, 29 set. 2016.

OLIVEIRA, Gabriella Novelli *et al.* Correlation between classification in risk categories and clinical aspects and outcomes. **Revista Latino-Americana de Enfermagem**, v. 24, n. 0, 2016.

PADRINI-ANDRADE, Lucio *et al.* EVALUATION OF USABILITY OF A NEONATAL HEALTH INFORMATION SYSTEM ACCORDING TO THE USER'S PERCEPTION. **Revista Paulista de Pediatria**, v. 37, n. 1, p. 90–96, jan. 2019.

PAGE LAUBHEIMER. **Wireflows: A UX Deliverable for Workflows and Apps**.

PAGE LAUBHEIMER. **Dashboards: Making Charts and Graphs Easier to Understand**.

PANESAR, Arjun. **Machine Learning and AI for Healthcare**. Berkeley, CA: Apress, 2021.

PAUL BOAG. **Why and How You Should Create Wireframes With Others**.

PINES, Jesse M. *et al.* International Perspectives on Emergency Department Crowding. **Academic Emergency Medicine**, v. 18, n. 12, p. 1358–1370, dez. 2011.

PREVEDELLO, Luciano M. *et al.* Automated Critical Test Findings Identification and Online Notification System Using Artificial Intelligence in Imaging. **Radiology**, v. 285, n. 3, p. 923–931, dez. 2017.

RAITA, Yoshihiko *et al.* Emergency department triage prediction of clinical outcomes using machine learning models. **Critical Care**, v. 23, n. 1, 22 fev. 2019.

RAKERS, Margot *et al.* **Cautiously optimistic about data-driven algorithms in paediatric critical care, nurses' perspectives in low-resource settings: a pre-implementation and human-centred design study in Malawi**. , 5 nov. 2024.

REICHHELD, Frederick F. **The One Number You Need to Grow**. *[S.l.: S.n.]*. Disponível em: <www.hbr.org>.

RICCALTON, Victoria *et al.* Modifications to the National Early Warning Score 2: a Scoping Review. **BMC Medicine**, v. 23, n. 1, p. 154, 11 mar. 2025.

RIES, Eric. **The lean startup : how today's entrepreneurs use continuous innovation to create radically successful businesses**. *[S.l.]*: Currency, 2017.

ROCHA, Hertaline Menezes do Nascimento *et al.* Usabilidade de um sistema de monitoramento das internações em pronto-socorro. **Revista de Saúde Pública**, v. 55, p. 113, 17 dez. 2021.

ROMERO-HALL, Enilda. Pain Assessment and Management in Nursing Education Using Computer-based Simulations. **Pain Management Nursing**, v. 16, n. 4, p. 609–616, ago. 2015.

RUBIN, Jeffrey.; CHISNELL, Dana. **Handbook of usability testing : how to plan, design, and conduct effective tests**. *[S.l.]*: Wiley Pub., 2008.

SAITO, Takaya; REHMSMEIER, Marc. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. **PLOS ONE**, v. 10, n. 3, p. e0118432, 4 mar. 2015.

SAKICHI TOYODA *et al.* **Toyota Production System: A production system based on the philosophy of achieving the complete elimination of waste in pursuit of the most efficient methods**.

SÁNCHEZ-SALMERÓN, Rocío *et al.* Machine learning methods applied to triage in emergency services: A systematic review. **International Emergency Nursing**, v. 60, p. 101109, jan. 2022.

SAX, Dana R. *et al.* Evaluation of Version 4 of the Emergency Severity Index in US Emergency Departments for the Rate of Mistriage. **JAMA Network Open**, v. 6, n. 3, p. e233404, 17 mar. 2023.

SCIKIT LEARN. **Machine Learning in Python**.

SCOTT, Halden; COLBORN, Kathryn. Machine Learning for Predicting Sepsis In-hospital Mortality: An Important Start. **Academic Emergency Medicine**, v. 23, n. 11, p. 1307–1307, 31 nov. 2016.

SCOTT RUPP. **How Nurses Are Using Health Informatics to Improve Patient Care**.

SIEGERT, Stefan. Variance estimation for Brier Score decomposition. **Quarterly Journal of the Royal Meteorological Society**, v. 140, n. 682, p. 1771–1777, 1 jul. 2014.

SILVA, Aline Teixeira *et al.* OS ENFERMEIROS E A SEGURANÇA DO PACIENTE NA PRÁXIS HOSPITALAR. **Cogitare Enfermagem**, v. 21, n. 5, 17 ago. 2016.

SILVA, Luana Matuella Figueira da *et al.* Performance of scores in the prediction of clinical outcomes in patients admitted from the emergency service. **Revista Latino-Americana de Enfermagem**, v. 29, 2021.

SMITH, Gary B. *et al.* The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. **Resuscitation**, v. 84, n. 4, p. 465–470, abr. 2013.

STEWART, Jonathon; SPRIVULIS, Peter; DWIVEDI, Girish. Artificial intelligence and machine learning in emergency medicine. **Emergency Medicine Australasia**, v. 30, n. 6, p. 870–874, 16 dez. 2018.

STEYERBERG, Ewout W. *et al.* Assessing the performance of prediction models: A framework for traditional and novel measures. **Epidemiology**, v. 21, n. 1, p. 128–138, jan. 2010.

STEYERBERG, Ewout W. **Clinical Prediction Models**. Cham: Springer International Publishing, 2019.

TAYLOR, R. Andrew *et al.* Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven, Machine Learning Approach. **Academic Emergency Medicine**, v. 23, n. 3, p. 269–278, 13 mar. 2016.

TEUBNER, David JO *et al.* Model to predict inpatient mortality from information gathered at presentation to an emergency department: The Triage Information Mortality Model (<scp>TIMM</scp>). **Emergency Medicine Australasia**, v. 27, n. 4, p. 300–306, 6 ago. 2015.

ULRICH, Karl T. ..; EPPINGER, Steven D. .. **Product design and development**. *[S.l.]*: McGraw-Hill Education, 2016.

USA; HEALTHIT. **Clinical Decision Support**.

WALTER CYBIS; ADRIANA HOLTZ BETIOL; RICHARD FAUST. **Ergonomia e Usabilidade**. 3ª Edição ed. São Paulo: Novatec, 2015.

WEI, Shengfeng *et al.* The accuracy of the National Early Warning Score 2 in predicting early death in prehospital and emergency department settings: a systematic

review and meta-analysis. **Annals of Translational Medicine**, v. 11, n. 2, p. 95–95, jan. 2023.

WEINSTOCK, David E. **The impact of healthcare {IT} on patient engagement and outcomes**.

WELCH, John; DEAN, John; HARTIN, Jillian. Using NEWS2: an essential component of reliable clinical assessment. **Clinical Medicine**, v. 22, n. 6, p. 509–513, nov. 2022.

WIBOWO, Arief; MASRURIYAH, Anis Fitri Nur; RAHMAWATI, Selly. Refining Diabetes Diagnosis Models: The Impact of SMOTE on SVM, Logistic Regression, and Naïve Bayes. **Journal of Electronics, Electromedical Engineering, and Medical Informatics**, v. 7, n. 1, p. 197–207, 11 jan. 2025.

WILLIAMS, Bryan *et al.* 2018 ESC/ESH Guidelines for the management of arterial hypertension. **European Heart Journal**, v. 39, n. 33, p. 3021–3104, 1 set. 2018.

WORLD HEALTH ORGANIZATION AND INTERNATIONAL TELECOMMUNICATION UNION. **National eHealth Strategy Toolkit**. Geneva: WHO Library Cataloguing-in-Publication Data, 2012.

WUNSCH, Guilherme; DA COSTA, Cristiano A.; RIGHI, Rodrigo R. A Semantic-Based Model for Triage Patients in Emergency Departments. **Journal of Medical Systems**, v. 41, n. 4, p. 65, 10 abr. 2017.

XIE, Kun *et al.* The different risk factors for isolated diastolic hypertension and isolated systolic hypertension: a national survey. **BMC Public Health**, v. 21, n. 1, p. 1672, 14 dez. 2021.

YANO, Yuichiro *et al.* Isolated Diastolic Hypertension and Risk of Cardiovascular Disease: Controversies in Hypertension - Pro Side of the Argument. **Hypertension**, v. 79, n. 8, p. 1563–1570, ago. 2022.

YI, Sojung *et al.* Designing and developing a digital equity dashboard for the emergency department. **JACEP Open**, v. 4, n. 4, p. e12997, ago. 2023.

YUH, Esther L. *et al.* Computer-Aided Assessment of Head Computed Tomography (CT) Studies in Patients with Suspected Traumatic Brain Injury. **Journal of Neurotrauma**, v. 25, n. 10, p. 1163–1172, out. 2008.

YURKOVA, Irina; WOLF, Lisa. Under-triage as a Significant Factor Affecting Transfer Time between the Emergency Department and the Intensive Care Unit. **Journal of Emergency Nursing**, v. 37, n. 5, p. 491–496, set. 2011.

ZACHARIASSE, Joany M. *et al.* Validity of the Manchester triage system in emergency care: A prospective observational study. **PLoS ONE**, v. 12, n. 2, 1 fev. 2017.

ZACHARIASSE, Joany M. *et al.* Multiple performance measures are needed to evaluate triage systems in the emergency department. **Journal of Clinical Epidemiology**, v. 94, p. 27–34, fev. 2018.

ZACHARIASSE, Joany M. *et al.* Improving the prioritization of children at the emergency department: Updating the Manchester Triage System using vital signs. **PLoS ONE**, v. 16, n. 2 February, 1 fev. 2021.

ZADROZNY, Bianca; ELKAN, Charles. Transforming classifier scores into accurate multiclass probability estimates. *In*: New York, NY, USA: ACM, 23 jul. 2002.

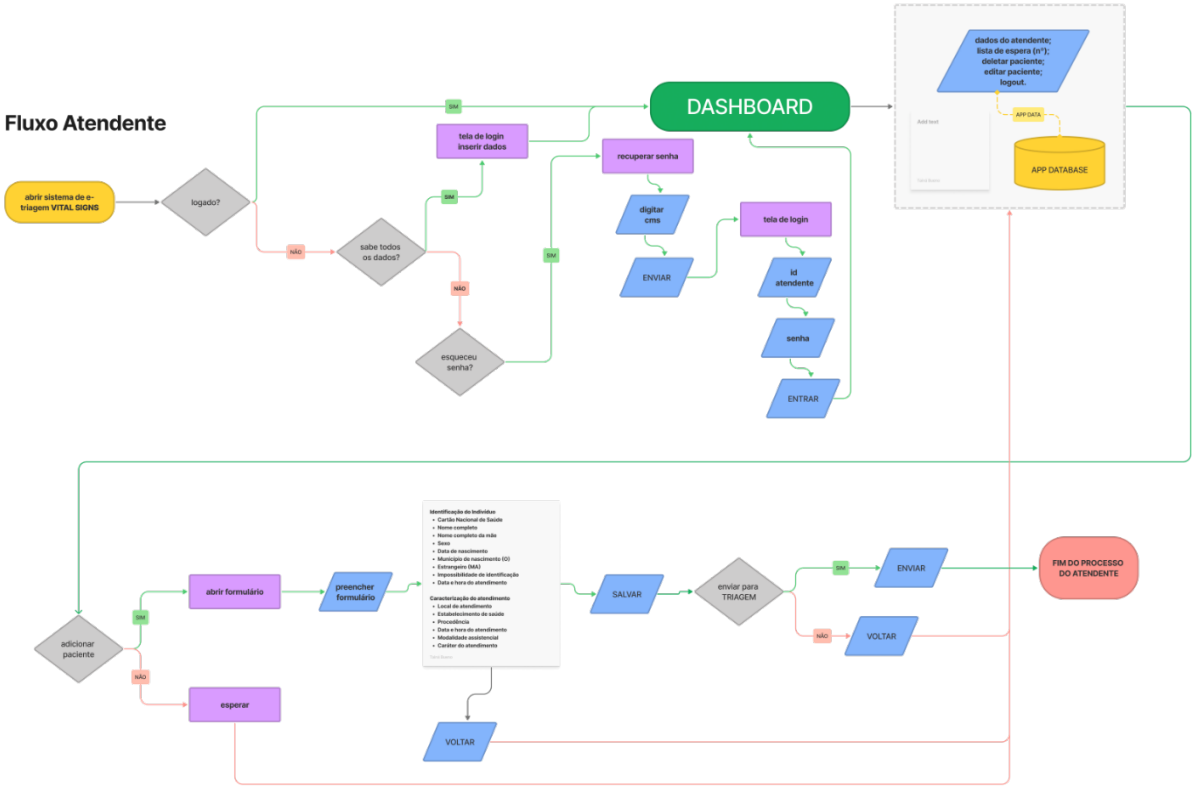
APÊNDICE A – QUESTIONÁRIO PRÉ-TESTE

1. Quais são as suas principais responsabilidades no início de um plantão na urgência e emergência?
2. Como é o processo de triagem dos pacientes na sua unidade? Que fatores você avalia primeiro?
3. Quais são os sinais de alerta que você observa para priorizar pacientes na emergência?
4. Pode descrever como se dá o uso da tecnologia ou softwares no seu dia a dia? Alguma ferramenta é usada para triagem ou monitoramento de pacientes?
5. Como você avalia o estado de saúde de um paciente ao longo do tempo enquanto ele permanece na unidade de urgência?
6. Quais são as ferramentas ou recursos que você gostaria que fossem melhorados para facilitar sua rotina de atendimento?
7. Como é feita a comunicação e a troca de informações com outros setores do hospital, como internação e UTI?
8. Quais são as principais causas de readmissão de pacientes na urgência, na sua experiência?
9. Quais são os sinais clínicos mais importantes que você monitora constantemente em pacientes de alto risco na emergência?
10. Como você acompanha e responde às mudanças súbitas no quadro clínico de pacientes?
11. Você já utilizou algum software para triagem eletrônica? Se sim, como isso impactou a rapidez e eficiência do atendimento?
12. Você sente que o uso de software ajuda a reduzir erros na administração de medicamentos ou no cuidado de pacientes?

APÊNDICE B – MAPA DA JORNADA DO ATENDENTE

O atendente de uma UE é a primeira pessoa de contato do paciente na hora que entra no hospital. O atendente tem a atribuição de coletar os dados de identificação do paciente conforme o que a Resolução nº 33/2018 que regulariza o Registro de Atendimento Clínico (BRASIL, 2018). Ao finalizar a conduta da coleta de dados, o sistema informa ao próximo agente que o paciente aguarda por atendimento. Entendendo que esta é a função do atendente, criou-se um mapa de sua jornada com as atribuições mínimas viáveis para o MVP (Figura x).

Figura 35 – Mapa da Jornada do Atendente



Fonte: O autor.

ANEXO A – CÓDIGO MACHINE LEARNING

A pasta com o código e as saídas dos modelos no link: machine-link-tese