

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

**APLICAÇÃO DE ALGORITMOS DE *MACHINE LEARNING* PARA ANÁLISE E
PREDIÇÃO DE DESFECHOS CLÍNICOS EM PACIENTES COM COVID-19**

BRUNO FONSECA REZENDE

PATOS DE MINAS
2025

BRUNO FONSECA REZENDE

**APLICAÇÃO DE ALGORITMOS DE *MACHINE LEARNING* PARA ANÁLISE E
PREDIÇÃO DE DESFECHOS CLÍNICOS EM PACIENTES COM COVID-19**

**Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Biotecnologia do
Instituto de Biotecnologia da Universidade
Federal de Uberlândia, como requisito parcial
para a obtenção do título de Mestre em
Biotecnologia.**

Área de concentração: Biociências

**Orientador: Prof. Dr. Laurence Rodrigues do
Amaral**

PATOS DE MINAS

2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

R467 2026	<p>Rezende, Bruno Fonseca, 1985- APLICAÇÃO DE ALGORITMOS DE MACHINE LEARNING PARA ANÁLISE E PREDIÇÃO DE DESFECHOS CLÍNICOS EM PACIENTES COM COVID-19 [recurso eletrônico] / Bruno Fonseca Rezende. - 2026.</p> <p>Orientador: PROF. DR. LAURENCE RODRIGUES DO AMARAL. Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Biotecnologia. Modo de acesso: Internet. DOI http://doi.org/10.14393/ufu.di.2026.12 Inclui bibliografia.</p> <p>1. Biotecnologia. I. AMARAL, PROF. DR. LAURENCE RODRIGUES DO, 1978-, (Orient.). II. Universidade Federal de Uberlândia. Pós- graduação em Biotecnologia. III. Título.</p> <p>CDU: 60</p>
--------------	---

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Biotecnologia				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	18/12/2025	Hora de início:	09:00	Hora de encerramento:	10:22
Matrícula do Discente:	42312BTC007				
Nome do Discente:	Bruno Fonseca Rezende				
Título do Trabalho:	Aplicação de Algoritmos de Machine Learning para Predição de desfechos Clínicos em pacientes com COVID-19				
Área de concentração:	Biociências				
Linha de pesquisa:	Bioinformática e Biologia Molecular aplicada à genômica, transcriptômica e proteômica				

Reuniu-se, por vídeo - conferência do Microsoft Teams, a Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Biotecnologia, composta pelos seguintes Professores Doutores: Pedro Luiz Lima Bertarini (FEELT/UFU), Lúcio Campos Queiroz (FHEMIG) e Laurence Rodrigues do Amaral (FACOM/UFU), presidente da banca.

Iniciando os trabalhos, o presidente da banca, Dr. Laurence Rodrigues do Amaral, apresentou a Comissão Examinadora e o candidato, agradeceu a presença dos membros. Em seguida, concedeu a palavra ao discente para a exposição do seu trabalho. A duração da apresentação seguiu as normas do Programa de Pós-Graduação em Biotecnologia.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente

ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Laurence Rodrigues do Amaral, Professor(a) do Magistério Superior**, em 18/12/2025, às 10:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Lúcio Campos Borges, Usuário Externo**, em 18/12/2025, às 11:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Pedro Luiz Lima Bertarini, Professor(a) do Magistério Superior**, em 30/12/2025, às 09:26, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6946409** e o código CRC **5E5A0B44**.

Referência: Processo nº 23117.089267/2025-92

SEI nº 6946409

BRUNO FONSECA REZENDE

**APLICAÇÃO DE ALGORITMOS DE *MACHINE LEARNING* PARA ANÁLISE E
PREDIÇÃO DE DESFECHOS CLÍNICOS EM PACIENTES COM COVID-19**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biotecnologia do Instituto de Biotecnologia da Universidade Federal de Uberlândia, com vistas à obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Laurence Rodrigues do Amaral

Aprovado em: 18/12/2025

BANCA EXAMINADORA

Prof. Dr. Laurence Rodrigues do Amaral

Prof. Dr. Pedro Luiz Lima Bertarini

Prof. Dr. Lucio Campos Borges

PATOS DE MINAS

2025

AUXÍLIO FINANCEIRO

O desenvolvimento desta pesquisa foi realizado com recursos próprios e por meio da análise de um banco de dados público, sem apoio financeiro de instituições públicas ou privadas.

Dedico este trabalho à minha família, fonte de amor, força e inspiração; aos meus amigos, pelo apoio em todos os momentos; aos meus pacientes, que motivam minha prática e meu compromisso diário com o cuidar; e aos meus colegas de profissão, pela parceria, aprendizado e incentivo ao longo desta jornada.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pela força, sabedoria e proteção em todos os momentos desta caminhada.

À minha esposa, Kymya Gonçalves Araújo Fonseca, pelo amor, paciência e apoio incondicional.

Aos meus filhos, Esther Gonçalves Fonseca, Maria Alice Gonçalves Fonseca e Emanuel Gonçalves Fonseca, que são minha motivação diária e a razão de muitos dos meus esforços.

Aos meus pais, Francisco Alves de Resende e Vera Lúcia Fonseca Rezende, por todo o carinho, educação, valores e incentivo que sempre nortearam minha vida.

À minha avó Luzia Alves da Fonseca e ao meu padrinho e avô Divino Alves de Resende, pela presença, exemplo e inspiração ao longo da minha trajetória.

Agradeço também aos meus sogros, Iracema Gonçalves da Silva Araújo e José Marcos de Araújo, pelo apoio constante.

Ao meu irmão, Guilherme Fonseca Rezende, e à sua família, pela parceria.

À minha cunhada Kelly Gonçalves Araújo Teotônio e à sua família, pelo carinho.

Registro minha profunda gratidão à professora e colega de profissão Roane Caetano de Faria, cuja motivação e incentivo foram fundamentais para que eu iniciasse esta jornada.

Ao meu orientador, Prof. Dr. Laurence Rodrigues do Amaral, agradeço pela orientação, apoio, pelas contribuições essenciais ao desenvolvimento deste trabalho e pela confiança depositada em minha trajetória acadêmica.

Aos meus amigos, em especial a Wálíce Rans, pela valiosa colaboração no desenvolvimento desta pesquisa, cuja presença e incentivo foram fundamentais.

A todos os meus colegas de trabalho, pela parceria, compreensão e incentivo.

Deixo registrado meu reconhecimento e minha gratidão a todos!

Que a inteligência artificial seja sempre uma ponte entre o conhecimento e o cuidado, entre o algoritmo e a empatia.

RESUMO

A pandemia de COVID-19 impôs desafios significativos aos sistemas de saúde, especialmente pela rápida evolução clínica e pela necessidade de decisões imediatas sobre suporte ventilatório e internação em UTI, revelando a lacuna no uso de ferramentas objetivas capazes de auxiliar a estratificação precoce do risco e a previsão de prognóstico em pacientes hospitalizados. Nesse contexto, modelos de Aprendizado de Máquina (ML), como árvores de decisão, apresentam potencial para apoiar a tomada de decisão clínica por permitirem analisar múltiplas variáveis simultaneamente e gerar resultados interpretáveis. O objetivo deste estudo foi aplicar algoritmos de ML a dados clínicos de pacientes com COVID-19 a fim de prever desfechos relevantes e identificar fatores associados à gravidade da doença. Trata-se de um estudo retrospectivo e quantitativo, baseado em dados secundários provenientes do banco público de Síndrome Respiratória Aguda Grave (SRAG) da cidade de Londrina (Paraná), contendo inicialmente 15.655 registros de pacientes hospitalizados entre janeiro de 2021 e fevereiro de 2022. Após filtragem e organização das variáveis, foram utilizados 5.704 registros no modelo de desfecho e 7.182 nos modelos complementares. As análises foram realizadas no software WEKA (v.3.8.6), utilizando o algoritmo J48 com validação cruzada, após seleção e eliminação de atributos incompletos. Os modelos desenvolvidos alcançaram acurácia aproximada de 80% na classificação dos casos. O tipo de suporte ventilatório emergiu como principal preditor tanto na etapa de triagem quanto na previsão do desfecho final, seguido por internação em UTI, presença de doença cardíaca, idade e tipo de hospital. Os resultados demonstraram que a ventilação mecânica invasiva e a internação em UTI estiveram fortemente associadas ao óbito, enquanto o uso de ventilação não invasiva e a ausência de suporte ventilatório relacionaram-se à cura. A idade avançada intensificou o risco de mortalidade nos pacientes sob ventilação invasiva, e o tipo de hospital influenciou o prognóstico entre os pacientes não admitidos em UTI, com maior mortalidade em hospitais públicos. Conclui-se que as árvores de decisão mostraram-se eficazes na identificação de fatores preditores de gravidade em pacientes hospitalizados com COVID-19, oferecendo modelos de fácil interpretação clínica e potencial aplicação na estratificação de risco, no direcionamento de recursos e na organização do cuidado em ambiente hospitalar, especialmente em situações de sobrecarga assistencial como a vivenciada durante a pandemia.

Palavras-chave: aprendizado de máquina; árvore de decisão; COVID-19; predição de desfechos clínicos; UTI; ventilação mecânica.

ABSTRACT

The COVID-19 pandemic imposed substantial challenges on healthcare systems worldwide, particularly due to the rapid clinical deterioration observed in many patients and the urgent need to allocate ventilatory support and intensive care resources efficiently, highlighting a gap in objective tools capable of assisting early risk stratification and outcome prediction. In this context, *Machine Learning* (ML) models, such as *decision trees*, offer the potential to support clinical decision-making by analyzing multiple variables simultaneously and generating easily interpretable predictive structures. This study aimed to apply ML algorithms to clinical data from hospitalized patients with COVID-19 to predict relevant clinical outcomes and identify factors associated with disease severity. This retrospective, quantitative study used secondary data from the public Severe Acute Respiratory Infection (SARI) database of Londrina, Paraná, Brazil, which initially contained 15,655 records of hospitalized patients between January 2021 and February 2022. After data cleaning and preprocessing, 5,704 records were used for the outcome prediction model and 7,182 for complementary analyses. Data were analyzed using the WEKA software (v.3.8.6) with the J48 decision tree algorithm and cross-validation, following attribute selection and removal of redundant or incomplete variables. The generated models achieved approximately 80% accuracy. Type of ventilatory support emerged as the most relevant predictor across analyses, followed by ICU admission, preexisting heart disease, age, and hospital type. Results showed that invasive mechanical ventilation and ICU admission were strongly associated with mortality, whereas non-invasive ventilation or absence of ventilatory support were associated with recovery. Advanced age substantially increased mortality risk among patients receiving invasive ventilation, and hospital type influenced outcomes among those not admitted to the ICU, with higher mortality in public hospitals. The study concludes that decision tree models are effective for identifying predictors of severity in hospitalized COVID-19 patients, providing clinically interpretable structures with potential application for risk stratification, resource allocation, and management of critical care pathways, especially in contexts of healthcare overload such as that experienced during the pandemic.

Keywords: *decision tree*; intensive care unit; *machine learning*; mechanical ventilation; clinical outcomes prediction.

LISTA DE SIGLAS E ABREVIATURAS

AMIB – Associação de Medicina Intensiva Brasileira

BiPAP ou BiLEVEL - Bilevel Positive Airway Pressure

CDC – Centros de Controle e Prevenção de Doenças

COVID-19 – Coronavirus Disease 2019

CPAP - Continuous Positive Airway Pressure

DPOC - Doença Pulmonar Obstrutiva Crônica

DL – Deep Learning

ECMO – Extracorporeal Membrane Oxygenation

FR – Frequência Respiratória

HFNO – High-Flow Nasal Oxygen

IA – Inteligência Artificial

IRpA – Insuficiência Respiratória Aguda

IRPM - Incursões Respiratórias por Minuto

LOOCV - Leave-One-Out Cross Validation

ML – *Machine Learning*

MERS – Middle East Respiratory Syndrome

NA – Not Available

P-SILI - Lesão Pulmonar Autoinfligida

PCR – Polymerase Chain Reaction

PaO₂ – Pressão Parcial de Oxigênio no sangue Arterial

PEEP – Positive End-Expiratory Pressure

RNA – Ribonucleic Acid

RT-PCR – Reverse Transcription Polymerase Chain Reaction

SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

SBPT – Sociedade Brasileira de Pneumologia e Tisiologia

SRAG – Síndrome Respiratória Aguda Grave

SaO₂ - Saturação Arterial de Oxigênio

SpO₂ – Saturação Periférica de Oxigênio

SUS – Sistema Único de Saúde

TC – Tomografia Computadorizada

UTI – Unidade de Terapia Intensiva

VILI – Ventilator-Induced Lung Injury

VMI – Ventilação Mecânica Invasiva

VNI – Ventilação Não Invasiva

WEKA – Waikato Environment for Knowledge Analysis

SUMÁRIO

1. INTRODUÇÃO	15
2. CAPÍTULO I: FUNDAMENTAÇÃO TEÓRICA	16
2.2 PANDEMIA DA COVID-19	20
2.3 INTELIGÊNCIA,INTELIGÊNCIA ARTIFICIAL, <i>MACHINE LEARNING E DEEP LEARNING</i>	21
2.4 ÁRVORES DE DECISÃO E FLORESTAS ALEATÓRIAS	23
2.5 VENTILAÇÃO MECÂNICA INVASIVA E NÃO INVASIVA	24
3. OBJETIVOS	24
3.1 OBJETIVO GERAL	25
3.2 OBJETIVOS ESPECÍFICOS	26
4. CAPITULO II: ARTIGO	26
4.1 Introdução	27
4.3 Objetivos	28
4.3.1 Objetivo Geral	28
4.3.2 Objetivos Específicos	28
4.4 Metodologia	28
4.5 Resultados	29
4.6 Discussão	33
4.7 Conclusão	34
4.8 Perspectivas Futuras	34
Referências	35

1. INTRODUÇÃO

A COVID-19, causada pelo Coronavírus, foi identificada pela primeira vez em Wuhan, na China, no final de 2019. Essa doença altamente contagiosa afeta principalmente o sistema respiratório humano e é transmitida principalmente por meio de gotículas respiratórias durante o contato entre as pessoas. Os sintomas incluem falta de ar, febre, tosse e perda de paladar, e alguns casos sendo assintomáticos (GUPTA et al., 2021). A doença se espalhou rapidamente pelo mundo, levando a uma pressão significativa sobre os sistemas de saúde e resultando em altas taxas de mortalidade, com mais de 280 milhões de infecções e 5,4 milhões de mortes até dezembro de 2021 (SHAKHOVSKA et al., 2022).

Diversos países adotaram medidas como isolamento, bloqueios, restrições em eventos sociais e incentivo ao uso de máscaras para diminuir a disseminação do vírus. Aproximadamente 80% dos infectados apresentaram sintomas leves que se resolveram em duas semanas, enquanto 20% dos casos necessitaram de hospitalização, com uma taxa de mortalidade de cerca de 13,4% entre os mais críticos (SHAKHOVSKA et al., 2022). A maioria dos pacientes se recuperaram em 10 a 15 dias, mas aqueles com infecções graves levaram em média até 45 dias e tinham um alto risco de desenvolver síndrome pós-Covid. Essa condição pode afetar diversos órgãos, aumentando as chances de complicações a longo prazo, e é influenciada por fatores como resposta imunológica inadequada e hospitalização prolongada (GUPTA et al., 2021).

Modelos de ML de excelência baseiam-se em conjuntos de dados abrangentes, minuciosamente organizados e criteriosamente selecionados. Quando são projetados e treinados de maneira eficaz, esses modelos conseguem identificar padrões relevantes e criar ferramentas preditivas robustas. Eles aprendem de maneira objetiva e, muitas vezes, geram variações mais precisas do que aquelas alcançadas na prática convencional. Recentemente, a utilização de métodos de inteligência artificial, *Deep Learning (DL)* e ML tem crescido significativamente na pesquisa biomédica, evidenciando seu impacto positivo em várias áreas do conhecimento (RUSIC et al., 2024).

Diante desse cenário, tornou-se crucial o desenvolvimento de métodos para avaliar e prever o risco de agravamento da doença de forma quantitativa, fundamental para otimizar o manejo clínico dos pacientes e garantir a distribuição eficiente dos recursos médicos disponíveis.

2. CAPÍTULO I: FUNDAMENTAÇÃO TEÓRICA

2.1 CORONAVÍRUS E COVID-19

O Coronavírus é um vírus de RNA envelopado de sentido positivo de fita simples, cujo nome foi devido por suas espículas superficiais de 9 a 12nm de comprimento que se assemelham a uma coroa (do latim corona) (ALSULIMAN, et al., 2020). Pertencem à subfamília *Coronavirinae*, da família *Coronaviridae*, incluídos na ordem *Nidovirales*. A ordem *Nidovirales* é composta pelas famílias *Coronaviridae*, *Arteriviridae*, *Mesovirididae* e *Roniviridae*. As principais características estruturais e funcionais dos *Nidovirales* incluem: genomas extremamente extensos para vírus de RNA, alta capacidade de replicação em razão de uma organização genômica conservada, presença de diversas atividades enzimáticas específicas e modificações complexas na estrutura ribossômica, decorrentes da expressão de múltiplos genes não estruturais. A família *Coronaviridae* é subdividida em duas subfamílias: *Coronavirinae* e *Torovirinae*. A subfamília *Coronavirinae* inclui os gêneros alfa-CoV, beta-CoV, gama-CoV e delta-CoV, classificados com base em sua estrutura genômica (UMAKANTHAN et al., 2020).

Embora a origem do SARS-CoV-2 ainda não tenha sido completamente esclarecida, há consenso de que o vírus apresenta origem zoonótica, possivelmente decorrente da transmissão entre espécies. Análises genômicas indicam que o SARS-CoV-2 provavelmente evoluiu a partir de uma cepa de coronavírus encontrada em morcegos, uma vez que sua sequência genômica apresenta cerca de 96% de similaridade com o betaCoV RaTG13, identificado em *Rhinolophus affinis*. De forma semelhante ao observado em surtos anteriores de SARS e MERS, propõe-se que o SARS-CoV-2 tenha sido transmitido de morcegos para hospedeiros intermediários, como pangolins ou visons, antes de alcançar os seres humanos (CASCELLA, et al., 2025).

Apesar dos avanços na pesquisa clínica e no conhecimento sobre o SARS-CoV-2, surtos continuam ocorrendo, principalmente devido ao surgimento de variantes com mutações genéticas que alteram características do vírus original. Como outros vírus de RNA, apresenta alta taxa de mutação, resultando em diversas linhagens virais, das quais algumas são classificadas como variantes com potencial aumentado de transmissão ou resistência a resposta imune (CASCELLA, et al., 2025).

O vírus causador da COVID-19 é transmitido predominantemente de pessoa para pessoa por meio de gotículas respiratórias eliminadas durante tosse, espirro ou fala. Essa

transmissão ocorre quando há contato próximo, permitindo que as gotículas atinjam as mucosas do hospedeiro, especialmente olhos, nariz e boca, promovendo a infecção (PARASHER, 2021).

A transmissão por fômites, decorrente da contaminação de superfícies inanimadas pelo SARS-CoV-2, foi demonstrada em diversos estudos que evidenciaram a presença do vírus em diferentes materiais. Em condições experimentais, o SARS-CoV-2 permaneceu estável por até 72 horas em aço inoxidável e plástico, e por até 28 dias a 20 °C em superfícies não porosas, como vidro e aço inoxidável. Em ambientes hospitalares, o vírus foi detectado em pisos, equipamentos e corrimãos, além de amostras de ar a até quatro metros dos pacientes. Apesar disso, os Centros de Controle e Prevenção de Doenças (CDC) consideram o risco de transmissão por superfícies baixo, não sendo essa a principal via de disseminação do vírus (CASCELLA, et al., 2025).

A transmissão aérea, por sua vez, não é considerada uma via comum de propagação da COVID-19, sendo descrita apenas em situações específicas que envolvem a geração de aerossóis, como durante procedimentos invasivos, incluindo intubação orotraqueal, broncoscopia, aspiração aberta, nebulização, ventilação com máscara eambu antes da intubação, traqueostomia e manobras de ressuscitação cardiopulmonar (PARASHER, 2021).

O período de incubação da COVID-19, que é o intervalo de tempo entre a exposição ao vírus e o início dos sintomas, é de 5 a 6 dias, mas pode chegar a 14 dias. Durante esse período, também conhecido como período pré-sintomático, os indivíduos infectados podem ser contagiosos e transmitir o vírus para indivíduos saudáveis. Pacientes com COVID-19 apresentam, mais comumente, os sintomas como febre, dores no corpo, falta de ar, mal-estar e tosse seca, alguns pacientes também podem apresentar sintomas gastrointestinais, como dor abdominal, vômitos e diarreia. Podem apresentar a doença de forma assintomática, leve, moderada ou grave; como descrito na Tabela 1 (PARASHER, 2021).

Tabela 1: Espectro clínico da COVID-19.

Gravidade da doença	Apresentação
Assintomático	<ul style="list-style-type: none"> • Sem sintomas clínicos • Teste de swab nasal positivo • Radiografia de tórax normal
Doença leve	<ul style="list-style-type: none"> • Febre, dor de garganta, tosse seca, mal-estar e dores no corpo ou náuseas, vômitos, dor abdominal, fezes amolecidas
Doença moderada	<ul style="list-style-type: none"> • Sintomas de pneumonia (febre persistente e tosse) sem hipoxemia • Lesões significativas na tomografia computadorizada de alta resolução do tórax.

Doença grave	<ul style="list-style-type: none"> • Pneumonia com hipoxemia ($SpO_2 < 92\%$)
Estado crítico	<ul style="list-style-type: none"> • Síndrome da Angústia Respiratória Aguda, juntamente com choque, distúrbios de coagulação, encefalopatia, insuficiência cardíaca e lesão renal aguda.

Fonte: PARASHER, 2021.

Devido a gravidade da COVID-19, a detecção precoce e precisa da infecção é fundamental para o manejo clínico eficaz. Os testes rápidos e sensíveis são essenciais nesse contexto, sendo os métodos baseados na detecção de ácidos nucleicos os mais confiáveis para a identificação do SARS-CoV-2. A Reação em Cadeia da Polimerase com Transcrição Reversa (RT-PCR) é considerada o padrão-ouro para o diagnóstico, devido à sua alta sensibilidade e especificidade. Essa técnica permite a detecção de material genético viral em amostras de vias respiratórias superiores e inferiores, como swabs nasofaríngeos e orofaríngeos, além de escarro (SAILUNAZ et al., 2023).

Além dos testes laboratoriais, exames de imagem torácica, especialmente a Tomografia Computadorizada (TC), são ferramentas valiosas para a identificação precoce, avaliação da gravidade e monitoramento da evolução da doença, uma vez que o vírus compromete predominantemente o parênquima pulmonar. Os achados radiológicos típicos incluem opacidades em vidro fosco (com distribuição bilateral, periférica, posterior e basal), opacidades reticulonodulares e subpleurais curvilíneas. Outras alterações incluem espessamento septal interlobular, consolidações alveolares, lesões cavitárias e o padrão de "pavimentação em mosaico", podendo ainda ser observados sinais como "árvore em brotamento" e halo (SAILUNAZ et al., 2023).

O Coronavírus pode causar uma série de doenças, incluindo infecções do trato respiratório, gastrointestinais, cardíacas e neurológicas, com gravidade variável entre animais e seres humanos (ALSULIMAN, et al., 2020). A atual doença, também chamada de Síndrome Respiratória Aguda Grave (SRAG) ou doença do coronavírus 2019 (COVID-19), representou uma ameaça global emergente à saúde (UMAKANTHAN et al., 2020).

Existe uma heterogeneidade significativa na apresentação clínica da infecção por COVID-19, variando desde pacientes assintomáticos até aqueles com doença grave. Identificar preditores de desfechos graves como: internação em UTI, necessidade de ventilação mecânica e óbito; é fundamental, pois muitos pacientes podem apresentar deterioração rápida, mesmo após iniciarem com sintomas leves. O reconhecimento desses preditores permite que os profissionais de saúde ofereçam assistência precoce e mais adequada aos pacientes, além de orientar intervenções capazes de reduzir o risco de morte (CHORON, et al., 2021).

Os fatores mais frequentemente associados à progressão para quadros graves incluem idade avançada, sexo masculino, obesidade e comorbidades, como diabetes e doença renal. Além disso, sabe-se que biomarcadores como Proteína C-Reativa (PCR) e dímero-D, podem estar relacionados a desfechos desfavoráveis. No entanto, muitos estudos conduzidos no início da pandemia possuíam amostras pequenas e buscavam identificar os preditores mais relevantes de gravidade e mortalidade a partir de um grande conjunto de variáveis. Ademais, parte desses estudos consistia em séries de casos hospitalares; portanto, seus achados podem não representar adequadamente todos os pacientes internados com COVID-19 (CHORON, et al., 2021).

A Síndrome Respiratória Aguda Grave (SRAG) foi reconhecida como a primeira infecção pandêmica associada ao coronavírus. Teve início na China entre os anos de 2002 e 2003, sendo causada por um novo tipo de coronavírus, o SARS-CoV-2. A doença espalhou-se rapidamente em 29 países ao longo do ano de 2003, devido a um grande fluxo de viagens internacionais, acometendo 8.098 pessoas e apresentando uma taxa de letalidade de 9,6%. Posteriormente, a circulação do vírus foi controlada e não houve registro de novos casos. O segundo surto associado ao coronavírus foi a Síndrome Respiratória do Oriente Médio (MERS), causada pelo MERS-CoV. Identificada pela primeira vez em humanos na Arábia Saudita, em abril de 2012, a infecção está relacionada ao contato com camelos ou seus derivados. Entre 2012 e dezembro de 2019, foram registrados 2.465 casos confirmados em 27 países, com 850 óbitos, resultando em uma taxa de mortalidade de 34,4% (ALSULIMAN, et al., 2020).

Como não existia vacina disponível para a COVID-19, o tratamento era principalmente sintomático e de suporte na maioria dos casos. Na admissão o paciente na emergência, era classificado como leve, moderado ou grave, de acordo com os sintomas apresentados. A maioria dos pacientes apresentavam sintomas leves a moderados, como febre, tosse seca persistente, dores no corpo e falta de ar ocasional. Uma pequena parcela dos pacientes também desenvolvia insuficiência respiratória aguda e síndrome da angústia respiratória aguda com sepse associada ou falência múltipla de órgãos (PARASHER, 2021).

O tratamento da insuficiência respiratória em pacientes com COVID-19 varia conforme a gravidade do quadro clínico. Nos casos leves, com Saturação Periférica de Oxigênio (SpO₂) entre 94% e 97%, recomenda-se oxigenoterapia com cânula nasal, máscara facial simples ou de Venturi, além de manejo sintomático com antipiréticos, hidratação e nutrição adequada. A reavaliação deve ser frequente, e, em caso de ausência de melhora, considera-se a Ventilação Não Invasiva (VNI). Já nos casos moderados (SpO₂ entre 90% e 94%), é indicado isolamento, monitoramento contínuo e uso de Oxigenoterapia de Alto Fluxo Nasal (HFNO) ou

VNI, conforme a resposta clínica. O tratamento pode incluir corticoides, anticoagulação profilática e antibióticos quando há infecção bacteriana associada (PARASHER, 2021).

Nos casos graves ($\text{SpO}_2 \leq 90\%$ ou Síndrome do Desconforto Respiratório Agudo), a oxigenoterapia deve ser iniciada imediatamente, podendo evoluir para Ventilação Mecânica Invasiva (VMI) com parâmetros protetores, como baixos volumes correntes e pressão de platô inferior a 30 cmH₂O. Pacientes com síndrome grave podem se beneficiar do posicionamento em decúbito ventral e de níveis mais elevados de PEEP. Em situações de hipoxemia refratária ou falência de múltiplos órgãos, recomenda-se considerar o uso de oxigenação por Membrana Extracorpórea (ECMO), quando disponível (PARASHER, 2021).

2.2 PANDEMIA DA COVID-19

Em dezembro de 2019, na cidade de Wuhan (China), teve início a Pandemia da COVID-19, com o surgimento de uma nova cepa do Coronavírus, o SARS-CoV-2, atingindo uma status pandêmico em Março de 2020, conforme declaração da Organização Mundial de Saúde. Foram necessárias adotar certas medidas de segurança a fim de conter sua rápida disseminação, incluindo o uso obrigatório de máscaras faciais, distanciamento social, cancelamento de eventos, restrições severas de viagens, fechamento de escolas e universidades, exceto os serviços essenciais e de saúde (BOTH et al., 2021).

O Brasil confirmou seu primeiro caso em 26 de fevereiro de 2020, em São Paulo, e em poucas semanas o vírus já havia se espalhado por todas as regiões do país. Com significativa heterogeneidade socioeconômica e climática entre os estados, a resposta à pandemia foi marcada por desigualdade na implementação de medidas sanitárias. A testagem sistemática de pacientes com SRAG iniciou-se em março de 2020, mas a rápida disseminação comunitária, aliada à dificuldade de rastreamento de contatos, contribuiu para o colapso do sistema de saúde em diversas unidades federativas (SERDAN et al., 2021).

A resposta nacional à crise sanitária foi prejudicada por conflitos entre esferas de governo e ausência de diretrizes unificadas, o que resultou em ações fragmentadas e de eficácia variável. Medidas como distanciamento social, lockdown e uso de máscaras foram adotadas de forma heterogênea, impactando a efetividade do controle epidemiológico. O estado de São Paulo, epicentro da pandemia na América Latina, apresentou altos índices de testagem e medidas restritivas, porém flexibilizações precoces influenciaram negativamente a dinâmica de transmissão (SERDAN et al., 2021).

A pandemia de COVID-19 exigiu a adoção de medidas de isolamento social, como o distanciamento físico entre as pessoas, restrição do convívio familiar e social. Embora eficazes no controle da disseminação viral, essas medidas geraram impactos psicológicos relevantes, incluindo medo, incertezas e sentimentos de desesperança. Juntamente, observou-se efeitos colaterais em múltiplas esferas, como economia, sistemas de saúde, relações familiares e uso excessivo de tecnologias (BOTH et al., 2021).

Durante a pandemia de COVID-19 evidenciou limitações estruturais no sistema hospitalar brasileiro, em especial no que se refere à oferta de leitos de UTI. Dados de julho de 2020 indicaram que o Brasil dispunha de aproximadamente 31.940 leitos de UTI adulto e 4.938 pediátricos, sendo que apenas 48% e 54,1%, respectivamente, estavam vinculados ao Sistema Único de Saúde (SUS). A ampliação emergencial de leitos de UTI para COVID-19, promovida por estados e municípios, representou um aumento de 58,7% sobre os leitos de UTI existentes antes da pandemia (CAMPOS et al., 2020).

Do total de leitos COVID-19 habilitados, 79,7% estavam inseridos em hospitais que já estavam em funcionamento, totalizando 2.278 unidades, das quais 9.761 leitos eram destinados a adultos e 201 à população pediátrica. Esses números refletiam com maior precisão a capacidade hospitalar disponível à população que dependia do SUS, evidenciando quais instituições podiam, de fato, ser mobilizadas para o enfrentamento da pandemia. Tais hospitais configuram-se como pontos estratégicos para investimentos prioritários, tanto no que se refere à sua sustentabilidade financeira quanto ao monitoramento contínuo de seu desempenho e de seu papel na organização da rede de atenção à saúde (CAMPOS et al., 2020).

2.3 INTELIGÊNCIA,INTELIGÊNCIA ARTIFICIAL, *MACHINE LEARNING* *E DEEP LEARNING*

A definição do termo inteligência tem se demonstrado desafiador, devido a sua complexidade e a várias interpretações. No entanto, com o objetivo de aprofundar a compreensão do conceito, foram reunidas múltiplas definições encontradas na literatura, de modo a delinear o nível atual de entendimento sobre o assunto (ALSULIMAN et al., 2020).

Inteligência é a capacidade de elaborar projetos bem estruturados, resolver problemas ou criar contribuições significativas em uma determinada cultura ou área de negócios. Utiliza associação, memorização, raciocínio, compreensão, abstração, conceituação, aproximação,

sistematização e inferência lógica. Esses elementos são aplicados para gerar novos conhecimentos com base em fatos já estabelecidos (ALSULIMAN et al., 2020).

Por outro lado, a Inteligência Artificial (IA) refere-se à capacidade do sistema de interpretar corretamente dados externos, aprender com esses dados e utilizar esses aprendizados para alcançar objetivos e realizar tarefas específicas, por meio de uma adaptação flexível (ALSULIMAN et al., 2020). Consiste na utilização de algoritmos e modelos que possibilitam que máquinas executem tarefas que normalmente exigem inteligência humana. As primeiras aplicações de IA na área da saúde baseavam em regras construídas a partir do conhecimento e da experiência de especialistas humanos (THEODOSIOU et al., 2023).

O ML é um tipo de IA que envolve o treinamento de um algoritmo usando dados com saídas ou características conhecidas. Ao ser treinado com esses dados, o algoritmo pode posteriormente prever uma saída ou encontrar características relevantes quando apresentado a um caso real. Como as saídas correspondem a previsões, quanto maior for o conjunto de dados utilizado para o treinamento, mais precisas e confiáveis tendem a ser as respostas do modelo quando aplicado a novos dados (PANESAR et al., 2020).

Pode ser dividido em quatro categorias: supervisionado, não supervisionado, semisupervisionado e por reforço. No aprendizado supervisionado, a máquina aprende a gerar saídas a partir de entradas utilizando amostras de entrada-saída rotuladas no treinamento, sendo aplicado em tarefas de classificação e regressão. Algoritmos como *Support Vector Machines*, regressão linear e árvores de decisão são comumente utilizados. O aprendizado semisupervisionado combina métodos supervisionados e não supervisionados, empregando dados rotulados e não rotulados para melhorar o desempenho da classificação. No aprendizado não supervisionado, programas de computador identificam padrões e relações nos dados sem que haja uma definição externa ou rótulos pré-estabelecidos. É comumente empregado em agrupamento, isto é, para revelar correlações implícitas ou não evidentes nos dados de entrada, de modo a constituir subconjuntos ou grupos que compartilham características ou propriedades semelhantes. Já o aprendizado por reforço é guiado pelo ambiente e baseia-se em recompensas e penalidades, aprendendo a partir de acertos e erros (ALSULIMAN et al., 2020).

O DL é uma técnica de aprendizado de máquina que utiliza sistemas computacionais baseados em redes neurais para identificar correlações nos dados por meio de testes evolutivos, com o objetivo de minimizar uma função de custo. O processo começa com valores aleatórios e ajusta os pesos até atingir a previsão mais precisa possível, permitindo que o sistema aprenda

continuamente a partir dos dados de entrada. Embora seja uma ferramenta poderosa para resolver problemas cognitivos complexos, a eficácia do aprendizado profundo pode ser limitada por desafios relacionados à escassez, volume insuficiente e baixa qualidade dos dados (ALSULIMAN et al., 2020).

Diferentemente dos algoritmos tradicionais de aprendizado de máquina, que dependem de classificadores previamente selecionados pelo usuário, os algoritmos de aprendizado profundo podem ser aplicados diretamente a conjuntos de dados brutos, sem necessidade de escolha prévia de características. Esses algoritmos são considerados autodidatas porque identificam e criam seus próprios classificadores a partir dos dados fornecidos, necessitando apenas dos dados brutos e da definição da saída desejada (PANESAR et al., 2020).

As previsões sobre o papel inovador da IA na área da saúde não são novas, com os primeiros relatos na década de 1950. No entanto, a velocidade de computação, memória, compactação, custo e as capacidades algorítmicas melhoraram extraordinariamente desde então, assim como a disponibilidade de dados clínicos digitais. Essa rápida taxa de desenvolvimento torna desafiador o planejamento do papel futuro da IA (ALSULIMAN et al., 2020).

A análise de grandes quantidades de dados médicos com velocidade, acurácia e precisão tornou as ferramentas de IA e sistemas automatizados baseados em IA amplamente utilizados por profissionais de saúde e integrados ao cuidado clínico do paciente. A maioria destes profissionais hoje em dia usa algum tipo de ferramenta de IA para todas as fases do diagnóstico e gerenciamento do paciente (SAILUNAZ et al., 2023).

2.4 ÁRVORES DE DECISÃO E FLORESTAS ALEATÓRIAS

Árvore de decisão é uma técnica de aprendizado supervisionado, utilizada principalmente em tarefas de classificação; mas também pode ser usada para regressão, quando o objetivo é prever um valor numérico contínuo. Começa com um ponto principal, o primeiro passo de decisão para dividir o conjunto de dados, e contém um único recurso que melhor divide os dados em suas respectivas classes. Cada divisão tem uma aresta que se conecta a um novo ponto de decisão que contém outro recurso para dividir ainda mais os dados em grupos homogêneos ou a um nó terminal que prevê a classe. Esse processo de separar os dados em duas partições binárias é conhecido como particionamento recursivo, representada na figura 1 (CHOI et al., 2020).

Uma floresta aleatória é uma extensão desse método, conhecido como método de conjunto, que produz múltiplas árvores de decisão. Em vez de usar todos os recursos para criar todas as árvores de decisão em uma floresta aleatória, uma subamostra de recursos é usada para criar cada árvore de decisão. As árvores então preveem um resultado de classe, e o voto majoritário entre as árvores é usado como a previsão final da classe do modelo (CHOI et al., 2020). É conhecida também por sua eficácia em problemas de classificação e regressão, além de fornecer uma estimativa de importância das características (ARAÚJO, 2024).

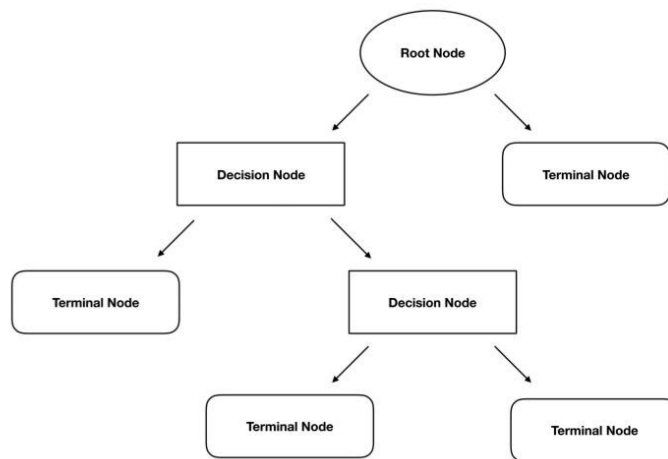


Figura 1 - Estrutura de uma árvore de decisão.

Fonte: CHOI et al., Introduction to *Machine Learning*, Neural Networks, and Deep Learning, 2020.

2.5 VENTILAÇÃO MECÂNICA INVASIVA E NÃO INVASIVA

A ventilação não invasiva (VNI) é definida como uma modalidade de assistência ventilatória sem o uso de uma prótese endotraqueal invasiva. Sua principal vantagem é evitar complicações associadas à intubação e à Ventilação Mecânica Invasiva (VMI). Ao evitar a intubação, a VNI deixa as vias aéreas superiores pervias, conservando as defesas das vias aéreas e, durante os intervalos, permite que os pacientes comam, vocalizem normalmente e expectorem secreções das vias aéreas. A principal desvantagem da VNI é o retardo da intubação nos pacientes nos quais ocorre falha da terapia, precoce ou tardiamente, causando aumento de morbimortalidade. A VNI pode ser aplicada utilizando diferentes interfaces e modalidades de ventilação, resultando em diferentes efeitos fisiológicos (AMIB, SBPT 2024).

Pode ser utilizada com um nível de pressão, CPAP (Continuous Positive Airway Pressure) e com dois níveis de pressão BiPAP ou BiLevel. O CPAP consiste na aplicação de uma pressão positiva contínua às vias aéreas tanto na inspiração quanto na expiração. Sua aplicação pode gerar diversos efeitos fisiológicos positivos sobre a troca gasosa, sobretudo na IRpA hipoxêmica, como melhora da troca gasosa, diminuição do esforço respiratório e melhora

da expansibilidade pulmonar. O BiPAP na oferta de pressão positiva com dois níveis de pressão: um maior na fase inspiratória, com disparo a partir do esforço muscular respiratório do paciente; e um menor na fase expiratória. Esse diferencial de pressões em sincronia com as fases do ciclo respiratório proporciona incremento de fluxo e volume corrente na inspiração, com impacto no trabalho muscular respiratório e na ventilação alveolar (AMIB, SBPT 2024).

A VMI, é ofertada através de um tubo endotraqueal ou traqueostomia, indicada na insuficiência respiratória aguda quando há hipoxemia grave, fadiga neuromuscular à demanda ventilatória, colapso circulatório, choque ou parada cardiorrespiratória e incapacidade de manter a permeabilidade da via aérea superior (AMIB, SBPT 2024).

As principais indicações de VMI incluem situações de falência da oxigenação, como hipoxemia grave ($\text{PaO}_2 < 60 \text{ mmHg}$ ou $\text{SaO}_2 < 90\%$ mesmo com $\text{FiO}_2 > 0,6$), e falência da ventilação, caracterizada por parada cardíaca ou respiratória iminente, taquipneia grave ($\text{FR} > 40 \text{ irpm}$) ou bradipneia significativa ($< 5 \text{ irpm}$), hipercapnia associada a escore de coma de Glasgow < 8 , obstrução da via aérea ou presença de doenças neuromusculares. A VMI também é indicada quando há incapacidade de proteção de via aérea, como em casos de coma (Glasgow < 8) ou hipersecreção importante das vias aéreas. Por outro lado, suas principais contraindicações incluem situações em que há benefício com o uso de VNI ou quando existe indicação de não intubação (AMIB, SBPT 2024).

De acordo com as orientações práticas de ventilação mecânica publicadas pela Associação de Medicina Intensiva Brasileira (AMIB) e pela Sociedade Brasileira de Pneumologia e Tisiologia (SBPT) em 2024, a intubação orotraqueal em pacientes com COVID-19 deve ser indicada quando houver falha das modalidades de suporte ventilatório não invasivo. Essa recomendação busca evitar atrasos na instituição da ventilação mecânica invasiva, uma vez que a persistência do esforço respiratório elevado está associada ao aumento da mortalidade hospitalar.

3. OBJETIVOS

3.1 OBJETIVO GERAL

Aplicar algoritmos de aprendizado de máquina, mais precisamente árvores de decisão, a dados clínicos de pacientes que tiveram diagnóstico de COVID-19, com intuito de predizer desfechos clínicos relevantes.

3.2 OBJETIVOS ESPECÍFICOS

- Pré-processar e organizar o banco de dados clínicos de pacientes que foram internados com COVID-19.
- Selecionar variáveis relevantes para os desfechos clínicos de interesse.
- Aplicar algoritmos de classificação, como J48, no Weka para prever os desfechos selecionados.
- Interpretar a estrutura das árvores de decisão geradas, destacando as variáveis com maior importância preditiva e sua aplicabilidade clínica.
- Comparar os resultados obtidos com evidências científicas, validando os achados com base em estudos sobre COVID-19 e seu prognóstico.

4. CAPÍTULO II: ARTIGO

Aplicação de algoritmos de *machine learning* para análise e predição de desfechos clínicos em pacientes com Covid-19

Bruno Fonseca Rezende^[1], Dr. Laurence Rodrigues do Amaral^[2]

[1] Programa de Pós-Graduação em Biotecnologia – PPGBIOTEC, Universidade Federal de Uberlândia (UFU), Uberlândia, Brasil.

[2] Professor do Programa de Pós-Graduação em Biotecnologia – PPGBIOTEC, Universidade Federal de Uberlândia (UFU), Uberlândia, Brasil.

[1] bruno.rezende@ufu.br

[2] laurence@ufu.br

Resumo A pandemia de COVID-19 impôs desafios significativos aos sistemas de saúde, especialmente pela rápida evolução clínica e pela necessidade de decisões imediatas sobre suporte ventilatório e internação em UTI, revelando a lacuna no uso de ferramentas objetivas capazes de auxiliar a estratificação precoce do risco e a previsão de prognóstico em pacientes hospitalizados. Nesse contexto, modelos de Aprendizado de Máquina (ML), como árvores de decisão, apresentam potencial para apoiar a tomada de decisão clínica por permitirem analisar múltiplas variáveis simultaneamente e gerar resultados interpretáveis. O objetivo deste estudo foi aplicar algoritmos de ML a dados clínicos de pacientes com COVID-19 a fim de prever desfechos relevantes e identificar fatores associados à gravidade da doença. Trata-se de um estudo retrospectivo e quantitativo, baseado em dados secundários provenientes do banco público de Síndrome Respiratória Aguda Grave (SRAG) da cidade de Londrina (Paraná), contendo inicialmente 15.655 registros de pacientes hospitalizados entre janeiro de 2021 e fevereiro de 2022. Após filtragem e organização das variáveis, foram utilizados 5.704 registros no modelo de desfecho e 7.182 nos modelos complementares. As análises foram realizadas no software WEKA (v.3.8.6), utilizando o algoritmo J48 com validação cruzada, após seleção e eliminação de atributos incompletos. Os modelos desenvolvidos alcançaram acurácia aproximada de 80% na classificação dos casos. O tipo de suporte ventilatório emergiu como principal preditor tanto na etapa de triagem quanto na previsão do desfecho final, seguido por internação em UTI, presença de doença cardíaca, idade e tipo de hospital. Os resultados demonstraram que a ventilação mecânica invasiva e a internação em UTI estiveram fortemente associadas ao óbito, enquanto o uso de ventilação não invasiva e a ausência de suporte ventilatório relacionaram-se à cura. A idade avançada intensificou o risco de mortalidade nos pacientes sob ventilação invasiva, e o tipo de hospital influenciou o prognóstico entre os pacientes não admitidos em UTI, com maior mortalidade em hospitais públicos. Conclui-se que as árvores de decisão mostraram-se eficazes na identificação de fatores preditores de gravidade em pacientes hospitalizados com COVID-19, oferecendo

modelos de fácil interpretação clínica e potencial aplicação na estratificação de risco, no direcionamento de recursos e na organização do cuidado em ambiente hospitalar, especialmente em situações de sobrecarga assistencial como a vivenciada durante a pandemia.

Palavras-chave: aprendizado de máquina; árvore de decisão; COVID-19; predição de desfechos clínicos; UTI; ventilação mecânica.

Abstract The COVID-19 pandemic imposed substantial challenges on healthcare systems worldwide, particularly due to the rapid clinical deterioration observed in many patients and the urgent need to allocate ventilatory support and intensive care resources efficiently, highlighting a gap in objective tools capable of assisting early risk stratification and outcome prediction. In this context, Machine Learning (ML) models, such as decision trees, offer the potential to support clinical decision-making by analyzing multiple variables simultaneously and generating easily interpretable predictive structures. This study aimed to apply ML algorithms to clinical data from hospitalized patients with COVID-19 to predict relevant clinical outcomes and identify factors associated with disease severity. This retrospective, quantitative study used secondary data from the public Severe Acute Respiratory Infection (SARI) database of Londrina, Paraná, Brazil, which initially contained 15,655 records of hospitalized patients between January 2021 and February 2022. After data cleaning and preprocessing, 5,704 records were used for the outcome prediction model and 7,182 for complementary analyses. Data were analyzed using the WEKA software (v.3.8.6) with the J48 decision tree algorithm and cross-validation, following attribute selection and removal of redundant or incomplete variables. The generated models achieved approximately 80% accuracy. Type of ventilatory support emerged as the most relevant predictor across analyses, followed by ICU admission, preexisting heart disease, age, and hospital type. Results showed that invasive mechanical ventilation and ICU admission were strongly associated with mortality, whereas non-invasive ventilation or absence of ventilatory support were associated with recovery. Advanced age substantially increased mortality risk among patients receiving invasive ventilation, and hospital type influenced outcomes among those not admitted to the ICU, with higher mortality in public hospitals. The study concludes that decision tree models are effective for identifying predictors of severity in hospitalized COVID-19 patients, providing clinically interpretable structures with potential application for risk stratification, resource allocation, and management of critical care pathways, especially in contexts of healthcare overload such as that experienced during the pandemic.

Keywords: decision tree; intensive care unit; machine learning; mechanical ventilation; clinical outcomes prediction.

4.1 Introdução

A COVID-19, causada pelo Coronavírus, foi identificada pela primeira vez em Wuhan, na China, no final de 2019. Trata-se de uma doença altamente contagiosa que afeta principalmente o sistema respiratório humano, sendo transmitida predominantemente por meio de gotículas respiratórias durante o contato entre as pessoas. Os sintomas incluem falta de ar, febre, tosse e perda de paladar, podendo ocorrer também casos assintomáticos [1]. A rápida disseminação global da doença resultou em significativa pressão sobre os sistemas de saúde, com mais de 280 milhões de infecções e 5,4 milhões de mortes até dezembro de 2021 [2].

Embora aproximadamente 80% dos indivíduos infectados apresentem sintomas leves, cerca de 20% evoluem com necessidade de hospitalização, podendo desenvolver insuficiência respiratória aguda e demandar suporte ventilatório não invasivo ou ventilação mecânica invasiva. Entre os pacientes críticos, observa-se elevada mortalidade, além de maior tempo de recuperação e risco aumentado de desenvolvimento de síndrome pós-COVID, condição associada a comprometimento multissistêmico e complicações a longo prazo [1,2].

A apresentação clínica heterogênea da COVID-19 dificulta a identificação precoce dos pacientes com maior risco de agravamento. Fatores como idade avançada, sexo masculino, obesidade, comorbidades e alterações laboratoriais, incluindo Proteína C-Reativa e dímero-D, têm sido associados a desfechos desfavoráveis, como internação em Unidade de Terapia Intensiva, necessidade de ventilação mecânica e óbito. Entretanto, estudos iniciais apresentaram limitações metodológicas, como amostras reduzidas e análises baseadas em grande número de variáveis, o que restringe a aplicabilidade clínica dos achados [10].

Nesse contexto, modelos baseados em inteligência artificial, especialmente técnicas de Machine Learning (ML), têm sido amplamente utilizados para apoiar a estratificação de risco e a tomada de decisão clínica. Quando treinados com conjuntos de dados abrangentes e criteriosamente selecionados, esses modelos são capazes de identificar padrões relevantes, construir ferramentas preditivas robustas e gerar estimativas mais precisas do risco de agravamento clínico. Entre essas abordagens, destacam-se as árvores de decisão e as florestas aleatórias, por sua interpretabilidade e capacidade de lidar com múltiplas variáveis clínicas simultaneamente [3].

Diante desse cenário, o desenvolvimento de modelos preditivos baseados em ML para identificar precocemente pacientes com maior risco de evolução desfavorável e necessidade de suporte ventilatório torna-se fundamental para otimizar o manejo clínico, reduzir atrasos na indicação de ventilação mecânica e promover a utilização mais eficiente dos recursos disponíveis nos serviços de saúde.

4.3 Objetivos

4.3.1 Objetivo Geral

Aplicar algoritmos de aprendizado de máquina, mais precisamente árvores de decisão, a dados clínicos de pacientes que tiveram diagnóstico de COVID-19, com intuito de prever desfechos clínicos relevantes

4.3.2 Objetivos Específicos

- Pré-processar e organizar o banco de dados clínicos de pacientes que foram internados com COVID-19.
- Selecionar variáveis relevantes para os desfechos clínicos de interesse.
- Aplicar algoritmos de classificação, como J48, no Weka para prever os desfechos selecionados.
- Interpretar a estrutura das árvores de decisão geradas, destacando as variáveis com maior importância preditiva e sua aplicabilidade clínica.
- Comparar os resultados obtidos com evidências científicas, validando os achados com base em estudos sobre COVID-19 e seu prognóstico

4.4 Metodologia

Esse é um estudo retrospectivo, quantitativo, com análise de dados secundários coletados durante a pandemia de COVID-19, de pacientes hospitalizados, registrados no banco de dados de infecções respiratórias agudas graves, coletados em janeiro de 2021 a fevereiro de 2022, na cidade de Londrina - Paraná – Brasil.

Para o levantamento bibliográfico, foi realizada uma busca na base de dados PubMed, utilizando as palavras-chave “decision tree; intensive care unit; machine learning; mechanical ventilation; clinical outcomes prediction.”. Foram selecionados 31 artigos considerados relevantes para o desenvolvimento do tema, publicados no período de 2019 até os dias atuais.

O dataset utilizado neste estudo foi gentilmente disponibilizado pelo pesquisador Hemanoel Passarelli Araújo [19], cujo trabalho integra as referências deste projeto de pesquisa. O contato foi realizado por meio de correio eletrônico (e-mail: hemanoell.passarelli@gmail.com) em 1º de novembro de 2024, tendo sido prontamente respondido em 6 de novembro do mesmo ano. Na ocasião, o pesquisador demonstrou disponibilidade e cordialidade ao compartilhar seu dataset, reconhecendo a relevância da colaboração para o desenvolvimento deste estudo.

Esta pesquisa não necessitou de submissão ao Comitê de Ética e Pesquisa, por se tratar de uma análise de dados secundários provenientes de um dataset previamente disponibilizado por outro pesquisador. O conjunto de dados utilizado já havia sido anonimizado e tornado público para fins científicos, não sendo possível a identificação dos indivíduos. Dessa forma, não houve risco à integridade ou à privacidade dos participantes.

O conjunto de dados contém informações sobre 12 comorbidades: asma, diabetes, síndrome de Down, doença cardíaca, doença hematológica, doença hepática, imunossupressão, doença renal, doença pulmonar, doença neurológica, obesidade e puerpério. Todas foram utilizadas no trabalho, os registros que contém “NA”, “ignored” e “óbito por outras causas” foram excluídos da pesquisa, onde foram interpretados como incompletos e não adequados à análise proposta. O dataset possui um total de 15.655 indivíduos. As linhas representam pacientes e colunas as variáveis avaliadas no estudo.

As colunas utilizadas como classe neste estudo foram: suporte ventilatório (invasivo, não invasivo, ausente), situação legal do hospital (1 = público, 2 = privado e 3 = instituição sem fins lucrativos), admissão em UTI (0 = sim ou 1 = não), comorbidades (0 = sim, 1 = não), e desfecho (considerado como óbito por COVID-19 e cura).

Inicialmente, foram criados 3 datasets distintos – Outcome, idade e tempo de internação - que foram salvos em arquivos no formato CSV, realizada uma triagem e limpeza dos dados com o objetivo de melhorar a qualidade e a consistência das informações utilizadas nas análises.

Foram excluídas do conjunto de dados as seguintes variáveis: age group, dt. outcome, total symptom, total morb, dt. notify, dt. symptom, mun. hospitalization, munic residence, dt. hospitalization, dt. admission, dt. discharge, is residence e UTI imp. A exclusão dessas colunas foi devido à presença de informações redundantes e irrelevantes para o modelo proposto ou por conterem alto índice de dados ausentes.

O dataset “Outcome” foi composto por 5.704 registros, após a remoção de 9.951 linhas com informações inconsistentes e/ou incompletas para a pesquisa. Nos datasets “idade” e “tempo de internação”, foram excluídas 8.473 linhas, resultando em 7.182 registros utilizados na análise.

Para tornar o modelo mais realista e clínico, o banco de dados foi estruturado em duas etapas analíticas complementares, denominadas Triagem e Desfecho:

- Triagem: corresponde à análise inicial de pacientes internados com COVID-19, com foco na caracterização do perfil clínico-epidemiológico e na identificação dos fatores associados à necessidade de suporte ventilatório (invasivo, não invasivo ou ausente). Nesta etapa, foram utilizadas variáveis relacionadas à presença de comorbidades e tipo de internação, buscando reproduzir o processo de decisão clínica na admissão hospitalar.
- Desfecho: após a triagem inicial, realizou-se uma segunda análise destinada à predição do desfecho clínico final (cura ou óbito por COVID-19). Nessa etapa, foram consideradas variáveis previamente filtradas na fase de triagem, além de informações referentes ao tipo de suporte ventilatório, idade, admissão em UTI e tempo de internação.

Essa separação permitiu criar modelos de aprendizado de máquina distintos, refletindo as etapas reais do cuidado hospitalar, identificando pacientes com maior risco óbito entre os casos mais graves.

Para a análise dos dados no software WEKA versão 3.8.6, inicialmente, o software foi aberto e, na interface principal, foi selecionada a opção Explorer, que dá acesso às funcionalidades de pré-processamento, classificação, agrupamento e avaliação de dados. Em seguida, o conjunto de dados previamente preparados foi carregado por meio da opção Open file, localizada na aba de pré-processamento. Após o carregamento, realizou-se o pré-processamento dos dados, com a seleção das variáveis de interesse que seriam utilizados na construção das árvores de decisão. Essa etapa foi fundamental para garantir que apenas as informações relevantes fossem analisadas, otimizando o desempenho do modelo e facilitando a interpretação dos resultados.

Com os dados devidamente preparados, acessou-se a aba Classify para dar início ao processo de classificação. Na seção Test options, foi selecionada a opção use training set, indicando que o próprio conjunto de dados seria utilizado para o treinamento e teste do modelo. No campo Classify, clicou-se em Choose e, dentro da categoria Trees, foi selecionado o algoritmo J48, uma implementação do algoritmo de árvore de decisão, reconhecido pela sua robustez e ampla utilização em problemas de classificação. Após confirmar a correta seleção do algoritmo, foi acionada a opção Start, dando início à construção e avaliação do modelo.

A avaliação do desempenho do modelo foi realizada por meio de validação cruzada, mais precisamente o Leave-One-Out Cross Validation (LOOCV), garantindo maior robustez e generalização dos resultados. Após a execução, o modelo gerado foi analisado com base na estrutura da árvore, nas métricas de desempenho (acurácia, sensibilidade, especificidade) e na matriz de confusão.

A árvore de decisão resultante permitiu identificar os principais atributos preditores associados ao desfecho de interesse, além de facilitar a interpretação clínica por sua estrutura lógica e hierárquica.

4.5 Resultados

Foram desenvolvidos dois modelos independentes de ML e análise estatística, triagem, destinado a identificar fatores preditores da necessidade de suporte ventilatório. As variáveis mais relevantes foram presença de doença cardíaca e admissão em UTI, alcançando acurácia aproximada de 80,2%. O outro modelo foi desfecho, voltado para prever o resultado clínico (cura ou óbito por COVID-19). Neste modelo, as variáveis com maior peso foram tipo de suporte ventilatório, internação em UTI e tipo de hospital, com acurácia média de 80%.

A separação entre triagem e desfecho permitiu avaliar o comportamento das variáveis em momentos distintos da hospitalização, melhorando a interpretabilidade clínica e a aplicabilidade prática dos modelos no contexto hospitalar.

Com base nas imagens das árvores de decisão geradas no software WEKA, utilizando o algoritmo J48, foi possível construir modelos preditivos para o desfecho clínico dos pacientes internados com COVID-19. A seguir, são descritas as análises das duas árvores geradas:

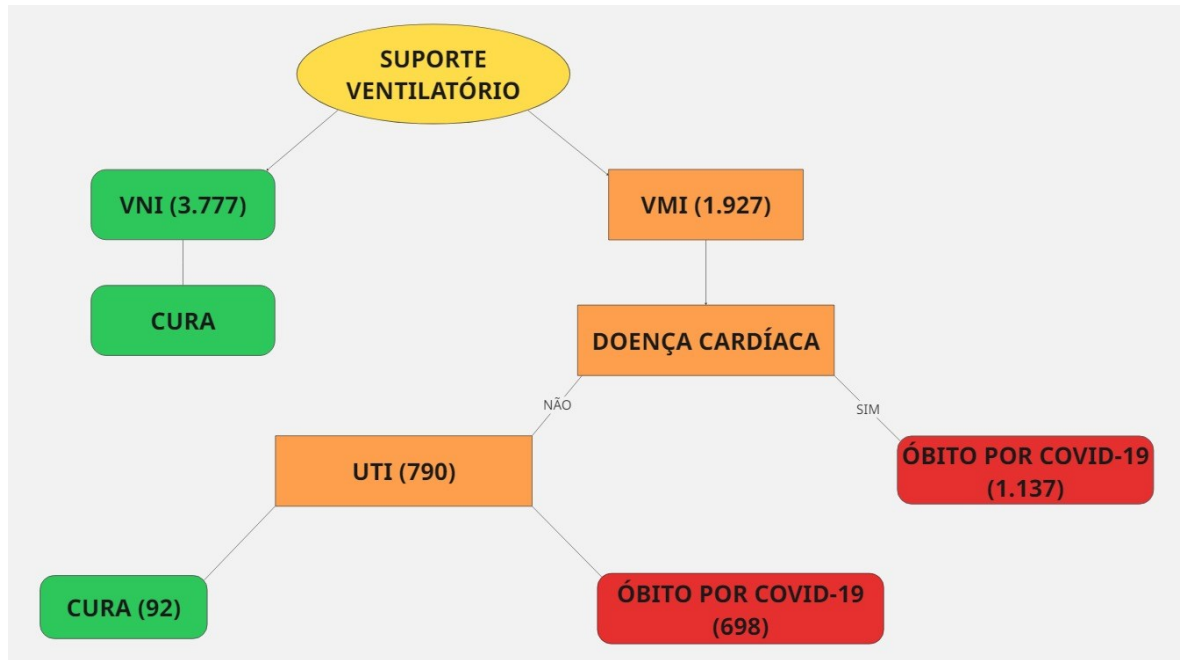


Figura 2: Árvore de decisão 1 – Ventilação de Suporte, doença cardíaca, UTI e desfecho.
 Legenda: valores entre parênteses: casos previstos; cura e óbito: desfecho predominante.

Na primeira árvore de decisão (Figura 2), construída a partir do dataset “Outcome”, composto por 5.704 registros e quatro atributos avaliados (suporte ventilatório, doença cardíaca, UTI e desfecho), observou-se uma acurácia de 80,2% na classificação dos dados. O suporte ventilatório foi identificado como o principal fator de separação entre os casos.

Os pacientes que receberam VNI foram, em sua maioria, classificados com o desfecho de cura (3.777 casos), correspondendo a 66,2% dos pacientes tratados com VNI. Já entre os pacientes submetidos à VMI, o principal fator de separação foi a presença de doença cardíaca. Dentre esses, 59% dos pacientes com doença cardíaca evoluíram para óbito por COVID-19. Para os pacientes intubados sem doença cardíaca, o próximo fator discriminante foi a permanência na UTI, sendo que 11,6% apresentaram cura, enquanto 88,3% foram a óbito.

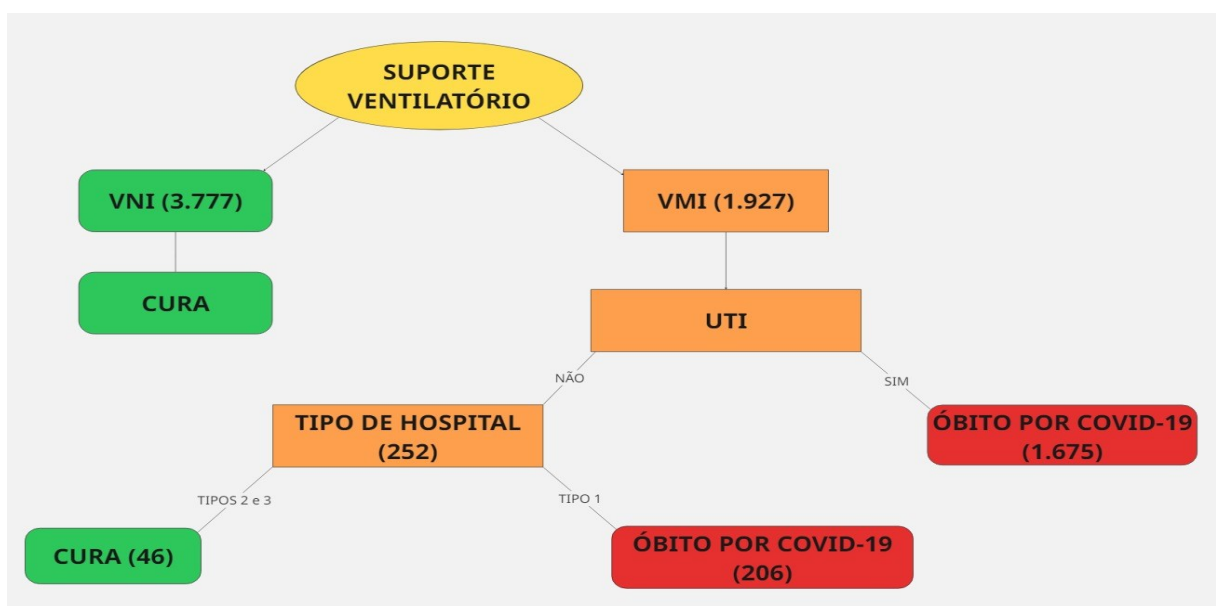


Figura 3: Árvore de decisão 2 – ventilação de suporte, UTI, tipo de hospital e desfecho.
 Legenda: valores entre parênteses: casos previstos; cura e óbito: desfecho predominante; tipo de hospital: 1 = público; 2 = privado; 3 = sem fins lucrativos.

Nessa árvore de decisão número 2 (Figura 3), gerada no dataset “Outcome”, contendo 4 atributos (suporte ventilatório, UTI, tipo de hospital e desfecho), com 80,2% dos dados classificados como corretamente, identificou-se o tipo de suporte ventilatório como o principal fator de classificação para o desfecho clínico dos pacientes. Aqueles que utilizaram VNI apresentaram maior probabilidade de cura, com 3.777 casos, ou seja, em 66,2% dos pacientes que usaram VNI tiveram cura, sendo 1.927 (33,7%) pacientes foram intubados e receberam VMI.

Para os pacientes submetidos a VMI, a variável subsequente mais relevante foi a admissão em UTI, onde pacientes que não foram admitidos na UTI foram então classificados com base no tipo de hospital.

Em hospitais classificados tipo 1 (hospital público), observou-se maior ocorrência de óbitos por COVID-19, com 206 casos classificados (81,7%). Já em hospitais tipo 2 e 3, predominou o desfecho de cura, 46 casos (18,3%).

Por outro lado, pacientes que utilizaram ventilação invasiva e foram admitidos na UTI apresentaram maior taxa de mortalidade, com 1.675 óbitos classificados (86,9%).

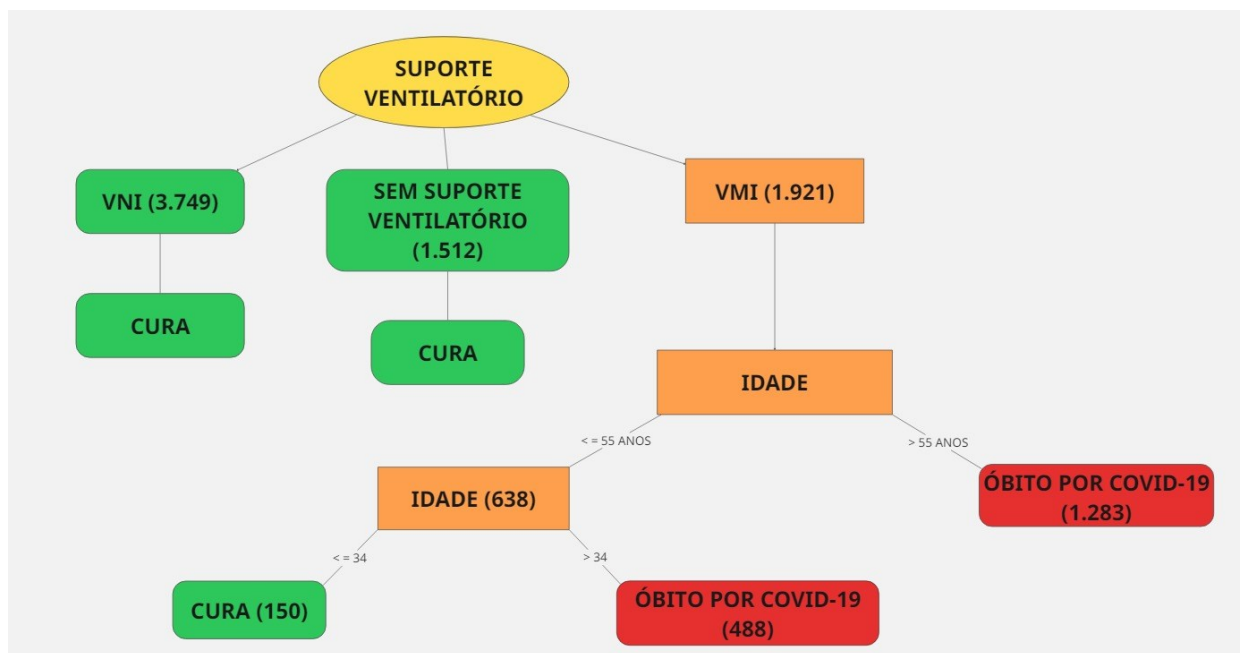


Figura 4: Árvore de decisão 3 – ventilação de suporte, idade e desfecho.

Legenda: valores entre parênteses: casos previstos; cura e óbito: desfecho predominante.

A terceira árvore de decisão gerada (Figura 4), através do dataset “age, cura, COVID-19, número total de 7.182 dados, três atributos (suporte ventilatório, idade e desfecho), identificou o tipo de suporte ventilatório como a principal variável determinante do desfecho clínico. Pacientes que utilizaram VNI foram predominantemente classificados com desfecho de cura 52,1 % (3.749 casos), assim como aqueles que não necessitaram de suporte ventilatório, 21,05% (1.512 casos). Entre os pacientes submetidos à VMI; 26,7% (1.921 casos), a variável idade emergiu como o segundo fator decisivo mais importante. Indivíduos com idade superior a 55 anos apresentaram maior probabilidade de óbito por COVID-19; 66,7% (1.283 casos). No grupo invasivo com idade menor ou igual a 55 anos, observou-se nova subdivisão: pacientes ≤ 34 anos foram majoritariamente classificados como cura; 23,5% (150 casos), enquanto aqueles com idade entre 35 e 55 anos apresentaram prevalência de óbito; 76,4% (488 casos). Esses achados demonstram a forte interação entre o tipo de suporte ventilatório e a idade, evidenciando que a VMI associada ao avanço da idade aumenta substancialmente o risco de mortalidade.

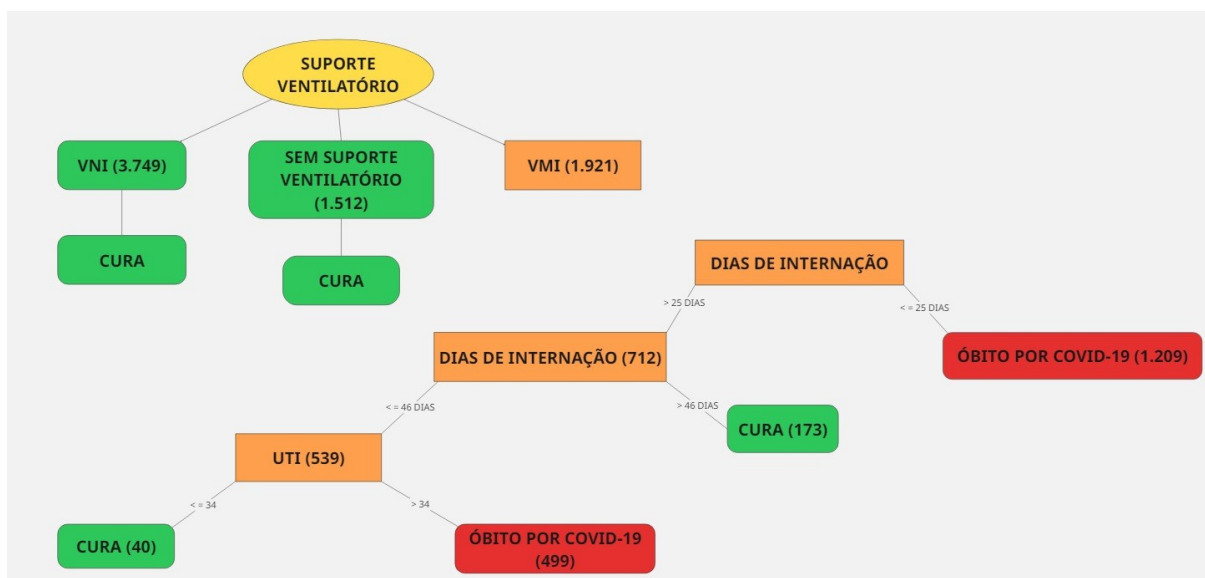


Figura 5: Árvore de decisão 4 – ventilação de suporte, tempo de internação e desfecho.
Legenda: valores entre parênteses: casos previstos; cura e óbito: desfecho predominante.

Os resultados obtidos na análise da árvore de decisão nº4 (Figura 5), demonstraram que o tipo de suporte ventilatório foi a variável mais determinante para o desfecho clínico. Pacientes que não necessitaram de suporte ventilatório (21,05%) ou que receberam VNI (52,1%) apresentaram predominância de cura, enquanto aqueles submetidos à VMI apresentaram maior risco de óbito. Entre os pacientes sob VMI (26,7%), observou-se que o tempo de internação até o desfecho e a necessidade de internação em UTI influenciaram significativamente o prognóstico. Pacientes que evoluíram para desfecho em até 25 dias de internação apresentaram maior probabilidade de óbito (62,9%), refletindo a gravidade do quadro e a falha da VMI. Por outro lado, indivíduos com internação prolongada, especialmente acima de 46 dias (24,2%), apresentaram maior taxa de cura, sugerindo recuperação progressiva após as fases mais críticas da doença. Além disso, entre aqueles com tempo intermediário de internação, entre 26 e 46 dias, a necessidade de UTI esteve associada a maior mortalidade (92,5%) e 7,42% de cura, reforçando o papel da gravidade clínica e do suporte intensivo como fatores preditores de pior desfecho.

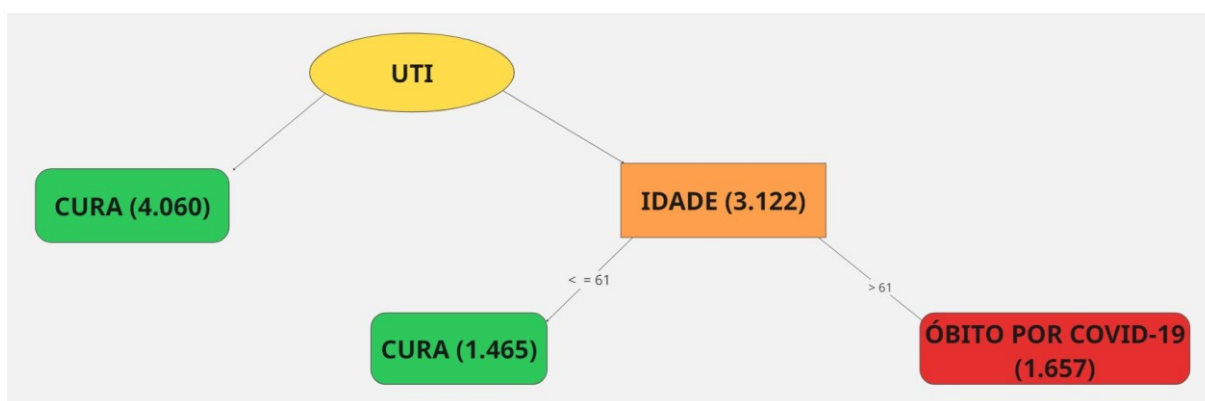


Figura 6: Árvore de decisão 5 – UTI, idade e desfecho.
Legenda: valores entre parênteses: casos previstos; cura e óbito: desfecho predominante.

A análise da árvore de decisão nº5 (Figura 6) mostrou que a internação em UTI foi o principal fator determinante do desfecho clínico. Pacientes que não necessitaram de cuidados intensivos apresentaram predominância de cura (56,5%), indicando menor gravidade e melhor evolução clínica. Entre aqueles internados na UTI, a idade foi identificada como a variável mais relevante para diferenciação dos resultados (43,4%). Indivíduos com até 61 anos tiveram maior probabilidade de recuperação (46,9%), enquanto os com idade superior a esse limite apresentaram maior risco de óbito (53,07%). Esses achados demonstram que a gravidade do quadro, refletida pela necessidade de cuidados intensivos, associada ao avanço da idade, exerce influência significativa sobre a mortalidade e a probabilidade de cura em pacientes hospitalizados com COVID-19.

4.6 Discussão

De acordo com a primeira árvore de decisão (Figura 2), 81% dos pacientes que apresentaram doença cardíaca e que foram intubados morreram, corroborando com o estudo de [20], conduzido como uma série de casos retrospectivos, onde analisou 5700 pacientes hospitalizados com COVID-19 em 12 hospitais do sistema Northwell Health, em Nova York, identificaram uma elevada taxa de mortalidade entre pacientes que necessitaram de VMI e apresentavam doenças cardíacas, sustentando a relevância clínica das variáveis selecionadas pela árvore de decisão construída.

Na revisão sistemática e meta-análise conduzida por [21], numa tentativa de coletar e avaliar sistematicamente as associações de fatores epidemiológicos e de comorbidade com a gravidade e o prognóstico da COVID-19 na população em geral, onde foi considerado com maior possibilidade de pior prognóstico, a admissão em UTI e o uso de VMI.

Já o estudo de [22], mostrou que indivíduos com comorbidades apresentaram a forma mais grave da doença, com maior indicação de internação em UTI, além da associação entre o uso de suporte ventilatório e internação em UTI para pacientes com doenças cardiovasculares e infecção confirmada para COVID-19, validando os desfechos deste estudo.

Segundo [23], doenças crônicas podem ser apontadas como fatores de risco à infecção por COVID-19 em função de sua predisposição à maior morbimortalidade associada. Sendo assim, indivíduos com doenças cardiovasculares prévias podem estar mais susceptíveis a quadros graves da infecção, considerando a fragilidade do sistema de cada indivíduo, favorecendo assim a ação potencial do vírus. Antes mesmo do surgimento da pandemia, as doenças cardiovasculares já eram condições clínicas frequentes em diagnósticos de SRAG, podendo elevar em até doze vezes o risco de mortalidade associada.

Embora os óbitos entre os participantes não tenham sido avaliados, um estudo desenvolvido por [24], em Wuhan, na China, demonstrou que a mortalidade por COVID-19 em pacientes com doenças cardiovasculares foi mais elevada (22,2%) quando comparada à amostra total analisada (9,8%).

Em uma revisão sistemática realizada por [25], aponta que comorbidades como hipertensão arterial, diabetes mellitus tipo 2 e doença arterial coronariana foram altamente prevalentes entre os pacientes internados, além de estarem associadas a maior risco de complicações e óbito. Além da elevada frequência dessas condições, o estudo destaca que a presença de doença cardiovascular preexistente contribui para a gravidade da COVID-19 por meio de múltiplos mecanismos fisiopatológicos, entre eles, destacam-se a resposta inflamatória exacerbada, a disfunção endotelial, a instabilidade de placas ateroscleróticas, a hipóxia causada pela Síndrome do Desconforto Respiratório Agudo que reduz a entrega de oxigênio ao miocárdio, podendo causar lesão e isquemia miocárdica especialmente em pessoas com doença arterial coronariana; e o estresse metabólico, todos esses fatores aumentam a sobrecarga cardíaca. Lesão miocárdica, insuficiência cardíaca e arritmias foram complicações comuns, reforçando a vulnerabilidade desse perfil de paciente.

Em relação ao suporte ventilatório, [26], destacam os aspectos fisiológicos da ventilação mecânica na COVID-19, na qual a ventilação mecânica invasiva seja essencial em casos graves, seu uso pode contribuir para piora do quadro clínico quando não há adaptação adequada aos parâmetros respiratórios individualizados para cada perfil de paciente, favorecendo a ocorrência de Lesão Pulmonar Induzida pela Ventilação (VILI).

Descrevem também a existência de diferentes perfis pulmonares na COVID-19, variando desde pacientes com complacência preservada, classificada como fenótipo L, nos quais a ventilação não invasiva pode ser eficaz, até casos mais graves com complacência reduzida e consolidação pulmonar acentuada, nomeada de fenótipo H, que frequentemente requerem ventilação invasiva. Nestes casos, os riscos de lesão pulmonar são maiores, principalmente com o uso de pressões elevadas ou volumes correntes inadequados. Apontam também que altos níveis de esforço inspiratório, mesmo durante suporte não invasivo, podem gerar Lesão Pulmonar Autoinfligida (P-SILI), comprometendo ainda mais o prognóstico.

Em um estudo conduzido por [27] identificaram que os pacientes que apresentaram falha na VNI tiveram uma taxa de mortalidade elevada, e esse perfil de pacientes tinham características de comprometimento multiorgânico contínuo, expressão de doença sistêmica, não relacionada apenas ao pulmão. [28], descreve que o comprometimento multiorgânico precoce devido à doença COVID-19 já estava presente na admissão à UTI e posteriormente piorou durante a internação na UTI, principalmente em não sobreviventes.

A segunda árvore de decisão (Figura 3) identificou também como principal fator de classificação para o desfecho clínico, o suporte ventilatório, admissão na UTI e outro dado importante, foi em relação aos pacientes que não foram admitidos na UTI, levando-se em consideração o tipo de hospital como pior desfecho, classificado como 1, que são hospitais públicos.

[13], destacou-se sobre a estrutura hospitalar do SUS antes e durante a Pandemia de COVID-19, revelou-se insuficiente, mal distribuída geograficamente, com limitações na estrutura física e operacional e afetada por subfinanciamento crônico. Além disso, evidenciaram que muitos hospitais considerados operantes apresentavam baixa taxa de ocupação real, infraestrutura precária, insuficiência de insumos e escassez de força de trabalho qualificada, especialmente em municípios menores. Essa fragilidade foi agravada pela ineficiência na gestão nacional e ausência de estratégias sólidas para direcionamento de leitos de UTI, além da concentração dos leitos

de alta complexidade em grandes centros urbanos e no setor privado, o que amplia as desigualdades no acesso ao cuidado intensivo.

Dessa forma, pacientes internados em hospitais públicos de menor porte, sem acesso à UTI, tiveram que enfrentar essas barreiras para um atendimento adequado, o que pode justificar a elevada associação com óbitos encontrada neste experimento.

Com base no trabalho publicado por [29], a VNI foi amplamente utilizada durante a pandemia de COVID-19, inclusive fora das UTI's, como estratégia para reduzir complicações associadas à intubação e à VMI. Os autores observaram que a VNI apresentou taxa de sucesso superior a 60%, especialmente quando aplicada precocemente, o que se alinha aos achados do presente estudo, nos quais pacientes com suporte não invasivo apresentaram maior probabilidade de cura. Além disso, o estudo identificou que fatores como idade avançada e menor relação ventilação/perfusão estavam associados ao insucesso da VNI e à necessidade de intubação, reforçando o papel da idade como determinante no prognóstico, também evidenciado no resultado deste trabalho, em que pacientes mais velhos submetidos à ventilação invasiva apresentaram maior risco de óbito.

Os achados desta pesquisa estão alinhados com o estudo de [10], que, ao analisar 103 pacientes com COVID-19 sob VMI, identificou a idade avançada como um dos principais preditores de mortalidade, juntamente com acidose e hipoxemia. No presente estudo, observou-se padrão semelhante: pacientes mais jovens (< 34 anos) submetidos à VMI apresentaram maior probabilidade de cura, enquanto a mortalidade aumentou progressivamente em indivíduos acima de 55 anos, reforçando a influência da idade na gravidade da doença. Além disso, conforme relatado por [10], a alta taxa de óbitos entre pacientes sob VMI demonstra o impacto da insuficiência respiratória grave e a limitação de reversibilidade dos quadros em estágios críticos, o que também foi evidenciado nos ramos correspondentes ao suporte ventilatório invasivo da árvore de decisão.

A idade e a necessidade de internação em UTI foram fatores determinantes para o desfecho clínico dos pacientes hospitalizados com COVID-19. Observou-se que indivíduos com idade superior a 61 anos apresentaram maior risco de óbito, especialmente quando necessitaram de cuidados intensivos, enquanto pacientes mais jovens apresentaram maior probabilidade de recuperação, mesmo sob tratamento em UTI. Esses achados são consistentes com os resultados descritos por [30] e [31] que identificaram a idade avançada como fator fortemente associado ao aumento da mortalidade e relataram maior risco de morte em pacientes idosos internados em UTI e submetidos à VMI. Em conjunto, esses resultados indicam que o envelhecimento está diretamente relacionado à pior evolução clínica, refletindo a redução da reserva fisiológica e a maior vulnerabilidade às complicações associadas à ventilação mecânica e à resposta inflamatória sistêmica. Além disso, a associação entre idade avançada e necessidade de suporte intensivo reforça a importância da estratificação de risco como instrumento para orientar condutas terapêuticas, otimizar recursos e aprimorar a tomada de decisão no manejo de pacientes com COVID-19 em ambiente hospitalar.

4.7 Conclusão

A aplicação dos algoritmos de ML demonstrou que o tipo de suporte ventilatório, a internação em UTI e a idade foram os principais fatores preditivos dos desfechos clínicos em pacientes hospitalizados com COVID-19. Observou-se que a VMI esteve associada a maior mortalidade, enquanto o uso de VNI e a ausência de suporte ventilatório correlacionaram-se com maior taxa de cura. Entre os pacientes internados em UTI, a idade surgiu como um importante determinante prognóstico, com indivíduos mais jovens apresentando melhores resultados clínicos. Além disso, o tempo de internação influenciou significativamente o prognóstico, indicando que evoluções mais rápidas até o óbito refletem quadros de maior gravidade.

De forma geral, os resultados obtidos evidenciam que modelos baseados em ML, como as árvores de decisão, podem contribuir de maneira relevante para a estratificação de risco e para a previsão de desfechos em pacientes críticos. Tais ferramentas oferecem suporte adicional à prática clínica, permitindo identificar precocemente grupos com maior probabilidade de óbito e orientar condutas terapêuticas mais direcionadas. A utilização desses modelos na rotina hospitalar pode favorecer a otimização de recursos, a individualização do cuidado e a melhoria da tomada de decisão, principalmente em UTI's.

4.8 Perspectivas Futuras

Existe uma crescente necessidade de desenvolvimento de ferramentas tecnológicas capazes de auxiliar no controle eficiente de agravos à saúde pública, especialmente em casos de alta complexidade clínica, como no enfrentamento à COVID-19. A utilização de modelos baseados em aprendizado de máquina representa uma estratégia promissora para combinar precisão e rapidez no uso de recursos hospitalares. O aprofundamento do conhecimento sobre os fatores preditivos de desfechos clínicos, a partir da análise de grandes volumes de dados clínicos, pode guiar pesquisas futuras voltadas à personalização do cuidado e à tomada de decisão clínica baseada em evidências.

Com base nas estratégias apresentadas neste estudo, espera-se contribuir para o desenvolvimento de modelos preditivos cada vez mais robustos, capazes de apoiar protocolos de estratificação de risco em tempo real, otimizando o direcionamento de suporte ventilatório e a alocação de leitos em unidades de terapia intensiva. Embora já existam ferramentas de apoio à decisão clínica, ainda são escassos os sistemas que integram múltiplas variáveis clínicas em modelos de fácil interpretação e ampla aplicabilidade no ambiente hospitalar.

Paralelamente, torna-se relevante o incentivo a projetos que explorem a integração entre ciência de dados, bioinformática clínica e inteligência artificial aplicada à saúde pública. A ampliação da base de dados com registros de qualidade, anonimizada e proveniente de diferentes realidades hospitalares, podendo proporcionar descobertas ainda mais relevantes e aplicáveis, contribuindo para uma resposta mais precisa e eficaz a futuras pandemias ou surtos epidemiológicos. Dessa forma, os achados deste estudo podem representar um marco inicial para o desenvolvimento de soluções tecnológicas que aliem inovação e equidade na gestão da saúde.

Referências

- [1] GUPTA, A.; et al. Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications. *New Generation Computing*, v. 40, n. 4, p. 987–1007, 2022. doi: 10.1007/s00354-021-00144-0. <https://doi.org/10.1007/s00354-021-00144-0>
- [2] SHAKHOVSKA, N. et al. A novel hybrid ensemble machine learning model for severity risk assessment and post-COVID prediction system. *Mathematical Biosciences and Engineering*, v. 19, n. 6, p. 6102–6123, 2022. <https://doi.org/10.3934/mbe.2022285>
- [3] RUSIC, D. et al. Tackling the antimicrobial resistance “pandemic” with machine learning tools: a summary of available evidence. *Microorganisms*, v. 12, n. 5, p. 842, 23 abr. 2024. doi: 10.3390/microorganisms12050842. <https://doi.org/10.3390/microorganisms12050842>
- [4] ALSULIMAN, Tamim et al. COVID-19 paraclinical diagnostic tools: Updates and future trends. *Current Research in Translational Medicine*, v. 68, n. 3, p. 83–91, ago. 2020. <https://doi.org/10.1016/j.retram.2020.06.001>
- [5] ALSULIMAN, Tamim; et al. Machine learning and artificial intelligence in the service of medicine: necessity or potentiality? *Current Research in Translational Medicine*, v. 68, n. 6, p. 245–251, 2020. <https://doi.org/10.1016/j.retram.2020.01.002>
- [6] UMAKANTHAN, S. et al. Origin, transmission, diagnosis and management of coronavirus disease 2019 (COVID-19). *Postgraduate Medical Journal*, v. 96, n. 1142, p. 753–758, dez. 2020. doi: 10.1136/postgradmedj-2020-138234. <https://doi.org/10.1136/postgradmedj-2020-138234>
- [7] CASCELLA, M. et al. Features, Evaluation, and Treatment of Coronavirus (COVID-19). In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. Atualizado em 18 ago. 2023. Disponível em: [https://www.ncbi.nlm.nih.gov/books/NBK554776/] (https://www.ncbi.nlm.nih.gov/books/NBK554776/).
- [8] PARASHER, Anant. COVID-19: Compreensão atual de sua fisiopatologia, apresentação clínica e tratamento, *Postgraduate Medical Journal*, Volume 97, Edição 1147, maio de 2021, Páginas 312–320, <https://doi.org/10.1136/postgradmedj-2020-138577>
- [9] SAILUNAZ, K. et al. A survey of machine learning-based methods for COVID-19 medical image analysis. *Medical & Biological Engineering & Computing*, v. 61, n. 6, p. 1257–1297, jun. 2023. doi: 10.1007/s11517-022-02758-y. <https://doi.org/10.1007/s11517-022-02758-y>
- [10] CHORON, R. L. et al. Fever in the ICU: a predictor of mortality in mechanically ventilated COVID-19 patients. *Journal of Intensive Care Medicine*, v. 36, n. 4, p. 484–493, abr. 2021. Disponível em: [https://pubmed.ncbi.nlm.nih.gov/33317374]. Acesso em: 21 de novembro de 2025. <https://doi.org/10.1177/0885066620979622>
- [11] BOTH, L. M. et al. COVID-19 pandemic and social distancing: economic, psychological, family, and technological effects. *Trends in Psychiatry and Psychotherapy*, v. 43, n. 2, p. 85–91, abr./jun. 2021. <https://doi.org/10.47626/2237-6089-2020-0085>
- [12] SERDAN, T. D. A. et al. COVID-19 Pandemic in Brazil: History, Characteristics, and Evolution. *Advances in Experimental Medicine and Biology*, v. 1327, p. 35–47, 2021. <https://doi.org/10.1007/978-3-030-71697-4>
- [13] CAMPOS, GW; et al. O Brasil na UTI: atenção hospitalar em tempos de pandemia. *Cadernos do Desenvolvimento*, Rio de Janeiro, v. 15, n. 26, p. 9–37, jul./dez. 2020. Disponível em: <https://www.inesc.org.br/wp-content/uploads/2021/03/Cadernos-do-Desenvolvimento-26.pdf>. Acesso em: 27 maio 2025.
- [14] THEODOSIOU, A. A. et al. Artificial intelligence, machine learning, and deep learning: potential resources for the infection clinician. *Journal of Infection*, v. 87, n. 4, p. 287–294, 2024. <https://doi.org/10.1016/j.jinf.2023.07.006>

- [15] PANESAR, S. S. et al. Promises and perils of artificial intelligence in neurosurgery. *Neurosurgery* [S.l.], v. 0, n. 0, p. 1-12, 2019. <https://doi.org/10.1093/neuros/nyz471>
- [16] CHOI, R. Y. et al. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science & Technology*, v. 9, n. 2, p. 14, fev. 2020. doi: 10.1167/tvst.9.2.14.
- [17] ARAÚJO, Murilo Freire Oliveira. Uso da aprendizagem de máquina supervisionada para predição de hipertensão arterial (HTA) e diabetes mellitus (DM) com base em dados sociodemográficos e de estilo de vida. 2024. 77 f. Tese (Doutorado em Biotecnologia em Saúde e Medicina Investigativa) – Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, 2024.
- [18] AMIB; SBPT. Orientações práticas de ventilação mecânica. São Paulo: Associação de Medicina Intensiva Brasileira; Sociedade Brasileira de Pneumologia e Tisiologia, 2024. Disponível em: <https://amib.org.br/amib-e-sbpt-lancam-edicao-atualizada-das-orientacoes-praticas-em-ventilacao-mecanica/>. Acesso em: 17 nov. 2025.
- [19] PASSARELLI ARAUJO, H. et al. Machine learning and comorbidity network analysis for hospitalized patients with COVID-19 in a city in Southern Brazil. *Smart Health* (Amsterdam), v. 26, p. 100323, dez. 2022. <https://doi.org/10.1016/j.smhl.2022.100323>
- [20] RICHARDSON, S. et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*, v. 323, n. 20, p. 2052–2059, 2020. <https://doi.org/10.1001/jama.2020.6775>
- [21] FANG, X. et al. Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. *Aging*, v. 12, n. 13, p. 12493–12503, 2020. <https://doi.org/10.18632/aging.103579>
- [22] WANG, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*, v. 323, n. 11, p. 1061–1069, 2020. <https://doi.org/10.1001/jama.2020.1585>
- [23] PAIVA, K. M. et al. Prevalence and Associated Factors of SARS by Covid-19 in Adults and Aged People with Chronic Cardiovascular Disease. *Arquivos Brasileiros de Cardiologia*, v. 117, n. 5, p. 968–975, nov. 2021. <https://doi.org/10.36660/abc.20200955>
- [24] ZHANG, J. et al. Do underlying cardiovascular diseases have any impact on hospitalised patients with COVID-19? *Heart*, v. 106, n. 15, p. 1148–1153, 2020. <https://doi.org/10.1136/heartjnl-2020-316909>
- [25] PELLICORI, P. et al. COVID-19 and its cardiovascular effects: a systematic review of prevalence studies. *Cochrane Database of Systematic Reviews*, Edição 3, Artigo CD013879, 2021. <https://doi.org/10.1002/14651858.CD013879>
- [26] CRONIN JN, et al. Mechanical ventilation in COVID-19: A physiological perspective. *Exp Physiol*. 2022 Jul;107(7):683-693. doi: 10.1113/EP089400. Epub 2021 Sep 27. PMID: 34541721; PMCID: PMC8667647. <https://doi.org/10.1113/EP089400>
- [27] ROMANELLI A, et al. Predictor factors for non-invasive mechanical ventilation failure in severe COVID-19 patients in the intensive care unit: a single-center retrospective study. *J Anesth Analg Crit Care*. 2022 Feb 15;2(1):10. <https://doi.org/10.1186/s44158-022-00038-7>
- [28] ZANNELLA A, et al. COVID-19 Italian ICU Network (2021) Time course of risk factors associated with mortality of 1260 critically ill patients with COVID-19 admitted to 24 Italian intensive care units. *Intensive Care Med* 47(9):995–1008. <https://doi.org/10.1007/s00134-021-06495-y>.
- [29] BELLANI, G. et al. Noninvasive ventilatory support of patients with COVID-19 outside the intensive care units (WARD-COVID). *Annals of the American Thoracic Society*, v. 18, n. 6, p. 1020-1026, jun. 2021. Disponível em: [\[https://pubmed.ncbi.nlm.nih.gov/33395553\]](https://pubmed.ncbi.nlm.nih.gov/33395553). Acesso em: 21 de novembro de 2025. <https://doi.org/10.1513/AnnalsATS.202008-1080OC>
- [30] LE BORGNE, P. et al. The impact of age on in-hospital mortality in critically ill COVID-19 patients: a retrospective and multicenter study. *Diagnostics*, v. 12, n. 666, p. 1-11, 2022. <https://doi.org/10.3390/diagnostics12030666>
- [31] CILLÓNIZ, C. et al. Risk factors associated with mortality among elderly patients with COVID-19: data from 55 intensive care units in Spain. *Pulmonology*, v. 29, n. 5, p. 362-374, 2023. <https://doi.org/10.1016/j.pulmoe.2023.01.007>