Leticia Araújo Silva

A Novel Framework for COVID-19 Detection and Clinical Triage Using Multimodal Physiological Signals on a Portable Medical Assistant

Vitória

2025

Leticia Araújo Silva

A Novel Framework for COVID-19 Detection and Clinical Triage Using Multimodal Physiological Signals on a Portable Medical Assistant

Doctoral thesis presented to the Postgraduate Program in Electrical Engineering (PPGEE), Federal University of Espírito Santo (UFES), in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering

Federal University of Espírito Santo

Postgraduate Program in Electrical Engineering

Supervisor: Prof. Teodiano Freire Bastos Filho, PhD (UFES, Brazil)

Co-supervisor: Prof. Sridhar Krishnan, PhD (Toronto Metropolitan University, Canada)

Vitória

2025

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

Silva, Leticia, 1994-
S586n A Novel Framework for COVID-19 Detection and Clinical Triage Using Multimodal Physiological Signals on a Portable Medical Assistant / Leticia Silva. - 2025.
136 f. : il.

Orientador: Teodiano Bastos Filho.
Coorientador: Sridhar Krishnan.
Tese (Doutorado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Inteligência artificial. 2. Diagnóstico. 3. COVID-19 (Doença). 4. Processamento de sinais. I. Bastos Filho, Teodiano. II. Krishnan, Sridhar. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

UFES

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Centro Tecnológico
Coordenação do Programa de Pós-Graduação em Engenharia Elétrica
Credenciamento/MEC 398 de 29/05/2025

176ª ATA DE DEFESA DE TESE DE DOUTORADO

Ata da sessão de defesa da 176ª Tese de Doutorado em Engenharia Elétrica do Coordenação do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Espírito Santo, da aluna Leticia Araújo Silva, candidato(a) ao grau de Doutor(a) em Engenharia Elétrica. Às 13:00 horas do dia 15/08/2025, no formato webconferência, o(a) presidente da Comissão Examinadora, Professor(a) Teodiano Freire Bastos Filho - UFES, iniciou a sessão apresentando a Comissão constituída, além dele(a) próprio(a), que é o(a) Orientador(a), pelo(s) membro(s) Eliete Maria De Oliveira Caldeira (Examinador Interno) - UFES, , , , e Sridhar Krishnan (Examinador externo)Adriano De Oliveira Andrade (Examinador externo)Eduardo Lazaro Martins Naves (Examinador externo)Denis Delisle Rodríguez (Examinador externo)Ana Cecilia Villa Parra (Examinador externo). A seguir, o(a) presidente passou a palavra ao candidato(a), que, em 50 minutos, apresentou a sua tese, intitulada “Design and Evaluation of a Multimodal Medical Device-Based Framework for COVID-19 Inference and Risk Classification Based on the Manchester Triage System”. Finda a apresentação, o(a) presidente passou a palavra aos membros da Comissão para procederem à arguição da candidata. Finda a arguição, o(a) presidente convidou a Comissão para dirigir-se a uma sala reservada, para deliberação. Após a deliberação, a Comissão retornou, e o(a) presidente informou aos presentes que a tese fora Aprovada. Logo após, o(a) presidente declarou encerrada a sessão, e eu, Aline Oliveira Amaral, lavrei a presente Ata,

Prof. Dr. Teodiano Freire Bastos Filho
Universidade Federal do Espírito Santo - Presidente

Prof. Dr. Eliete Maria De Oliveira Caldeira
Universidade Federal do Espírito Santo - Examinador Externo

Sridhar Krishnan
- Coorientador

Adriano De Oliveira Andrade
- Examinador Externo

Eduardo Lazaro Martins Naves
- Examinador Externo

Denis Delisle Rodríguez
- Examinador Externo

Ana Cecilia Villa Parra
- Examinador Externo

Documento assinado eletronicamente nos moldes do art. 10 da MP 2200/01 e Lei 14063/20
[Hash SHA256] f10a6ce09d3e7e64017ba9e20338d4f2222ed6226ec44367304276e915d4d1a8

Datas e horários baseados em Brasília, Brasil
Sincronizado com o NTP.br e Observatório Nacional (ON) em 15/08/2025 às 16:03:34

176 Ata de defesa pública - Letícia Araújo Silva

Data e Hora de Criação: 15/08/2025 às 14:04:16

Documentos que originaram esse envelope:
- 176 Ata de defesa pública - Letícia Araújo Silva.pdf (Arquivo PDF) - 1 página(s)

Hashs únicas referente à esse envelope de documentos

[SHA256]: f10a6ce09d3e7e64017ba9e20338d4f2222ed6226ec44367304276e915d4d1a8
[SHA512]: 920cc223921caa61fb3a2b3156b606e6c41480deb4f9335431a8950156a02374058477cff56173cf4bb97c1e27e2b232476255597cf9a726285bf8bec30dfe05

Lista de assinaturas solicitadas e associadas à esse envelope

ASSINADO - Teodiano Freire Bastos Filho (teodiano.bastos@ufes.br)
Data/Hora: 15/08/2025 - 15:47:07, IP: 149.102.233.37, Geolocalização: [-20.273449, -40.306205]
[SHA256]: 683bdde91a98c4e30c9cd544e3315387455c9caa1236f70db0919487a13a7f88
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Sridhar Krishnan (krishnan@torontomu.ca)
Data/Hora: 15/08/2025 - 15:50:57, IP: 200.137.65.104, Geolocalização: [-20.276857, -40.305023]
[SHA256]: 2e6430c2c956da9969c868c5dc55f46810fb0596de6011e56efee0028b4c3065
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Eduardo Lázaro Martins Naves (eduardonavesufu@gmail.com)
Data/Hora: 15/08/2025 - 15:52:36, IP: 200.137.65.103, Geolocalização: [-20.273494, -40.306014]
[SHA256]: eaea59bee5c67caff94d185e049f7dc4e5c008a63bc758e527900a48f713de96
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Eliete Maria de Oliveira Caldeira (eliete.caldeira@ufes.br)
Data/Hora: 15/08/2025 - 15:53:56, IP: 200.137.65.100, Geolocalização: [-20.273487, -40.306157]
[SHA256]: b1780f01b2ae2d8142908b8012bc76e033f029c80e0fe3d27d289ef289ee384e
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Denis Delisle Rodríguez (denis.rodriguez@isd.org.br)
Data/Hora: 15/08/2025 - 15:57:13, IP: 179.102.138.0
[SHA256]: e0ee3d5e8eff632acee2dd975f78c03cbc5616723de2c90c6277a80f73eb24fe
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Ana Cecilia Villa Parra (avilla@ups.edu.ec)
Data/Hora: 15/08/2025 - 15:58:34, IP: 200.137.65.107
[SHA256]: 0be8704f5ab133edb04e5c1e544628574cd03648f6df80a7d38502414f48b572
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

ASSINADO - Adriano de Oliveira Andrade (adriano@ufu.br)
Data/Hora: 15/08/2025 - 16:03:34, IP: 177.50.58.226, Geolocalização: [-20.273531, -40.305990]
[SHA256]: c507a30159a2da364b7fd4207ae3a74b254854f90437b030d72a3745da0ef4c7
Assinatura Eletrônica Avançada (Conforme Lei nº 14.063/20, art. 4º, II)

Histórico de eventos registrados neste envelope

15/08/2025 16:03:34 - Envelope finalizado por adriano@ufu.br, IP 177.50.58.226
15/08/2025 16:03:34 - Assinatura realizada por adriano@ufu.br, IP 177.50.58.226
15/08/2025 16:01:05 - Envelope visualizado por adriano@ufu.br, IP 177.50.53.252
15/08/2025 15:58:34 - Assinatura realizada por avilla@ups.edu.ec, IP 200.137.65.107
15/08/2025 15:57:13 - Assinatura realizada por denis.rodriguez@isd.org.br, IP 179.102.138.0
15/08/2025 15:53:56 - Assinatura realizada por eliete.caldeira@ufes.br, IP 200.137.65.100
15/08/2025 15:53:53 - Envelope visualizado por eliete.caldeira@ufes.br, IP 200.137.65.100
15/08/2025 15:52:36 - Assinatura realizada por eduardonavesufu@gmail.com, IP 200.137.65.103
15/08/2025 15:50:57 - Assinatura realizada por krishnan@torontomu.ca, IP 200.137.65.104

Documento assinado digitalmente em conformidade com o padrão ICP-Brasil e validado segundo as diretrizes do Instituto Nacional de Tecnologia da Informação (ITI), em atendimento à Medida Provisória nº 2.200-2/2001 e à Lei nº 14.063/2020.

Os registros de assinatura presentes nesse documento pertencem única e exclusivamente a esse envelope.
Documento final gerado e certificado por **Universidade Federal do Espírito Santo**

Datas e horários baseados em Brasília, Brasil
Sincronizado com o NTP.br e Observatório Nacional (ON) em 15/08/2025 às 16:03:34

176 Ata de defesa pública - Letícia Araújo Silva

Data e Hora de Criação: 15/08/2025 às 14:04:16

Documentos que originaram esse envelope:

- 176 Ata de defesa pública - Letícia Araújo Silva.pdf (Arquivo PDF) - 1 página(s)

Hashs únicas referente à esse envelope de documentos

[SHA256]: f10a6ce09d3e7e64017ba9e20338d4f2222ed6226ec44367304276e915d4d1a8
[SHA512]: 920cc223921caa61fb3a2b3156b606e6c41480deb4f9335431a8950156a02374058477cff56173cf4bb97c1e27e2b232476255597cf9a726285bf8bec30dfe05

Histórico de eventos registrados neste envelope

15/08/2025 15:50:28 - Envelope visualizado por krishnan@torontomu.ca, IP 200.137.65.104
15/08/2025 15:47:07 - Assinatura realizada por teodiano.bastos@ufes.br, IP 149.102.233.37
15/08/2025 15:47:00 - Envelope visualizado por teodiano.bastos@ufes.br, IP 149.102.233.37
15/08/2025 14:04:44 - Envelope registrado na Blockchain por aline.amaral@ufes.br, IP 187.36.175.82
15/08/2025 14:04:44 - Envelope encaminhado para assinaturas por aline.amaral@ufes.br, IP 187.36.175.82
15/08/2025 14:04:16 - Envelope criado por aline.amaral@ufes.br, IP 187.36.175.82

Documento assinado digitalmente em conformidade com o padrão ICP-Brasil e validado segundo as diretrizes do Instituto Nacional de Tecnologia da Informação (ITI), em atendimento à Medida Provisória nº 2.200-2/2001 e à Lei nº 14.063/2020.

Os registros de assinatura presentes nesse documento pertencem única e exclusivamente a esse envelope.
Documento final gerado e certificado por **Universidade Federal do Espírito Santo**

Dedicated to my parents, Valnete and Lorentino; my grandparents, Paulina (in memoriam), Moacir (in memoriam), Maria and Venâncio; and my aunt Arlete and her family.

Acknowledgements

Esta seção está em português para facilitar o entendimento de todos os citados ao longo do texto.

Agradeço a Deus por mais esta conquista. Foi Ele quem sustentou meus passos e estendeu a mão quando, como Pedro ao sentir o vento forte (Mt 14,22–36), o medo tentou me fazer desanimar. Sua presença me ergueu, me guiou e me deu coragem para continuar, mesmo quando tudo parecia incerto.

Agradeço à minha querida e amada mãe, Valnete. Sem sua incessante oração, seu amor incondicional e sua força silenciosa, esta caminhada teria sido muito mais árdua. Sua fé me sustentou nos dias em que a minha parecia vacilar. Assim como a Virgem Maria, que guardava tudo em seu coração e confiava mesmo sem compreender plenamente os caminhos de Deus, minha mãe me ensinou que a entrega verdadeira se vive no silêncio, com o coração voltado para Deus.

Agradeço ao meu querido e amado pai, Lorentino que, ao seu modo, sempre me apoiou e proporcionou que eu estudasse, e chegasse aonde estou. Muito obrigada por tudo.

Agradeço aos meus avós maternos, Paulina e Moacir, por tudo que eles fizeram por mim. Infelizmente vocês não estão aqui presentes para compartilhar e comemorar este momento, mas estão comigo em meu coração para todo lugar que vou. Agradeço aos meus avós paternos, Maria e Venâncio, por estarem presentes em todos os momentos importantes de minha vida. Obrigada por todo carinho que vocês têm comigo.

Agradeço à minha família, em especial aos meus tios, Arlete e Helder, e também à sua família — Anna Luisa, Cecília, Lucas e Elizângela. O apoio, o carinho e a presença de vocês foram fundamentais nesta jornada.

Agradeço também ao meu amor, Leonardo. Sem você neste último ano, tudo teria sido mais difícil. Obrigada por estar ao meu lado, por me apoiar e por me ouvir.

Agradeço aos meus amigos Lorana, Drielle, Mariana, Larissa, Maiara, Dayane, Sheila, Jéssica, Ledy e a todos aqueles que, mesmo não citados aqui, estiveram comigo de alguma forma. Obrigada pelas conversas, pelas risadas, pelos conselhos e pela companhia nos dias em que tudo parecia pesado demais. Cada gesto, cada palavra e cada momento com vocês me ajudaram a seguir em frente com mais leveza. Minha gratidão também à Angélica, por sua escuta atenta e presença acolhedora, que fizeram diferença em momentos de grande desafio. Sou igualmente grata aos amigos que fiz no LRTA, em especial Carlos e Alan, pela parceria, amizade e pelo apoio técnico fundamental ao longo da pesquisa. A contribuição de vocês foi essencial em momentos decisivos. Agradeço também aos amigos

que a vida me presenteou no ambiente acadêmico da UCL, por todo apoio e incentivo ao longo do caminho. Ao grupo Signal Analysis Research (SAR), em Toronto, minha sincera gratidão pela acolhida, pelos aprendizados e pela convivência durante esse período tão importante da minha formação. Aos amigos da Paróquia Bom Pastor, na Praia da Costa, minha gratidão por iluminarem meus dias com fé.

Agradeço ao meu orientador, Prof. Dr. Teodiano, e aos professores doutores Denis, Eliete e Sri Krishnan, pela orientação ao longo da minha pesquisa. Sou grata pela constante disponibilidade, pelo apoio técnico, pelas palavras de incentivo e por me ajudarem a encontrar os melhores caminhos.

Finalmente, agradeço ao Fundo de Apoio a Ciência e Tecnologia (FACITEC) por financiar minha bolsa de estudos durante o curso do doutorado, aos programas *Emerging Leaders in the Americas Program* (ELAP) e *Global Affairs Canada*. Também agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Código de Financiamento: 88887.506772/2020-00) pelo apoio ao projeto de pesquisa.

"Onde estaria eu se não fosse o teu amor, Senhor?"

Toca de Assis

Resumo

Sistemas de atendimento de urgência e emergência enfrentam desafios crescentes para oferecer triagem rápida e precisa, especialmente em ambientes com recursos limitados, onde subjetividade, falta de infraestrutura e o excesso de pacientes comprometem as decisões clínicas. Essas limitações tornaram-se ainda mais evidentes durante a pandemia de *Coronavirus Disease 2019* (COVID-19), que expôs lacunas críticas na capacidade diagnóstica e evidenciou a falta de ferramentas acessíveis, não invasivas e capazes de realizar avaliações autônomas. Esta pesquisa investiga se sinais fisiológicos multimodais — tosse, fala, respiração e sinais vitais — coletados por um equipamento portátil denominado Assistente Médico Portátil Integrado (IPMA), podem apoiar a triagem inteligente e a inferência de COVID-19 com o uso de modelos de Aprendizado de Máquina (ML). Para isso, foi conduzido um experimento dividido em duas partes. A primeira concentrou-se na detecção de COVID-19 com uso de bases públicas e dados reais coletados com o IPMA. Espectrogramas Mel foram extraídos dos sinais de áudio, seguidos pela extração de características de textura com os descritores Padrões Binários Locais (LBP) e Padrões Ternários Locais (LTP). O LBP superou consistentemente o LTP nas tarefas de classificação, com a fala apresentando o maior poder discriminativo e SpO_2 e temperatura emergindo como os sinais fisiológicos mais informativos. Embora treinados com dados públicos, os modelos apresentaram generalização moderada para os dados do IPMA, especialmente para fala e respiração. A segunda parte avaliou a classificação de risco com base no Protocolo de Triagem de Manchester, por meio de uma abordagem estruturada que envolveu pré-processamento dos dados, comparação de modelos de ML e Aprendizado Profundo (DL) e avaliação de usabilidade. Utilizando uma base pediátrica pública, classificadores como XGBoost e Stacking alcançaram F1-scores superiores a 0,99 com variáveis clínicas abrangentes. Resultados promissores também foram obtidos com sinais vitais e variáveis de baixa subjetividade, com F1-scores em torno de 0,74, demonstrando o potencial de dados objetivos para estratificação de risco com baixo viés em sistemas autônomos. No entanto, ao serem testados com dados de adultos coletados com o IPMA, os modelos apresentaram desempenho limitado, indicando desafios na generalização entre populações e contextos distintos. A usabilidade foi componente central do estudo, e avaliações padronizadas com os questionários *System Usability Scale* (SUS) (e *Post-Study System Usability Questionnaire* — PSSUQ — na tarefa de COVID-19) indicaram alta aceitação do IPMA por pacientes e profissionais de saúde. As pontuações refletem sua facilidade de uso e integração aos fluxos clínicos, reforçando o potencial do sistema para triagem em cenários reais.

Palavras-Chave: Detecção de COVID-19, triagem clínica, protocolo de triagem de Manchester, aprendizado de máquina, equipamento médico portátil, sinais fisiológicos, avaliação de usabilidade.

Abstract

Emergency and urgent care systems face growing challenges in providing timely and accurate triage, especially in resource-constrained environments where subjectivity, lack of infrastructure, and high patient volumes compromise clinical decisions. These limitations became even more evident during the Coronavirus Disease 2019 (COVID-19) pandemic, which exposed critical gaps in diagnostic capacity and highlighted the absence of scalable, non-invasive tools for autonomous assessment. This research investigates whether multimodal physiological signals — cough, speech, breath, and vital signs — collected through a portable equipment called Integrated Portable Medical Assistant (IPMA) may support intelligent triage and COVID-19 inference via Machine Learning (ML) models. To address this, a two-part experimental design was conducted. The first part focused on COVID-19 detection using public datasets and real-world data collected with the IPMA. Mel-spectrograms were extracted from audio signals, followed by texture-based feature extraction using Local Binary Pattern (LBP) and Local Ternary Pattern (LTP). LBP consistently outperformed LTP across classification tasks, with speech showing the highest discriminative power, and SpO_2 and temperature emerging as the most informative physiological indicators. Although trained on public datasets, models achieved moderate generalization to IPMA data, particularly for speech and breath signals. The second part evaluated clinical risk classification based on the Manchester Triage System through a structured approach that included data preprocessing, comparison of ML and Deep Learning (DL) models, and usability assessment. Using a public pediatric dataset, ensemble classifiers such as XGBoost and Stacking achieved F1-scores above 0.99 when trained on comprehensive clinical features. Additionally, promising results were obtained using primarily vital signs and low-subjectivity variables, with models reaching F1-scores around 0.74, demonstrating the potential of objective data for low-bias risk stratification in autonomous systems. However, when tested on adult data collected with IPMA, the models showed limited performance, indicating challenges in generalizing across different populations and clinical contexts. Usability was also a central component of this study. Standardized evaluations using the System Usability Scale (SUS) (and Post-Study System Usability Questionnaire — PSSUQ — for the COVID-19 task) indicated high user acceptance of the IPMA by both patients and healthcare professionals. Reported scores reflect the system's ease of use and perceived integration into clinical workflows, reinforcing its potential for deployment in real-world triage and screening scenarios.

Keywords: COVID-19 detection, clinical triage, Manchester triage system, machine learning, portable medical equipment, physiological signals, usability evaluation.

List of Figures

List of Tables

List of abbreviations and acronyms

ACC	Accuracy
AI	Artificial Intelligence
ANOVA	One-way Analysis of Variance
API	Application Programming Interface
CCS	COVID-19 Cough Sub-Challenge
COVID-19	Coronavirus Disease 2019
CSS	COVID-19 Speech Sub-Challenge
DBP	Diastolic Blood Pressure
DT	Decision Tree
DL	Deep Learning
EDA	Exploratory Data Analysis
FFT	Fast Fourier Transform
HSD	Honestly Significant Difference
INFOQUAL	Information Quality
INTERQUAL	Interface Quality
IoT	Internet of Things
IPMA	Integrated Portable Medical Assistant
KDD	Knowledge Discovery and Data Mining
kNN	K-Nearest Neighbors
LBP	Local Binary Pattern
LTP	Local Ternary Pattern
ML	Machine Learning
MLP	Multilayer Perceptron

MICE	Multivariate Imputation by Chained Equations
MTS	Manchester Triage System
PA	Urgent Care Unit
PSSUQ	Post-Study System Usability Questionnaire
RF	Random Forest
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SBP	Systolic Blood Pressure
SpO_2	Oxygen Saturation
STD	Standard Deviation
STFT	Short-Time Fourier Transform
SUS	System Usability Scale
SVM	Support Vector Machine
SYSUSE	System Usefulness
UVC	Ultraviolet-C
UFES	Federal University of Espírito Santo
UBS	Unidade Básica de Saúde
WHO	World Health Organization
XGBoost	eXtreme Gradient Boosting

Contents

1 Introduction

This chapter provides an introduction to the research by first discussing the motivation and the problem statement. Next, the hypotheses are presented, followed by the research aim, the contributions and the justification. Finally, the thesis structure is presented.

1.1 Motivation

Emergency and urgent care services are essential for managing acute health conditions and preventing clinical deterioration (Zachariasse et al., 2021). In high-demand environments, medical triage serves as a critical tool to prioritize care based on clinical urgency. However, in Brazil, the volume of patients imposes significant pressure on these systems. A typical Emergency Care Unit (Unidade de Pronto Atendimento, UPA) attends to approximately 115 patients per day, based on data from ten UPAs over a 28-day period in February 2021 (Costa et al., 2022). In the city of Vitória/Brazil alone, the Municipal Health Department (SEMUS) recorded over 9.6 million healthcare actions in 2024, including consultations, diagnostic procedures, and therapeutic interventions — more than double the 4 million actions reported in 2021 (Vitória, 2025).

This growing demand exposes the fragilities of triage practices, which are often affected by the subjectivity of clinical judgment, lack of standardization, and insufficient training of professionals (Fekonja et al., 2023; Zaboli et al., 2025; Ingielewicz; Rychlik; Sieminski, 2024). While structured protocols such as the Manchester Triage System (MTS) have been implemented to mitigate these issues, their effective application depends on the systematic and accurate collection of clinical data. In practice, this requirement proves difficult to meet in overcrowded settings or those with limited infrastructure (Zaboli et al., 2025; Fekonja et al., 2023; Da'Costa et al., 2025).

The Coronavirus disease 2019 (COVID-19) pandemic further exposed the structural fragilities of emergency care systems, particularly regarding triage and diagnostic capacity. The World Health Organization (WHO) officially declared COVID-19 a global pandemic on March 11, 2020 (World Health Organization, 2020). As the disease spread rapidly across continents, healthcare networks worldwide were forced to respond to an overwhelming volume of cases in a short period. By 2023, COVID-19 had resulted in millions of infections and deaths globally (Johns Hopkins University, 2022), with Brazil alone reporting over 38 million confirmed cases and more than 700,000 deaths (Worldometer, 2024). This public health emergency increased the demand for emergency care, revealing critical limitations in the capacity of healthcare systems to manage sudden and sustained surges in patient

volume — particularly in low-resource settings.

Beyond the immediate strain on emergency services, the pandemic also exposed structural barriers to effective large-scale diagnostic testing. Although Reverse Transcription Polymerase Chain Reaction (RT-PCR) remains the molecular gold standard for COVID-19 detection, its widespread application was hindered by high costs, long processing times, and the need for specialized infrastructure and trained personnel (Filip et al., 2022; Haldane et al., 2021; Gupta-Wright et al., 2021; Williams et al., 2023; Teymouri et al., 2021; Minhas et al., 2023). These operational constraints severely restricted testing coverage in remote or underserved areas. As a consequence, many cases went undetected or were diagnosed too late, undermining both clinical management and public health control efforts. This scenario emphasized the urgent need for rapid, accessible, and scalable alternatives for initial disease screening (Haldane et al., 2021).

In response to the growing strain on healthcare systems, telemedicine has emerged as a strategic tool to expand access to care. By enabling remote consultations, it has helped reduce geographic disparities and increase coverage, particularly during periods of mobility restrictions (Doraiswamy et al., 2020). However, the effectiveness of telemedicine depends on more than just virtual communication. For it to serve as a reliable clinical interface, it must be supported by robust technological infrastructure capable of acquiring, transmitting, and analyzing physiological data in real time (Kobeissi; Ruppert, 2022). In practice, the lack of integrated and objective remote assessment tools continues to limit telemedicine's ability to respond to complex and urgent medical demands (Kobeissi; Ruppert, 2022; Farzandipour; Nabovati; Sharif, 2024).

Recent advances in biomedical engineering have opened new possibilities for decentralizing healthcare delivery. Developments in wearable sensors, low-power microcontrollers, and wireless data transmission allow for the non-invasive, real-time capture of physiological signals with minimal user intervention (Shajari et al., 2023; Tao et al., 2023; Acosta et al., 2022). Parallel progress in Machine Learning (ML) and Deep Learning (DL) has further enabled the automated interpretation of these signals, leading to the development of advanced clinical decision support systems. These technologies hold the potential to reduce response times, increase diagnostic consistency, and standardize triage procedures across different care settings (Tao et al., 2023; Cai; Ma; Ge-Zhang, 2025).

Despite these promising developments, key challenges remain before such systems can be broadly adopted in real-world contexts. The integration of heterogeneous sensors with intelligent algorithms requires not only technical compatibility but also validation in diverse clinical environments. Furthermore, ensuring high usability for non-specialized users, maintaining data reliability, and adapting to local infrastructure constraints are critical to deployment at scale (Tao et al., 2023; Alzghaibi, 2025). Without addressing these barriers, the practical implementation of portable and intelligent screening tools

remains limited, especially in the resource-constrained settings that would benefit most from them.

1.2 Problem Statement

Despite significant progress in digital health, a persistent gap remains between technological potential and its effective application in clinical practice—particularly in resource-constrained environments. Traditional triage and diagnostic processes often rely on subjective clinical judgment, leading to variability and potential mistakes in patient prioritization (Porto, 2024). Although structured protocols have been developed, their consistent implementation is challenging in overcrowded or under-resourced settings (Ftouni et al., 2022).

Existing digital solutions, including telemedicine platforms and automated symptom-checkers, often face limitations in supporting objective, real-time decision-making. During the COVID-19 pandemic, for instance, many triage and diagnostic decisions were made without access to physiological data, resulting in delayed or incorrect assessments (Gupta-Wright et al., 2021). In addition, while telemedicine expanded access to care, it often lacks the capacity to remotely acquire and interpret biomedical signals (Kobeissi; Ruppert, 2022). Automated symptom-checker tools have also shown inconsistent performance and low diagnostic accuracy, raising safety concerns in high-stakes environments (Wallace et al., 2022).

A current limitation in the field is the lack of validated systems capable of autonomously acquiring, processing, and interpreting multimodal physiological data to support clinical decision-making based on objective measurements (Resende et al., 2023). For such systems to be effective in real-world healthcare settings — particularly in remote or underserved areas — they must combine clinical reliability, operational autonomy, usability by non-specialized personnel, in addition to adaptability to infrastructure variability. Moreover, evaluating these systems under realistic conditions is important to verify their safety, usability, and applicability in diverse healthcare contexts.

To bridge this gap, this thesis proposes the development and validation of algorithms based on Artificial Intelligence (AI) for autonomous triage and diagnostic inference, using multimodal physiological signals collected by the Integrated Portable Medical Assistant (IPMA), a portable equipment designed for real-world clinical environments.

This leads to the following research questions:

1. How can AI-based algorithms be developed to enable a portable system to autonomously acquire, process, and interpret multimodal physiological signals for clinical triage and COVID-19 diagnostic inference in resource-constrained environments?

2. What methodologies and evaluation protocols are required to clinically validate and assess the usability of such a system for safe and reliable deployment in remote or unsupervised healthcare settings?

1.3 Hypotheses

Recent advancements in AI-based diagnostic and triage systems have demonstrated promising results in controlled environments. However, their applicability in real-world contexts remains limited, particularly when dealing with multimodal physiological signals collected in decentralized or resource-constrained settings. Furthermore, few studies have addressed the usability and integration of such systems into clinical workflows. In this context, the present research is guided by the following hypotheses:

Main Hypothesis: It is feasible to develop and validate ML and DL algorithms capable of objectively performing COVID-19 detection and clinical risk triage using multimodal physiological signals collected by portable equipment (IPMA) and data obtained in real-world settings.

Secondary Hypotheses:

1. Vital signs (SpO_2, blood pressure, heart rate, and temperature) can contribute to COVID-19 detection and clinical risk stratification when used together with structured triage protocols such as the MTS.
2. ML/DL algorithms trained on data from controlled environments can generalize to physiological signals collected by a portable equipment, such as the IPMA, supporting COVID-19 detection and clinical triage in real-world scenarios.
3. The combined use of the IPMA equipment and the developed algorithms demonstrates good usability and acceptance among healthcare professionals and patients in contexts of remote or assisted triage.

1.4 Research Aim and Objectives

The aim of this research was to develop and validate ML and DL algorithms to support autonomous clinical triage and COVID-19 detection based on multimodal physiological signals collected by a portable equipment (IPMA). The focus is on real-world applicability, particularly in resource-constrained healthcare environments.

To achieve this aim, the following research objectives were established:

1. Develop and evaluate ML/DL algorithms for COVID-19 detection using multimodal physiological signals, including audio (cough, speech, breathing) and vital signs.

2. Investigate the individual and combined contributions of different physiological modalities (audio and vital signs) to the performance of COVID-19 detection models.
3. Develop and evaluate ML/DL algorithms for clinical risk classification using vital signs, following the MTS protocol.
4. Assess the generalizability of models trained on public datasets when applied to physiological signals collected by the IPMA in real-world scenarios.
5. Evaluate the usability and acceptance of the IPMA in clinical triage scenarios.

1.5 Justification

Recent challenges in emergency care and pandemic response have underscored the need for intelligent, accessible systems to support clinical triage and diagnosis. Despite advances in AI and biomedical technologies, integrated and deployable solutions remain scarce. In light of these challenges, there is a clear need for research focused on integrated, intelligent systems that can support both diagnosis and triage in practical settings.

In this context, clinical triage in emergency and urgent care settings became even more vital. When patient prioritization is delayed or inadequate, the consequences may include increased clinical risks, inefficient care delivery, and intensified pressure on limited resources (Hwang; Lee, 2022; Chang et al., 2024; Porto, 2024). These limitations underscored the need for alternative solutions capable of making triage and diagnostic procedures more accessible, standardized, and objective — reducing dependence on subjective evaluations or resource-intensive laboratory testing (Porto, 2024). In this context, there is a growing interest in scalable strategies that can support both diagnostic and triage workflows, particularly in scenarios marked by operational overload, limited infrastructure, or remote care demands.

While laboratory-based tests remain central for COVID-19 diagnosis, several studies have investigated the feasibility of automatic detection based on acoustic biomarkers such as cough, breathing, and speech, given their accessibility and non-invasive nature (Despotovic et al., 2021; Aly; Rahouma; Ramzy, 2022; Husain et al., 2022; Dang et al., 2022; Pahar et al., 2022). These physiological signals, often altered by respiratory infections, exhibit distinctive acoustic patterns that can be leveraged by ML and DL models for screening and early detection (Laguarta; Hueto; Subirana, 2020; Husain et al., 2022; Dang et al., 2022; Pahar et al., 2022). A wide range of handcrafted features has been employed, including Mel-Frequency Cepstral Coefficients (MFCCs), Zero Crossing Rate (ZCR), and Spectral Roll-off (SR) (Brown et al., 2020; Pramono; Imtiaz; Rodriguez-Villegas, 2016; Verde et al., 2021), with promising results using traditional classifiers such as Logistic Regression (LR), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and ensemble-based

methods (Brown et al., 2020; Sharma; Umapathy; Krishnan, 2022; Verde et al., 2021). More recent approaches have explored DL models with Convolutional Neural Networks (CNNs), Long Short-Term Memories (LSTMs), and ResNet architectures, as well as transfer learning and dimensionality reduction techniques (Pahar et al., 2022; Brown et al., 2020). Although performance varies across datasets and signal types, cough has frequently emerged as the most informative modality (Aly; Rahouma; Ramzy, 2022; Pahar et al., 2022). These findings reinforce the relevance of developing intelligent, low-cost, and scalable tools for accessible screening in settings where conventional diagnostics are not readily available.

In addition to diagnostic support, AI has also gained attention as a tool for enhancing clinical triage. Multiple studies have proposed ML and DL models trained on structured patient data — such as age, vital signs, and reported symptoms — to predict outcomes like hospital admission, need for critical care, or clinical urgency (Hong; Haimovich; Taylor, 2018; Joseph et al., 2020; Chang et al., 2024). For instance, Cicolo and Peres (2019) compared the performance of manual and electronic MTS-based triage, reporting modest improvements in both accuracy and efficiency. Other approaches have adopted fuzzy logic to assist in risk scoring (Jatobá et al., 2018), while recent models have employed physiological and symptomatic data for automated severity classification (Almulhim et al., 2025). Reviews highlight the growing role of explainable ML in triage (Doorn et al., 2024), and prediction systems have been applied to early mortality estimation and alert mechanisms in emergency settings (Hsieh et al., 2021; Kim et al., 2024). These contributions emphasize the potential of intelligent systems to optimize patient flow, enhance risk assessment, and support clinical decision-making.

While these AI-based models show promising results — often rivaling traditional protocols such as the Emergency Severity Index (ESI) and MTS (Chang et al., 2024) — several challenges remain. Performance across diverse populations, infrastructure constraints, and data quality may limit model generalization. Furthermore, many studies rely on retrospective datasets or simulations, which restrict their practical applicability. Usability considerations are often overlooked, despite their importance for implementation in real-world environments.

Several technologies have been proposed to automate aspects of medical triage and COVID-19 detection, including health pods for vital sign measurement (Arslan et al., 2025), wearable Bluetooth and Internet of Medical Things (IoMT)-enabled biosensors (Alqudah; Aloqaily; Karray, 2022), robotic platforms for imaging and triage (Gabriel et al., 2023), and AI-powered decision support systems (Lyell et al., 2021). In the context of COVID-19, advances include smartphone-based cough and speech analysis for rapid, non-invasive screening (Stasak et al., 2021; Tena; Clarià; Solsona, 2021) and DL models for interpreting medical images (Ozturk et al., 2020). These systems have demonstrated high diagnostic and triage accuracy, in some cases outperforming clinicians in controlled

environments (Meral et al., 2024). Integration into telemedicine platforms has also improved care access and enabled large-scale remote monitoring (Torrente-Rodríguez et al., 2020). However, despite these advances, most existing solutions address specific components of the triage or diagnostic process in isolation and have not been integrated into cohesive, portable systems suitable for deployment in real-world clinical workflows. Key barriers to broader adoption include challenges in Information Technology (IT) infrastructure integration, lack of interoperability, concerns about data privacy, and limited validation across diverse healthcare settings (Townsend et al., 2023; Channa et al., 2021).

Finally, there remains a critical gap in the development of integrated systems that combine autonomous data acquisition with intelligent inference to support both diagnostic screening and clinical triage. Advancing such solutions requires not only technical validation, but also attention to usability, portability, and adaptability across diverse healthcare contexts.

In response to these challenges, this thesis presents a study on the integration of AI-based diagnostic and triage models into a portable equipment (in the case, the IPMA) capable of collecting physiological signals such as audio, SpO_2, blood pressure, heart rate, and temperature. The focus lies on validating these models with real-world data and assessing their usability in clinical triage scenarios, particularly in remote or resource-constrained environments.

1.6 Structure of the Thesis

In Chapter 1, the context of the study is introduced. The hypotheses, research objectives, and contributions are identified, and the motivation for this research is discussed.

Chapter 2 presents the theoretical background of this research, introducing key concepts related to clinical triage, COVID-19 diagnosis, and the use of acoustic and vital signs for patient evaluation. The chapter also covers fundamental topics in signal processing, feature extraction, and supervised learning, which support the development of intelligent triage and diagnostic systems.

Chapter 3 presents the IPMA equipment used in this research, including both the Brazilian and Ecuadorian versions. It describes the technical specifications and functionalities.

Chapter 4 presents the materials and methods used in the development of the COVID-19 detection system based on acoustic and vital signs. It describes the public datasets and the dataset acquired using the IPMA equipment, as well as the preprocessing steps, feature extraction methods, and classification models. This chapter also presents and discusses the results related to COVID-19 prediction, including model performance

for both the public datasets and the dataset acquired using the IPMA equipment. In addition, it includes the usability evaluation of the IPMA equipment based on standardized questionnaires.

Chapter 5 presents the materials and methods used in the development of the risk classification system based on the MTS. It describes the public dataset and the dataset collected with the IPMA equipment, the preprocessing steps, and the implementation of the classification models. This chapter also discusses the results related to risk prediction, including model performance and feature relevance for both the public dataset and the dataset collected with the IPMA equipment, and details its usability evaluation.

Finally, Chapter 6 provides the research conclusion. The key research findings are summarized, the main contributions achieved are discussed, and the limitations of the study are highlighted. The chapter also proposes recommendations for future work and presents the scientific publications derived from this research.

2 Theoretical Background

2.1 Clinical and Operational Context for Patient Evaluation

This section presents the clinical and technological foundations that support the development of signal-based triage and diagnostic systems. It begins with an overview of COVID-19 and its clinical features, emphasizing the relevance of physiological signals for early detection. It then examines how structured triage protocols operate in emergency care and where their limitations emerge in real-world application. Finally, it discusses how telemedicine and portable technologies are reshaping patient assessment, with particular relevance to settings where infrastructure and access are limited.

2.1.1 Clinical and Diagnostic Aspects of COVID-19

COVID-19 is an acute infectious disease caused by the SARS-CoV-2 virus, first identified in late 2019 in Wuhan, China (Wiersinga et al., 2020; Hu et al., 2021). It is primarily transmitted through respiratory droplets and aerosols during close interpersonal contact. Notably, transmission may occur from asymptomatic and presymptomatic individuals, which greatly facilitated its global spread (Wiersinga et al., 2020).

Clinically, COVID-19 manifests primarily as a respiratory illness with systemic involvement. The most frequently reported symptoms include fever, dry cough, fatigue, and shortness of breath (Wiersinga et al., 2020; Guan et al., 2020). In a study involving over 1,000 hospitalized patients in China, cough was reported in approximately 68 % of cases, while fever was initially present in 44 %, eventually rising to 88.7 % during hospitalization (Guan et al., 2020). Other common symptoms include fatigue (38 %), sputum production (33 %), and dyspnea (19 %), with gastrointestinal manifestations being relatively uncommon (Gomes, 2020). Importantly, many infected individuals, particularly young or otherwise healthy, may remain asymptomatic or experience only mild symptoms.

Although most cases are mild to moderate (around 80 %), approximately 14 % develop severe disease (e.g., pneumonia with hypoxemia), and 5–6 % progress to critical conditions, such as respiratory failure, septic shock, or multi-organ dysfunction (Hu et al., 2021). Severe COVID-19 is frequently associated with viral pneumonia and may lead to Acute Respiratory Distress Syndrome (ARDS), often requiring mechanical ventilation and intensive care. Intensive Care Unit (ICU) support is needed in about 5 % of all cases and up to 20 % of hospitalized patients (Wiersinga et al., 2020). Risk factors for disease progression include advanced age and comorbidities such as cardiovascular disease, diabetes, and chronic respiratory conditions. The wide variability in clinical presentation

has made early recognition and risk stratification difficult, especially in high-demand care settings.

In clinical settings, certain physiological signals – such as peripheral oxygen saturation (SpO_2), respiratory rate, heart rate, and body temperature — are frequently altered in COVID-19 and serve as key indicators of disease progression (Tobin; Laghi; Jubran, 2020). These signals are accessible through non-invasive sensors and can support early detection and monitoring, even outside hospital environments. Additionally, changes in respiratory acoustics, such as cough, breathing effort, and speech, have been explored as digital biomarkers in diagnostic support systems (Laguarta; Hueto; Subirana, 2020; Pahar et al., 2022).

As previously mentioned, RT-PCR is the reference standard for confirming SARS-CoV-2 infection, which is based on amplification of viral Ribonucleic Acid (RNA) from respiratory samples (Oliveira; al, 2020). While highly specific and sensitive when performed correctly, RT-PCR faces practical limitations, as: it requires specialized laboratory infrastructure, trained personnel, and multiple reagent-dependent steps (Minhas et al., 2023). Additionally, false-negative results — particularly in early stages of infection or due to sampling issues — may reach rates exceeding 30 % (Wiersinga et al., 2020). These limitations reduce its scalability, especially in emergency, remote, or resource-limited scenarios.

Given these challenges, alternative approaches that leverage physiological and acoustic signals have gained attention as complementary tools for COVID-19 screening. Their non-invasive nature, low cost, and compatibility with portable technologies make them particularly valuable for integration into telemedicine platforms and automated triage systems.

2.1.2 Medical Triage in Emergency Departments

Triage is a fundamental process in emergency care, designed to assess and prioritize patients according to the urgency of their clinical condition (Joseph et al., 2020; Fernandes et al., 2020a). In high-demand environments, it plays a key role in organizing patient flow, minimizing delays for critically ill individuals, and ensuring efficient resource allocation (Wegen et al., 2025; Fekonja et al., 2023). Several studies have demonstrated that effective triage directly influences outcomes such as hospital admission rates, mortality, and Emergency Department (ED) overcrowding (Wegen et al., 2025).

To standardize this process and reduce reliance on unstructured clinical judgment, modern triage protocols such as the MTS, ESI, and Canadian Triage and Acuity Scale (CTAS) have been widely implemented (Karjala; Eriksson, 2017; Association, 2023). These systems assign patients to predefined urgency levels based on clinical criteria. The MTS,

extensively used in Europe and Latin America, employs 53 complaint-based flowcharts and clinical discriminators to assign patients to one of five priority levels, each with a target maximum waiting time: red (immediate), orange (10 minutes), yellow (60 minutes), green (120 minutes), or blue (240 minutes) (Mackway-Jones; Marsden; Windle, 2014; Zaboli et al., 2025; Seiger et al., 2014).

A defining feature of the MTS is that it uses clinical discriminators such as abnormal vital signs, level of consciousness or age group to guide the triage nurse's decision. For example, in the pediatric protocol, a neonate presenting with general malaise and SpO_2 below 92 % on room air must be assigned to the highest urgency level (red) (Triagenet.net, 2022). This structured, rule-based logic offers a consistent and reproducible approach that aligns with algorithmic reasoning, making the MTS a suitable framework for intelligent triage support systems.

Nevertheless, practical limitations remain. Despite the protocol's structure, final decisions still depend on subjective clinical judgment. Studies have shown that triage outcomes can vary due to professional experience, workload, environmental stress, and even cognitive bias (Fekonja et al., 2023; Jatobá et al., 2018; Ingielewicz; Rychlik; Sieminski, 2024; Jachmann et al., 2025). Inter-rater agreement on MTS assessments is often moderate (Zaboli et al., 2025), and both undertriage (assigning too low a priority) and overtriage (assigning too high a priority) are common. These errors can compromise patient safety or overload the system (Porto, 2024; Sax et al., 2023), with reported undertriage rates reaching up to 25 % among patients later admitted to intensive care (Fekonja et al., 2023).

Another critical challenge is the requirement for continuous training. Correct application of triage protocols demands not only technical knowledge but also clinical experience and regular reinforcement, especially in high-pressure environments (Jatobá et al., 2018; Fekonja et al., 2023). As EDs become increasingly crowded, the cognitive demands on triage nurses intensify, raising the risk of mistake and inconsistency (Porto, 2024).

Given these limitations, there is a growing interest in developing automated triage systems that use physiological signals and ML to enhance decision-making. The structured logic of protocols like the MTS, particularly due to their reliance on discrete and measurable discriminators, offers a natural bridge to algorithmic modeling. By mapping vital signs such as SpO_2, temperature, and heart rate to triage categories, intelligent systems may improve consistency, reduce cognitive burden, and extend triage capabilities to remote or resource-constrained settings.

2.1.3 Telemedicine and Remote Patient Assessment

The COVID-19 pandemic accelerated the use of telemedicine as a strategy to maintain access to healthcare services in the face of restricted mobility, resource constraints, and overwhelming demand (Portnoy; Waller; Elliott, 2020; Ftouni et al., 2022). In this context, remote patient assessment tools, including the home-based collection of physiological and audio data, emerged as viable alternatives to in-person evaluations. These tools support triage, clinical monitoring, and early identification of clinical deterioration (Smith et al., 2020; Po et al., 2024).

Several studies have highlighted the effectiveness of telemonitoring systems in reducing unnecessary emergency visits and hospital admissions. For example, remote transmission of vital signs such as oxygen saturation, heart rate, and body temperature has been implemented through dashboards monitored by clinical teams, with high rates of patient adherence and professional satisfaction (Simonetti et al., 2023). These systems are especially relevant for diseases like COVID-19, in which patients may deteriorate rapidly despite initially mild presentations.

The evolution of wearable sensors, mobile health applications, and integrated digital platforms has expanded the scope of measurable clinical data beyond vital signs. Acoustic signals, including cough, breathing, and speech, can now be captured and analyzed remotely using smartphones and embedded microphones (Laguarta; Hueto; Subirana, 2020; Kong et al., 2024; Topol, 2020). As a result, telemedicine has evolved from simple video consultations into a data-rich environment that enables real-time signal analysis and automated clinical decision-making.

Together, these developments highlight the potential of portable, intelligent systems to support signal-based triage and diagnosis, particularly in contexts where conventional infrastructure is limited.

2.2 Physiological Signals for Health Assessment Fundamentals

This section presents the theoretical background on the use of physiological signals in clinical decision-making and respiratory assessment. It introduces key concepts and types of signals used for health monitoring and risk classification, including vital signs and acoustic signals such as cough, breathing, and speech. The clinical relevance of these modalities is examined, along with their integration into diagnostic and triage support systems, to provide a foundation for the signal-based methods explored in this research.

2.2.1 Vital Signs in Clinical Assessment

Blood Pressure

Blood Pressure (BP) is a fundamental physiological parameter that reflects the force exerted by circulating blood on the arterial walls. It is typically expressed as Systolic Blood Pressure (SBP), the peak pressure during ventricular contraction, and Diastolic Blood Pressure (DBP), the lowest pressure during ventricular relaxation (Clark; Kruse, 1990). Maintaining normal BP levels (between 120/80 mmHg) is essential to ensure adequate tissue perfusion and preserve cardiovascular homeostasis (Zhou et al., 2021a).

During the COVID-19 pandemic, emerging evidence raised interest in the relationship between SARS-CoV-2 and cardiovascular regulation. Several studies reported measurable changes in BP among infected individuals, particularly in those with pre-existing hypertension (Gotanda et al., 2022; Ikram; Pillay, 2022). For instance, a large-scale study identified significant increases in both SBP (1.79 mmHg) and DBP (1.30 mmHg) during the early stages of the pandemic (Gotanda et al., 2022). In hospital settings, low DBP at admission has been independently associated with increased mortality in COVID-19 patients (Ikram; Pillay, 2022). These findings underscore the potential of BP as both a cardiovascular outcome and a prognostic marker in infectious diseases.

In clinical triage settings, BP remains a key parameter used to support rapid risk stratification. Initial readings may reflect a wide range of acute conditions, from hypertensive crises to hypovolemic (low blood volume) or septic shock. Abnormal BP values trigger specific clinical pathways and often influence the urgency level assigned to the patient (Erwander; Agvall; Ivarsson, 2025). Accurate and timely BP assessment at the point of entry is therefore critical to identify high-risk individuals, guide early intervention, and improve outcomes (Yazici et al., 2024).

In this study, BP is automatically acquired by the IPMA and included among the physiological variables used for triage and diagnostic modeling. Its clinical relevance and prognostic value make it a key component of the multimodal dataset analyzed.

Oxygen Saturation

Oxygen saturation (SpO_2) reflects the percentage of hemoglobin molecules bound to oxygen and serves as a core indicator of respiratory efficiency and peripheral tissue oxygenation (Tobin; Laghi; Jubran, 2020; Swenson; Hardin, 2022). In healthy individuals, values typically range between 95 and 99 %, while levels below 92 % are considered clinically significant and may indicate hypoxemia (reduced oxygen levels in the blood) requiring prompt medical intervention (Fawzy et al., 2023).

SpO_2 monitoring gained particular importance during the COVID-19 pandemic, when clinicians began to report cases of “silent hypoxemia”, characterized by severe

oxygen desaturation in the absence of perceived breathing difficulty (Tobin; Laghi; Jubran, 2020). In response, home-based SpO_2 monitoring became a key strategy for outpatient care. In a study involving 105 COVID-19 patients, Simonetti et al. (2023) demonstrated the feasibility and safety of telemonitoring SpO_2, body temperature, and heart rate, achieving over 85 % adherence and resulting in only three hospitalizations. This approach enabled early detection of clinical deterioration and helped avoid unnecessary admissions, reinforcing the value of SpO_2 as a readily measurable and clinically informative parameter for remote risk assessment.

Beyond the pandemic, SpO_2 remains a critical component of emergency triage systems, particularly for identifying early signs of respiratory failure or circulatory compromise. Triage protocols such as the MTS incorporate SpO_2 thresholds — commonly below 92 % — as criteria for high-priority classification, given the elevated risk of tissue hypoxia and clinical deterioration (Mackway-Jones; Marsden; Windle, 2013). Moreover, low SpO_2 at admission has been independently associated with adverse outcomes in several conditions, such as pneumonia acquired outside the hospital setting (community-acquired pneumonia) (Majumdar et al., 2011). Accurate and timely measurement of oxygen saturation is therefore essential for informed clinical decision-making.

In this study, SpO_2 is automatically acquired by the IPMA and incorporated into the multimodal dataset used for both triage and diagnostic inference. Its non-invasive nature, strong physiological relevance, and widespread clinical use makes it a key input for the development of intelligent risk classification models.

Heart Rate

Heart Rate (HR) refers to the number of cardiac contractions per minute and is a key indicator of cardiovascular integrity and autonomic nervous system function. In healthy adults, resting HR typically ranges from 60 to 100 beats per minute, with sustained deviations potentially reflecting cardiac pathology, systemic inflammation, metabolic imbalance, or pharmacological influence (Sapra; Malik; Bhandari, 2023).

In the context of COVID-19, HR alterations have been frequently observed, particularly among patients with severe or critical illness (Chen et al., 2020). Sinus tachycardia emerged as the most common arrhythmia, often attributed to direct myocardial injury, systemic hyperinflammation, and autonomic dysregulation (Chen et al., 2020; Sharma et al., 2021b).

In emergency care, HR is typically one of the first parameters measured upon arrival and serves as a rapid indicator of circulatory and systemic stress. Triage protocols such as the MTS incorporate HR thresholds to help determine the urgency of care (Mackway-Jones; Marsden; Windle, 2014). Elevated HR has been associated with a higher likelihood of Medical Emergency Conditions (MECs) (Yazici et al., 2024) and increased mortality

risk, particularly among older adults (Erwander; Agvall; Ivarsson, 2025), reinforcing its importance in identifying high-risk patients early.

Given its clinical relevance and ease of acquisition, HR is included in the dataset collected by the IPMA and serves as a key input for the AI-based models developed in this study.

Body Temperature

Body temperature (BT) is a fundamental physiological parameter that reflects the balance between heat production and loss, with normal core BT typically maintained within a narrow range of approximately $36 - 37$ °C (Yazici et al., 2024). Fever, defined as a core BT above 38 °C, is a key host defense mechanism against infection and a widely recognized indicator of systemic illness (Yazici et al., 2024; Shen et al., 2021).

In the assessment of COVID-19 patients, fever is among the most common presenting symptoms, reported in approximately 88.7 % of cases after hospitalization (Guan et al., 2020). However, the clinical relevance of BT extends beyond the presence or absence of fever. Temperature trajectories have been identified as valuable predictors of disease severity and outcomes. In a cohort of 5,903 hospitalized COVID-19 patients, four distinct BT patterns were observed: 25 % were slow to recover from high fever, 25 % recovered quickly, 36 % remained normothermic, and 15 % were hypothermic (Bhavani et al., 2022).

In the ED setting, BT remains a critical vital sign for triaging patients and estimating prognosis. Deviations from normal BT (hypothermia or hyperthermia) are associated with significant physiological stress and have been strongly linked to increased mortality (Yu et al., 2012; Erwander; Agvall; Ivarsson, 2025). For example, in a study of acutely poisoned patients, an initial BT below 34 °C or above 38 °C at triage was significantly associated with higher in-hospital mortality. Optimal predictive thresholds were identified as BT < 36 °C or > 37 °C (Yu et al., 2012). Although individual vital signs may have limited predictive power alone, their combined evaluation, including BT, is essential for effective risk stratification, especially in vulnerable groups such as elderly patients (Erwander; Agvall; Ivarsson, 2025).

In this study, BT is automatically acquired by the IPMA and used as part of the multimodal input for predictive modeling. Its dynamic behavior and strong association with clinical deterioration enhance the system's ability to stratify patient risk in real time.

2.2.2 Acoustic Signals in Respiratory Health Assessment

Acoustic signals produced during coughing, breathing, and speech are physiological manifestations with significant potential for health analysis, particularly in clinical scenarios involving acute respiratory diseases (Landry et al., 2025). These signals are generated by

natural processes of the human body, can be recorded using simple acoustic devices (such as conventional microphones), and are suitable for digital processing and analysis (Brown et al., 2020; Gupta et al., 2021). The COVID-19 pandemic highlighted their relevance as accessible, non-invasive, and low-cost markers for clinical evaluation and rapid triage, especially in high-demand settings such as emergency departments (Landry et al., 2025).

Cough is a protective reflex triggered by airway irritation and is commonly classified as dry or productive, continuous or episodic, depending on the underlying pathology (Murgia et al., 2020). In the context of COVID-19, dry cough was one of the most frequently reported symptoms, often appearing in the early stages of infection. While some machine learning studies have reported high classification performance using forced cough sounds for COVID-19 detection (Topol, 2020; Laguarta; Hueto; Subirana, 2020), clinical data indicate that only 59% of infected individuals present with dry cough (Benisek, 2023). This discrepancy reinforces the importance of multimodal approaches for robust screening and diagnosis.

Respiratory sounds, in turn, are generated by the movement of air through the lungs and airways. Normal respiratory sounds are soft and continuous, whereas abnormal or adventitious sounds — such as crackles, wheezes, rhonchi, and stridor — indicate alterations in respiratory mechanics and are often associated with inflammation, obstruction, or secretion accumulation (Gupta et al., 2021). In COVID-19, such abnormal sounds have been documented in patients with pulmonary involvement and have shown potential for the early detection of viral pneumonia (Furman et al., 2022). In triage settings, recognizing these sounds can help identify clinical severity markers such as hypoxemia or impending respiratory failure (Gupta et al., 2021).

Speech is a complex acoustic signal resulting from the coordinated action of respiration, phonation, and articulation (Waage; Iwarsson, 2024). Alterations in speech may reflect respiratory muscle fatigue, dyspnea, or neurological impairment (Alvarado et al., 2023). Parameters such as fundamental frequency, phonation time, pause duration, and intonation patterns are sensitive to changes in physiological state and have been explored as input features in predictive models for COVID-19 and other respiratory conditions (Sharma; Umapathy; Krishnan, 2022). For instance, patients with respiratory compromise often exhibit interrupted speech, reduced vocal intensity, or abnormal breathing pauses (Waage; Iwarsson, 2024). In clinical practice, these alterations may serve as indirect indicators of respiratory distress.

In this study, cough, breathing, and speech signals are acquired using the IPMA's embedded microphone and analyzed as part of the multimodal dataset used for triage and diagnostic inference. Their non-invasive nature and sensitivity to respiratory dysfunction make them valuable inputs for intelligent, signal-based screening systems applicable in both hospital and remote care scenarios.

2.3 Time-Frequency Image Representation of Audio Signals

The analysis of physiological sounds in medical contexts, particularly for respiratory conditions such as COVID-19, requires signal processing techniques capable of capturing both temporal and spectral characteristics of acoustic signals (Tena; Claria; Solsona, 2022; Aytekin et al., 2023; Kumar; Alphonse, 2022). This section examines the use of Mel spectrograms as a biologically inspired time-frequency representation method for transforming audio recordings of cough, speech, and breathing into two-dimensional visual representations that retain perceptually and diagnostically relevant features.

2.3.1 Theoretical Foundation of Mel Spectrograms

The Mel spectrogram represents a biologically-inspired time-frequency representation of audio signals that incorporates psychoacoustic principles of human auditory perception. The fundamental basis of this approach lies in the mel scale, a perceptual scale of pitches that reflects how humans naturally perceive sound frequencies (Rabiner; Schafer, 2010). The mel scale is defined by the relationship between physical frequency and perceptual pitch.

The mathematical transformation from frequency F in Hertz to the mel scale follows the natural logarithmic relationship:

$$\text{pitch}_{\text{mel}} = 1127 \cdot \ln\left(1 + \frac{F}{700}\right) \tag{2.1}$$

This transformation reflects the auditory system's critical band structure, wherein frequency resolution decreases at higher frequencies. As a result, the mel scale aligns signal analysis with human hearing, allowing for better emphasis on perceptually relevant features (Rabiner; Schafer, 2010; Nanni et al., 2021).

2.3.2 Mel spectrogram Construction

The process of generating Mel spectrogram begins with the segmentation of the audio signal into overlapping frames. Each segment is multiplied by a windowing function to reduce spectral leakage (Aytekin et al., 2023). The Short-Time Fourier Transform (STFT) is then applied to convert each frame from the time domain to the frequency domain:

$$A[k, n] = \sum_{m=-\infty}^{\infty} a[m] \cdot w[m - n] \cdot \exp\left(-j\frac{2\pi km}{N}\right) \tag{2.2}$$

where $a[m]$ is the discrete-time signal, $w[n]$ is the window function of length N, and k denotes the frequency bin. A commonly used window function is the Hanning window, given by:

$$w[n] = 0.5 - 0.5 \cdot \cos\left(\frac{2\pi n}{N}\right) \tag{2.3}$$

After computing the magnitude spectrum for each frame, the output is passed through a mel filter bank composed of triangular filters distributed according to the mel scale (Aytekin et al., 2023). Each filter aggregates energy from specific frequency bands, and the result is compressed using a logarithmic function to approximate the nonlinearity of human loudness perception:

$$\text{Mel spectrogram}[mel, n] = \log\left(\left|A\left[\frac{f(mel)N}{f_s}, n\right]\right|^2\right) \tag{2.4}$$

Here, $f(mel)$ maps the mel-scale value back to its corresponding frequency in Hertz, and f_s denotes the sampling frequency of the signal (Rabiner; Schafer, 2010). The resulting Mel spectrogram encodes the temporal and spectral energy distribution of the signal, emphasizing perceptually relevant components and providing a robust input for diagnostic models.

In our study, Mel spectrograms derived from cough, speech, and breathing signals serve as the basis for image-based feature extraction using texture descriptors, as detailed in the following section.

2.4 Audio Texture Feature Extraction

Texture-based analysis of Mel spectrograms has shown strong potential for the automatic classification of cough sounds, particularly in the context of COVID-19 detection. By transforming audio signals into time-frequency images, it becomes possible to apply well-established image processing techniques to extract discriminative features. Among the available methods, Local Binary Pattern (LBP) and Local Ternary Pattern (LTP) stand out for their robustness and simplicity. These descriptors, originally developed for texture recognition in images (Sharma et al., 2020), have been adapted to extract local structural information from spectrograms in a compact and noise-tolerant manner. These characteristics make LBP and LTP particularly suitable for representing the subtle acoustic variations found in respiratory sounds. The following sections detail the theoretical foundations and applications of both descriptors in the context of audio-based classification.

2.4.1 Local Binary Patterns

LBP is a widely used descriptor in image processing for texture classification, known for its computational simplicity and effectiveness in capturing local spatial patterns. It has been successfully applied in diverse domains, including lung sound classification (Sengupta; Sahidullah; Saha, 2017), pathological speech analysis (Sharma et al., 2020; Sharma et al., 2021a), COVID-19 screening (Sharma; Umapathy; Krishnan, 2022), scene classification (Abidin; Togneri; Sohel, 2018), and snore detection (Demir et al., 2018).

The original LBP operator, commonly referred to as "uniform LBP", operates on a 3×3 neighborhood and encodes the local texture based on grayscale intensity comparisons (Ojala; Pietikainen; Maenpaa, 2002). Each neighboring pixel is compared to the center: values greater than or equal to the center are assigned 1; otherwise, 0. The resulting binary pattern — typically ordered clockwise starting from the top-left neighbor — is then converted to its decimal equivalent and used to label the center pixel in the LBP-transformed image.

Figure 1 illustrates this process. In the example, the center pixel has an intensity of 51. After comparing with its eight neighbors and applying the binary thresholding rule, the resulting binary pattern is 11000011, which corresponds to the decimal value 195. This value replaces the center pixel in the resulting LBP image.

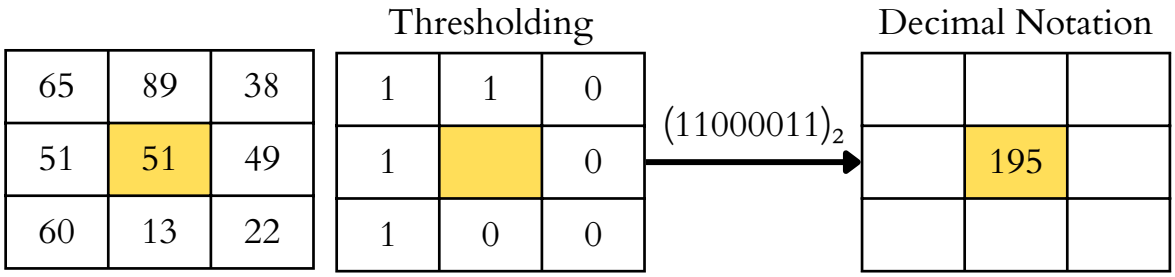


Figure 1 – Example of LBP encoding: thresholding neighbors of the center pixel (51), forming the binary pattern 11000011 and its decimal value 195.

The LBP-transformed image is subsequently divided into non-overlapping blocks, and a histogram of pattern occurrences is computed for each block. These histograms are then concatenated to form the final feature vector, which compactly represents the texture distribution of the spectrogram and can be directly used as input to a classification model.

2.4.2 Local Ternary Patterns

LTP extend the LBP operator by introducing a third quantization level, increasing robustness to noise in near-uniform image regions (Tan; Triggs, 2010a). It has been applied in various contexts such as speech emotion recognition (Sönmez; Varol, 2020), fall detection (Adnan et al., 2018), and heart sound classification (Er, 2021).

Instead of encoding neighbors as strictly greater or smaller than the center pixel, LTP introduces a user-defined threshold τ to create a ternary pattern (Tan; Triggs, 2010a). For a neighbor pixel value i_p and center pixel i_c, the ternary encoding function is defined as:

$$s(i_p, i_c) = \begin{cases} 1, & \text{if } i_p \geq i_c + \tau \\ 0, & \text{if } |i_p - i_c| < \tau \\ -1, & \text{if } i_p \leq i_c - \tau \end{cases} \tag{2.5}$$

This produces a three-level code, which is then split into two binary patterns: the upper pattern (positive values mapped to 1, others to 0), and the lower pattern (negative values mapped to 1, others to 0). Each binary pattern is converted to a decimal value, just like in LBP, and the final feature vector can be represented as a concatenation of both parts (Tan; Triggs, 2010a; Turan; Lam, 2018).

Figure 2 illustrates the LTP encoding process using a 3×3 neighborhood. Unlike the LBP example, the input matrix (leftmost) already represents the result of a prior thresholding step applied to a grayscale image — typically a Mel spectrogram — yielding ternary values (-1, 0, and 1). In the first row, positive values are retained and converted into the upper binary pattern, resulting in 11000010, which corresponds to 194 in decimal. In the second row, the lower binary pattern encodes the negative values, producing 00101100, equivalent to 44 in decimal. These two values together form the LTP representation of the local neighborhood.

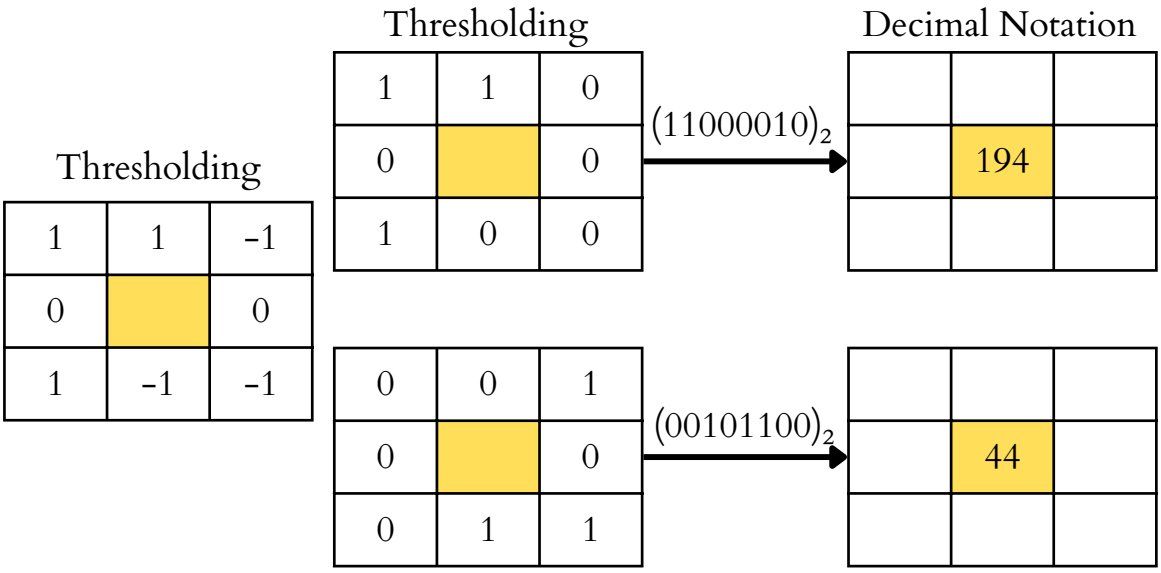


Figure 2 – Example of LTP encoding from a ternary image. The ternary pattern is split into upper and lower binary components, yielding decimal values 194 and 44, respectively.

After applying the LBP or LTP encoding to the Mel spectrogram, the resulting descriptor image is divided into a non-overlapping grid of cells. For each cell, a histogram of code occurrences is computed. All histograms are L_2-normalized and concatenated to form a single feature vector, which compactly encodes the local time–frequency texture of the signal while preserving spatial layout. This feature vector is then used as input to the classification stage.

2.5 Missing Data and Imputation Techniques

Incomplete data is a common and often unavoidable challenge in biomedical datasets, particularly in real-world clinical settings. If not properly addressed, missing values may compromise the validity of statistical analyses and the reliability of predictive models. A widely used strategy to handle this issue is known as imputation, in which the analyst replaces the missing data with estimated values based on observed information. Such an approach is called "imputation", because one is imputing a value of the variable for those subjects with missing data on that variable (Austin et al., 2021). This section reviews the main types of missing data mechanisms and introduces commonly used imputation techniques suitable for biomedical data.

2.5.1 Missing Data in Biomedical Contexts

Missing data is a prevalent issue in biomedical research, significantly impacting the integrity of research findings, particularly in epidemiological and clinical studies. Three fundamental mechanisms categorize missing data: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) (Pedersen et al., 2017; Bennett, 2001). MCAR occurs when the likelihood of missingness is entirely independent of both observed and unobserved variables, whereas MAR is dependent on observed variables but independent of the unobserved ones. In contrast, MNAR occurs when missingness depends on the values of the unobserved data themselves (Pedersen et al., 2017; Carpenter; Smuk, 2021; Bennett, 2001).

In clinical epidemiology, missing data typically fall under the MAR category. For instance, younger individuals may be less likely to have weight or BP recorded simply because they visit healthcare facilities less frequently, making age an observable factor related to missingness (Pedersen et al., 2017). This scenario indicates that MAR assumptions may be plausible but still require careful consideration and justification within the research design and statistical analysis.

The presence of missing data poses significant methodological challenges as it may introduce bias, compromise statistical power, and potentially distort associations and conclusions drawn from the analysis (Pedersen et al., 2017; Carpenter; Smuk, 2021).

Traditional methods for addressing missing data, such as complete-case analysis, single imputation, and worst-case/best-case scenario analyses, have been widely used but are generally suboptimal as they either waste valuable data or underestimate variability (Pedersen et al., 2017; Bennett, 2001). For instance, complete-case analysis may lead to biased estimates and loss of statistical power if the missingness is related to key study variables (Pedersen et al., 2017; Hegde et al., 2019).

Advanced statistical approaches, such as Multiple Imputation (MI) and maximum likelihood methods, have emerged as more robust and efficient methods for handling missing data (Carpenter; Smuk, 2021; Pedersen et al., 2017; Hegde et al., 2019). MI, in particular, is beneficial as it creates several plausible datasets by imputing missing values based on the distribution of observed data, thus adequately accounting for the uncertainty inherent in missing data (Pedersen et al., 2017; Carpenter; Smuk, 2021). Techniques like Multiple Imputation by Chained Equations (MICE) and Probabilistic Principal Component Analysis (PPCA) have demonstrated improved performance in preserving data integrity and analytical validity compared to traditional methods (Hegde et al., 2019).

Given the importance of proper handling and transparent reporting, best practices recommend explicit documentation of missing data patterns, justification of the chosen imputation strategy, and use of sensitivity analyses to assess robustness (Pedersen et al., 2017; Carpenter; Smuk, 2021; Hegde et al., 2019). In the following subsection, the MICE algorithm, its underlying assumptions, and its applicability to the biomedical datasets used in this work are described in detail.

2.5.2 Multiple Imputation by Chained Equations

MICE, also known as fully conditional specification or sequential regression multiple imputation, is a powerful statistical method for handling missing data, particularly in biomedical research contexts. Unlike traditional imputation methods that assume a joint multivariate distribution, MICE operates by iteratively fitting conditional distributions for each variable with missing values. Each incomplete variable is modeled as a function of the other variables, allowing flexibility in handling different types of data such as continuous, binary, ordinal, or count data (Azur et al., 2011; Junaid et al., 2025).

The MICE algorithm proceeds in a structured sequence of steps. Initially, missing values are filled using simple methods such as mean or median imputation. Then, for each variable with missing data, a regression model is built using the other variables as predictors (Azur et al., 2011). The missing values for this variable are then imputed based on predictions from this model. This process is repeated for all incomplete variables in the dataset. The entire cycle — imputing each variable one-by-one — is then iterated multiple times until the imputations stabilize and convergence is achieved (Azur et al., 2011).

Formally, the missing values of each variable are iteratively imputed by sampling from the conditional distribution:

$$X_j^{(mis)} \sim P\left(X_j | X_{-j}^{(obs)}, X_{-j}^{(mis)}, \theta_j\right), \tag{2.6}$$

where X_{-j} represents all variables except X_j, and θ_j denotes parameters specific to the model for X_j.

Several simulation studies highlight the efficacy of MICE compared to traditional single-value imputation methods, emphasizing reduced bias and improved efficiency in parameter estimates (Deng et al., 2016; Hegde et al., 2019; Junaid et al., 2025). MICE has been shown to maintain robust performance even with missing proportions up to 50 %, though caution is advised for higher proportions, as accuracy significantly declines beyond 70 % missingness (Junaid et al., 2025).

Despite its strengths, MICE has certain limitations. It assumes MAR, meaning that the probability of a value being missing depends only on the observed data and not on the value that is missing. Violations of this assumption, common in clinical datasets, may potentially introduce bias into the imputed datasets (Azur et al., 2011; Hegde et al., 2019). To mitigate this issue, sensitivity analyses are recommended to assess the robustness of the imputation results against deviations from the MAR assumption (Azur et al., 2011; Carpenter; Smuk, 2021).

2.6 Machine Learning Techniques

Machine learning (ML) techniques have become increasingly central to the development of predictive models in healthcare, enabling the extraction of meaningful patterns from complex and high-dimensional biomedical data. Their ability to support automated diagnosis, clinical decision-making, and real-time triage has driven their widespread adoption across a range of medical applications. This section introduces the supervised ML algorithms employed in this study, detailing their underlying mechanisms, advantages, and relevance to biomedical classification tasks. Particular attention is given to decision trees, ensemble methods, distance-based classifiers, SVMs, and neural networks, all of which have demonstrated effectiveness in handling structured physiological and clinical data.

2.6.1 Supervised Machine Learning Algorithms

Decision Tree

A Decision Tree (DT) is a non-parametric, supervised ML algorithm that creates a hierarchical tree-like structure to model decisions and their potential outcomes. It is commonly employed for classification and regression tasks due to its interpretability and

straightforward representation. Each internal node of a DT corresponds to a test on an attribute, with branches representing the outcomes of these tests, and leaf nodes indicating final decisions or class labels (Learning, 1997; Gupta et al., 2017). Figure 3 illustrates a typical DT applied to a weather classification task. At each internal node, a condition is evaluated, and branches lead to further tests or final classification labels at the leaves.

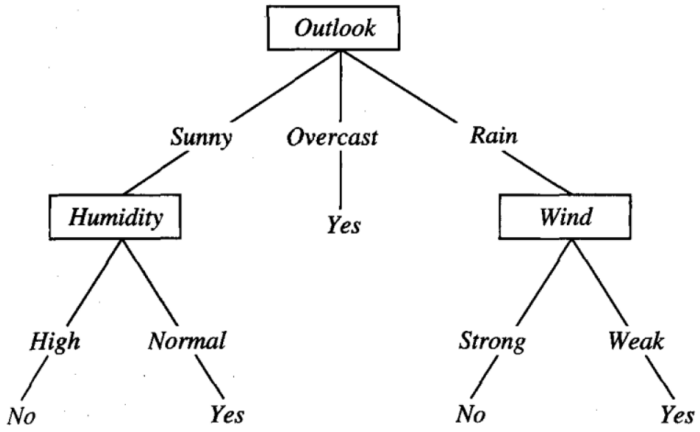


Figure 3 – Example of a DT for weather-based classification. Internal nodes denote decision criteria (e.g., Outlook, Humidity, Wind), and leaf nodes represent the predicted class labels (e.g., Yes or No). Source: Learning (1997).

Classification and Regression Tree (CART), introduced by Breiman et al. (1984), represent a specific implementation of DTs capable of handling both classification and regression problems. CART generates binary trees by recursively partitioning the dataset based on the attributes that maximize purity or minimize impurity within subsets (Breiman et al., 1984; Gupta et al., 2017). To determine the optimal splits during tree construction, CART employs the Gini Index. The Gini Index measures node impurity, reflecting the likelihood of an incorrect classification of an observation if randomly labeled according to class distribution within the node. Formally, for a node t with classes $i = 1, 2, ..., C$, the Gini Index is defined as:

$$\text{Gini}(t) = 1 - \sum_{i=1}^{C} p_i^2, \tag{2.7}$$

where p_i denotes the proportion of instances in node t that belong to class i (Breiman et al., 1984; Learning, 1997). A Gini impurity of 0 indicates a perfectly pure node, meaning that all observations within the node belong to a single class — the ideal scenario for classification.

Advantages of CART include its interpretability, ease of implementation, robustness in handling missing data, and inherent feature selection capabilities (Gupta et al., 2017). However, CART can be sensitive to data variations, potentially resulting in unstable trees (Gupta et al., 2017). Despite this, its transparent decision-making process and

effective management of diverse data types make it highly valuable in biomedical and clinical applications (Breiman et al., 1984; Gupta et al., 2017).

In this study, DT was selected for its ability to provide interpretable decision paths and for its frequent application in biomedical prediction tasks, including triage systems (Lu et al., 2023; Liu et al., 2025; Fernandes et al., 2020b).

Random Forest

Random Forest (RF) is an ensemble ML algorithm introduced by Breiman (2001), which constructs multiple decision trees and aggregates their results to enhance classification accuracy and robustness against noise. Unlike a single DT, which relies on one hierarchical tree structure for prediction, RF builds a multitude of trees using bootstrapped samples of the original dataset, a process known as bagging (bootstrap aggregating). Each tree within the RF is constructed by considering a random subset of available attributes at each split, which increases diversity among the trees and reduces the overall variance, typically leading to improved generalization performance (Altman; Krzywinski, 2017).

The main difference between RF and DT lies in the aggregation process. While DT predictions come from a single hierarchical structure prone to instability with small variations in data, RF mitigates this limitation by combining outputs from multiple trees. The final classification in RF is typically decided by majority voting across the individual tree predictions, significantly reducing variance without substantially increasing bias (Breiman, 2001; Altman; Krzywinski, 2017). Furthermore, RF utilizes the Gini Index, as in CART, but benefits from decreased sensitivity to noise and overfitting due to the ensemble strategy. Consequently, RFs are often more robust, more accurate, and provide more reliable generalizations compared to single DTs, particularly in complex and noisy biomedical datasets.

RF is selected in this study for its proven performance in structured healthcare data and its ability to manage feature interactions without requiring prior transformation. RF is applied because of its strong generalization and ensemble strength that is supposed to contribute to classification under real-world data variability (Levin et al., 2018; Lu et al., 2023; Liu et al., 2025).

XGBoost

XGBoost (eXtreme Gradient Boosting) is a scalable and optimized implementation of the gradient tree boosting technique, which builds predictive models by iteratively combining weak learners through functional gradient descent. This method is grounded in the work of Friedman (2001), who formalized boosting as a gradient-based optimization process in function space. In this framework, the model is constructed additively, where each successive tree is trained to approximate the negative gradient (also known as the

pseudo-residual) of a differentiable loss function with respect to the current ensemble's prediction. This approach enables the model to progressively correct mistakes made by previous learners, thereby minimizing the overall loss function.

Chen and Guestrin (2016) extended this paradigm by developing a highly efficient and scalable system tailored to large-scale and real-world ML tasks. XGBoost introduces several algorithmic enhancements, including a sparsity-aware split-finding mechanism capable of efficiently handling missing values and sparse input features; a regularized learning objective that penalizes model complexity to mitigate overfitting; and a weighted quantile sketch algorithm, which facilitates fast and accurate histogram-based split selection in approximate tree learning (Chen; Guestrin, 2016). These innovations allow XGBoost to operate effectively in memory-constrained or distributed environments, achieving high predictive performance across diverse datasets.

In this research, XGBoost is employed to evaluate its performance on structured clinical features and explore its advantages over traditional bagging approaches. It is selected due to its consistent ranking among top performers in structured data competitions and medical ML tasks (Lu et al., 2023).

k-Nearest Neighbors

The k-Nearest Neighbors (kNN) algorithm is a simple, intuitive, and widely-used non-parametric classification method. It classifies a new data point based on the majority label among its k closest neighbors in the training dataset (Zhang, 2016). The concept is straightforward: for a given instance, the algorithm calculates the distance from this point to all other points in the training set.

Let $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ represent two feature vectors in an n-dimensional space. Commonly used distance measures include the Euclidean distance, given by:

$$\text{Euclidean}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}, \tag{2.8}$$

and the Manhattan distance, defined as:

$$\text{Manhattan}(x, y) = \sum_{i=1}^{n} |x_i - y_i|. \tag{2.9}$$

The choice of k significantly impacts the classifier's accuracy: small values of k tend to result in noisy classifications, whereas larger values create smoother decision boundaries. Although computationally intensive for large datasets, the kNN algorithm remains highly popular in biomedical applications due to its simplicity, interpretability, and minimal assumptions about data distribution (Gupta et al., 2017; Zhang, 2016; Uddin et al., 2022).

In this Thesis, kNN was employed due to its compatibility with structured numerical features and its suitability as a baseline model for benchmarking more complex algorithms. Its non-parametric nature and simplicity provided a valuable reference for assessing the added predictive value of ensemble and boosting methods (Elhaj et al., 2023; Liu et al., 2025).

Support Vector Machine

Support Vector Machines (SVMs) are a powerful and robust classification and regression algorithm, widely recognized for their generalization capabilities and optimal solutions. Introduced by Vapnik, SVMs aim to find an optimal separation hyperplane that maximizes the margin between different classes in a high-dimensional feature space (Cervantes et al., 2020; Awad; Khanna, 2015). This approach minimizes classification errors on training data and enhances the model's ability to generalize to unseen data, aligning with the Structural Risk Minimization (SRM) principle.

For linearly separable data, the decision function is defined by a hyperplane $w^T x + b = 0$. The goal is to maximize the margin between classes, which can be formulated as:

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{2.10}$$

subject to:

$$y_i(w^T x_i + b) \geq 1 \quad \forall i \tag{2.11}$$

In cases where data are not linearly separable, the concept of soft margin is introduced, allowing for some misclassifications by incorporating non-negative slack variables $\xi_i \geq 0$ (Cervantes et al., 2020). The objective then becomes to minimize $||w||^2 + C\sum \xi_i$, where C is a regularization parameter that balances the trade-off between maximizing the margin and minimizing classification errors.

In this study, SVM was employed due to its strong generalization performance and success in prior studies involving cough and respiratory sounds classification (Shati; Hassan; Datta, 2023).

Multilayer Perceptron

A Multilayer Perceptron (MLP), also known as a feedforward neural network, is a foundational model in deep learning. It is structured as a directed graph where information flows strictly forward — from the input layer, through one or more hidden layers, to the output layer — without cycles or feedback connections (Goodfellow; Bengio; Courville,

2016; Haykin, 2009). This architecture distinguishes MLPs from recurrent neural networks, which incorporate temporal dynamics via feedback loops.

An MLP consists of an input layer, one or more hidden layers, and an output layer, as shown in Figure 4. Each layer is composed of multiple artificial neurons, or units. The connections between units in successive layers are associated with weights, and each unit (except those in the input layer) has a bias term. The output of each unit in a hidden or output layer is computed by applying an activation function to the weighted sum of its inputs plus the bias (Haykin, 2009). Common activation functions include the Rectified Linear Unit (ReLU), sigmoid, and hyperbolic tangent (tanh) (Goodfellow; Bengio; Courville, 2016). Mathematically, the output of a neuron j is given by:

$$a_j = g\left(\sum_i w_{ji} x_i + b_j\right), \tag{2.12}$$

where w_{ji} are the weights, x_i are the inputs, b_j is the bias, and $g(\cdot)$ is the activation function.

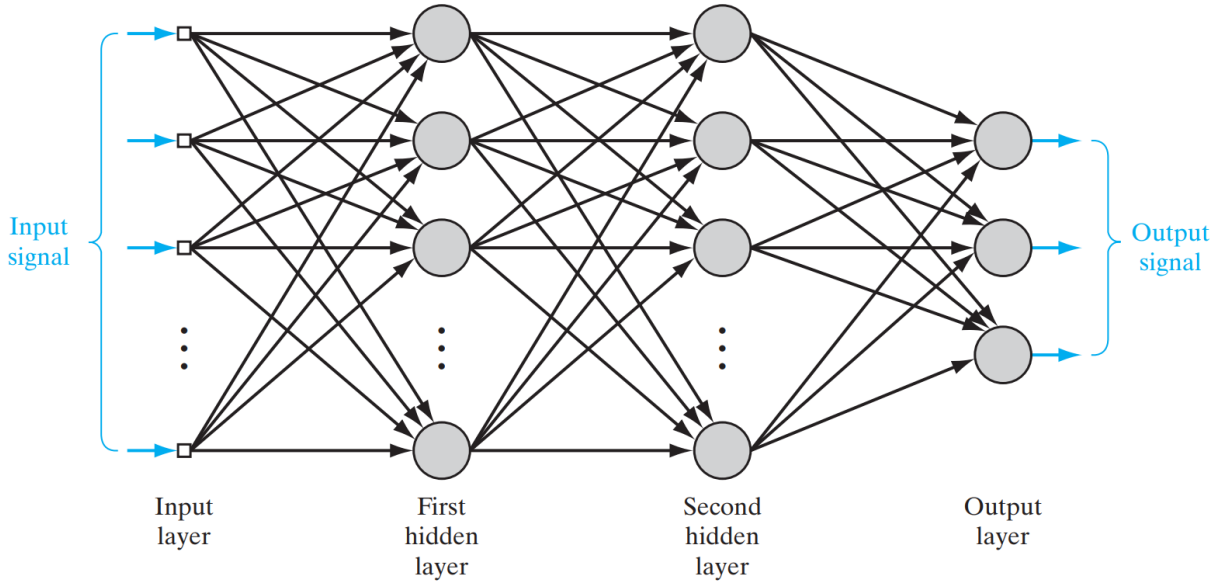


Figure 4 – Illustration of an MLP architecture with an input layer, two hidden layers, and an output layer. Each unit in a layer is fully connected to the units in the next layer. Source: Haykin (2009).

Training an MLP involves adjusting the network's weights and biases to minimize a predefined cost (or loss) function. This is typically done using gradient-based optimization algorithms such as Stochastic Gradient Descent (SGD), which iteratively update parameters in the direction opposite to the gradient of the cost function (Haykin, 2009). The backpropagation algorithm, introduced by Rumelhart, Hinton and Williams (1986), is central to training MLPs. It efficiently computes the gradients of the cost function with

respect to all network parameters by propagating error signals backward through the network using the chain rule of calculus.

The approach used in this Thesis used MLP in triage models, leveraging its capacity to capture complex non-linear patterns for improved classification results (Porto, 2024; Liu et al., 2025).

Meta-Ensemble Strategies

Meta-ensemble methods, such as voting and stacking, aim to enhance predictive performance by leveraging the diversity and complementary strengths of multiple individual classifiers (Zhou, 2025; Başer; Evran; Cifci, 2025).

In the case of voting, the final decision is derived from the predictions of base learners, using either a majority rule (hard voting) or aggregated probabilities (soft voting). Hard voting selects the class that receives the highest number of individual votes (Zhou, 2025). This approach is especially effective when classifiers are diverse and individually competent. Formally, for a sample x, the final decision $H(x)$ is given by:

$$H(x) = \begin{cases} c_j, & \text{if } \sum_{i=1}^{T} h_i^j(x) > \frac{1}{2} \sum_{k=1}^{L} \sum_{i=1}^{T} h_i^k(x) \\ \text{rejection}, & \text{otherwise} \end{cases} \tag{2.13}$$

In this formulation, j denotes the index of the candidate class being evaluated, and c_j is the corresponding class label. The variable T represents the total number of base classifiers in the ensemble. The term $h_i^j(x)$ indicates the output of the i-th classifier for class c_j, which in hard voting is typically binary — equal to 1 if the classifier assigns x to c_j, and 0 otherwise. Similarly, $h_i^k(x)$ refers to the output of the i-th classifier for class c_k, where k is another class index. Finally, L denotes the total number of possible classes in the classification problem.

Soft voting, on the other hand, aggregates the predicted class probabilities across classifiers, assigning x to the class with the highest average (or weighted) probability. This method generally provides superior performance when the classifiers are well-calibrated (Zhou, 2025). The unweighted and weighted formulations are as follows, respectively:

$$H^j(x) = \frac{1}{T} \sum_{i=1}^{T} h_i^j(x) \tag{2.14}$$

$$H^j(x) = \sum_{i=1}^{T} w_i h_i^j(x) \tag{2.15}$$

where w_i reflects the relative importance or trust assigned to classifier h_i, with j indexing the class being evaluated and T denoting the number of classifiers in the ensemble. The term $h_i^j(x)$ represents the probability, predicted by classifier i, that sample x belongs to class c_j, where $h_i^j(x) \in [0, 1]$. Thus, h_i refers to the i-th classifier, and $h_i^j(x)$ specifies its output for a particular class.

Stacking, or stacked generalization, offers a more sophisticated ensemble approach. Instead of relying on fixed aggregation rules, stacking trains a meta-learner to optimally combine the predictions of base models (Başer; Evran; Cifci, 2025). First, base classifiers are trained on the original data. Their predictions (usually obtained through cross-validation to avoid overfitting) form a new dataset of meta-features (Zhou, 2025). This dataset is used to train a meta-learner — typically a simple but robust model, such as Logistic Regression (LR) — to learn the best combination of base outputs. The effectiveness of stacking lies in its ability to capture complex interactions between base learners and exploit their individual strengths, often outperforming single models and simpler ensemble strategies.

2.6.2 Model Evaluation Metrics

In the context of clinical classification tasks, proper evaluation metrics are essential for assessing the effectiveness and safety of predictive models. This is particularly critical when models are intended to support diagnostic or triage decisions, where misclassification can lead to delayed treatment or unnecessary interventions.

A wide range of studies in medical AI adopts classical metrics derived from the confusion matrix, including True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) (Fernandes et al., 2020a; Spooner et al., 2025). From these, two of the most commonly reported metrics are accuracy (ACC) and F1-score, defined respectively as:

$$\text{ACC} = \frac{TN + TP}{TN + TP + FN + FP} \tag{2.16}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{2.17}$$

ACC represents the proportion of correctly classified instances out of the total number of samples (Spooner et al., 2025; Başer; Evran; Cifci, 2025). However, in class — imbalanced scenarios—which are common in medical datasets — ACC can be misleading, since it can hide inadequate sensitivity to less frequent outcomes. To address this, many studies also report the F1-score, which balances precision and sensitivity:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.18}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2.19}$$

The F1-score is particularly valued in healthcare-related AI research for its robustness to imbalance and its emphasis on correct identification of positive cases (Elhaj et al., 2023; Spooner et al., 2025). Its adoption has become standard in benchmarking efforts for COVID-19 detection, vital sign analysis, and other critical care applications, underscoring the need for metrics that reflect both clinical relevance and statistical rigor.

2.7 Usability Assessment

Evaluating usability is a critical component in the development and validation of healthcare technologies. It provides insights into how effectively, efficiently, and satisfactorily users interact with a given system. To this end, standardized questionnaires such as the System Usability Scale (SUS) and the Post-Study System Usability Questionnaire (PSSUQ) have been widely adopted for quantifying user experience. The following subsections describe the theoretical basis, structure, and scoring procedures of these two instruments.

2.7.1 System Usability Scale

The System Usability Scale (SUS) is a widely adopted method for assessing the usability of technological systems, including software, hardware, and interactive tools. Developed by John Brooke in 1986, it offers a quick yet robust way to capture user perceptions of effectiveness, efficiency, and satisfaction through a standardized questionnaire (Brooke, 2013).

In this context, the central question is how to quantitatively measure a user's experience with the system under evaluation. SUS addresses this challenge using a ten-item Likert-scale questionnaire, where users rate their level of agreement with specific statements on a scale from 1 (strongly disagree) to 5 (strongly agree). Importantly, the questionnaire is structured with alternating polarity: odd-numbered items express negative sentiment about the system, while even-numbered items are positively phrased, providing internal consistency and balance to the evaluation.

The score is computed using a specific formula that accounts for the polarity of each question. For odd-numbered items, 1 is subtracted from the user's response; for even-numbered items, the response is subtracted from 5. The adjusted values are summed and multiplied by 2.5 to yield a final SUS score ranging from 0 to 100, as described in Equation 2.20:

$$T = 2.5 \left[\left(\sum_{\text{odd}} p - 1 \right) + \left(\sum_{\text{even}} 5 - p \right) \right] \tag{2.20}$$

Although SUS scores range between 0 and 100, this scale is not a percentile. According to Lewis (2018), the average usability score across systems typically falls around 68, which corresponds to the 50th percentile. Therefore, interpretation of SUS scores should consider benchmarking graphs rather than raw numerical values. The full list of SUS items is provided in Appendix A.

2.7.2 Post Study System Usability Questionnaire

The Post Study System Usability Questionnaire (PSSUQ) is a standardized instrument designed to evaluate user satisfaction and system usability immediately after task execution (Lewis, 1995). Developed in 1988 at IBM as part of the System Usability Metrics (SUMs) project, the PSSUQ has since become a widely accepted method for assessing user experience in computing environments.

The main objective of this tool is to quantify user perceptions regarding different aspects of system interaction. It consists of 19 items rated on a 7-point Likert scale, where 1 corresponds to "Strongly Agree", 4 is "Neutral", and 7 means "Strongly Disagree". The overall usability score is obtained by calculating the mean of the responses, with lower values indicating more favorable usability ratings.

To enhance interpretability, the questionnaire is divided into three subscales. The System Usefulness (SYSUSE) score reflects responses to questions 1 through 6; the Information Quality (INFOQUAL) score covers questions 7 to 12; and the Interface Quality (INTERQUAL) score corresponds to the average of questions 13 to 15. Although question 16 is not included in any subscale, it is considered in the global usability score. Questions 17 to 19, while part of the original version, are often excluded in practice due to their optional nature (Lewis, 2018).

In both global and subscale scores, lower values indicate better usability, with 1 representing the most favorable evaluation and 7 the least. Since the neutral midpoint is 4, scores closer to 1 suggest a more satisfactory user experience. The complete set of PSSUQ items can be found in Appendix B.

2.8 Statistical Methods

A range of statistical methods was used to analyze the data, which where selected based on the type of variables, distribution characteristics, and sample size. Both parametric and nonparametric approaches were applied, depending on whether assumptions such as normality and homogeneity of variances were met. When applicable, post-hoc tests were conducted to explore group differences following significant results. The following sections summarize the main statistical tests used.

2.8.1 Chi-squared Test

The Chi-squared (χ^2) test assesses independence between two categorical variables by comparing observed and expected frequencies (Kim, 2017). It requires an adequately large sample (no more than 20 % of cells with expected frequencies < 5, no cell < 1). Significant results with multiple levels require post-hoc pairwise comparisons with Bonferroni correction (Kim, 2017).

2.8.2 Shapiro-Wilk Test

The Shapiro-Wilk test is a powerful formal test for normality. A significant p-value indicates non-normality. However, its sensitivity to sample size is important: for very large samples ($n > 300$), it may detect negligible deviations, while for small samples, it may lack power to detect true non-normality. Interpretation should always consider sample size and the robustness of subsequent analyses (Kim, 2012).

2.8.3 Parametric Tests

Parametric tests assume specific population distribution parameters, typically normality, homogeneity of variances, and independence of observations. When these assumptions are met, parametric tests generally offer higher statistical power.

Student's t-test

The Student's t-test compares the means of two groups (Kim, 2019). It can be applied to independent groups (independent samples t-test) or correlated data (paired samples t-test). For independent groups, two variations exist: the classical Student's t-test assumes equal population variances, whereas Welch's t-test is more robust for unequal variances or sample sizes (Kim, 2019). A core assumption is data normality, which should be assessed using tests like Shapiro-Wilk (Kim, 2012; Kim, 2019). If normality is violated, especially with small samples ($n < 25$), nonparametric alternatives like the Mann-Whitney U test are preferred (Kim, 2019). The t-distribution, central to the t-test, approximates the z-distribution as degrees of freedom increase (Kim, 2019).

Analysis of Variance

Analysis of Variance (ANOVA) extends the comparison of means to three or more groups, determining if significant differences exist among them (Kim, 2016). One-way ANOVA is used for a single independent variable. Key assumptions include normality and homogeneity of variances. When ANOVA yields a significant result, post-hoc multiple comparison procedures are necessary to identify specific group differences while controlling the family-wise error rate (Kim, 2015). Among post-hoc tests, Tukey's Honestly Significant

Difference (HSD) is widely used and considered one of the most preferable methods for pairwise comparisons when all possible comparisons between groups are of interest. It uses the studentized range statistic to determine critical values and is appropriate when group sizes are equal (Kim, 2015).

2.8.4 Nonparametric Tests

Nonparametric methods, or "distribution-free tests", are used when parametric assumptions (especially normality) are not met (Kim, 2014a). They are suitable for skewed data, outliers, small samples, and ordinal data (Kim, 2014a). While offering flexibility, they may have less power than parametric tests when assumptions are met, and handling tied values can be complex (Kim, 2014a).

Wilcoxon Rank-Sum Test (Mann-Whitney U Test)

This test is a nonparametric alternative to the independent samples t-test, comparing two independent groups with at least ordinal data when normality is not assumed (Kim, 2014a). It assesses if two sets of scores differ systematically by comparing ranks.

Wilcoxon Signed-Rank Test

The nonparametric equivalent to the paired samples t-test, is compares two related or correlated data sets (e.g., pre-post measurements) with at least ordinal data and non-normal distribution (Kim, 2014a). It determines if there is a systematic difference between paired scores by analyzing signed ranks of differences.

Kruskal-Wallis Test

This nonparametric method compares three or more independent groups, serving as an alternative to one-way ANOVA when assumptions are violated (Kim, 2014b). It uses ranks to test whether the samples differ systematically. When the Kruskal-Wallis test results in statistical significance, post-hoc tests are necessary to identify which specific groups differ from each other. The Dunn's test is a commonly used post-hoc method following a significant Kruskal-Wallis result, as it adjusts p-values for multiple comparisons, controlling the family-wise Type I error rate (Dunn, 1964).

Friedman Test

The Friedman Test is a nonparametric alternative to repeated measures ANOVA. It is used in within-subject designs when there are three or more related measurements, and the data do not follow a normal distribution (Kim, 2014b). It assesses differences across

multiple measurements from the same subjects. Significant results necessitate post-hoc paired comparisons (e.g., Wilcoxon Signed-Rank with Bonferroni correction) (Kim, 2014b).

3 Integrated Portable Medical Assistant

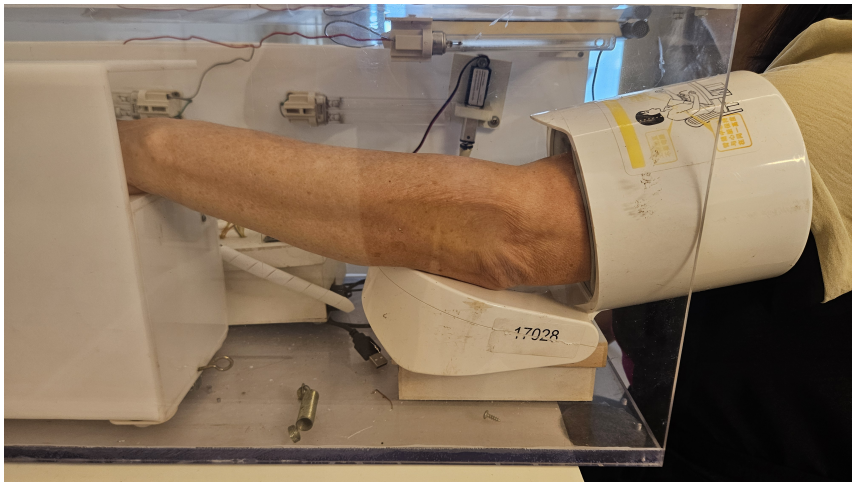
In the context of global health crises and constrained healthcare infrastructure, the demand for autonomous, reproducible, and biosafe tools for initial patient assessment has become increasingly urgent. A central challenge is to design equipment capable of collecting reliable physiological and acoustic data without requiring specialized personnel or compromising sanitary conditions. To address this challenge, the Integrated Portable Medical Assistant (IPMA) was developed and validated (Villa-Parra et al., 2022; Ormaza-Siguenza et al., 2024), a multimodal system that automates data acquisition while preserving medical certification standards. This chapter introduces the Brazilian and Ecuadorian versions of the IPMA, detailing their hardware architecture, operational workflow, and design considerations.

3.1 Brazilian Version of the IPMA

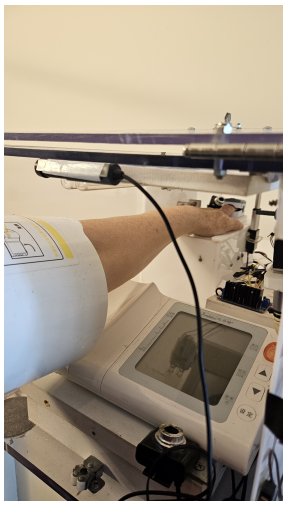
As part of our study, the Brazilian version of the IPMA (designed to autonomously acquire physiological and audio data for COVID-19 screening and triage) was used (Villa-Parra et al., 2022). This version served as the reference throughout all experimental stages and usability evaluations presented in this thesis. The equipment was constructed to prioritize safety, automation, and reproducibility in constrained environments, ensuring data quality while avoiding human manipulation.

The Brazilian IPMA consists of a 70cm×30cm×33cm transparent polycarbonate box. Internally, a white acrylic roller-based structure supports the insertion of arms of varying sizes and ensures proper positioning of the sensors relative to the skin. Housed within this internal structure are the core measurement components: a non-contact thermometer, a fingertip pulse oximeter, and a fully automatic sphygmomanometer. All devices used are medically certified and remain unmodified to preserve their regulatory approval. Instead of altering their internal electronics — which could compromise medical certification — the system initiates measurements by mechanically simulating manual activation through linear actuators. The readings displayed on each device are captured by embedded miniature cameras aligned with the screens. The actuators are powered by a 12V battery located in the system's base, while the ultraviolet-C (UVC) disinfection lamps are connected directly to the Alternative Current (AC) mains supply. Figures 5 and 6 illustrate the equipment's physical layout and sterilization mechanism.

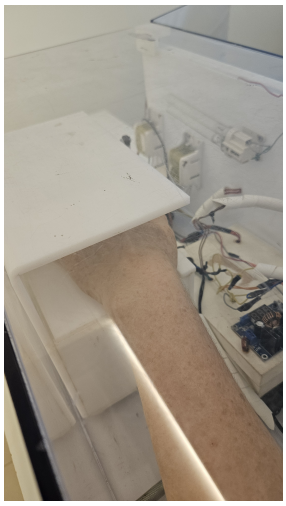
To acquire physiological and acoustic data, the IPMA employs a fingertip pulse oximeter (model FS10K, Hunan Accurate Bio-Medical Technology Co.), a non-contact infrared thermometer (model E122, Bioland), and a fully automatic wrist sphygmomano-



(a)

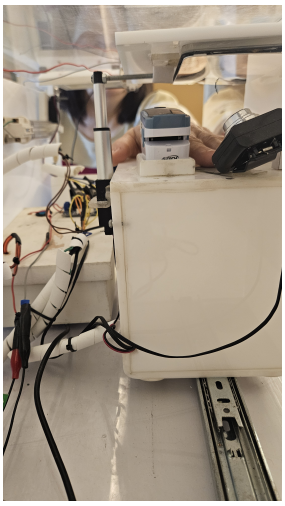


(b)



(c)

(d)



(e)

Figure 5 – Experimental setup of the Brazilian IPMA equipment. (a) Arm insertion into the device. (b) Internal view of arm in the blood pressure monitor. (c) Top view of hand placement. (d) Internal view of hand on the oximeter. (e) Front view of hand on the oximeter and cart guide rail.

meter (Banglijian), which eliminates the need for manual cuff inflation. Additionally, a unidirectional condenser microphone (KNUP KP-911) is integrated into the enclosure to capture voice and forced cough sounds. Once the procedure is initiated via the user interface, the complete sequence — including data entry, audio recording, vital sign acquisition, and automated disinfection — is executed autonomously. This design ensures compliance with medical device regulations, minimizes user error, and enhances operational safety.

To guarantee biosafety, especially given the infectious nature of SARS-CoV-2, three 9W UVC TUV PL9 lamps (Philips) are mounted inside the IPMA. These lamps are controlled by a Raspberry Pi using a relay and are activated for 2.5 minutes following each measurement cycle. A redundant safety mechanism was implemented using a limit switch on the IPMA's door: the UVC circuit is physically interrupted if the door is

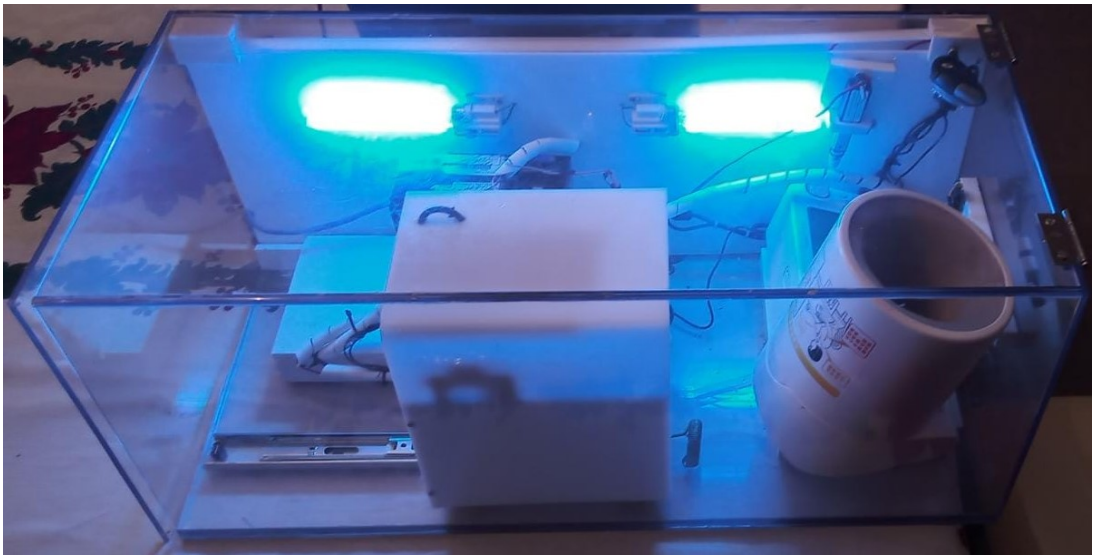


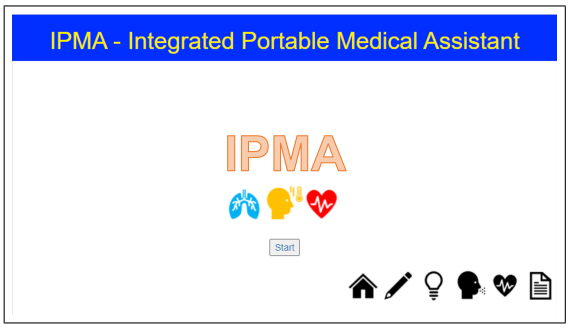
Figure 6 – Application of UVC light within the Brazilian IPMA to ensure surface sterilization.

open, preventing exposure to UVC radiation. While the blue glow of the lamps is visible through the polycarbonate, the material blocks UVC wavelengths, thus ensuring the user is not exposed to harmful radiation, in accordance with biosafety guidelines (University of Washington, 2017). Figure 6 shows the equipment during active disinfection.

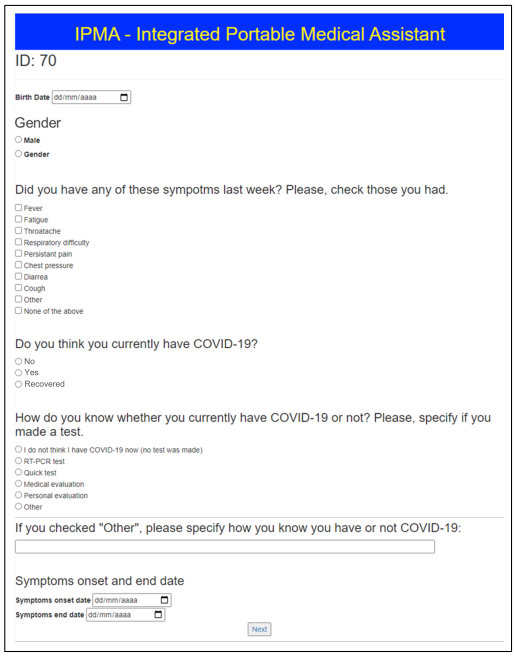
The system's user interface was developed using Python, Flask, and JavaScript. Once powered, the IPMA creates a Wi-Fi access point and serves its web application through an embedded Flask server. The interface guides the user through all steps, as shown in Figure 7, starting with a form that collects personal and health-related information relevant to COVID-19, such as date of birth, gender, and symptom history. Each session is assigned a unique test identifier to maintain data traceability while preserving individual privacy.

Following the form submission, the interface prompts the user to perform a sequence of speech and respiratory sound tasks in a predefined order. Upon opening the IPMA door, the subject is first instructed to read a predefined phonetically balanced sentence in Portuguese, providing consistent audio features across subjects. Next, the participant performs five forced breathing cycles toward the microphone, followed by five forced coughs, both guided by auditory and visual cues. The entire sequence lasts approximately 15 seconds, and all audio signals — sentence, breathing, and cough — are stored for subsequent signal processing and classification. These strategies align with prior studies that leverage respiratory acoustics for COVID-19 detection (Sharma; Umapathy; Krishnan, 2022), which explores sound-based diagnostics through ML.

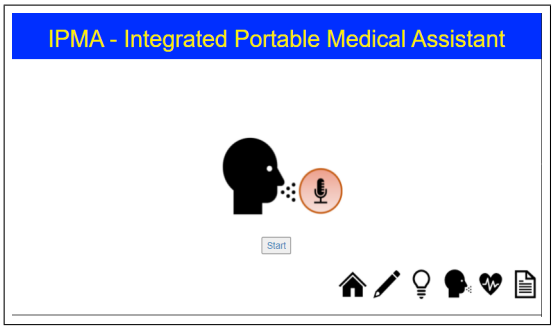
Once the audio capture is complete, the physiological measurement phase is initiated. Upon inserting their arm, the participant triggers a fully autonomous sequence in which each medical instrument is actuated, readings are captured via camera, and automatically



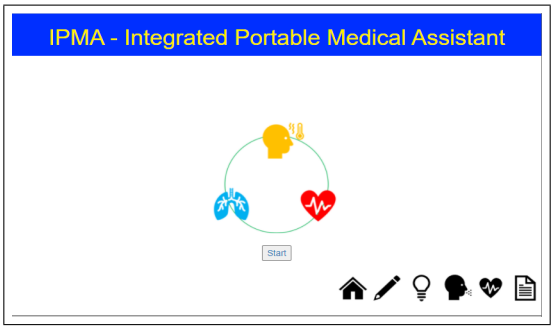
(a)



(b)

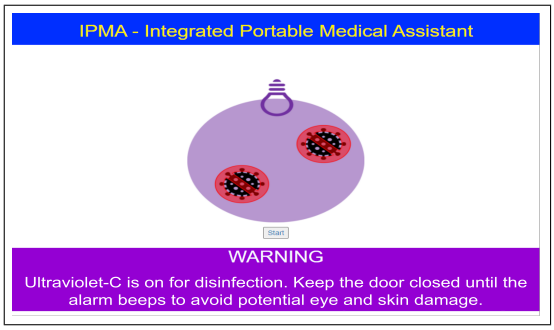


(c)



(d)

(e)



(f)

Figure 7 – User interface of the IPMA equipment. The sequence illustrates key screens presented during the interaction: (a) initial welcome screen; (b) user registration form; (c) instruction screen guiding the user to biomedical sounds (speech, and forced breathing and cough); (d) screen to initiate data acquisition; (e) thank you message displayed upon completion; and (f) final disinfection step.

identified using an Optical Character Recognition (OCR) program. The results are then presented on a summary screen (Figure 8). This closed-loop workflow significantly reduces

Figure 8 – Result screen showing captured physiological measurements from the Brazilian IPMA.

the risk of user error and supports high reproducibility, which is essential for reliable classification in health screening systems.

4 COVID-19 Respiratory Sound and Vital Signs Analysis and Classification

This chapter presents the development and evaluation of a multimodal classification system for COVID-19 inference using biomedical audio and physiological signals. The approach combines public datasets with data collected through the IPMA, allowing for the comparison of feature extraction methods, signal modalities, and classification strategies.

The complete experimental pipeline, including audio representation via Mel-spectrograms, texture-based feature extraction (LBP and LTP), classification using ML models, and late fusion of decision scores are described here. In addition to performance metrics, statistical analyses were applied to validate the results and assess the system's usability through standardized questionnaires. The following sections detail the materials, methods, experiments, and findings that support the feasibility of COVID-19 detection using low-cost, non-invasive signals.

4.1 Materials and Methods

4.1.1 Overview of the Proposed System for COVID-19 Inference

Figure 9 illustrates the proposed workflow for inferring COVID-19 from biomedical signals, integrating both physiological data and audio modalities — namely cough, breathing, and speech. The pipeline is designed to process multimodal biomedical data and generate discriminative features for COVID-19 detection.

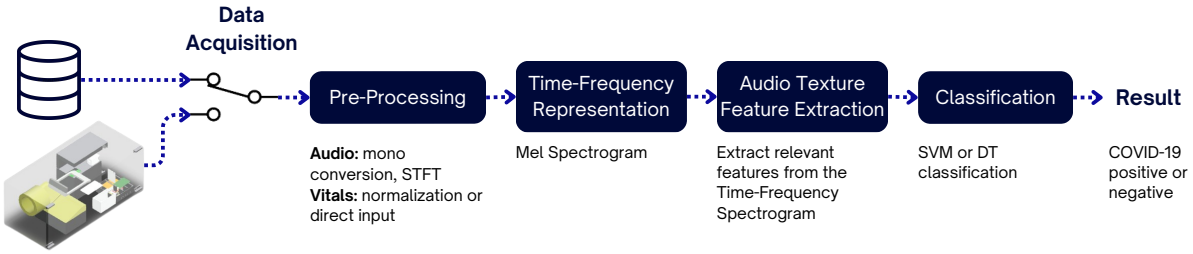


Figure 9 – Proposed schematic of the processing pipeline for COVID-19 screening using cough, speech, breath sounds, and physiological signals.

The process begins with data acquisition using the IPMA equipment, which simultaneously collects audio and vital sign measurements. Audio signals are converted into Mel-spectrograms, yielding time-frequency image representations. Texture-based features are then extracted from these images using LBP or LTP descriptors. These features are combined with physiological measurements, forming a multimodal feature vector. This

vector is used to train ML models — such as SVM or DT classifiers — to infer the subject's COVID-19 infection status. Figure 9 provides an overview of each processing stage, which will be detailed in the following sections.

4.1.2 Subjects and Ethical Aspects of Public Datasets

CCS and CSS Datasets

The COVID-19 Cough Sub-Challenge (CCS) and the COVID-19 Speech Sub-Challenge (CSS) are part of the INTERSPEECH 2021 Computational Paralinguistics Challenge (ComParE), which aims to advance speech-based health diagnostics (Schuller et al., 2021). These datasets were selected in the present work due to their focus on respiratory-related audio signals for COVID-19 inference.

The CCS dataset comprises cough recordings, while the CSS dataset includes speech samples. In both cases, data were collected from COVID-19 positive and negative individuals through the "COVID-19 Sounds App", available via a web interface and mobile applications. Participants were instructed to provide one to three forced coughs and to repeat the phrase "*I hope my data can help to manage the virus pandemic*" up to three times. After submission, all recordings were manually reviewed, resampled, and standardized to 16 kHz mono/16-bit format (Schuller et al., 2021).

The datasets were provided by Cambridge University under a mutual research agreement. Their use in this study was approved by the Department of Computer Science and Technology, in accordance with institutional ethical standards and committee guidelines. In total, the CCS dataset contains recordings from 725 subjects, and the CSS dataset comprises 893 subjects. It is worth noting that in both datasets, the number of COVID-19 negative samples is substantially higher than the number of positive ones. This class imbalance was taken into account during model design and performance evaluation.

Cambridge KDD Dataset

The Cambridge Knowledge Discovery and Data Mining (KDD) dataset was also included in this study due to its relevance for respiratory sound analysis in the context of COVID-19 (Brown et al., 2020). Developed as part of the same initiative as CCS and CSS, this dataset was made available by Cambridge University under a research agreement and in accordance with ethical guidelines.

Data were collected via the "COVID-19 Sounds App" and originally included cough, speech, and breathing recordings. However, only cough and breath signals were available for analysis in this work (Brown et al., 2020). Unlike CCS and CSS, the Cambridge KDD dataset is organized into web-based and Android subsets; here, only the Android partition was selected, given its consistency in recording protocols and the widespread

use of smartphones (Sharma; Umapathy; Krishnan, 2022). Participants were instructed to produce three voluntary coughs and perform three to five deep mouth breaths.

The dataset comprises recordings from 115 subjects. As with the previous datasets, COVID-19 negative individuals are more prevalent than positive ones. Moreover, the dataset includes both symptomatic and asymptomatic individuals in each class, along with a distinct group of asthma patients who are COVID-19 negative and contributed cough and breath recordings. Although the number of subjects is limited, the total number of samples is larger due to data augmentation procedures applied by the original authors (Brown et al., 2020). In this study, we used the dataset as provided by the original authors, including the augmented samples available in the official release.

Wearable Device Dataset

The Wearable Device dataset was selected for this study due to its inclusion of real-time physiological data relevant to COVID-19 monitoring. It was developed using an Internet of Things (IoT)-based health monitoring prototype designed to collect clinical parameters from patients under quarantine conditions (Hussain, 2021). This system records heart rate, body temperature, and peripheral SpO_2, transmitting the data through an Application Programming Interface (API) that enables centralized storage and remote access (Bassam et al., 2021). The API functions as a repository for continuous monitoring and infection level assessment. The dataset used in this study consists of recordings from 1,084 individuals. Unlike the other datasets employed, the majority of samples in the Wearable Device dataset correspond to COVID-19 positive individuals, accounting for approximately 87 % of the total. This strong class imbalance in favor of positive cases was considered during experimental design and model evaluation.

4.1.3 Data Collection and Ethical Aspects of the IPMA COVID-19 Dataset

To support the development and validation of the proposed system, a real-world dataset was collected using the IPMA equipment. This study was approved by the Federal University of Espírito Santo (UFES) Research Ethics Committee (CAAE: 64800116.9.0000.5542) and by the Technical Research Commission of the Municipal Health Department of Vitória (SEMUS/PMV), Brazil. The official authorization letter issued by SEMUS/PMV is included in Appendix C in Portuguese.

The experimental protocol was carried out at the Basic Health Unit (Unidade Básica de Saúde, UBS) of Jardim da Penha and the UFES (campus Goiabeiras), both located in Vitória, Brazil. Subjects were selected based on specific inclusion and exclusion criteria. Volunteers were eligible if they were 18 years of age or older, while individuals under 18 were excluded. Participation was conditional on the voluntary signing of the informed consent form, in accordance with ethical guidelines.

The IPMA was used in this study to collect audio signals and clinical data from both COVID-19 positive and negative individuals. Before data acquisition, each participant received a verbal explanation of the protocol, supported by a visual poster displayed at the UBS and UFES (Figure 10). Participants who agreed to take part filled out a questionnaire addressing their symptoms and medical history.

Figure 10 – Posters (in Portuguese) displayed at the UBS and UFES to support the explanation of the proposed system and the biomedical data collection process.

The data collection took place in March 2023 as part of the validation process for the IPMA. To ensure subject privacy and comfort, a hospital-grade folding three-part ward screen was used during all recordings (Figure 11). Throughout the protocol, strict biosafety measures were observed: all participants and researchers wore protective face masks, with researchers specifically using N95 respirators in accordance with WHO guidelines (World Health Organization, 2023). These precautions contributed to creating a safe and controlled environment for accurate data acquisition.

Each participant was seated comfortably, with feet flat on the ground, and positioned their arm into the equipment, placing their index finger in the integrated oximeter (Figure 12). The procedure began with the recording of respiratory sounds. Participants were instructed to read a phonetically balanced sentence in Portuguese displayed on the screen: "*É de fundamental importância encontrar uma solução comum*" ("It is of funda-

Figure 11 – Example of participant interaction during the data collection phase.

mental importance to find a common solution"). Then, they performed five deep mouth breaths and five voluntary coughs (Figure 13a). Simultaneously, physiological parameters were recorded, including SpO_2, heart rate, body temperature, and systolic and diastolic blood pressure (Figure 13b). The synchronization of all signals during the same acquisition session ensured temporal alignment among modalities, enabling integration of audio and physiological data for subsequent analysis.

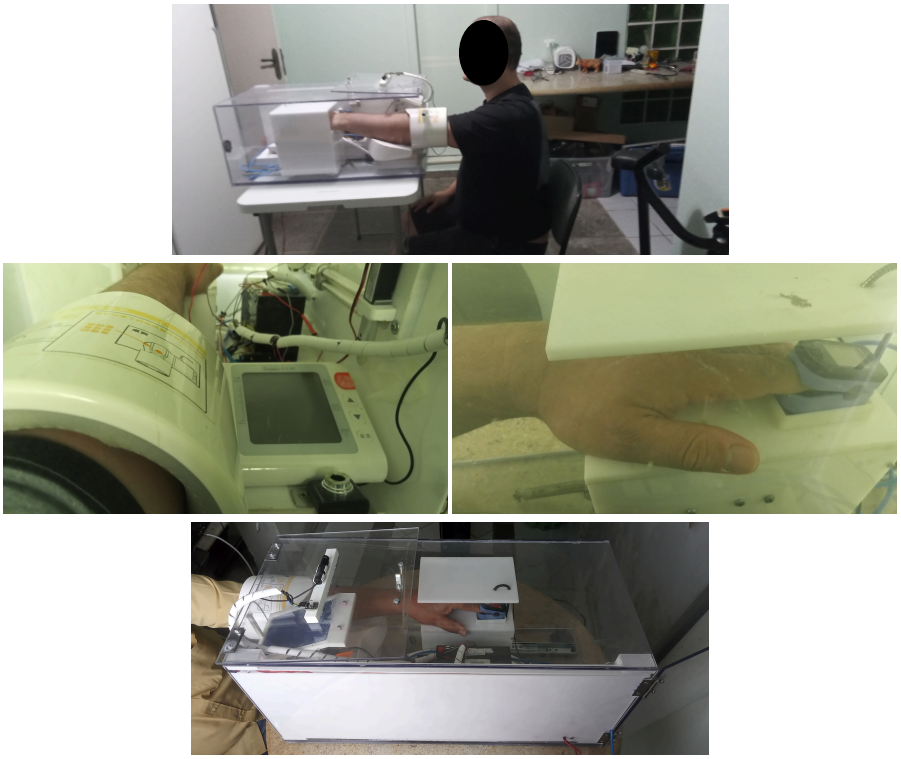


Figure 12 – Illustration of arm and hand placement for data collection using the IPMA equipment.

To mitigate the risk of cross-contamination, UVC surface disinfection was applied between participants (Figure 13c). At the end of the procedure, subjects completed a

second questionnaire to assess their experience with the IPMA equipment. A total of eleven individuals participated in the study. This limited sample size was primarily due to the timing of data collection, which occurred at the tail end of the COVID-19 pandemic, when the prevalence of positive cases was significantly reduced. Consequently, recruitment efforts focused on individuals with confirmed infection, further restricting the sample size. This protocol enabled the collection of synchronized audio and physiological data under controlled conditions, serving as the basis for developing and evaluating ML models aimed at real-time medical triage.

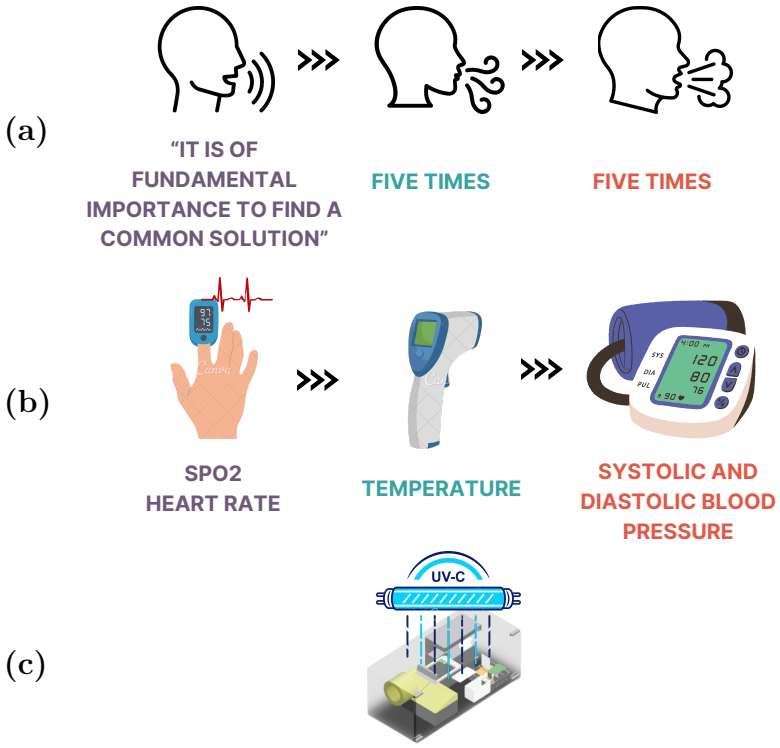


Figure 13 – Sequence of data collection: (a) speech, breath, and cough; (b) SpO_2, heart rate, temperature, and systolic and diastolic blood pressure; and (c) UVC disinfection.

4.1.4 Audio Representation and Texture-Based Feature Extraction

To ensure comparability across different audio modalities, each raw signal (cough, speech, and breath) was first standardized by converting it to mono. Then a Mel-spectrogram was computed for each signal to obtain a perceptually motivated time-frequency representation. This process involved applying the Short-Time Fourier Transform (STFT) using 20 ms Hamming windows with 50 % overlap and 1024 Fast Fourier Transform (FFT) bins obtained through zero-padding. The resulting complex spectrum $S(n, f) = |S(n, f)|e^{j\theta(n,f)}$ was converted into magnitude values and subsequently passed through a Mel filterbank composed of 80 triangular filters spaced linearly in the Mel scale. Fi-

nally, the magnitude spectrum was transformed into the log scale as $S(n, f) = \log(|S(n, f)|)$, as commonly applied in auditory perception modeling (Himawan; Towsey; Roe, 2018; Zhou et al., 2021b). Figure 14 illustrates examples of COVID-19 cough, speech, and breath signals both in the time domain and as Mel-spectrograms.

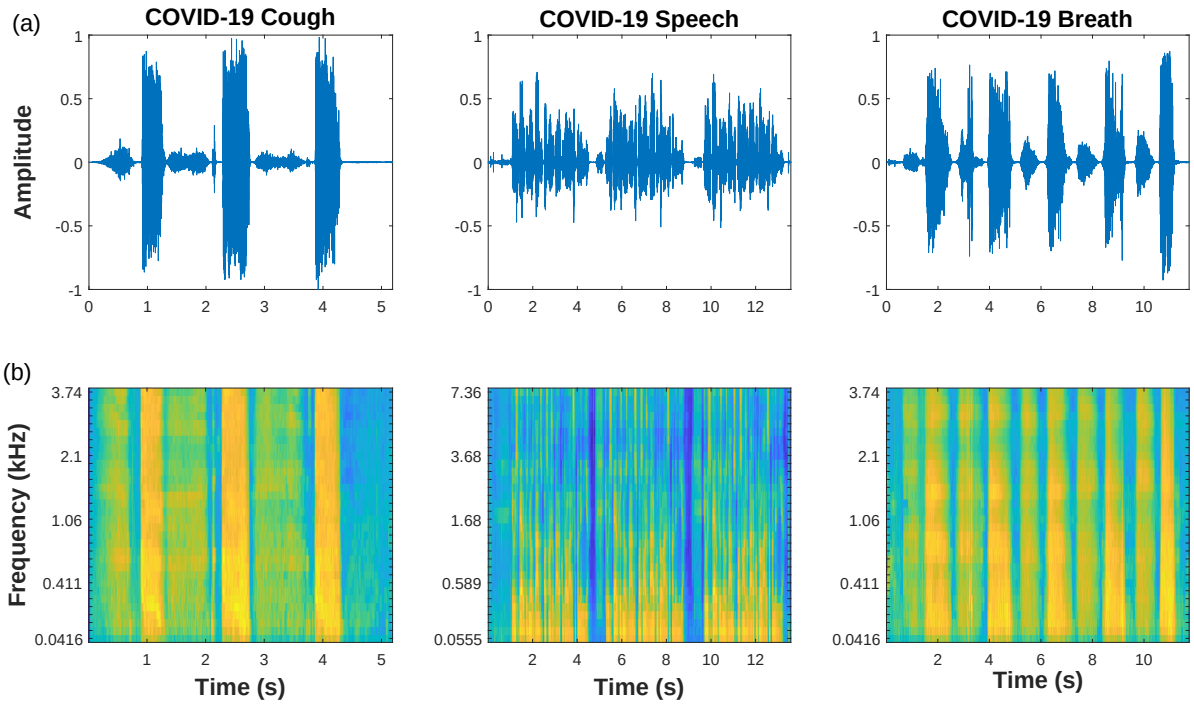


Figure 14 – Cough, speech, and breath signal representation. (a) shows the signals in the time domain; and (b) shows the time-frequency Mel-spectrogram representation.

To extract discriminative patterns from these time-frequency images, texture-based descriptors were applied. The Mel-spectrograms were first converted from Red, Green and Blue (RGB) to grayscale. Then, LBP features were extracted using a radius of 3 pixels and 8 sampling points, following the configuration proposed in (Sharma; Umapathy; Krishnan, 2022). This resulted in a 59-dimensional feature vector, normalized using the L_2 norm to ensure rotational invariance. Similarly, LTP features were extracted using the same parameters, yielding a 512-dimensional feature vector. These descriptors capture local intensity transitions in the spectrograms, providing texture-based representations for classification. Once extracted, these vectors served as input to the ML models described in Section 4.1.6, which enabled a consistent classification process, independent of the signal modality.

4.1.5 Experimental Design

To investigate the feasibility of inferring COVID-19 infection using audio and physiological data, a series of binary classification experiments across four datasets (CCS,

CSS, Cambridge KDD, Wearable Device) were design using data from the IPMA. Each experiment sought to investigate the classification performance associated with distinct signal modalities (cough, speech, breath, vital signs) and population subsets in detecting COVID-19 infection.

CCS and CSS Datasets

- **C1 – COVID-19 vs. Non-COVID (Cough)**
  Objective: Assess whether cough sounds can distinguish between individuals who tested positive for COVID-19 and those who did not.
- **C2 – COVID-19 vs. Non-COVID (Speech)**
  Objective: Evaluate the potential of speech recordings to identify COVID-19 positive individuals.
- **C3 – COVID-19 vs. Non-COVID (Cough + Speech)**
  Objective: Investigate whether combining cough and speech improves the discrimination of COVID-19 status.

Cambridge KDD Dataset

- **K1 – COVID-19 vs. Healthy Controls (Cough)**
  Objective: Assess whether cough sounds can distinguish COVID-positive individuals from asymptomatic, non-smoking, healthy controls.
- **K2 – COVID-19 with Cough vs. Non-COVID with Cough (Cough)**
  Objective: Evaluate whether cough patterns differ between COVID-positive and COVID-negative individuals who both report coughing.
- **K3 – COVID-19 vs. Non-COVID (Breath)**
  Objective: Investigate the potential of breath sounds to differentiate between COVID-positive and COVID-negative subjects.
- **K4 – COVID-19 with Cough vs. Non-COVID with Cough (Breath)**
  Objective: Determine whether breath sounds vary between COVID-positive and COVID-negative individuals who share the cough symptom.
- **K5 – COVID-19 with Cough vs. Non-COVID with Cough (Cough + Breath)**
  Objective: Test if combining cough and breath signals improves discrimination between COVID-positive and COVID-negative individuals with cough.

- **K6 – COVID-19 with Cough vs. Asthmatic with Cough (Cough)**
  Objective: Assess whether coughs from COVID-19 patients differ from those of asthmatic individuals with cough.

Wearable Device Dataset

- **W1 – COVID-19 vs. Non-COVID (Vital Signs)**
  Objective: Evaluate whether physiological parameters (heart rate, SpO_2, and body temperature) can be used to infer COVID-19 infection.

IPMA COVID-19 Dataset

- **A1 – COVID-19 vs. Non-COVID (Cough)**
  Objective: Assess whether cough sounds collected using the IPMA equipment can distinguish COVID-positive from COVID-negative individuals.
- **A2 – COVID-19 vs. Non-COVID (Speech)**
  Objective: Investigate the discriminative potential of speech audio collected in a clinical setting.
- **A3 – COVID-19 vs. Non-COVID (Breath)**
  Objective: Evaluate whether breath sounds can be used for COVID-19 detection under standardized acquisition conditions.
- **A4 – COVID-19 vs. Non-COVID (Vital Signs)**
  Objective: Assess whether clinical measurements (SpO_2, heart rate, and temperature) collected via the IPMA equipment are predictive of COVID-19 status.
- **A5 – COVID-19 vs. Non-COVID (Multimodal)**
  Objective: Determine whether combining all available modalities (cough, speech, breath, and physiological signals) enhances COVID-19 detection performance.

4.1.6 Classification and Evaluation

To evaluate the classification performance across different datasets and modalities, three sets of experiments were conducted using the CCS, CSS, Cambridge KDD, Wearable Device, and IPMA COVID-19 datasets. The objective was to assess whether features extracted from audio and physiological data could accurately distinguish COVID-19 positive and negative individuals under different experimental conditions.

In the first analysis, we evaluated experiments C1–C3 (CCS and CSS), K1—K6 (Cambridge KDD), and W1 (Wearable). For audio-based experiments, the feature vectors derived from cough, speech, and breath signals — processed via LBP or LTP — were

used as input to a SVM classifier. For W1, which included only physiological data, a DT classifier was used instead.

All experiments employed binary classification. A holdout strategy with stratified random sampling was used: 80% of the data was allocated for training and 20 % for testing. To ensure model robustness, 5-fold cross-validation was applied during training, and repeated the entire evaluation process 30 times using different random splits (with stratified holdout). This approach aimed to mitigate sampling variability and yield more stable performance estimates. Hyperparameter tuning was performed exclusively on the training data, and the configuration with the highest F1-score was selected for final evaluation on the test set. This evaluation strategy ensured fair comparisons across tasks while reducing overfitting and enhancing generalization.

In the second analysis, the generalization capability of the trained models was evaluated by applying them to the IPMA COVID-19 dataset (experiments A1–A4). For each modality in IPMA (cough, speech, breath, and physiological signals), it was used the classifier trained on the corresponding public dataset. For instance, the model from experiment C1 (trained on CCS cough data) was used to classify IPMA cough samples (Figure 15a).

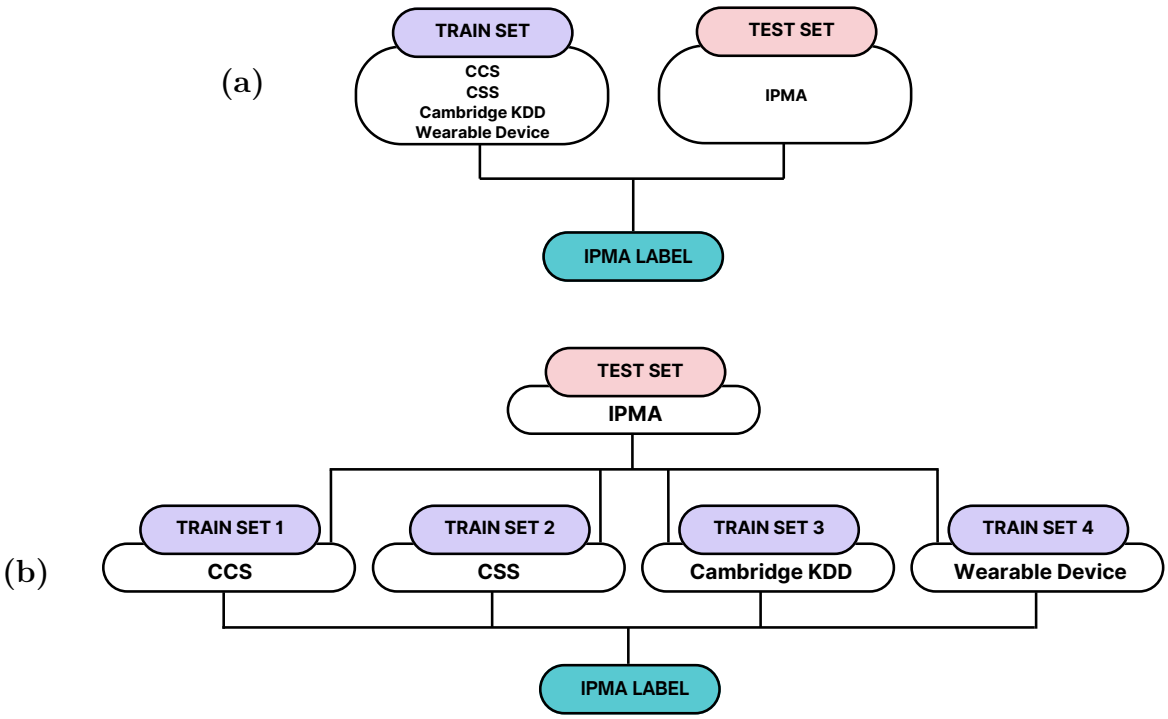


Figure 15 – Block diagram of the proposed IPMA data evaluation: (a) evaluating IPMA signals individually; (b) evaluating all signals together.

The third analysis investigated the effectiveness of late fusion by combining the decision scores from models trained on different modalities, as illustrated in Figure 15b.

For each IPMA sample, it was extracted the score vectors $\mathbf{w}^{(i)} = [w_{\text{neg}}^{(i)}, w_{\text{pos}}^{(i)}]$ from four public models: cough (CCS), breath (KDD), speech (CSS), and vital signs (Wearable). These vectors were summed element-wise to produce a global score vector $\mathbf{w}$:

$$\mathbf{w} = \sum_{i=1}^{4} \mathbf{w}^{(i)} = [w_{\text{neg}}, w_{\text{pos}}] \tag{4.1}$$

This approach corresponds to unweighted soft voting, in which the final prediction is determined by comparing the aggregated class scores:

$$\text{Decision rule: } \begin{cases} w_{\text{pos}} > w_{\text{neg}} & \text{classified as COVID-19 } (+) \\ w_{\text{pos}} \leq w_{\text{neg}} & \text{classified as COVID-19 } (-) \end{cases}$$

To evaluate performance, ACC and F1-score were used for all experiments, as they provide complementary views of classification effectiveness. Due to the presence of class imbalance in some datasets — especially the IPMA data — precision and sensitivity were also reported to better characterize classifier behavior under skewed distributions. This comprehensive set of metrics enabled a robust assessment of the models' predictive capabilities.

To assess the statistical significance of performance differences, normality tests (Shapiro–Wilk) were conducted followed by appropriate comparative tests. For normally distributed data, the paired Student's t-test was applied; otherwise, the Wilcoxon signed-rank test was used. These tests were applied to compare the performance of LBP and LTP feature extraction methods within each experiment. When evaluating multiple groups (e.g., across datasets or signal modalities), One-way ANOVA or the Kruskal–Wallis test was applied, depending on the distribution. Post-hoc analyses were conducted using Tukey's HSD (Honestly Significant Difference) (for parametric tests) or Dunn's test (for non-parametric tests), with all statistical tests adopting a significance level of $p < 0.05$.

4.1.7 Usability Evaluation

Assessing usability is essential to determine how effectively users interact with the proposed system and to identify potential barriers to adoption in real-world settings. To this end, a usability evaluation was conducted focused on the IPMA esquipment applied to COVID-19 inference tasks. Two standardized questionnaires — SUS and PSSUQ — were employed, both widely used to capture subjective user perceptions. Participants completed these questionnaires after their interaction with the system. The collected responses were used to quantify perceived usability, identify strengths and limitations in the user experience, and guide future design improvements.

4.2 Results of the COVID-19 Inference Analysis

To investigate the viability of using biomedical audio and physiological signals for COVID-19 detection, a set of experiments were prepared using four public datasets (CCS, CSS, Cambridge KDD, and Wearable Device) and one real-world dataset collected with the IPMA esquipment. The goal was twofold: (i) to compare the performance of two audio texture descriptors (LBP and LTP) across different modalities — cough, speech, and breath — and (ii) to assess the generalization of models trained on public datasets when applied to constrained clinical data from IPMA. In the following sections, it is presented a detailed overview of the results, highlighting how each signal type and experimental configuration influenced model performance, and which conditions favored reliable COVID-19 inference.

4.2.1 Characteristics of the Study Subjects

CCS Dataset

As previously mentioned, the CCS dataset comprises cough audio signals along with demographic and clinical information, including sex, age, smoking status, medical history, and symptom occurrence. To investigate the relationship between these variables and COVID-19 infection, the Chi-squared test was applied. The analysis included 725 subjects, as shown in Table 1, of whom 158 (21.79 %) tested positive (COVID-19 (+)) and 567 (78.21 %) tested negative for COVID-19 (COVID-19 (-)).

No significant association was found between infection status and smoking habits or medical history ($p > 0.05$), suggesting that these factors did not meaningfully differentiate positive and negative cases in this dataset. A statistically significant association was identified for sex ($p < 0.05$), with a slightly higher proportion of females among positive cases (54.43 %) compared to males (43.67 %). Age group also showed a significant association with infection status, with differences in the distribution of subjects across age ranges between positive and negative groups. Among all variables, symptom presence showed the strongest association with infection: 81.6 % of COVID-19 (+) individuals reported symptoms, compared to only 46.7 % among COVID-19 (-) cases ($p \ll 0.05$). Together, these findings suggest that while demographic variables such as sex and age show some association with infection status, the presence of symptoms remains the most informative feature for distinguishing between COVID-19 positive and negative individuals in the CCS dataset.

CSS Dataset

Following the same approach applied to the CCS dataset, a statistical analysis of the CSS dataset was conducted to explore potential associations between clinical and demographic variables and COVID-19 infection status. It is worth noting that this dataset

Table 1 – Association between sex, age, clinical characteristics, and COVID-19 infection using the CCS database.

	COVID-19 (+)	**COVID-19 (−)**	**p-value**
	n = 158	**n = 567**	
Sex			**0.0326**^{*,a}
F	86	242	
M	69	322	
Unspecified	3	3	
Age			**0.0034**^{*,a}
0 – 19	4	10	
20 – 29	33	50	
30 – 39	37	131	
40 – 49	37	120	
50 – 59	33	106	
60 – 69	5	104	
70 – 79	8	42	
80 – 89	0	2	
>90	0	1	
Unspecified	1	1	
Smoking			**0.9815**^{a}
Never	90	352	
Ex	44	123	
1 to 10	13	50	
11 to 20	5	23	
21+	1	4	
Unspecified	5	15	
Medical History			**0.0793**^{a}
No	115	358	
Yes	43	209	
Symptoms			$\ll$ **0.05**$^{*,\mathbf{a}}$
No	29	302	
Yes	129	265	

* Significant values $p < 0.05$.
^{a} Chi-squared test.

includes only speech audio recordings, which distinguishes it from the cough-based signals used in the CCS dataset. To evaluate whether any of these variables are associated with COVID-19 infection, the Chi-squared test was applied to sex, age, smoking status, medical history, and symptom presence.

As shown in Table 2, 308 subjects (34.49 %) tested positive, while 585 (65.51%) tested negative. No significant association was found for sex or smoking status ($p > 0.05$), indicating that these variables did not meaningfully differentiate between positive and negative cases. In contrast, medical history showed a statistically significant association with infection ($p < 0.05$), suggesting that pre-existing conditions may influence infection susceptibility — differently from the CCS dataset, where no such association was observed. A significant association was also found for age group ($p \ll 0.05$), with COVID-19 positive individuals concentrated in the 30—59 age range. Finally, symptom presence was the most strongly associated variable: 79.2 % of COVID-19 positive subjects reported symptoms, compared to only 46.5 % among negatives ($p \ll 0.05$). Overall, the results suggest that while age and clinical history contribute to characterizing infection patterns, the presence

of symptoms continues to be the most robust indicator for distinguishing COVID-19 cases in the CSS dataset.

Table 2 – Association between sex, age, clinical characteristics, and COVID-19 infection using the CSS database.

	COVID-19 (+)	**COVID-19 (−)**	**p-value**
	n = 308	**n = 585**	
Sex			**0.9751**[a]
F	134	251	
M	173	330	
Unspecified	1	4	
Age			$\ll$ **0.05**[*,a]
0 – 19	8	9	
20 – 29	35	50	
30 – 39	53	137	
40 – 49	114	124	
50 – 59	80	112	
60 – 69	10	105	
70 – 79	8	43	
80 – 89	0	3	
>90	0	1	
Unspecified	0	1	
Smoking			**0.2680**[a]
Never	219	367	
Ex	65	126	
1 to 10	12	51	
11 to 20	7	21	
21+	1	4	
Unspecified	4	16	
Medical History			$\ll$ **0.05**[*,a]
No	257	373	
Yes	51	212	
Symptoms			$\ll$ **0.05**[*,a]
No	64	312	
Yes	244	273	

[*] Significant values $p < 0.05$.
[a] Chi-squared test.

Cambridge KDD Dataset

As previously mentioned, the Cambridge KDD dataset comprises audio signals related to cough and breath. Unlike the CCS and CSS datasets, however, it does not include explicit information on clinical variables. Instead, it only indicates the presence or absence of cough as a symptom in both positive and negative cases, based on the cough and breath recordings. Furthermore, it includes recordings from negative cases involving asthmatic subjects who present cough as a symptom. Given these characteristics, the Chi-squared test was applied to evaluate whether the presence of cough symptoms in each signal type is associated with COVID-19 infection.

A total of 115 subjects recorded their signals, as illustrated in Table 3. In the COVID-19 positive group, the same individuals who recorded breathing signals also recorded coughing signals, resulting in consistent subject counts across both modalities (35

subjects, 30.43 %). In contrast, the COVID-19 negative group includes an additional set of 11 asthmatic individuals who are not part of the other subgroups, which explains the total of 80 subjects (i.e., 69 non-asthmatic + 11 asthmatic, 69.57 %). A slight discrepancy can be observed between the number of subjects in the cough and breath categories, particularly within the "no symptom – COVID-19 (–)" subgroup. This occurs due to the dataset structure: although some individuals share the same symptom profile, they appear only in the breathing signal folders, and not in the coughing ones. It is also important to note that Table 3 reports both the number of subjects and the total number of audio samples (in brackets), which includes augmented data, as detailed in Section 4.1.2.

Table 3 – Number of unique subjects and associated samples (in brackets) based on cough as a symptom and COVID-19 status, for both cough and breathing signals, using the Cambridge KDD dataset.

Type of Signal	**Asthma**	**Reported Cough**	**COVID-19 (+)**	**COVID-19 (–)**	**p-value**
			n = 35 [110]	**n = 80 [304]**[‡]	
Cough	No	No	20 [64]	60 [138]	$\ll 0.05^{\dagger,a}$
	No	Yes	15 [46]	5 [56]	
	Yes	Yes	–	11 [104]	
Breath	No	No	20 [64]	64 [144]	$\ll 0.05^{\dagger,a}$
	No	Yes	15 [46]	5 [56]	
	Yes	Yes	–	11 [104]	

[*] Number of unique subjects in each group. The value in brackets represents the total number of samples, including those obtained by data augmentation.
[‡] The discrepancy between cough and breath sample counts in the COVID-19 (–) group without symptoms arises from dataset structure: some users appear only in the breath folders, despite matching symptom profiles.
[†] Significant values $p < 0.05$.
[a] Chi-squared test for independence between cough symptoms and group classification.

The results of the Chi-squared test revealed a statistically significant association between the presence of cough symptoms and COVID-19 status for both signal types ($p \ll 0.05$). In the cough recordings, most non-asthmatic COVID-19 negative subjects (60 out of 69) reported no symptoms, whereas 15 out of 35 COVID-19 positive subjects presented cough symptoms. A similar distribution was observed in the breath recordings, with 64 non-symptomatic individuals in the negative group and 15 symptomatic individuals in the positive group. Asthmatic individuals (n = 11), all COVID-19 negative and symptomatic, contributed equally to both signal types and were analyzed separately to avoid bias in comparisons with non-asthmatic subjects. These findings reinforce the relationship between symptomatic audio patterns and infection status, supporting the potential of respiratory sounds as a non-invasive screening tool for COVID-19.

Wearable Device Dataset

To assess whether physiological signals recorded by IPMA are associated with COVID-19 infection, the Wearable Device dataset was analyzed, which includes body temperature, heart rate, SpO_2, and basic demographic variables such as sex and age. It

was aimed to investigate the extent to which these variables contribute to distinguishing between COVID-19 positive and negative individuals, using the Chi-squared test for statistical comparison.

As shown in Table 4, no significant association was found for sex or age group ($p > 0.05$), indicating limited discriminative power for these demographic features. In contrast, all three physiological variables showed statistically significant differences between groups. Temperature was elevated in most positive cases, with the majority falling within the $38.0-38.9\,^{\circ}\mathrm{C}$ range. Heart rate was also higher among positives, particularly in the 100–119 bpm range. Notably, SpO_2 presented the most prominent difference, with substantial desaturation observed in COVID-19 positive individuals, reflecting the respiratory impact of the infection. These findings highlight the role of physiological signals — especially SpO_2 and temperature — as non-invasive indicators of infection when acquired through external systems such as the IPMA.

IPMA COVID-19 Dataset

Finally, the demographic and clinical characteristics of subjects included in the IPMA COVID-19 dataset were analyzed to assess their relationship with COVID-19 infection. Although the small sample size ($n = 11$) limits statistical power, the Chi-squared test was applied to explore potential associations between variables and infection status. As shown in Table 5, the dataset presents a balanced distribution between female and male participants, and most individuals were between 20 and 49 years old. No significant differences ($p > 0.05$) were found in sex, age, smoking habits, medical history, or symptom occurrence between individuals with and without COVID-19. Of note, although both COVID-19 positive individuals reported symptoms, the majority of negative cases did as well (8 out of 9), suggesting that symptom presence alone was not a reliable discriminator of infection status within this small cohort.

Having described the characteristics of the study subjects, the classification results obtained using the proposed methods are now presented.

4.2.2 Performance of the Classification Model

Results on the CCS and CSS Dataset

In this section, the classification performance of audio-based features are analyzed for COVID-19 inference using two datasets: CCS (cough recordings) and CSS (speech recordings). Both datasets were processed to evaluate which audio texture representation — LBP or LTP — achieves better classification results across three experimental setups: C1 (cough-only), C2 (speech-only), and C3 (combined signals). Features were extracted using both LBP and LTP methods and classifiers were trained accordingly. The performance

Table 4 – Association between sex, age, physiological variables, and COVID-19 infection using the Wearable Device dataset.

	COVID-19 (+)	**COVID-19 (−)**	**p-value**
	n = 945	**n = 139**	
Sex			**0.6012**[a]
F	336	46	
M	447	72	
Unspecified	162	21	
Age			**0.5459**[a]
0 – 19	27	5	
20 – 29	77	18	
30 – 39	128	21	
40 – 49	108	19	
50 – 59	163	20	
60 – 69	124	11	
70 – 79	72	9	
80 – 89	33	5	
>90	2	0	
Unspecified	211	31	
Temperature			**0.0003**[*,a]
36.0 – 36.9	37	15	
37.0 – 37.9	129	31	
38.0 – 38.9	667	77	
39.0 – 39.9	103	16	
40.0 – 40.9	2	0	
Unspecified	7	0	
Heart Rate			**0.0006**[*,a]
40 – 59	53	2	
60 – 79	283	51	
80 – 99	224	53	
100 – 119	383	33	
120 >	2	0	
SpO_2			$\ll$ **0.05**[*,a]
30 – 39	359	8	
40 – 49	18	0	
50 – 59	14	0	
60 – 69	18	0	
70 – 79	109	0	
80 – 89	335	5	
90 – 99	92	109	
100	0	1	
Unspecified	0	16	

[*] Significant values $p < 0.05$.
[a] Chi-squared test.

on the test set was assessed using ACC and F1-score metrics. Statistical significance was tested using the paired t-test applied to the F1-score due to its robustness in evaluating the balance between precision and recall. This approach aimed to determine whether one feature extraction method consistently outperforms the other across distinct audio contexts.

As shown in Figure 16, LBP consistently outperformed LTP in terms of F1-score across all experimental configurations (C1–C3), with statistically significant differences observed ($p < 0.05$). In the cough-only scenario (C1, CCS dataset), although LTP slightly outperformed LBP in ACC, LBP showed superior F1-score, indicating better overall balance between precision and recall. In the speech-only scenario (C2, CSS dataset),

Table 5 – Association between sex, age, clinical characteristics, and COVID-19 infection using the IPMA database.

	COVID-19 (+)	**COVID-19 (−)**
	n = 2	**n = 9**
Sex[a]		
F	2	2
M	0	4
Unspecified	0	3
Age[a]		
20 – 29	2	1
30 – 39	0	2
40 – 49	0	2
70 – 79	0	1
Unspecified	0	3
Smoking[b]		
No	2	8
Yes	0	1
Medical History[b]		
No	1	4
Yes	1	5
Symptoms[b]		
No	0	1
Yes	2	8

[a] Chi-squared test.
[b] Fisher's exact test.

LBP demonstrated better results in both ACC and F1-score, confirming its robustness in handling speech signals. The combined scenario (C3), which includes both cough and speech signals, also showed a consistent advantage for LBP over LTP in both metrics. These findings highlight that, despite minor fluctuations in ACC, LBP provides a more balanced and stable performance in COVID-19 detection from respiratory and vocal signals.

After identifying LBP as the most promising feature extraction method, it was further investigated whether the type of input signal — cough, speech, or a combination of both — affects classification performance when using this representation. This analysis aimed to determine which acoustic source provides the most discriminative information for COVID-19 inference. To this end, ANOVA was applied to the F1-scores obtained from each experimental configuration (C1, C2, and C3) using LBP. This approach was chosen because the F1-score distributions satisfied the assumption of normality, and ANOVA is appropriate for comparing the means of three or more independent groups. The ANOVA results indicated a statistically significant difference among the groups ($p < 0.001$), prompting a post hoc Tukey's HSD (Honestly Significant Difference) test to identify specific differences. The post hoc analysis revealed that the speech-only scenario (C2) produced significantly higher F1-scores than both the cough-only (C1) and combined (C3) setups ($p < 0.001$), while no significant difference was observed between C1 and C3 ($p = 0.075$). These results indicate that speech signals, when processed with LBP features, yield the most effective discrimination of COVID-19 cases, outperforming both cough and combined

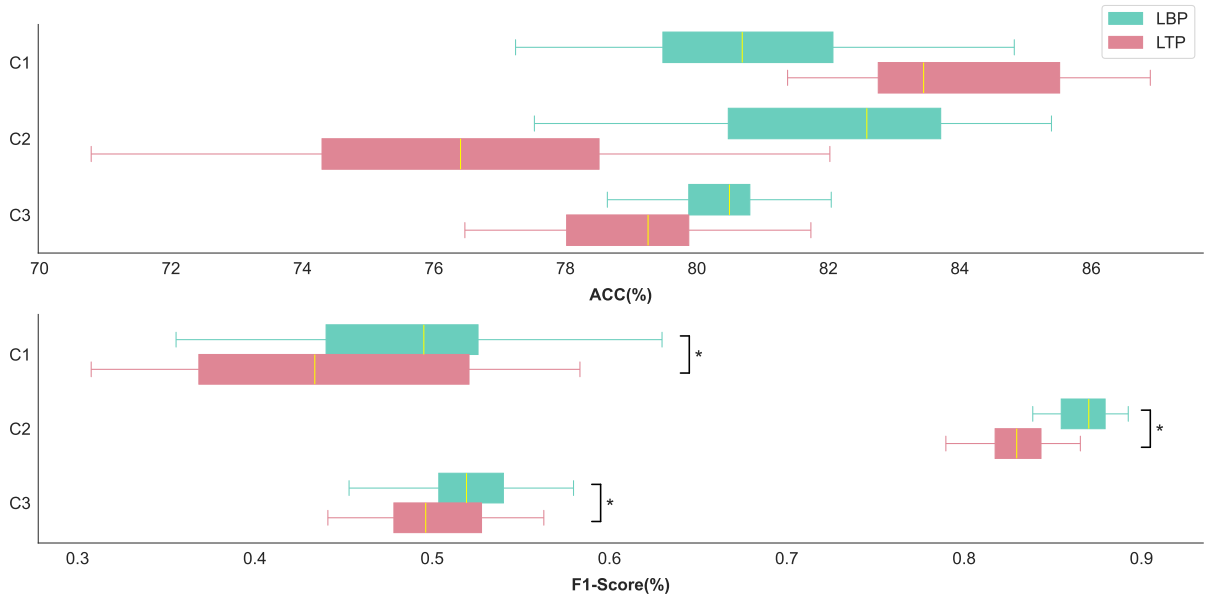


Figure 16 – Performance on the test set in terms of ACC and F1-score for the C1–C3 experiments, using LBP and LTP. Asterisks (*) indicate statistically significant differences between the methods ($p < 0.05$).

inputs.

Results on the Cambridge KDD Dataset

To explore how audio texture features perform across more diverse respiratory conditions, the Cambridge KDD dataset was analyzed, which includes cough and breath recordings from subjects with varying clinical profiles. The primary objective was to evaluate which representation method — LBP or LTP — yields better classification performance in six diagnostic scenarios (K1–K6), ranging from comparisons with healthy controls to subjects with asthma. Each task involved extracting features from the respective audio signals and evaluating classifier performance using ACC and F1-score. Paired t-tests were applied to the F1-score to assess statistical significance. This evaluation was designed to determine the robustness of each method across multiple audio contexts and clinical challenges.

Figure 17 shows that LBP generally yielded higher F1-scores than LTP across the classification tasks, with statistically significant differences observed in tasks K2, K4, K5, and K6 ($p < 0.05$). In tasks K1, K3, and K4, LBP consistently outperformed LTP across both ACC and F1-score, presenting higher medians and narrower interquartile ranges, which reflect more stable and reliable performance in scenarios based on either cough or breath signals. In contrast, tasks K2 and K5 demonstrated superior performance by LTP, which not only achieved superior ACC but also significantly outperformed LBP in F1-score based on statistical analysis and quartile distribution. Finally, in task K6 — which involved discriminating COVID-19 from asthma-related coughs — LBP once again

surpassed LTP in F1-score, with lower variability, whereas LTP showed slightly higher ACC but more dispersed results. These findings reinforce the overall advantage of LBP in delivering more consistent and discriminative performance across diverse clinical and acoustic conditions.

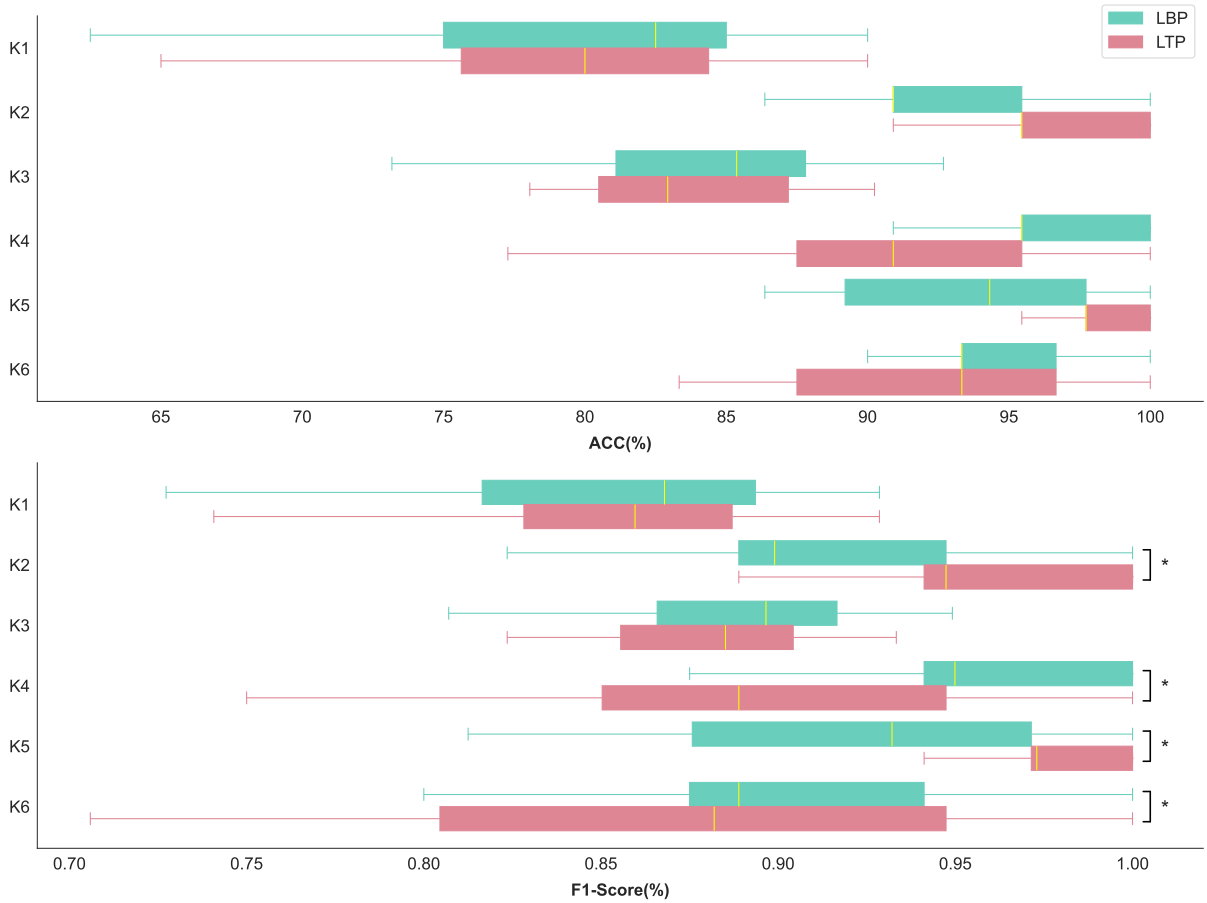


Figure 17 – Performance on the test set in terms of ACC and F1-score for the K1–K6 experiments, using LBP and LTP. Asterisks (*) indicate statistically significant differences between the methods ($p < 0.05$).

To further explore how clinical context and signal type affect classification performance using LBP, comparisons pairwise were conducted between specific diagnostic tasks using Dunn's post hoc test. The non-parametric Kruskal–Wallis test was first applied to the F1-scores obtained across the six KDD tasks, confirming a statistically significant difference among them ($H = 88.10$, $p < 0.001$). Figure 18 highlights the task pairs in which performance differences were found to be statistically significant. Comparisons between K1 and K3, and between K1 and K2, indicate that distinguishing COVID-19 from healthy individuals is more straightforward than from symptomatic non-COVID cases. This suggests that the presence of symptoms introduces acoustic variability that challenges classification. In the respiratory tasks, K2 vs. K4 and K6 vs. K4 showed that symptoms such as cough alter breathing patterns in a way that impacts model performance, reinforcing the value of breath signals in symptomatic scenarios. The comparisons K2

vs. K5 and K3 vs. K5 demonstrated that combining cough and breath sounds leads to better results than using either modality alone. Finally, the contrast between K6 and K5 revealed that asthma-related cough carries distinct acoustic features, which can interfere with classification and require special attention in real-world deployment. These findings offer practical insights into how signal type and symptom presence influence classification outcomes, which is essential for developing reliable and generalizable diagnostic tools.

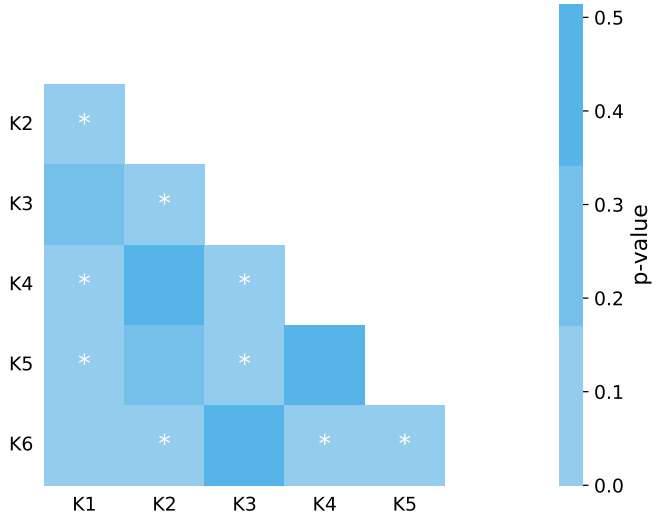


Figure 18 – Pairwise p-values from Dunn's test comparing F1-scores across KDD tasks (K1–K6) using LBP. Asterisks (*) mark significant differences ($p < 0.05$).

Results on the Wearable Device Dataset

To evaluate the predictive value of physiological signals in identifying COVID-19 cases, a classification task (W1) using the Wearable Device dataset was conducted. This task aimed to distinguish between COVID-19 positive and negative individuals based on heart rate, SpO_2, and temperature measurements, all of which showed significant differences between groups in the population analysis. A DT classifier was applied to the test set to explore whether these variables, when combined, could infer infection status. This model was selected for its simplicity, interpretability, and compatibility with structured biomedical data. The classifier achieved high performance, with a mean ACC of 97.16 % ($std = 0.84$) and F1-score of 0.98 ($std = 0.01$). These findings indicate that temperature and SpO_2, in particular, hold potential as non-invasive markers for predicting COVID-19 outcomes in systems such as IPMA.

Results on the Audio Subset of the IPMA Dataset

To assess the generalization capacity of models trained on public datasets when applied to real-world scenarios, the IPMA dataset was evaluated exclusively as a test set. This analysis aimed to determine whether classifiers trained on CCS, CSS, and Cambridge

KDD datasets could distinguish COVID-19 positive and negative individuals using audio signals collected under the IPMA equipment. Three experiments were designed: A1 (cough), A2 (speech), and A3 (breath), each corresponding to a specific signal modality. In all cases, the models were trained on the respective public datasets and evaluated solely on IPMA data, allowing to test real-world generalization without retraining. Given the limited number of IPMA samples and the potential for misleading interpretation when precision or sensitivity is low, it is reported the results using ACC, precision, and sensitivity instead of the F1-score. The classifiers were based on SVMs, with performance compared across LBP and LTP feature extraction methods.

The evaluation on the IPMA dataset revealed clear differences in classification performance depending on the signal type and the feature extraction method used. As shown in Table 6, LBP consistently outperformed LTP across all three experiments. In A1 (cough), LBP achieved an ACC of 84.05 % with high sensitivity (0.90), but low precision (0.30), suggesting many false positives. In contrast, LTP failed to detect any true positives (sensitivity = 0.00), making it impractical in this setting. In A2 (speech), both ACC and precision were considerably higher with LBP (ACC = 76.21 %, precision = 0.88), while sensitivity remained balanced (0.78), indicating a more favorable trade-off between false positives and false negatives. For A3 (breath), LBP again showed strong performance (ACC = 83.93 %, precision = 0.78, sensitivity = 0.76), whereas LTP showed limited detection capability (ACC = 18.18 %, sensitivity = 0.18). These findings confirm that, although classification performance may be affected by signal modality and class imbalance, LBP-based models trained on external datasets retain good generalization capacity when applied to real-world data collected in constrained clinical conditions such as those encountered with IPMA.

Table 6 – Classification performance on the test set of the IPMA dataset using LBP and LTP features. Results are reported in terms of ACC (%), precision, and sensitivity. Values in parentheses indicate the standard deviation.

Experiment	**ACC(%)**		**Precision**		**Sensitivity**	
	LBP	LTP	LBP	LTP	LBP	LTP
A1	84.05 (0.45)	81.82 (0.00)	0.30 (0.02)	0.00 (0.00)	0.90 (0.03)	0.00 (0.00)
A2	76.21 (1.43)	73.94 (4.61)	0.88 (0.01)	0.90 (0.06)	0.78 (0.01)	0.80 (0.01)
A3	83.93 (0.91)	18.18 (0.00)	0.78 (0.03)	1.00 (0.00)	0.76 (0.02)	0.18 (0.00)

When comparing the results obtained in A1–A3 with those from previous experiments using public datasets (C1–C3 and K1–K6), a consistent pattern emerges regarding the influence of signal type on classification performance. As observed in previous analyses, speech and breath signals tend to produce more balanced results than cough, in terms of F1-score. This trend is also evident in the IPMA evaluation: A2 (speech) and A3 (breath) yielded better overall performance than A1 (cough), corroborating the findings previously

reported for the public datasets.

Results on the Vital Signs Subset of the IPMA Dataset

To complement the audio-based experiments, the experiment A4 was conducted to evaluate whether physiological signals — specifically heart rate, SpO_2, and temperature — could support the distinction between COVID-19 positive and negative individuals. A DT classifier was trained using the Wearable Device dataset and then applied to the IPMA data for testing, following the same configuration described in previous sections. On the IPMA test set, the model achieved an ACC of 81.82 %, but both precision and sensitivity were 0.00 across all runs. This outcome suggests that, although the model correctly classified most negative cases, it failed to identify any true positives. The result reflects the challenges of generalizing to real-world scenarios when using only physiological data and highlights the limitations imposed by small and imbalanced datasets such as IPMA.

Results on Audio and Vital Signs from the IPMA Dataset

Finally, to evaluate whether combining all available signal types could improve COVID-19 classification performance, experiment A5 was conducted using the full set of audio and physiological signals collected by the IPMA equipment — cough, speech, breath, heart rate, SpO_2, and temperature. The objective was to assess the benefit of multimodal integration in a real-world testing scenario. For this experiment, each signal type was mapped to a corresponding public dataset for training: cough signals were trained using the CCS dataset, speech using CSS, breath using KDD, and physiological signals using the Wearable Device dataset. Additionally, different classifiers were selected based on prior performance: SVMs were used for the audio-based signals, while a DT classifier was used for the physiological data. The final classification was based on the aggregation of class probabilities from individual models trained on each signal type, with predictions applied to the multimodal IPMA dataset.

The model achieved an ACC of 79.09 % ($std = 5.92$) on the IPMA test set, but both precision and sensitivity were near zero, indicating that it failed to identify any true positive cases. This outcome is consistent with the limitations observed in experiments A1 and A4, where the use of isolated cough or physiological signals also led to poor sensitivity. Despite incorporating multiple modalities and classifiers, the system did not generalize well to the IPMA dataset. These findings suggest that simply increasing the number of input signals does not necessarily improve detection performance in the presence of limited and imbalanced real-world data. They also reinforce the importance of training with representative and diverse datasets when deploying multimodal systems for COVID-19 inference.

4.3 Results of the IPMA Usability for COVID-19 Screening

Usability is a critical factor in the effectiveness and practical implementation of clinical decision support systems. In the context of IPMA, evaluating how users interact with the system is essential to ensure that it is not only technically sound, but also intuitive and aligned with user needs. To assess these aspects, two standardized and widely validated instruments were employed: the SUS and the PSSUQ. These instruments were selected for their complementary perspectives on usability, capturing dimensions such as perceived ease of use, interface quality, and clarity of information. The following subsections present the results obtained from each instrument.

4.3.1 Results on the SUS Scale

The SUS questionnaire was applied to evaluate the usability of the IPMA equipment from the perspective of end users. The complete list of SUS items can be found in Appendix A. This evaluation aimed to identify not only the general usability level but also specific aspects that could be improved. The analysis involved individual scores for each of the ten SUS questions, with values ranging from 1 to 5, as illustrated in Figure 19(a). Questions with a positive formulation, such as 1, 3, and 5, received some of the highest average ratings (4.29, 4.52, and 4.52, respectively), reflecting favorable user perceptions regarding system usability and integration. In contrast, negatively formulated questions related to complexity and inconsistency, such as 2, 6, and 8, showed lower mean scores (1.67, 1.76, and 1.67), indicating areas of difficulty and user dissatisfaction. Figure 19(b) presents the distribution of overall SUS scores among participants. These results indicate that, although usability is generally positive, efforts should be made to reduce inconsistency, clarify interface elements, and ensure that common tasks can be performed with minimal effort.

4.3.2 Results on the PSSUQ Scale

To complement the SUS evaluation, a second usability assessment using the PSSUQ was conducted, which measures user satisfaction across specific dimensions of system interaction. The complete list of PSSUQ items is provided in Appendix B. The objective was to obtain a more detailed understanding of how users perceived the system's usefulness, informational support, and interface design.

Figure 20 presents the individual scores for each of the 16 questions on the PSSUQ scale. Most responses concentrated between values 1 and 3, indicating positive user perceptions, especially for items related to ease of use, learnability, and interface satisfaction. In addition, Figure 21 shows the average scores for the three subscales and the overall score. The global PSSUQ score was 2.34, reflecting a generally favorable evaluation of

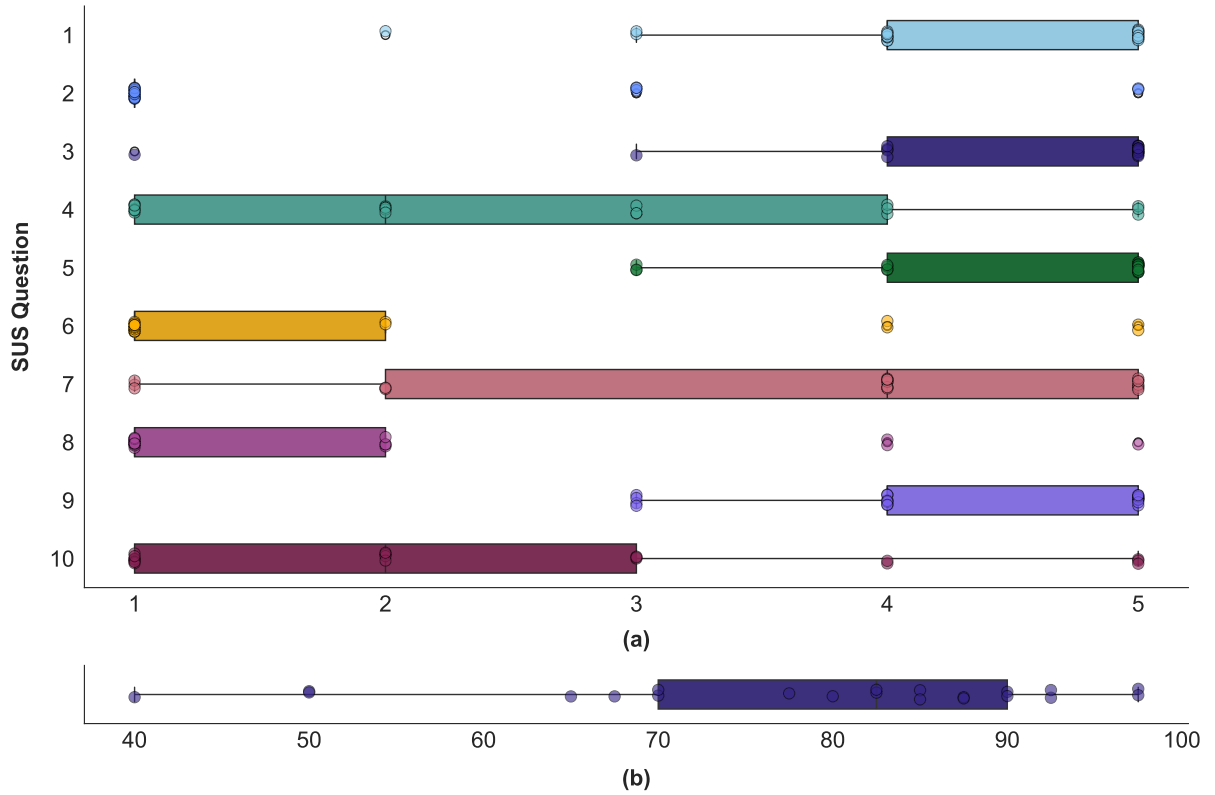


Figure 19 – SUS results for the IPMA equipment. (a) Individual question scores. (b) Overall SUS score distribution across participants.

the system. Among the subscales, INTERQUAL received the lowest average score (2.16), suggesting that users found the interface intuitive and visually coherent. SYSUSE and INFOQUAL obtained slightly higher means of 2.33 and 2.46, respectively, indicating overall satisfaction with system functionality and informational content, while also suggesting that enhancements in clarity and supportiveness could further improve the user experience. These findings reinforce the high level of satisfaction observed in the SUS evaluation and emphasize the system's strengths in interface design, while also pointing to opportunities for refinement in content presentation and functional efficiency.

4.4 Discussion

This study investigated whether non-invasive physiological signals, including audio from cough, speech, and breath, as well as vital signs, could support autonomous COVID-19 detection in real-world environments using a portable equipment (IPMA). The research addressed a critical gap in digital health by proposing and validating ML models capable of operating with multimodal inputs under constrained clinical conditions.

The presence of symptoms emerged as the most consistent variables associated with COVID-19 status across all public datasets analyzed, including CCS, CSS, and KDD. This finding reinforces the relevance of symptomatic expression as a proxy for infection,

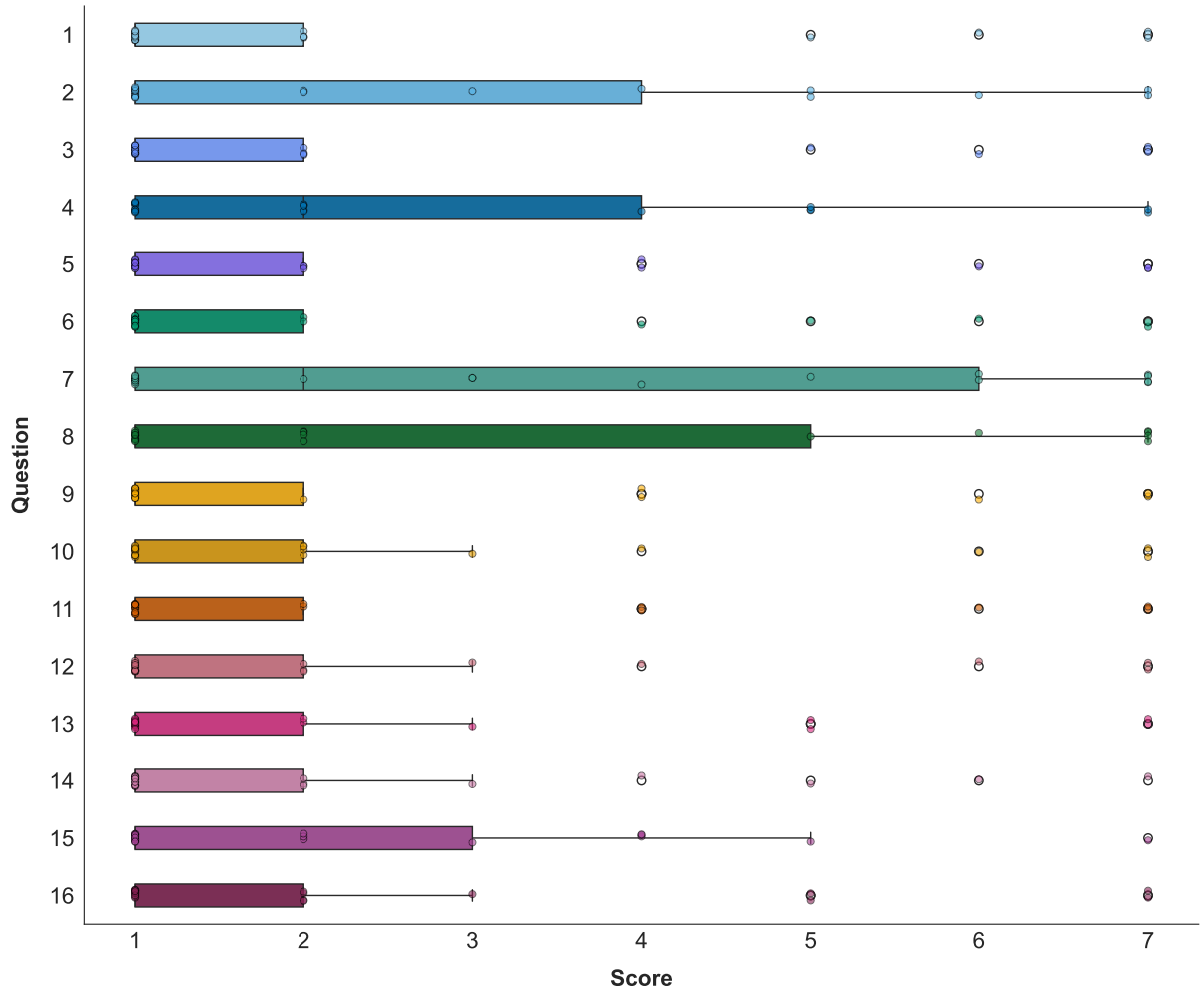


Figure 20 – Distribution of user responses to the 16 individual items of the PSSUQ. Lower scores (closer to 1) indicate higher satisfaction with specific aspects of the system, including ease of use, information quality, and interface design.

particularly in datasets composed of self-reported audio and demographic information. Previous studies have similarly identified symptoms such as cough, fever, and fatigue as early and strong indicators of SARS-CoV-2 infection (Despotovic et al., 2021; Husain et al., 2022; Laguarta; Hueto; Subirana, 2020). For instance, Husain et al. (2022) reported symptom presence as the most predictive clinical variable in models based on crowdsourced cough data. However, other studies have shown that acoustic features may capture COVID-19-related alterations even in the absence of overt symptoms. Laguarta, Hueto and Subirana (2020), for example, showed that even in the absence of self-reported symptoms, cough signals could be used to infer infection status, indicating that acoustic patterns might reflect early or subclinical manifestations of the disease.

Symptom-based discrimination draws support from viral effects on respiration and inflammation, which lead to detectable changes in vocal and respiratory features (Pahar et al., 2022). Still, relying exclusively on self-reported symptoms introduces limitations. Such data may be affected by reporting bias and differences in how symptoms are interpreted, and

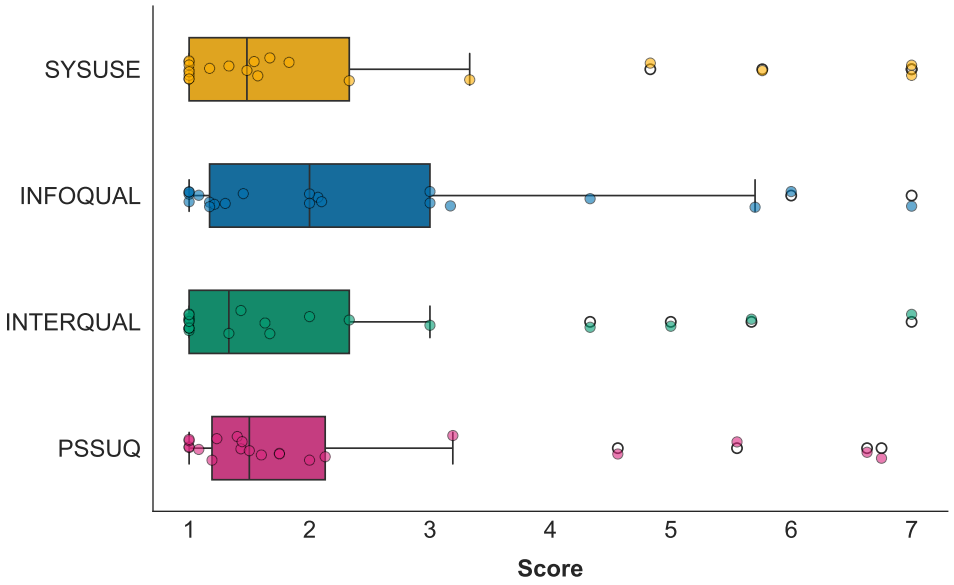


Figure 21 – Average scores obtained for each PSSUQ subscale — SYSUSE, INFOQUAL, and INTERQUAL — along with the overall score. All scores range from 1 (best) to 7 (worst), with lower values reflecting higher perceived usability.

it becomes less useful for diagnosing asymptomatic or minimally symptomatic individuals, which is a known challenge in early screening (Gupta-Wright et al., 2021).

Beyond symptom reporting, physiological signals, especially temperature and SpO_2, also showed strong associations with infection status in the Wearable Device dataset. These indicators reflect objective physiological changes such as fever and silent hypoxemia, both characteristic of COVID-19 (Swenson; Hardin, 2022; Tobin; Laghi; Jubran, 2020). Simonetti et al. (2023) similarly highlighted SpO_2 and temperature as the most effective markers in a remote monitoring system for COVID-19 patients, reinforcing the relevance of vital signs for non-invasive, real-time screening. In contrast, the IPMA dataset presented challenges for symptom-based inference. The small sample size and high prevalence of symptoms among COVID-19 negative participants limited the discriminative power of clinical characteristics. This underscores the importance of balanced datasets and the integration of multiple information sources for robust screening.

In parallel to clinical variable analysis, in this Thesis it was evaluated how audio-based feature extraction methods contributed to the detection performance of COVID-19 classifiers. LBP consistently outperformed LTP across most classification experiments using audio signals, particularly in speech-based configurations. This pattern was observed in both the CCS and CSS datasets (Figure 16), as well as in multiple diagnostic scenarios evaluated with the Cambridge KDD dataset (Figure 17). LBP yielded higher F1-scores in nearly all tasks, and the difference was statistically significant in several comparisons ($p < 0.05$, Figure 18). These findings suggest that LBP may offer a more robust representation of acoustic textures for COVID-19 inference.

Previous studies have demonstrated the utility of LBP in audio-based health monitoring. For instance, Sharma, Umapathy and Krishnan (2022) used LBP to extract texture features from spectrograms of respiratory sounds, achieving high classification ACC for COVID-19. LBP's superior performance may be related to its lower sensitivity to noise and simpler binary encoding, which tend to produce more stable descriptors in small-sample scenarios. Although LTP improves robustness to minor intensity variations through the use of a neutral zone (Tan; Triggs, 2010b), the resulting increase in feature complexity may not translate into consistent classification gains, particularly in datasets with limited samples and substantial intra-class variation. In the experiments realized in this Thesis, LTP showed inconsistent performance across tasks, often underperforming compared to LBP, especially in configurations involving speech signals. Its consistent performance, simplicity, and widespread applicability in lightweight systems suggest that LBP is a practical option for real-time COVID-19 screening.

Building on the consistent advantage of LBP, it was further examined how its performance was affected by the nature of the audio signal, particularly when distinguishing between cough, speech, and breath. Among the three modalities tested, speech signals consistently achieved the highest F1-scores, outperforming both isolated cough signals and combined configurations (Figure 16). This trend was statistically confirmed by ANOVA and Tukey's post hoc tests (Figure 18), which indicated that speech-based models performed significantly better than other configurations. These results suggest that speech may carry more stable and discriminative acoustic patterns for COVID-19 detection, especially when processed with LBP features. Unlike cough, speech involves continuous and voluntary phonation (Waage; Iwarsson, 2024), which may capture subtle alterations in articulation, respiration, and vocal tract coordination associated with respiratory infection.

Several studies have highlighted the diagnostic potential of cough signals for COVID-19 detection. For instance, Laguarta, Hueto and Subirana (2020) demonstrated high classification performance using forced coughs from asymptomatic individuals. Similarly, Sharma, Umapathy and Krishnan (2022) and Pahar et al. (2022) reported superior results for cough compared to other modalities in controlled datasets. Cough reflects direct alterations in the respiratory tract caused by inflammation, congestion, and irritation, which are common manifestations of COVID-19 (Laguarta; Hueto; Subirana, 2020). However, in the experiments realized in this Thesis, speech consistently outperformed cough across most tasks. This discrepancy may be explained by the variability of cough signals in intensity, duration, and intentionality, as well as by the lack of standardization in self-reported or real-world recordings. While cough remains a relevant acoustic marker, its effective application appears to depend strongly on recording conditions, dataset quality, and analysis pipeline design.

Despite these advantages, it is important to note that speech-based performance

may be influenced by the demographic and linguistic composition of the datasets. In real-world conditions, factors such as language, accent, and vocal effort may affect the generalizability of speech models (Hickman et al., 2024; Lou; Ren, 2021). Moreover, while combining speech and cough signals might intuitively appear beneficial, results obtained in this work show that multimodal fusion did not outperform speech-only configurations. These findings reinforce the potential of speech as a primary modality for acoustic COVID-19 screening and highlight the importance of signal quality and task-specific design when integrating multiple inputs.

In addition to audio signals, the contribution of physiological data to COVID-19 detection using the Wearable Device dataset was investigated. Among the recorded variables, temperature and SpO_2 emerged as the most informative indicators, both individually and in combination. A DT classifier trained on these features achieved high performance, with accuracy near 97 % and an F1-score of 0.98. These findings align with previous studies that have emphasized the relevance of fever and silent hypoxemia as characteristic manifestations of COVID-19 (Swenson; Hardin, 2022; Tobin; Laghi; Jubran, 2020). For example, Simonetti et al. (2023) demonstrated that remote monitoring of SpO_2 and temperature enabled timely clinical responses and effective home management of infected patients. Together, these results support the use of simple, non-invasive vital signs as viable markers for scalable and real-time screening systems.

To investigate the feasibility of using intelligent algorithms in real-world, resource-constrained environments, it was evaluated how models trained on public datasets performed when applied to signals acquired by the IPMA equipment. This step directly tested the core hypothesis of the study: whether a portable system could autonomously acquire, process, and interpret multimodal physiological signals for COVID-19 inference. Among all configurations, models using LBP features continued to outperform their LTP counterparts, offering better balance between sensitivity and precision across all tasks.

Speech and breathing emerged as the most reliable modalities in this scenario (A2 and A3), achieving moderate performance despite real-world data collection constraints. In contrast, cough-based models (A1) showed high sensitivity but very low precision, leading to a large number of false positives. The model based solely on physiological signals (A4) failed to detect positive cases, and the multimodal fusion (A5) did not improve results compared to the best unimodal inputs. These findings suggest that combining multiple signals did not improve performance in this case, likely because the data came from different sources and the model was not specifically trained for multimodal input.

Several factors likely contributed to the reduced performance observed in the IPMA experiments. Although the dataset was collected following a consistent internal protocol (e.g., controlled environment, standardized instructions, and consistent hardware setup), its recording conditions and structure differed considerably from those of the public datasets

used for model training. Notably, differences in language, vocal effort, and recording context likely impacted signal comparability and reduced the models' ability to generalize. This highlights the challenge of applying AI models to data collected under heterogeneous real-world conditions.

Additionally, the limited number of COVID-19 samples in the IPMA dataset must be acknowledged as a critical constraint. Data collection took place during 2023, a period in which the circulation of SARS-CoV-2 had significantly decreased due to widespread vaccination and changes in public health policies. As a result, the availability of positive cases was restricted, reducing the statistical power of classification models and increasing the risk of overfitting or performance fluctuations. This limitation affects the robustness of the findings and restricts the extent to which the system's diagnostic capacity can be generalized to broader populations or emerging respiratory conditions.

To complement the performance evaluation, it was investigated how users perceived the IPMA equipment in terms of usability, a critical factor for deployment in remote or unsupervised healthcare scenarios. Quantitative results from the SUS questionnaire indicated a generally positive experience, with average scores suggesting acceptable system usability (Figure 19). However, users raised concerns about interface clarity and consistency, highlighting areas for improvement. This feedback was reinforced by the PSSUQ results (Figure 21), which showed favorable evaluations in the INTERQUAL subscale, reflecting good acceptance of the interface design.

These findings confirm that while the equipment is functional and well-received in many aspects, its design may still benefit from improvements that promote clarity, reduce cognitive load, and ensure a more intuitive experience. Importantly, the use of standardized tools like SUS and PSSUQ enabled a structured and replicable assessment of usability, providing insights that go beyond subjective impressions. Altogether, the usability results reinforce the importance of human-centered design in the development of intelligent healthcare systems and emphasize that technical performance must be matched by ease of use to ensure use in real-world settings. Ultimately, these findings support the potential of the system for application in remote triage and preliminary COVID-19 screening, especially in scenarios with limited clinical resources.

5 Risk Classification System Based on Manchester Protocol Triage

Following its initial use for COVID-19 screening, the IPMA equipment was adapted for a second clinical application: automated risk classification using the MTS. The central question guiding this investigation was whether physiological and structured data collected by IPMA may reproduce MTS clinical decisions through supervised ML models. To address this, we designed and implemented a complete pipeline that integrates data acquisition, feature selection, classifier comparison, late fusion techniques, and usability assessment via the SUS questionnaire. Both public and real-world datasets were used to validate the system's ability to generalize across triage scenarios. The following sections detail the materials, methods, experiments, and findings that support the feasibility of risk classification based on the MTS using low-cost, non-invasive multimodal signals acquired through a portable medical equipment.

5.1 Materials and Methods

5.1.1 Overview of the Proposed System for MTS-Based Risk Classification

Figure 22 presents the overall structure of the proposed system for classifying patient risk levels based on the MTS. The pipeline begins with the acquisition of clinical data, including vital signs and clinical symptoms, which may be collected through the IPMA equipment. These data are then preprocessed to handle missing values, rebalance class distributions, and encode categorical variables. A feature selection step follows, aiming to retain variables of clinical relevance while mitigating potential sources of bias. Classification models are then employed to infer the patient's risk category according to MTS guidelines. This modular pipeline was designed to reflect real-world triage conditions while enabling the integration of data-driven methods into clinical decision-making. The next sections describe each component of the system in detail.

5.1.2 Subjects and Ethical Aspects of MTS Pediatric Dataset

To evaluate the performance of the proposed classification system, a publicly available dataset originally published by (Seiger et al., 2014) was used. This dataset[1] comprises clinical records from 60,735 pediatric patients (under 16 years old), collected between 2006 and 2010 across four hospital institutions in different countries (The Netherlands, Portugal

[1] The dataset is available at the PLOS ONE repository: DOI:10.1371/journal.pone.0083267.

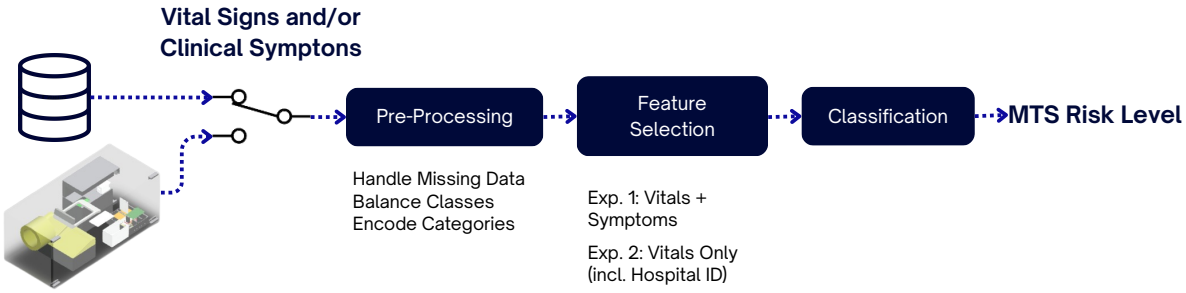


Figure 22 – Overview of the proposed system for MTS-based risk classification, comprising data preprocessing, feature selection, and classification modules.

and United Kingdom): Sophia Children's Hospital, Juliana Children's Hospital, Fernando Fonseca Hospital, and St. Mary Hospital, respectively. Each record includes MTS triage information such as the assigned category, the specific flowchart applied, and the positive discriminator selected by trained nurses. Additional data include vital signs, admission status, and follow-up outcomes.

To capture the clinical context of each triage case, the dataset provides categorical fields such as the "Positive discriminator" — with over 150 specific entries (e.g., abdominal pain, airway compromise, inability to walk) — and the "Presented problem", which groups patient complaints into eight categories (e.g., dyspnea, injuries, local infections). These categorical variables were later encoded and used in combination with numerical data during model training. In total, a set of 10 features was selected for modeling, combining both objective measurements and categorical clinical descriptors, enabling comprehensive patient risk stratification based on the MTS protocol.

5.1.3 Data Collection and Ethical Aspects of the IPMA MTS Dataset

To support the development and validation of the proposed MTS-based triage system, a real-world dataset was collected using the IPMA equipment. This study was approved by the Research Ethics Committee of the UFES, under protocol number CAAE: 77830524.9.0000.5542, and also by the Technical Research Commission of the Municipal Health Department of Vitória (SEMUS/PMV), Brazil. A formal authorization letter from SEMUS/PMV is provided in Appendix D.

The study was conducted in February 2025 at the Urgent Care Unit (Pronto Atendimento, PA) of Praia do Suá„ located in Vitória, Brazil. Participants were selected based on predefined inclusion and exclusion criteria: individuals aged 18 years or older were eligible to participate, while those under 18 were excluded. All volunteers provided informed consent prior to enrollment, in full compliance with applicable ethical standards.

Data collection was conducted directly in the triage room, allowing the IPMA equipment to be used in the same setting as routine emergency evaluations, as shown in

Figure 23. To ensure participant privacy and minimize interference with clinical operations, sessions were carried out with care to preserve the standard workflow. Researchers wore protective face masks throughout all interactions, and standard biosafety measures were followed to maintain a safe and hygienic environment for both participants and staff (see Figure 23a).

(a) (b) (c)

Figure 23 – Participant interaction and UV sterilization of the IPMA equipment. (a) Participant seated, responding and receiving instructions from researchers, (b) participant during data collection, and (c) UV sterilization of the IPMA equipment.

The data acquisition protocol followed the same structure previously adopted for the COVID-19 study. Each participant was seated comfortably, with their feet flat on the ground, and positioned their arm into the IPMA equipment, placing their index finger in the integrated oximeter (Figure 23b). Participants first performed a set of respiratory tasks, beginning with the reading of a phonetically balanced sentence: "*It is of fundamental importance to find a common solution*". They then completed five deep breaths and five voluntary coughs. Throughout the session, physiological signals (SpO_2, heart rate, body temperature, and blood pressure) were captured concurrently. Synchronizing all data streams within the same session allowed for integrated analysis of audio and physiological parameters.

To reduce the risk of cross-contamination between sessions, UVC surface disinfection was applied to the equipment after each participant (Figure 23c). At the end of the protocol, participants completed the SUS questionnaire (Appendix A) evaluating their experience with the system. A total of 20 individuals participated in this study. This protocol enabled

the controlled collection of synchronized audio and physiological signals, forming the basis for evaluating the performance and usability of the IPMA equipment in supporting autonomous MTS-based triage.

5.1.4 Data Preprocessing for MTS Pediatric Dataset

During the Exploratory Data Analysis (EDA) phase of the MTS Pediatric dataset, the presence of incomplete data was observed, particularly among the vital signs variables. As performed in (Zachariasse et al., 2021), it was assumed these missing values to be missing at random. An initial assessment revealed that categorical variables (4 out of 10 features), such as "Hospital" and "Positive discriminator", exhibited no missing values. In contrast, vital signs, including heart rate and temperature, showed a substantial proportion of missing data.

To address these missing values,a multiple imputation approach was employed using the MICE method, a strategy also utilized in (Zachariasse et al., 2021). The MICE algorithm models each feature with missing values as a function of other features, leveraging these estimates for imputation (Buuren; Groothuis-Oudshoorn, 2011). In our implementation, missing values in the numerical columns were imputed using an iterative multivariate approach, with a maximum of 100 iterations to ensure convergence. This method allowed handling of missing numerical values, ensuring the integrity and completeness of the dataset for subsequent analyses.

Furthermore, during the EDA, a significant class imbalance was identified within the target variable. To mitigate the potential for overfitting and to enhance the model's generalization capabilities, a strategy of combining less represented classes was employed, as performed in (Zachariasse et al., 2021). Specifically, the "Non urgent" and "Standard" classes were merged into a single category, as both typically imply less immediate clinical intervention. Similarly, the "Very urgent" and "Emergent" classes were grouped, given that both represent conditions necessitating rapid clinical response. This strategy not only yielded a more balanced dataset, thereby reducing the risk of a complex model, but also preserved clinical coherence and interpretability. Consequently, the redistribution of classes enhanced the potential for more robust and clinically meaningful predictive modeling.

Finally, categorical variables were transformed into appropriate numerical representations. This encoding step was essential to ensure compatibility with ML algorithms, while maintaining the semantic content of each feature and enabling their integration into the modeling pipeline.

5.1.5 Experimental Design

To evaluate the proposed system for MTS-based risk classification, three experiments were designed based on two datasets: a public triage database and real-world data collected with the IPMA equipment. Experiments #1 and #2 used the MTS Pediatric dataset to test different feature configurations, whereas Experiment #3 focused on validating the approach using prospectively acquired data in a clinical setting. The next subsections describe the setup and variables used in each case.

Experiment #1 on the MTS Pediatric Dataset Using Vital Signs and Symptoms

In EDs, initial triage decisions are typically based on a combination of objective physiological measurements and patient-reported symptoms. This dual-source approach enables a more comprehensive assessment of clinical urgency. To reflect this real-world scenario, Experiment #1 was designed to evaluate the performance of ML models trained on both types of information. The primary goal was to investigate whether routinely collected data could effectively discriminate between different MTS risk categories. The input feature set consisted of ten variables, including four numerical vital signs — respiratory rate, heart rate, temperature, and oxygen saturation — and six categorical variables representing key clinical descriptors such as pain, bleeding, dizziness, hospital of origin, presented problem, and positive discriminator. Blood pressure was not available in this dataset. This experimental setting provides a realistic basis for evaluating the feasibility of automated risk classification using comprehensive triage data.

Experiment #2 on the MTS Pediatric Dataset Using Vital Signs

Bias in MTS-based patient risk assessment may have significant implications for the effectiveness of triage protocols, particularly in terms of undertriage and overtriage. To minimize such effects, this experiment was designed with a primary focus on numerical variables — arrival date, age, respiratory rate, heart rate, temperature, and SpO_2 — as the core predictors, aiming to provide a more objective and standardized basis for risk classification.

In addition to these vital signs, the "Hospital" variable was included to account for differences in medical equipment and measurement procedures across institutions, which may introduce subtle discrepancies in the recorded values. Within the specific operational context of the IPMA equipment — which is designed for autonomous use by patients — the "Positive discriminator" variable was also considered, and its individual categories were evaluated using mutual information to identify the most relevant for prediction. Among the most informative categories, those that introduced minimal patient-related bias were "Recent problem" and "Significant history" (the latter including important conditions such as diabetes). Some other categories, such as "Hot" or "Low SaO_2", were found to be

redundant with the numerical vital signs but were retained due to their clinical relevance and potential to reinforce the consistency and interpretability of the model's predictions.

Experiment #3 on the IPMA MTS Dataset

Following the experiments conducted on the MTS Pediatric dataset, a third experiment was designed to evaluate the performance of ML models using real-world data collected directly with the IPMA equipment. This experiment aimed to assess the feasibility of deploying the proposed risk classification approach in autonomous triage scenarios, where patients interact with the equipment without direct professional supervision.

The dataset used in this experiment comprises synchronized physiological measurements and symptom information acquired from participants at the Praia do Suá PA, as detailed in Section 5.1.3. The selected feature set included heart rate, body temperature, and SpO_2. Although systolic and diastolic blood pressure were collected using the IPMA equipment, these parameters were excluded from the analysis to maintain compatibility with the MTS Pediatric dataset, which does not include blood pressure measurements.

This experimental setting enabled the evaluation of model performance under real-world operational conditions, accounting for potential noise in signal acquisition, user interaction variability, and the limited sample size inherent to on-site deployments. By leveraging data collected in situ, this experiment served as a important step toward validating the IPMA equipment's capability to support reliable, low-bias MTS-based risk classification in practical healthcare environments.

5.1.6 Classification and Evaluation

To evaluate the proposed models for MTS-based triage, an experimental protocol was used combining repeated data partitioning, hyperparameter optimization, and stratified evaluation. Specifically, the holdout method was used to divide the dataset into training (80 %) and testing (20 %) subsets via stratified random sampling, preserving the proportion of class labels in both sets. To ensure that results were not dependent on a specific split, this entire pipeline — including data splitting, preprocessing, model training, and evaluation — was repeated 30 times. In each iteration, a new random split was generated, enabling a broader assessment of model generalization and reducing the influence of partition-related bias. Performance was quantified using two widely adopted metrics: ACC and F1-score, both calculated on the test set.

Prior to modeling, preprocessing steps were applied as described in previous sections. Importantly, data normalization was performed after the train-test split, using statistics computed solely from the training set to prevent information leakage. Hyperparameter tuning was then carried out on the training set using a stratified K-Fold cross-validation

strategy, aiming to maximize the F1-score. The best-performing configuration was selected and used to train the final model, which was subsequently evaluated on the corresponding test fold in each repetition.

A diverse set of supervised ML algorithms was use, comprising both individual and ensemble-based approaches. The individual classifiers included DT, MLP and kNN. Ensemble methods included RF, XGBoost, Stacking, and two Voting variants (hard and soft). In the implementation of the ensemble strategies, DT, MLP, XGBoost, kNN, and RF were used as individual models in both Stacking and Voting configurations. For the Stacking model, predictions from the base classifiers were combined via a Logistic Regression meta-learner, which was selected for its simplicity and effectiveness in learning decision boundaries across heterogeneous learners. Among all models, MLP was the only classifier trained using epoch-based optimization, with the number of epochs treated as a key hyperparameter. The remaining models, including the ensemble techniques, relied on non-iterative training schemes.

Finally, statistical analyses were conducted to compare model performance and assess potential differences in subject characteristics. A significance level of $p < 0.05$ was adopted for all statistical tests. For categorical variables, associations with the outcome were evaluated using the Chi-squared (χ^2) test. For numerical variables, the Shapiro–Wilk test was applied to assess normality. Given that the variables followed a normal distribution, parametric testing was employed. One-Way ANOVA was used to compare means across groups and Tukey's HSD post hoc test was used for pairwise comparisons between classifiers while controlling the family-wise error rate.

5.1.7 Usability Evaluation

Given the autonomous nature of the IPMA equipment and its intended use in real-time triage, assessing usability is essential to ensure that users can interact with the equipment efficiently and comfortably. In this study, the focus was on evaluating usability in the context of MTS-based risk assessment, where participants provided vital signs and symptom information without professional assistance. As in the previous COVID-19 study, a post-interaction questionnaire was administered to capture subjective perceptions of system usability. However, it was observed in the earlier protocol that applying both the SUS and PSSUQ questionnaires resulted in respondent fatigue, particularly in time-constrained clinical environments. To address this, the strategy was revised for the current study and opted to use only the SUS, due to its brevity and well-established reliability in usability research. Participants completed the SUS immediately after using the system. Their responses were analyzed to assess perceived usability, identify interaction challenges, and guide future refinements in the design and deployment of the IPMA platform for autonomous triage support. In addition to patient feedback, SUS questionnaires were also

administered to healthcare professionals involved in the study, aiming to capture their perspective on the system's usability and potential integration into clinical workflows.

5.2 Results of the Risk Classification System

This section presents the main findings from the evaluation of the proposed risk classification system, based on clinical triage data and physiological measurements. The results are organized to provide a comprehensive view of subject characteristics, model performance, and statistical comparisons across experimental conditions. The section first outlines the clinical and physiological profiles of the study subjects, highlighting patterns relevant to automated triage. It then details the performance of multiple classification algorithms across three experimental setups: (1) a full-feature model trained and tested on the MTS Pediatric dataset, (2) a reduced-feature model using vital signs also for the MTS Pediatric dataset, and (3) a cross-dataset generalization test using IPMA-collected data. Results are presented in terms of accuracy and F1-score, followed by post hoc statistical analysis to assess significant differences among classifiers.

5.2.1 Characteristics of the Study Subjects

MTS Pediatric Dataset

Automated triage systems require input variables that not only reflect clinical severity but also exhibit consistent patterns across urgency levels. Identifying which features most effectively distinguish risk groups is essential for building robust classification models. To prioritize the most informative clinical features for real-time risk prediction, associations between patient variables and Manchester Triage categories were analized, aiming to balance statistical significance with physiological and practical relevance.

Using a dataset of 60,735 ED visits, Chi-squared tests were applied to evaluate the association between clinical variables and triage urgency, as summarized in Table 7. Age and body temperature were significantly associated with urgency levels ($p \ll 0.05$), but their distributions overlapped substantially across categories, limiting their standalone discriminative power. Heart rate and respiratory rate showed more distinct changes across triage categories, especially in higher-risk groups, and were better indicators of clinical deterioration. Higher urgency levels were also associated with increased hospitalization rates, further supporting the validity of the triage classification. Among all variables, SpO_2 demonstrated the strongest association with urgency levels. The proportion of patients with oxygen saturation below 95 % increased substantially in the Urgent and Very urgent + Emergent groups, indicating a clear relationship between lower SpO_2 and clinical severity. This consistent and clinically meaningful pattern confirms SpO_2 as a valuable single predictor for identifying critical cases in automated triage systems.

Table 7 – Clinical profile of patients according to triage urgency classification.

	Non urgent + Standard	**Urgent**	**Very urgent + Emergent**	**p-value**
	$n = 36{,}577$	$n = 16{,}817$	$n = 7{,}341$	
Age				
0–3	18247	8321	4542	$\ll 0.05^{*,a}$
4–7	8282	3415	1221	
8–11	5577	2641	757	
12–15	4465	2437	821	
16–18	4	2	0	
Invalid	2	1	0	
Respiratory Rate				
0–29	21290	9058	2617	$\ll 0.05^{*,a}$
30–59	15150	7565	4299	
60–89	125	190	419	
90–119	11	4	6	
≥ 120	1	0	0	
Invalid	0	0	0	
Heart Rate				
0–59	53	40	35	$\ll 0.05^{*,a}$
60–149	33977	14293	5416	
150–199	2522	2406	1797	
200–220	24	74	83	
Invalid	1	4	10	
Temperature				
0–33.9	3	1	11	$\ll 0.05^{*,a}$
34.0–35.9	343	220	140	
36.0–37.9	33820	12663	5529	
38.0–39.9	2338	3487	1493	
40.0–41.9	73	446	168	
Invalid	0	0	0	
SpO_2				
< 70	4	5	15	$\ll 0.05^{*,a}$
70–79	6	14	16	
80–89	9	39	141	
90–94	107	250	540	
95–100	36399	16468	6616	
Invalid	52	41	13	
Hospitalization				
No	34974	14115	4751	$\ll 0.05^{*,a}$
Yes	1603	2702	2590	

[*] Significant values $p < 0.05$.
[a] Chi-squared test.

It is worth noting that physiologically implausible values, such as negative readings or measurements far outside clinically acceptable ranges, were intentionally retained in the dataset. This decision reflects real-world operational conditions, particularly in autonomous triage scenarios, where user input errors, sensor noise, or device malfunctions may lead to such anomalies. By preserving these values, the analysis captures the variability and noise present in emergency care settings, allowing the model to become more robust to imperfect data rather than relying on idealized inputs.

IPMA MTS Dataset

To describe the physiological characteristics of patients classified as "Non urgent + Standard" by the MTS, the dataset collected with the IPMA equipment (Table 8) was used. Given that all individuals belonged to the same triage category, the goal was to identify potential variability or borderline findings within this supposedly low-risk group. The majority of participants presented vital signs within expected ranges, including normal body temperature (mean 36.52 °C) and SpO_2 levels predominantly between 96–98 %. However, specific cases raised clinical attention: one individual exhibited a SBP of 221 mmHg, and two others had heart rates of 120 bpm, values that may carry potential clinical relevance even in non-critical patients. These outliers suggest that some cases within the "Non urgent + Standard" class may still exhibit physiological signs warranting closer monitoring.

Table 8 – Clinical and physiological characteristics of the patients evaluated with IPMA.

	Interval	**n = 20**	**mean (std)**
Age			41.58 (17.39)
	0–19	1	
	20–29	5	
	30–39	4	
	40–49	2	
	50–59	3	
	60–69	3	
	≥ 70	1	
	Missing	1	
SBP (mmHg)			126.55 (27.89)
	80–99	2	
	100–119	7	
	120–139	7	
	140–159	3	
	≥ 200	1	
DBP (mmHg)			78.00 (17.27)
	≤ 59	1	
	60–69	8	
	70–79	3	
	80–89	3	
	90–99	3	
	≥ 100	2	
Heart Rate (bpm)			83.50 (17.11)
	≤ 69	5	
	70–79	5	
	80–89	4	
	90–99	3	
	≥ 100	3	
SpO_2 (%)			97.25 (0.89)
	96	4	
	97	9	
	98	5	
	≥ 99	2	
Temperature (°C)			36.52 (0.45)
	36.0–36.4	12	
	36.5–36.9	5	
	≥ 37.0	2	
	Missing	1	

5.2.2 Performance of the Classification Model

Results on the MTS Pediatric Dataset – Experiment #1

Effective triage automation depends on models that may convert diverse inputs — vital signs and symptom reports — into consistent, accurate risk categories, even under real-world variability. To determine the most reliable classification strategies for this task, a range of individual and ensemble models was evaluated using a comprehensive clinical feature set from the MTS Pediatric dataset. Figures 24 and 25 present the ACC and F1-score results for eight classifiers: DT, RF, MLP, XGBoost, kNN, and three meta-ensemble methods (Stacking, Soft Voting, Hard Voting). Among individual models, XGBoost achieved the highest mean ACC (99.06 % $\pm$ 0.07) and F1-score (0.99 $\pm$ 0.00), slightly outperforming RF (98.81 % $\pm$ 0.09) and DT (98.04 % $\pm$ 0.12). Stacking and Soft Voting matched XGBoost's performance nearly identically, both reaching 99.04 % and 98.87 % ACC, respectively, and F1-scores of 0.99, with minimal variance ($std \leq 0.09$). In contrast, MLP and kNN underperformed, with lower ACC (87.70 % and 89.55 %) and greater variability, especially in MLP ($std = 0.85$ in ACC), suggesting less robustness in handling structured clinical data. These findings reveal a clear performance advantage for tree-based models — particularly XGBoost and ensemble strategies — under the current experimental conditions. Their high ACC and F1-score and low variability suggest potential for triage classification tasks.

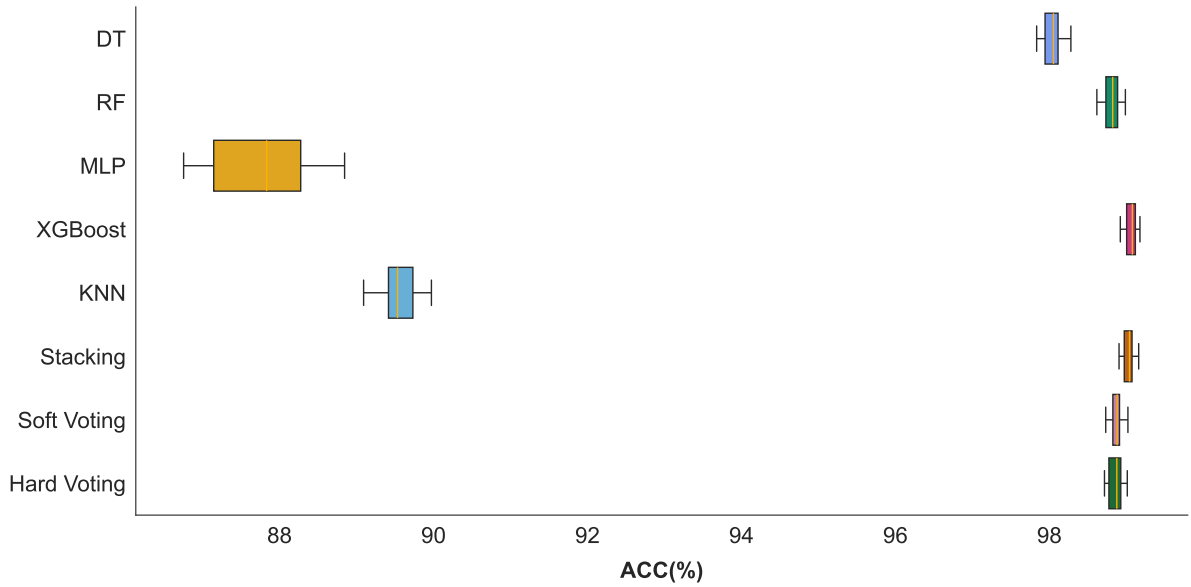


Figure 24 – Performance in terms of ACC on the test set using both numerical and categorical features

Figure 26 shows the results of the post hoc analysis following the One-Way ANOVA test, applied to assess pairwise differences in classifier performance. Statistically significant differences ($p < 0.05$) were observed between several models. Notably, DT differed significantly from all other classifiers, indicating a distinct performance profile despite

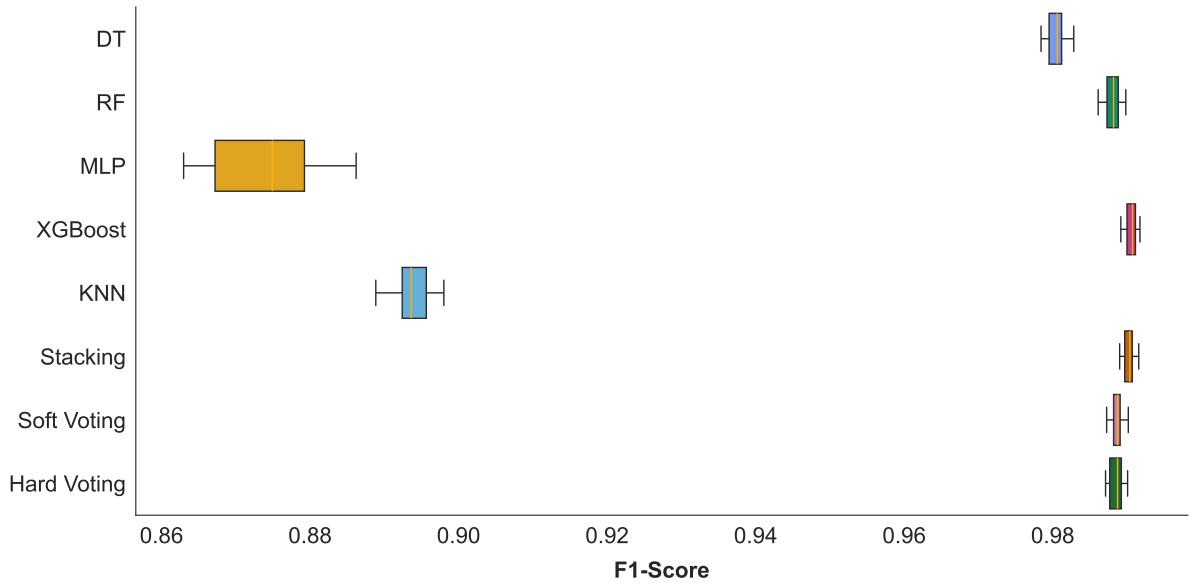


Figure 25 – Performance in terms of F1-score on the test set using both numerical and categorical features

maintaining high average ACC and F1-score. kNN also exhibited significant differences when compared to all other models, including MLP, reflecting its relative distance from the top-performing group. MLP, in turn, showed statistically lower performance when compared to RF, Stacking, Soft Voting, Hard Voting, and XGBoost.

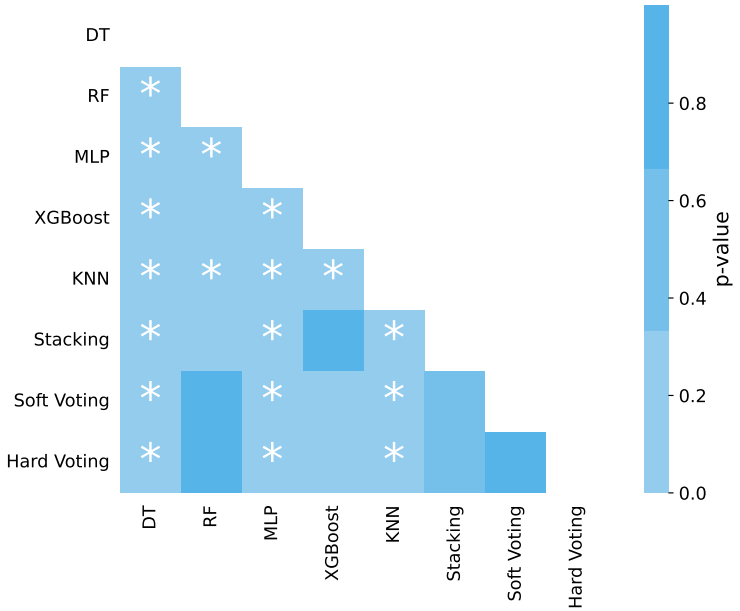


Figure 26 – Pairwise p-values from Tukey's post hoc test following One-Way ANOVA comparing classifier performance on the MTS Pediatric dataset. Asterisks (*) indicate statistically significant differences ($p < 0.05$).

XGBoost demonstrated significantly higher performance than DT, kNN, and MLP, aligning with its strong average metrics. RF showed differences when compared to DT,

MLP, and kNN, but not with the ensemble methods or XGBoost. The ensemble classifiers — Stacking, Soft Voting, and Hard Voting — presented statistically significant differences only when compared to DT and MLP, highlighting their overall consistency and robustness. Although several pairwise differences reached statistical significance, the absolute mean metrics among the top five classifiers (RF, XGBoost, Stacking, Soft Voting, and Hard Voting) remained tightly clustered around 99 %, with low standard deviations. This underscores the need to interpret statistical differences alongside practical performance, especially in clinical contexts where minimal gains may not justify added complexity.

Results on the MTS Pediatric Dataset – Experiment #2

As previously discussed, undertriage or overtriage may result from inaccuracies in risk classification, particularly when relying on subjective or categorical assessments prone to biases from healthcare professionals or patient-reported symptoms (Mackway-Jones; Marsden; Windle, 2014; Jatobá et al., 2018). To mitigate these biases, Experiment #2 was conducted utilizing primarily numerical biomedical signals directly measured from patients, alongside the hospital identifier as an additional categorical variable to control for measurement variability across institutions.

Stacking, XGBoost, and Soft Voting emerged as the top-performing classifiers, exhibiting the highest median scores in both ACC and F1-score, as shown in Figures 27 and 28. Stacking slightly outperformed the others, with a median ACC of 74.94 % and an F1-score of 0.74, closely followed by XGBoost (74.88 %, 0.73) and Soft Voting (74.17 %, 0.74). Hard Voting and kNN formed an intermediate tier, with median ACC values of 73.85 % and 72.64 %, and F1-scores of 0.73 and 0.71, respectively. Although MLP and RF showed comparable ACC values (71.70 % and 70.91 %), MLP's F1-score (0.69) was notably lower than RF's (0.72), suggesting reduced consistency in class-wise performance. DT consistently ranked lowest across both metrics, with a median ACC of 66.19 % and an F1-score of 0.68. These results confirm the superior predictive capacity of ensemble-based approaches while also underscoring the limitations of simpler tree-based models like DT in this classification task.

To statistically assess performance differences between classifiers, a pairwise comparison was performed using One-Way ANOVA (Figure 29), consistent with the analysis conducted in Experiment #1 (see Section 5.1.5). The results show statistically significant differences ($p < 0.05$) in most pairwise comparisons. In particular, ensemble methods such as Stacking and Soft Voting consistently outperformed DT and MLP, and showed statistically superior performance to kNN despite closer median values. These findings reinforce the effectiveness of ensemble strategies in this classification task. Although it did not outperform the full feature set, the use of quantitative biomedical signals combined with institutional identifiers yielded consistent and promising results, suggesting potential

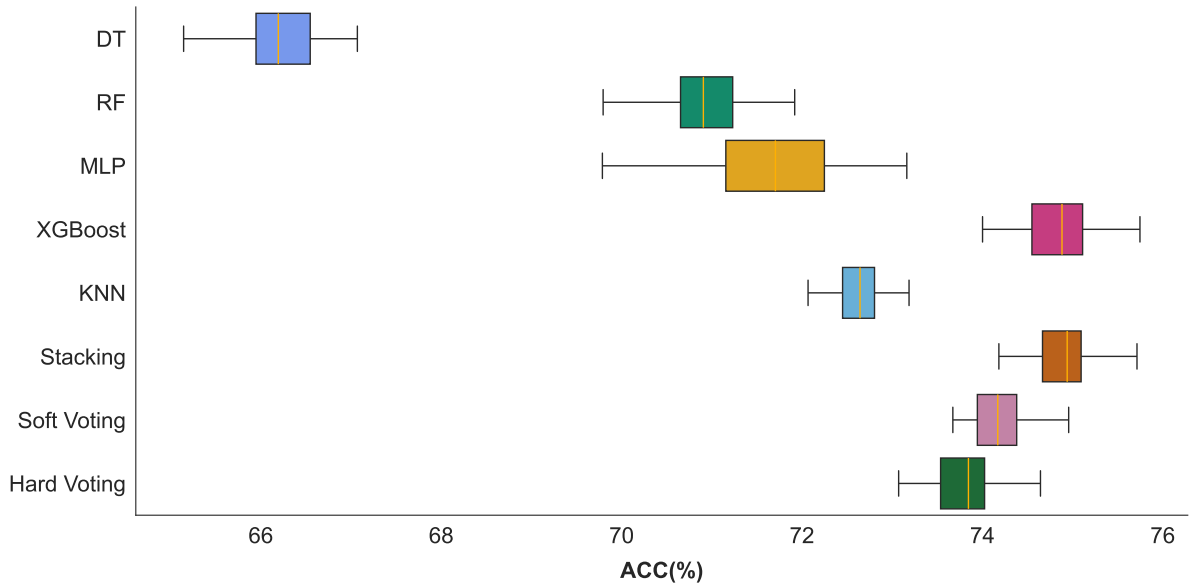


Figure 27 – Performance in terms of ACC on the test set for Experiment #2

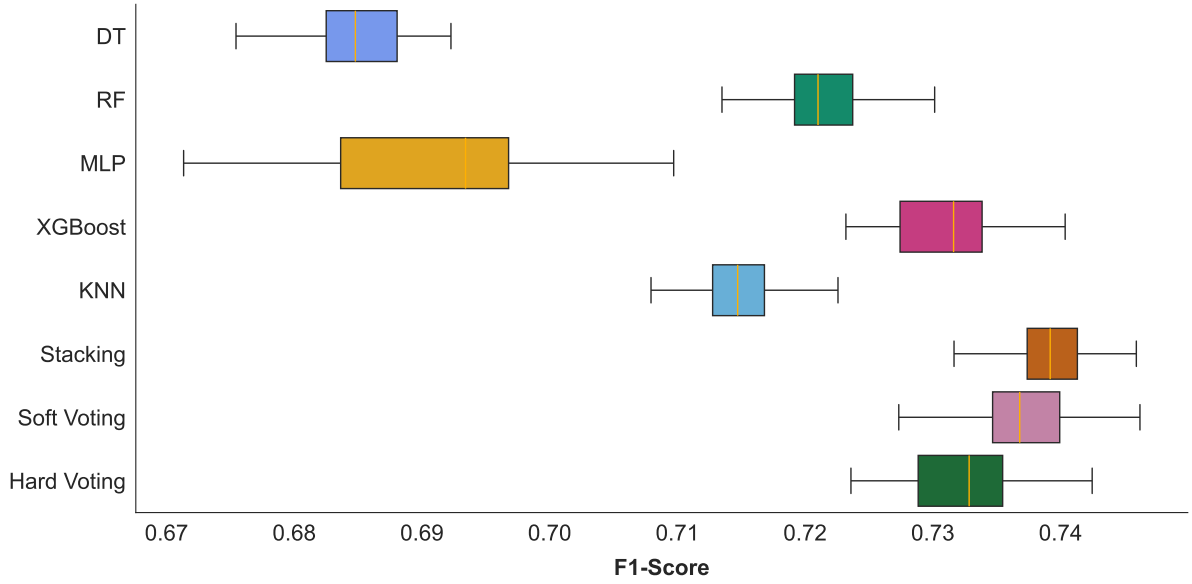


Figure 28 – Performance in terms of F1-score on the test set for Experiment #2

value in certain triage scenarios.

Results on the IPMA MTS Dataset

Following the cross-dataset experimental design, this final experiment aimed to evaluate whether models trained on pediatric data could accurately classify patient risk levels when tested on data collected through the IPMA equipment in an adult population. The challenge was to assess model performance under real-world operational conditions, including signal variability, autonomous usage, and physiological differences between datasets. All models exhibited extremely low performance, with ACC and F1-scores

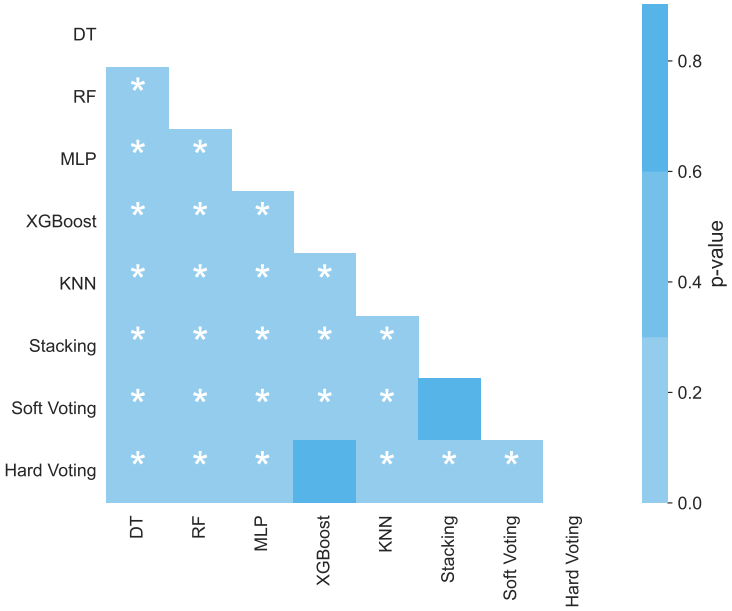


Figure 29 – Pairwise statistical comparison of classifier performance for Experiment #2. The symbol $*$ denotes a statistically significant difference ($p < 0.05$)

equal to 0 % across the board. MLP and kNN showed slightly higher mean ACCs (17 % and 12.17 %, respectively), along with mean F1-scores of 0.21 and 0.15. However, their large standard deviations (29.67 % for ACC and 0.34 for F1-score in MLP) indicate inconsistent behavior across executions. The remaining classifiers yielded minimal predictive performance, highlighting the difficulty of applying pediatric-trained models to adult data. These findings underscore the need for domain-specific training or the use of domain adaptation strategies to ensure reliable performance when deploying autonomous triage systems across distinct clinical populations.

5.2.3 Inference Time Analysis on the MTS Pediatric Dataset

To complement the evaluation of predictive performance, the computational efficiency of each model for MTS Pediatric dataset was analyzed. Table 9 presents the normalized response time of the evaluated models for both experimental settings. The values were scaled between 0 and 1, using the slowest model in each experiment as reference. As expected, the DT classifier achieved the fastest response time in both scenarios, 2.41 ms, with the normalized value of 0.01 indicated in Table 9. kNN and XGBoost also showed very fast response times (0.02). MLP exhibited intermediate times (0.16). Ensemble strategies, including Hard Voting and Soft Voting, demonstrated relatively fast response times (0.05–0.06). In contrast, Stacking was the slowest model in both experiments, reaching 1.00 in Experiment #1 and 0.79 in Experiment #2. RF was consistently the second slowest, with normalized times of 0.81 and 0.69, respectively. Overall, Experiment #2, which considered only numerical features, resulted in faster inference for most models, except for RF and

Stacking.

Table 9 – Performance of classifiers under different feature configurations

Classifier	Experiment	ACC (%)	F1-Score	Normalized Time
DT	# 1	98.04 (± 0.12)	0.98 (± 0.00)	0.01
	# 2	66.18 (± 0.54)	0.68 (± 0.00)	0.01
RF	# 1	98.81 (± 0.09)	0.99 (± 0.00)	0.81
	# 2	70.93 (± 0.49)	0.72 (± 0.00)	0.69
MLP	# 1	87.69 (± 0.85)	0.87 (± 0.01)	0.15
	# 2	71.62 (± 0.91)	0.69 (± 0.01)	0.15
XGBoost	# 1	99.06 (± 0.07)	0.99 (± 0.00)	0.02
	# 2	74.85 (± 0.35)	0.73 (± 0.00)	0.02
KNN	# 1	89.55 (± 0.21)	0.89 (± 0.00)	0.03
	# 2	72.63 (± 0.35)	0.71 (± 0.00)	0.02
Hard Voting	# 1	98.86 (± 0.09)	0.99 (± 0.00)	0.06
	# 2	73.81 (± 0.41)	0.73 (± 0.00)	0.06
Soft Voting	# 1	98.87 (± 0.09)	0.99 (± 0.00)	0.05
	# 2	74.14 (± 0.37)	0.74 (± 0.00)	0.06
Stacking	# 1	99.04 (± 0.07)	0.99 (± 0.00)	1.00
	# 2	74.91 (± 0.37)	0.74 (± 0.00)	0.79

5.3 Results of the IPMA Usability for Risk Classification System

During the IPMA COVID-19 usability evaluation, in which the same equipment was previously applied for COVID-19 screening, both the SUS and PSSUQ questionnaires were used to assess user experience. However, it was observed that the combination of both instruments was excessively long and led to participant fatigue, especially in clinical settings where time and patient availability were limited. This limitation prompted the investigation group to revise usability assessment strategy. For the current study involving risk classification, it was opted to apply only the SUS, given its concise format and established reliability. The complete list of SUS items is available in Appendix A in Portuguese. This adjustment enabled a more agile and feasible usability evaluation without compromising the quality of the collected feedback.

As shown in Figure 30(a), the analysis of individual SUS questions revealed a consistent trend of positive responses, particularly for items 1, 3, 5, 7, and 9, which reflect perceived ease of use and user confidence. Among the negatively worded items, Question 6 — related to the perceived need for technical support — showed slightly more dispersed responses, suggesting that a few participants might have felt unsure about using the system independently. In contrast, responses to Question 2 (complexity) were consistently low,

indicating that most users did not find the system unnecessarily complex. Despite these variations, overall usability perception remained high. As shown in Figure 30(b), total SUS scores ranged from 62.5 to 95.0, with a mean of 82.20, indicating that participants generally found the system effective, intuitive, and appropriate for use in clinical environments.

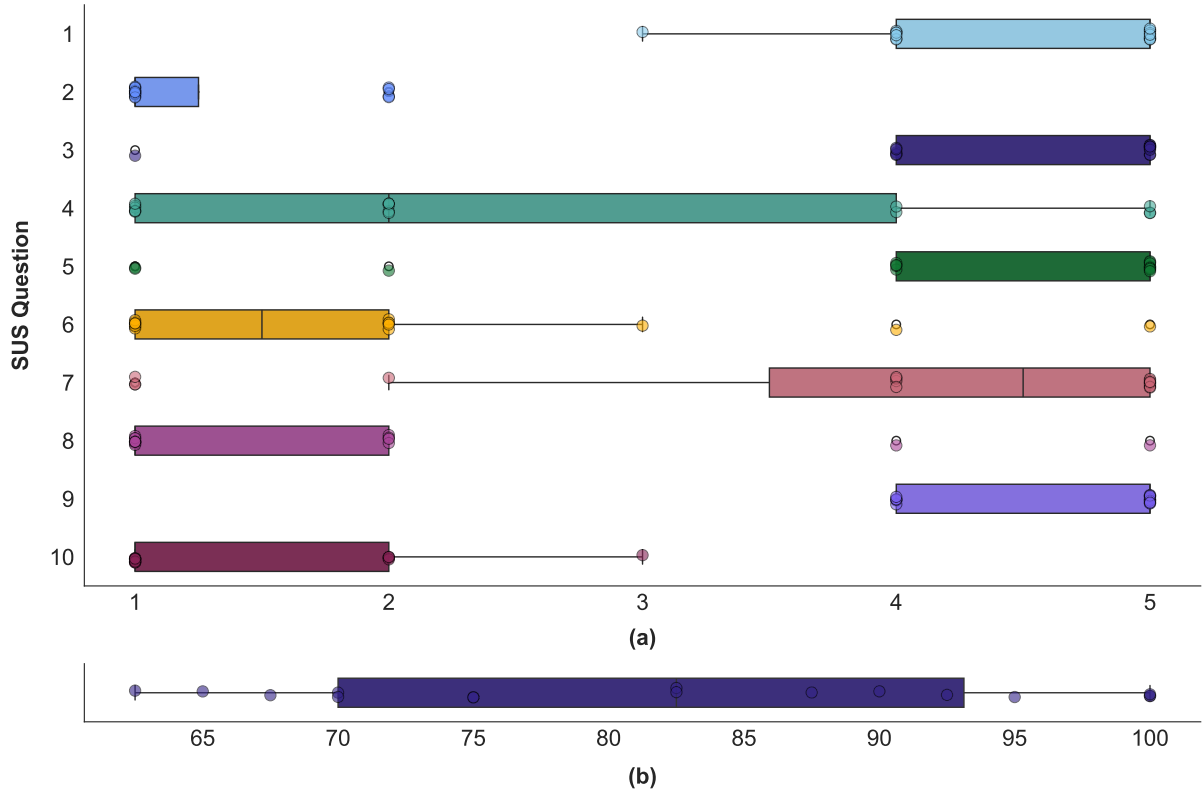


Figure 30 – SUS results for the IPMA equipment. (a) Individual question scores. (b) Overall SUS score distribution across participants.

To complement the evaluation from the patients' perspective, the perception of healthcare professionals regarding the usability of the IPMA equipment was also assessed. This step aimed to understand how the system was perceived by healthcare professionals, who play a key role in its potential integration into clinical workflows. Two professionals from the Praia do Suá PA completed the SUS questionnaire after interacting with the equipment. As shown in Table 10, both healthcare workers provided highly favorable ratings, with total SUS scores of 87.5 and 100, respectively. These results reinforce the system's perceived ease of use and its potential acceptance by clinical staff, an essential factor for successful deployment in real-world healthcare environments.

Table 10 – SUS results for the IPMA equipment based on responses from healthcare professionals at the Praia do Suá PA.

ID	01	02	03	04	05	06	07	08	09	10	SUS
P01	5	1	5	4	5	1	4	1	5	2	87.5
P02	5	1	5	1	5	1	5	1	5	1	100

5.4 Discussion

Automated decision making in clinical triage remains a challenge, particularly in ED where rapid and accurate risk stratification is essential to avoid under or overtriage. To address this need, this study investigated whether low cost, noninvasive signals collected via a portable equipment, the IPMA, could replicate the clinical decisions defined by the MTS. A three stage experimental protocol combining public pediatric triage data and real world adult data acquired using the IPMA was developed, comparing different classification strategies under both ideal and realistic conditions.

To contextualize model performance, this work initially evaluated the clinical and physiological profiles of the study subjects, aiming to identify the most informative variables for risk stratification. By analyzing the MTS Pediatric dataset, it was observed that SpO_2 was the feature most clearly associated with triage urgency levels. Patients in higher-risk categories exhibited significantly lower SpO_2 values, reinforcing its clinical importance in the early identification of deterioration. Similar associations have been reported in the literature. Seiger et al. (2014) included SpO_2 as a key variable in their improved triage models and showed that reduced SpO_2 was associated with increased risk of hospitalization in pediatric emergency visits, whereas Zachariasse et al. (2021) emphasized its value in pediatric early warning scores. The underlying reason for SpO_2's discriminative power lies in its physiological role as a marker of respiratory efficiency and tissue oxygenation — both of which are often compromised in acute illness. These associations are not limited to pediatric populations. In an adult triage study conducted by Jesus et al. (2021), over 63 % of patients in the red priority group presented with altered SpO_2 values (< 95 %), while significant alterations in blood pressure, heart rate, and respiratory rate were also concentrated among higher urgency levels. These findings confirm that vital signs remain reliable indicators of clinical severity across age groups, and their integration into automated triage systems may help identify high-risk patients even when categorical labels suggest otherwise.

Given the strong association between vital signs and clinical urgency, it was investigated whether these variables, when used alongside structured categorical data, could support automated classification of triage priority levels. In Experiment #1, it was found that models such as XGBoost and meta-ensembles like Stacking, achieved near-perfect performance when trained with both categorical and numerical features from the MTS Pediatric dataset. This result demonstrates the potential of structured clinical data to support highly accurate and consistent triage classification, even in large and heterogeneous datasets. Similar findings have been reported by Porto (2024), who identified ensemble models as top performers in clinical triage tasks, and by Cicolo and Peres (2019), who observed accuracy gains when transitioning from manual to electronic MTS implementations. The results obtained in this Thesis go further, showing that automated

models not only match but may surpass traditional triage strategies, reaching F1-scores above 0.99. However, such high performance may reflect dataset-specific characteristics, particularly the presence of strongly predictive categorical variables. Features such as "Positive discriminator" and "Presented problem" may have captured information closely tied to the triage label, potentially increasing performance in ways not reproducible in autonomous or unsupervised contexts. Moreover, the exclusive use of pediatric records in the training set further limits the model's applicability to broader populations.

Another important limitation concerns the quality and completeness of the structured numerical features. The MTS Pediatric dataset presented non-negligible proportions of missing data in vital sign variables, particularly heart rate and temperature. To address this, it was adopted the MICE strategy, which is widely accepted for handling missing clinical data. However, the extent of missingness in these key predictors raises concerns about the robustness of the final model. In high-missingness settings, imputed values may carry limited individual-level ACC, potentially introducing bias or artificial correlations into the training process. Although MICE has shown strong performance in healthcare applications Zachariasse et al. (2021), its application in this study was limited to a few vital signs with moderate levels of missingness.

In Experiment #2, it was evaluated whether ML models could classify triage priority levels using only numerical features such as vital signs, arrival date, and hospital identifier. While some low-subjectivity categorical variables — including "Positive discriminator", "Recent problem", and "Significant history" — were retained due to their relevance and structured nature, the configuration simulated a more constrained scenario by excluding most free-text and highly subjective descriptors. The best-performing models, including Stacking and XGBoost, achieved ACC and F1-scores around 74 %, while DT presented the weakest performance (ACC = 66.19 %, F1-score = 0.68), indicating reduced performance with numerical-only inputs.

Despite the performance drop compared to Experiment #1, the results from Experiment #2 remain promising. Even when trained primarily on objective data, the models were able to identify clinically meaningful patterns and maintain satisfactory predictive performance. For instance, F1-scores around 0.74 indicate that vital signs alone may support risk stratification with reasonable ACC. This aligns with findings from Chang et al. (2024), who demonstrated that structured variables such as blood pressure and respiratory rate contributed substantially to predicting clinical outcomes, even in the absence of richer contextual information. Studies such as Nsubuga et al. (2025) and Masanneck et al. (2024) also reinforce that simplified feature sets, including physiological variables alone, may still support triage decisions with moderate-to-high accuracy, especially when deployed in resource-limited environments. Moreover, recent evaluations of ML-based scoring tools indicate that even limited models may outperform

traditional triage scales like Kampala trauma score (KTS) or early warning scores when trained appropriately (Chang et al., 2024).

Nonetheless, relying exclusively on numerical inputs introduces certain limitations. Categorical fields such as "Presented problem" and "Positive discriminator", although potentially subjective, often encode important clinical reasoning that complements vital signs. Their exclusion may partially explain the performance plateau observed in Experiment #2. Still, the results remain relevant: achieving F1-scores around 0.74 with mostly objective features reinforces the feasibility of using simplified inputs in autonomous triage applications. Moreover, the practical implications of model efficiency must be considered. While complex strategies like Stacking achieved high performance, they exhibited higher computational demands. In contrast, ensemble methods like Soft Voting maintained consistent performance while demonstrating low inference cost. For instance, Soft Voting achieved a median normalized inference time of 0.06 (relative to the slowest model), compared to 0.90 for Stacking. This balance between ACC and efficiency reinforces the practical value of lightweight models in fast-paced or resource-limited settings. As highlighted by Masanneck et al. (2024), AI-based systems may contribute to reducing workload and triage variability, particularly when designed to support timely and consistent clinical decision-making.

To assess the generalizability of models trained on pediatric data, their performance were evaluated on an adult dataset collected using the IPMA equipment in real-world conditions. As anticipated, most classifiers failed to generalize across domains, with accuracy and F1-scores near zero. Even high-performing models like XGBoost and Stacking could not replicate their previous success. Only MLP and kNN reached slightly higher values (around 17 % and 12 %, respectively), but with high variability across runs. This outcome reinforces known challenges in cross-population inference due to physiological and demographic differences, combined with the lack of exposure to adult signal patterns during training. While limited, this evaluation provides empirical evidence of the risks associated with applying models trained on structured datasets to real-world data from different populations.

Despite the value of this cross-domain evaluation, several critical limitations must be acknowledged. First, the data acquisition at the Praia do Suá PA was restricted to patients classified in the lowest triage priority level, as higher-risk individuals were not included to avoid disrupting clinical workflows. This constraint introduces a strong sampling bias, as the model was never exposed to the physiological profiles or signal variability associated with more severe conditions. Consequently, the observed failure to generalize across populations may reflect not only demographic differences but also the overfitting of models to a narrow and homogeneous training distribution. Such exclusion compromises the robustness and applicability of the IPMA-based inference in broader clinical contexts.

These findings reinforce a critical requirement for autonomous triage systems: population-specific training and adaptation are essential. For portable equipment like the IPMA to function reliably in real-world, resource-constrained environments, algorithms must be trained with representative data and account for variability in signal quality, demographic profiles, and acquisition conditions. Otherwise, performance may degrade substantially during deployment, undermining clinical reliability.

To complement the technical analysis, the usability of the IPMA equipment was evaluated in autonomous triage scenarios. Usability plays a central role in the real-world success of AI-based systems, particularly when used by non-specialists. The SUS score averaged 82.20, reflecting high acceptance among users. Participants reported that the system was easy to use, predictable, and inspired confidence. Healthcare professionals gave particularly high scores (87.5 and 100), indicating clinical approval. Although variation was observed in responses to negatively worded items, such as perceived need for technical support, this is common in SUS assessments and did not significantly impact the overall positive impression. Compared to the COVID-19 approach, in which the combined use of SUS and PSSUQ led to participant fatigue and inconsistent responses, focusing solely on the SUS in this study proved more effective and interpretable. These results suggest that focusing on usability may help make autonomous triage systems more ready for real-world use.

6 Conclusion and Recommendations

This chapter concludes the work by summarizing the key findings in relation to the research objectives and hypotheses, and discusses the main contributions provided by this research. It also describes the limitations of the study and proposes recommendations for future work.

6.1 Major Findings

This research addressed a gap in digital health by proposing and validating an AI-based framework designed to support COVID-19 detection and clinical triage using multimodal physiological signals acquired by the IPMA equipment. The findings indicate that developing ML and DL models based on real-world physiological data is a promising approach, even in non-controlled environments. This demonstrates the potential of integrating objective physiological measurements and AI methods to assist clinical decision-making in settings with limited resources.

This work focused on the development and evaluation of AI models trained to process and analyze various physiological signals, including audio recordings such as cough, speech, and breathing, as well as vital signs such as SpO_2, heart rate, and temperature. These signals were explored individually and in combination to assess their contribution to COVID-19 inference. The results suggest that speech and breath signals offered more consistent performance compared to cough, while SpO_2 and temperature emerged as informative indicators among the vital signs. Models were also designed to perform risk classification based on the MTS, using both public datasets and physiological data collected with the IPMA. The ability to apply models trained on external datasets to real-world clinical signals highlights the relevance of this approach for practical deployment.

In parallel to the classification tasks, the usability of the IPMA equipment was evaluated to ensure its suitability for deployment in clinical workflows. Feedback from patients and healthcare professionals was gathered using standardized instruments, including the SUS and PSSUQ. The results indicated high levels of user satisfaction, particularly regarding interface clarity, ease of use, and perceived reliability. While a few aspects related to consistency and user confidence revealed opportunities for refinement, the overall perception was positive and aligned with expectations for medical technologies intended for unsupervised or assisted use.

This work demonstrates the feasibility of designing and validating a clinically relevant AI-based system for COVID-19 detection and risk classification using low-cost,

non-invasive signals acquired in real-world conditions. By integrating ML techniques with physiological and audio data collected through the IPMA, the system addresses a gap in remote and decentralized healthcare. The results respond directly to the central research question and contribute toward the development of intelligent triage strategies that balance technical performance with practical usability. Rather than presenting a final solution, this study offers a structured and adaptable foundation that may support future efforts to expand access to early and objective clinical assessment, particularly in resource-limited settings.

6.2 Limitations

This research has some potential limitations that should be acknowledged. The dataset collected with the IPMA included a small number of participants, particularly among COVID-19 positive individuals, which limited the statistical robustness of the experiments. In the risk classification task, the use of a pediatric dataset for training and adult data for testing introduced population mismatches that may have affected generalization. Furthermore, some clinical categories were underrepresented, leading to class imbalance in both tasks. Usability was assessed in a controlled setting with a small group of users, which may not fully reflect performance in busy clinical environments. Additionally, some datasets, particularly the MTS Pediatric dataset, contained missing or incomplete values in key physiological variables. Although imputation methods were applied to address this issue, the presence of missing data may have influenced model robustness and introduced bias, especially in limited sample scenarios.

Another important limitation concerns the availability of public datasets. There is a lack of openly available databases that include all the modalities captured by the IPMA — such as synchronized audio, vital signs, and structured symptoms — which made it necessary to train models on separate datasets and apply them individually. Similarly, very few datasets contain physiological signals labeled according to the MTS, restricting the development of fully integrated triage models. This remains a limiting factor for training and evaluating models that truly reflect the complexity of real-world triage, especially when multiple modalities and structured protocols are involved.

6.3 Future Work

Building upon the limitations identified and the promising results achieved with the IPMA equipment, future research should prioritize the expansion and diversification of the dataset through multicentric studies involving larger and more heterogeneous populations. This will enhance the statistical power and generalizability of AI models, capturing a wider range of demographic, clinical, and environmental variability. Additionally, it

will be crucial to validate the IPMA's applicability in diverse healthcare infrastructures, ensuring robustness across real-world scenarios. In the development of a new version of the IPMA, the circuit assembly can be redesigned to incorporate dedicated sensors for direct measurement, rather than relying on certified standalone equipment. This approach would not only simplify the architecture but also reduce the overall size of the system, facilitating portability and integration into diverse healthcare environments. The integration of additional physiological signals, such as BP, and the development of a comprehensive, synchronized multimodal database aligned with clinical triage protocols like the MTS will further support the training of more effective and unified models.

From a methodological and technological standpoint, future work may explore advanced algorithms tailored for multimodal data fusion, such as convolutional and recurrent neural networks combined with attention mechanisms, to better capture complex patterns across physiological and audio signals. The implementation and optimization of real-time inference on embedded hardware platforms will be essential to enable practical deployment in clinical environments, including stress testing under varying noise and connectivity conditions. Expanding the IPMA's capabilities to support the detection and triage of other respiratory and systemic conditions, as well as adaptation for use in occupational health settings and different clinical specialties, represent important avenues to extend its impact and utility.

6.4 Publications

The following publications resulted from this doctoral research:

1. SILVA, Leticia et al. COVID-19 respiratory sound analysis and classification using audio textures. Frontiers in signal processing, v. 2, p. 986293, 2022. DOI: <https://doi.org/10.3389/frsip.2022.986293>.

2. SILVA, Leticia et al. Triagem automática em centros de saúde: avaliação de técnicas de aprendizado de máquina baseadas no protocolo de Manchester para um dispositivo multimodal. In: SIMPÓSIO DE ENGENHARIA BIOMÉDICA – XV SEB, 15., 2023, Uberlândia. Anais do Simpósio de Engenharia Biomédica. Uberlândia: XV SEB, 2023. DOI: <https://doi.org/10.5281/zenodo.10140042>.

3. SILVA, Leticia et al. Ensemble Learning Approach to Support Manchester Protocol Triage. *Multimedia Tools and Applications*, [Status: under review].

4. ORMAZA-SIGUENZA, Leonardo et al. Towards a Multimodal Automatic Equipment for Predictive Health Monitoring of Industrial Workers: A Usability Perspective. In: 2024 IEEE International Symposium on Medical Measurements and Applications

(MeMeA). IEEE, 2024. p. 1-6. DOI: <https://doi.org/10.1109/MeMeA60663.2024.10596775>.

5. VILLA-PARRA, Ana Cecilia et al. Towards multimodal equipment to help in the diagnosis of COVID-19 using machine learning algorithms. Sensors, v. 22, n. 12, p. 4341, 2022. DOI: <https://doi.org/10.3390/s22124341>.

6. VALADÃO, Carlos et al. IPMA — An Automated System to Capture Biomedical Signals and Help the Diagnosis of Respiratory Diseases. In: Latin American Conference on Biomedical Engineering. Cham: Springer Nature Switzerland, 2022. p. 409-419. DOI: <https://doi.org/10.1007/978-3-031-49407-9_42>.

Additional publications from research collaborations in other projects:

1. DE SOUZA, Fernanda et al. Análise dos Efeitos da Ativação Cerebral por Eletroencefalografia em Pacientes Pós-AVC Durante Imagética Motora, Estimulação Elétrica Funcional e Luva Robótica. Brazilian Journal of Biological Sciences, v. 12, n. 26, p. e132-e132, 2025. DOI: <https://doi.org/10.21472/bjbs.v12n26-002>.

2. MEHRPOUR, Sheida et al. A review about synergistic effects of transcranial direct current stimulation (tdcs) in combination with motor imagery (mi)-based brain computer interface (bci) on post-stroke rehabilitation. Research on Biomedical Engineering, v. 40, n. 1, p. 43-67, 2024. <https://doi.org/10.1007/s42600-023-00329-0>.

3. LAMPIER, Lucas et al. A deep learning approach for estimating SpO 2 using a smartphone camera. IEEE Transactions on Instrumentation and Measurement, v. 72, p. 1-8, 2023. <https://doi.org/10.1109/TIM.2023.3306832>.

4. DELISLE-RODRIGUEZ, Denis; SILVA, Leticia; BASTOS-FILHO, Teodiano. EEG changes during passive movements improve the motor imagery feature extraction in BCIs-based sensory feedback calibration. Journal of Neural Engineering, v. 20, n. 1, p. 016047, 2023. <https://doi.org/10.1088/1741-2552/acb73b>.

5. CORREIA, Iamara et al. Desenvolvimento de uma plataforma interativa para reabilitação motora fina em membros superiores com jogos sérios. In: SIMPÓSIO DE ENGENHARIA BIOMÉDICA – XV SEB, 15., 2023, Uberlândia. Anais do Simpósio de Engenharia Biomédica. Uberlândia: XV SEB, 2023. <https://doi.org/10.5281/zenodo.10157141>.

6. SILVA, Leticia et al. Analysis of Brain Excitability After Transcranial Direct Current Stimulation and Brain-Computer Interface Based on Motor Imagery on a Post-stroke Patient. In: Latin American Conference on Biomedical Engineering. Cham: Springer

Nature Switzerland, 2022. p. 211-217. <https://doi.org/10.1007/978-3-031-49407-9_22>.

7. LAMPIER, Lucas et al. A deep learning approach to estimate pulse rate by remote photoplethysmography. Physiological Measurement, v. 43, n. 7, p. 075012, 2022. <https://doi.org/10.1088/1361-6579/ac7b0b>.

8. CARDOSO, Vivianne et al. Effect of a brain–computer interface based on pedaling motor imagery on cortical excitability and connectivity. Sensors, v. 21, n. 6, p. 2020, 2021. <https://doi.org/10.3390/s21062020>.

References

Abidin, S.; Togneri, R.; Sohel, F. Spectrotemporal analysis using local binary pattern variants for acoustic scene classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 26, n. 11, p. 2112–2121, 2018.

Acosta, J. N.; Falcone, G. J.; Rajpurkar, P.; Topol, E. J. Multimodal biomedical ai. *Nature medicine*, Nature Publishing Group US New York, v. 28, n. 9, p. 1773–1784, 2022.

Adnan, S. M. et al. Fall detection through acoustic local ternary patterns. *Applied Acoustics*, Elsevier, v. 140, p. 296–300, 2018.

Al-qudah, R.; Aloqaily, M.; Karray, F. Computer vision-based architecture for iomt using deep learning. *2022 International Wireless Communications and Mobile Computing (IWCMC)*, p. 931–936, 2022.

Almulhim, M. et al. Using machine learning technique in managing emergency triage flow. *Acta Informatica Medica*, v. 33, n. 2, p. 152, 2025.

Altman, N.; Krzywinski, M. Ensemble methods: bagging and random forests. *Nature Methods*, Nature Publishing Group, v. 14, n. 10, p. 933–935, 2017.

Alvarado, E. et al. Dyspnea severity assessment based on vocalization behavior with deep learning on the telephone. *Sensors*, MDPI, v. 23, n. 5, p. 2441, 2023.

Aly, M.; Rahouma, K. H.; Ramzy, S. M. Pay attention to the speech: Covid-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings. *Alexandria Engineering Journal*, Elsevier, v. 61, n. 5, p. 3487–3500, 2022.

Alzghaibi, H. Adoption barriers and facilitators of wearable health devices with ai integration: a patient-centred perspective. *Frontiers in Medicine*, Frontiers Media SA, v. 12, p. 1557054, 2025.

Arslan, B.; Nuhoglu, C.; Satici, M.; Altinbilek, E. Evaluating llm-based generative ai tools in emergency triage: A comparative study of chatgpt plus, copilot pro, and triage nurses. *The American Journal of Emergency Medicine*, Elsevier, v. 89, p. 174–181, 2025.

Association, E. N. *Emergency Severity Index Handbook Fifth Edition*. 2023. <https://www.ena.org/>. [Accessed: 2025-07-18].

Austin, P. C.; White, I. R.; Lee, D. S.; Buuren, S. van. Missing data in clinical research: a tutorial on multiple imputation. *Canadian Journal of Cardiology*, Elsevier, v. 37, n. 9, p. 1322–1331, 2021.

Awad, M.; Khanna, R. *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. : Springer nature, 2015.

Aytekin, I. et al. Covid-19 detection from respiratory sounds with hierarchical spectrogram transformers. *IEEE journal of biomedical and health informatics*, IEEE, v. 28, n. 3, p. 1273–1284, 2023.

Azur, M. J.; Stuart, E. A.; Frangakis, C.; Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, Wiley Online Library, v. 20, n. 1, p. 40–49, 2011.

Başer, H. Y.; Evran, T.; Cifci, M. A. Machine learning-augmented triage for sepsis: Real-time icu mortality prediction using shap-explained meta-ensemble models. *Biomedicines*, MDPI, v. 13, n. 6, p. 1449, 2025.

Bassam, N. A. et al. Iot based wearable device to monitor the signs of quarantined remote patients of covid-19. *Informatics in medicine unlocked*, Elsevier, v. 24, p. 100588, 2021.

Benisek, A. *Symptoms of Coronavirus*. 2023. Available on: <https://www.webmd.com/lung/covid-19-symptoms\#1>.

Bennett, D. A. How can i deal with missing data in my study? *Australian and New Zealand journal of public health*, Elsevier, v. 25, n. 5, p. 464–469, 2001.

Bhavani, S. V. et al. Coronavirus disease 2019 temperature trajectories correlate with hyperinflammatory and hypercoagulable subphenotypes. *Critical Care Medicine*, LWW, v. 50, n. 2, p. 212–223, 2022.

Breiman, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

Breiman, L.; Friedman, J.; Olshen, R. A.; Stone, C. J. *Classification and regression trees*. : Chapman and Hall/CRC, 1984.

Brooke, J. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, v. 189, 11 1995.

Brooke, J. SUS : A Retrospective. *Journal of Usability Studies*, v. 8, n. 2, p. 29–40, 2013. ISSN 1931-3357. Available on: <http://www.usabilityprofessionals.org/upa_publications/jus/2013february/brooke1.html%5Cnhttp://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>.

Brown, C. et al. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: Association for Computing Machinery, 2020. (KDD '20), p. 3474–3484. ISBN 9781450379984. Available on: <https://doi.org/10.1145/3394486.3412865>.

Buuren, S. V.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, v. 45, p. 1–67, 2011.

Cai, T.; Ma, J.; Ge-Zhang, S. Smart cities and smart health: Innovations in home medical devices for efficient healthcare delivery. *Results in Engineering*, Elsevier, p. 105739, 2025.

Carpenter, J. R.; Smuk, M. Missing data: A statistical framework for practice. *Biometrical Journal*, Wiley Online Library, v. 63, n. 5, p. 915–947, 2021.

Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, Elsevier, v. 408, p. 189–215, 2020.

Chang, Y.-H. et al. Using machine learning and natural language processing in triage for prediction of clinical disposition in the emergency department. *BMC Emergency Medicine*, Springer, v. 24, n. 1, p. 237, 2024.

Channa, A.; Popescu, N.; Skibinska, J.; Burget, R. The rise of wearable devices during the covid-19 pandemic: A systematic review. *Sensors (Basel, Switzerland)*, v. 21, 2021.

Chen, Q. et al. Cardiovascular manifestations in severe and critical patients with covid-19. *Clinical cardiology*, Wiley Online Library, v. 43, n. 7, p. 796–802, 2020.

Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016. p. 785–794.

Cicolo, E. A.; Peres, H. H. C. Electronic and manual registration of manchester system: reliability, accuracy, and time evaluation. *Revista Latino-Americana de Enfermagem*, SciELO Brasil, v. 27, 2019.

Clark, V. L.; Kruse, J. A. Clinical methods: the history, physical, and laboratory examinations. *Jama*, American Medical Association, v. 264, n. 21, p. 2808–2809, 1990.

Costa, H. G.; Silva, M. H. T. da; Santos, G. N.; Bonamigo, A.; Callado, R. D. Clustering brazilian public emergency healthcare units. *IFAC-PapersOnLine*, Elsevier, v. 55, n. 10, p. 566–571, 2022.

Dang, T. et al. Exploring longitudinal cough, breath, and voice data for covid-19 progression prediction via sequential deep learning: model development and validation. *Journal of medical Internet research*, JMIR Publications Toronto, Canada, v. 24, n. 6, p. e37004, 2022.

Da'Costa, A. et al. Ai-driven triage in emergency departments: A review of benefits, challenges, and future directions. *International Journal of Medical Informatics*, Elsevier, p. 105838, 2025.

Demir, F.; Sengur, A.; Cummins, N.; Amiriparian, S.; Schuller, B. Low level texture features for snore sound discrimination. In: IEEE. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2018. p. 413–416.

Deng, Y.; Chang, C.; Ido, M. S.; Long, Q. Multiple imputation for general missing data patterns in the presence of high-dimensional data. *Scientific reports*, Nature Publishing Group UK London, v. 6, n. 1, p. 21689, 2016.

Despotovic, V.; Ismael, M.; Cornil, M.; Call, R. M.; Fagherazzi, G. Detection of covid-19 from voice, cough and breathing patterns: Dataset and preliminary results. *Computers in Biology and Medicine*, Elsevier, v. 138, p. 104944, 2021.

Doorn, W. P. van et al. Explainable machine learning models for rapid risk stratification in the emergency department: a multicenter study. *The Journal of Applied Laboratory Medicine*, Oxford University Press US, v. 9, n. 2, p. 212–222, 2024.

Doraiswamy, S.; Abraham, A.; Mamtani, R.; Cheema, S. Use of telehealth during the covid-19 pandemic: scoping review. *Journal of medical Internet research*, JMIR Publications Toronto, Canada, v. 22, n. 12, p. e24087, 2020.

Dunn, O. J. Multiple comparisons using rank sums. *Technometrics*, Taylor & Francis, v. 6, n. 3, p. 241–252, 1964.

Elhaj, H.; Achour, N.; Tania, M. H.; Aciksari, K. A comparative study of supervised machine learning approaches to predict patient triage outcomes in hospital emergency departments. *Array*, Elsevier, v. 17, p. 100281, 2023.

Er, M. B. Heart sounds classification using convolutional neural network with 1d-local binary pattern and 1d-local ternary pattern features. *Applied Acoustics*, Elsevier, v. 180, p. 108152, 2021.

Erwander, K.; Agvall, B.; Ivarsson, K. The role of vital signs in predicting mortality risk in elderly patients visiting the emergency department. *BMC Emergency Medicine*, v. 25, n. 1, p. 144, Aug 2025. ISSN 1471-227X. Available on: <https://doi.org/10.1186/s12873-025-01307-8>.

Farzandipour, M.; Nabovati, E.; Sharif, R. The effectiveness of tele-triage during the covid-19 pandemic: A systematic review and narrative synthesis. *Journal of telemedicine and telecare*, SAGE Publications Sage UK: London, England, v. 30, n. 9, p. 1367–1375, 2024.

Fawzy, A. et al. Clinical outcomes associated with overestimation of oxygen saturation by pulse oximetry in patients hospitalized with covid-19. *JAMA network open*, American Medical Association, v. 6, n. 8, p. e2330856–e2330856, 2023.

Fekonja, Z. et al. Factors contributing to patient safety during triage process in the emergency department: A systematic review. *Journal of clinical nursing*, Wiley Online Library, v. 32, n. 17-18, p. 5461–5477, 2023.

Fernandes, M. et al. Predicting intensive care unit admission among patients presenting to the emergency department using machine learning and natural language processing. *PloS one*, Public Library of Science San Francisco, CA USA, v. 15, n. 3, p. e0229331, 2020.

Fernandes, M. et al. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. *Artificial Intelligence in Medicine*, Elsevier, v. 102, p. 101762, 2020.

Filip, R.; Puscaselu, R. G.; Anchidin-Norocel, L.; Dimian, M.; Savage, W. K. Global challenges to public health care systems during the covid-19 pandemic: a review of pandemic measures and problems. *Journal of personalized medicine*, MDPI, v. 12, n. 8, p. 1295, 2022.

Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.

Ftouni, R.; AlJardali, B.; Hamdanieh, M.; Ftouni, L.; Salem, N. Challenges of telemedicine during the covid-19 pandemic: a systematic review. *BMC medical informatics and decision making*, Springer, v. 22, n. 1, p. 207, 2022.

Furman, G. et al. Remote analysis of respiratory sounds in patients with covid-19: Development of fast fourier transform–based computer-assisted diagnostic methods. *JMIR Formative Research*, JMIR Publications Toronto, Canada, v. 6, n. 7, p. e31200, 2022.

Gabriel, J. et al. A robotic platform for assistance in the medical triage process. *JOURNAL OF BIOENGINEERING, TECHNOLOGIES AND HEALTH*, 2023.

Gomes, C. Report of the who-china joint mission on coronavirus disease 2019 (covid-19). *Brazilian Journal of Implantology and health sciences*, v. 2, n. 3, 2020.

Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*. : MIT Press, 2016. <http://www.deeplearningbook.org>.

Gotanda, H. et al. Changes in blood pressure outcomes among hypertensive individuals during the covid-19 pandemic: a time series analysis in three us healthcare organizations. *Hypertension*, Lippincott Williams & Wilkins Hagerstown, MD, v. 79, n. 12, p. 2733–2742, 2022.

Guan, W.-j. et al. Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine*, Mass Medical Soc, v. 382, n. 18, p. 1708–1720, 2020.

Gupta, B.; Rawat, A.; Jain, A.; Arora, A.; Dhami, N. Analysis of various decision tree algorithms for classification in data mining. *International Journal of Computer Applications*, v. 163, n. 8, p. 15–19, 2017.

Gupta, P.; Wen, H.; Francesco, L. D.; Ayazi, F. Detection of pathological mechano-acoustic signatures using precision accelerometer contact microphones in patients with pulmonary disorders. *Scientific Reports*, Nature Publishing Group UK London, v. 11, n. 1, p. 13427, 2021.

Gupta-Wright, A. et al. False-negative rt-pcr for covid-19 and a diagnostic risk score: a retrospective cohort study among patients admitted to hospital. *BMJ open*, British Medical Journal Publishing Group, v. 11, n. 2, p. e047110, 2021.

Haldane, V. et al. Health systems resilience in managing the covid-19 pandemic: lessons from 28 countries. *Nature medicine*, Nature Publishing Group US New York, v. 27, n. 6, p. 964–980, 2021.

Haykin, S. *Neural networks and learning machines, 3/E*. : Pearson Education India, 2009.

Hegde, H. et al. Mice vs ppca: Missing data imputation in healthcare. *Informatics in medicine unlocked*, Elsevier, v. 17, p. 100275, 2019.

Hickman, L.; Langer, M.; Saef, R. M.; Tay, L. Automated speech recognition bias in personnel selection: The case of automatically scored job interviews. *Journal of Applied Psychology*, American Psychological Association, 2024.

Himawan, I.; Towsey, M.; Roe, P. 3d convolution recurrent neural networks for bird sound detection. In: Detection and Classification of Acoustic Scenes and Events. *Proceedings of the 3rd Workshop on Detection and Classification of Acoustic Scenes and Events*. 2018. p. 1–4.

Hong, W. S.; Haimovich, A. D.; Taylor, R. A. Predicting hospital admission at emergency department triage using machine learning. *PloS one*, Public Library of Science San Francisco, CA USA, v. 13, n. 7, p. e0201016, 2018.

Hsieh, M.-J. et al. Developing and validating a model for predicting 7-day mortality of patients admitted from the emergency department: an initial alarm score by a prospective prediction model study. *BMJ open*, British Medical Journal Publishing Group, v. 11, n. 1, p. e040837, 2021.

Hu, B.; Guo, H.; Zhou, P.; Shi, Z.-L. Characteristics of sars-cov-2 and covid-19. *Nature reviews microbiology*, Nature Publishing Group UK London, v. 19, n. 3, p. 141–154, 2021.

Husain, M. et al. Artificial intelligence for detecting covid-19 with the aid of human cough, breathing and speech signals: Scoping review. *IEEE Open Journal of Engineering in Medicine and Biology*, IEEE, v. 3, p. 235–241, 2022.

Hussain, S. A. *Assessing the infection status of COVID-19 patients using a wearable prototype*. Zenodo, 2021. Available on: <https://doi.org/10.5281/zenodo.4766192>.

Hwang, S.; Lee, B. Machine learning-based prediction of critical illness in children visiting the emergency department. *Plos one*, Public Library of Science San Francisco, CA USA, v. 17, n. 2, p. e0264184, 2022.

Ikram, A. S.; Pillay, S. Admission vital signs as predictors of covid-19 mortality: a retrospective cross-sectional study. *BMC Emergency Medicine*, Springer, v. 22, n. 1, p. 68, 2022.

Ingielewicz, A.; Rychlik, P.; Sieminski, M. Drinking from the holy grail—does a perfect triage system exist? and where to look for it? *Journal of Personalized Medicine*, MDPI, v. 14, n. 6, p. 590, 2024.

Jachmann, A. et al. Burnout, depression, and stress in emergency department nurses and physicians and the impact on private and work life: A systematic review. *JACEP Open*, Elsevier, v. 6, n. 2, p. 100046, 2025.

Jatobá, A. et al. Supporting decision-making in patient risk assessment using a hierarchical fuzzy model. *Cognition, Technology & Work*, Springer, v. 20, p. 477–488, 2018.

Jesus, A. P. S. d.; Okuno, M. F. P.; Campanharo, C. R. V.; Lopes, M. C. B. T.; Batista, R. E. A. Sistema de triagem de manchester: avaliação em um serviço hospitalar de emergência. *Revista Brasileira de Enfermagem*, SciELO Brasil, v. 74, p. e20201361, 2021.

Johns Hopkins University. *COVID-19 Dashboard*. 2022. [Accessed July 1, 2022].

Joseph, J. W. et al. Deep-learning approaches to identify critically ill patients at emergency department triage using limited information. *Journal of the American College of Emergency Physicians Open*, Wiley Online Library, v. 1, n. 5, p. 773–781, 2020.

Junaid, K.; Kiran, T.; Gupta, M.; Kishore, K.; Siwatch, S. How much missing data is too much to impute for longitudinal health indicators? a preliminary guideline for the choice of the extent of missing proportion to impute with multiple imputation by chained equations. *Population health metrics*, Springer, v. 23, n. 1, p. 2, 2025.

Karjala, J.; Eriksson, S. Inter-rater reliability between nurses for a new paediatric triage system based primarily on vital parameters: the paediatric triage instrument (peti). *BMJ open*, British Medical Journal Publishing Group, v. 7, n. 2, p. e012748, 2017.

Kim, H.-J. et al. Sepsis alert systems, mortality, and adherence in emergency departments: a systematic review and meta-analysis. *JAMA network open*, American Medical Association, v. 7, n. 7, p. e2422823–e2422823, 2024.

Kim, H.-Y. Statistical notes for clinical researchers: assessing normal distribution (1). *Restorative dentistry & endodontics*, The Korean Academy of Conservative Dentistry, v. 37, n. 4, p. 245–248, 2012.

Kim, H.-Y. Statistical notes for clinical researchers: Nonparametric statistical methods: 1. nonparametric methods for comparing two groups. *Restorative dentistry & endodontics*, v. 39, n. 3, p. 235, 2014.

Kim, H.-Y. Statistical notes for clinical researchers: Nonparametric statistical methods: 2. nonparametric methods for comparing three or more groups and repeated measures. *Restorative Dentistry & Endodontics*, v. 39, n. 4, p. 329, 2014.

Kim, H.-Y. Statistical notes for clinical researchers: post-hoc multiple comparisons. *Restorative dentistry & endodontics*, The Korean Academy of Conservative Dentistry, v. 40, n. 2, p. 172–176, 2015.

Kim, H.-Y. Statistical notes for clinical researchers: Sample size calculation 3. comparison of several means using one-way anova. *Restorative dentistry & endodontics*, v. 41, n. 3, p. 231, 2016.

Kim, H.-Y. Statistical notes for clinical researchers: Chi-squared test and fisher's exact test. *Restorative dentistry & endodontics*, v. 42, n. 2, p. 152, 2017.

Kim, H.-Y. Statistical notes for clinical researchers: the independent samples t-test. *Restorative Dentistry & Endodontics*, v. 44, n. 3, p. e26, 2019.

Kobeissi, M. M.; Ruppert, S. D. Remote patient triage: shifting toward safer telehealth practice. *Journal of the American Association of Nurse Practitioners*, LWW, v. 34, n. 3, p. 444–451, 2022.

Kong, F.; Zou, Y.; Li, Z.; Deng, Y. Advances in portable and wearable acoustic sensing devices for human health monitoring. *Sensors*, MDPI, v. 24, n. 16, p. 5354, 2024.

Kumar, L. K.; Alphonse, P. Covid-19 disease diagnosis with light-weight cnn using modified mfcc and enhanced gfcc from human respiratory sounds. *The European Physical Journal Special Topics*, Springer, v. 231, n. 18, p. 3329–3346, 2022.

Laguarta, J.; Hueto, F.; Subirana, B. Covid-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, IEEE, v. 1, p. 275–281, 2020.

Landry, V. et al. Audio-based digital biomarkers in diagnosing and managing respiratory diseases: a systematic review and bibliometric analysis. *European Respiratory Review*, European Respiratory Society, v. 34, n. 176, 2025.

Learning, M. Tom mitchell. *Publisher: McGraw Hill*, p. 31, 1997.

Levin, S. 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*, Elsevier, v. 71, n. 5, p. 565–574, 2018.

Lewis, J. R. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, v. 7, n. 1, p. 57–78, jan 1995. ISSN 1044-7318. Available on: <http://www.tandfonline.com/doi/abs/10.1080/10447319509526110>.

Lewis, J. R. The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, Taylor Francis, v. 34, n. 7, p. 577–590, 2018. Available on: <https://doi.org/10.1080/10447318.2018.1455307>.

Liu, Z.; Shu, W.; Liu, H.; Zhang, X.; Chong, W. Development and validation of interpretable machine learning models for triage patients admitted to the intensive care unit. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 20, n. 2, p. e0317819, 2025.

Lou, Z.; Ren, Y. Investigating issues with machine learning for accent classification. In: IOP Publishing. *Journal of Physics: Conference Series*. 2021. v. 1738, n. 1, p. 012111.

Lu, T.-C. et al. Machine learning to predict in-hospital cardiac arrest from patients presenting to the emergency department. *Internal and Emergency Medicine*, Springer, v. 18, n. 2, p. 595–605, 2023.

Lyell, D.; Coiera, E.; Chen, J. A.; Shah, P.; Magrabi, F. How machine learning is embedded to support clinician decision making: an analysis of fda-approved medical devices. *BMJ Health Care Informatics*, v. 28, 2021.

Mackway-Jones, K.; Marsden, J.; Windle, J. *Emergency triage: Manchester triage group.* : John Wiley & Sons, 2013.

Mackway-Jones, K.; Marsden, J.; Windle, J. *Emergency triage: Manchester triage group.* London: John Wiley & Sons, 2014.

Majumdar, S. R.; Eurich, D. T.; Gamble, J.-M.; Senthilselvan, A.; Marrie, T. J. Oxygen saturations less than 92% are associated with major adverse events in outpatients with pneumonia: a population-based cohort study. *Clinical infectious diseases*, The University of Chicago Press, v. 52, n. 3, p. 325–331, 2011.

Masanneck, L. et al. Triage performance across large language models, chatgpt, and untrained doctors in emergency medicine: comparative study. *Journal of medical Internet research*, JMIR Publications Toronto, Canada, v. 26, p. e53297, 2024.

Meral, G.; Ateş, S.; Günay, S.; Öztürk, A.; Kuşdoğan, M. Comparative analysis of chatgpt, gemini and emergency medicine specialist in esi triage assessment. *The American journal of emergency medicine*, v. 81, p. 146–150, 2024.

Minhas, N. et al. Cost-analysis of real time rt-pcr test performed for covid-19 diagnosis at india's national reference laboratory during the early stages of pandemic mitigation. *Plos one*, Public Library of Science San Francisco, CA USA, v. 18, n. 1, p. e0277867, 2023.

Murgia, V. et al. Upper respiratory tract infection-associated acute cough and the urge to cough: New insights for clinical practice. *Pediatric allergy, immunology, and pulmonology*, Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New . . . , v. 33, n. 1, p. 3–11, 2020.

Nanni, L.; Maguolo, G.; Brahnam, S.; Paci, M. An ensemble of convolutional neural networks for audio classification. *Applied Sciences*, MDPI, v. 11, n. 13, p. 5796, 2021.

Nsubuga, M.; Kintu, T. M.; Please, H.; Stewart, K.; Navarro, S. M. Enhancing trauma triage in low-resource settings using machine learning: a performance comparison with the kampala trauma score. *BMC Emergency Medicine*, Springer, v. 25, n. 1, p. 14, 2025.

Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 24, n. 7, p. 971–987, 2002.

Oliveira, B. A.; al et. Sars-cov-2 and the covid-19 disease: a mini review on diagnostic methods. *Rev. Inst. Med. Trop. S. Paulo*, Epub, v. 62, 2020.

Ormaza-Siguenza, L.; Silva, L.; Villa-Parra, A. C.; Bastos-Filho, T. Towards a multimodal automatic equipment for predictive health monitoring of industrial workers: A usability perspective. In: IEEE. *2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2024. p. 1–6.

Ozturk, T. et al. Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, v. 121, p. 103792 – 103792, 2020.

Pahar, M.; Klopper, M.; Warren, R.; Niesler, T. Covid-19 detection in cough, breath and speech using deep transfer learning and bottleneck features. *Computers in biology and medicine*, Elsevier, v. 141, p. 105153, 2022.

Pedersen, A. B. et al. Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, Taylor & Francis, p. 157–166, 2017.

Po, H.-W. et al. Efficacy of remote health monitoring in reducing hospital readmissions among high-risk postdischarge patients: prospective cohort study. *JMIR Formative Research*, JMIR Publications Toronto, Canada, v. 8, p. e53455, 2024.

Portnoy, J.; Waller, M.; Elliott, T. Telemedicine in the era of covid-19. *The Journal of Allergy and Clinical Immunology: In Practice*, Elsevier, v. 8, n. 5, p. 1489–1491, 2020.

Porto, B. M. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. *BMC Emergency Medicine*, Springer, v. 24, n. 1, p. 219, 2024.

Pramono, R. X. A.; Imtiaz, S. A.; Rodriguez-Villegas, E. A cough-based algorithm for automatic diagnosis of pertussis. *PloS one*, Public Library of Science San Francisco, CA USA, v. 11, n. 9, p. e0162128, 2016.

Rabiner, L.; Schafer, R. *Theory and applications of digital speech processing.* : Prentice Hall Press, 2010.

Resende, C. Brandao-de et al. A machine learning system to optimise triage in an adult ophthalmic emergency department: a model development and validation study. *EClinicalMedicine*, Elsevier, v. 66, 2023.

Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986.

Sapra, A.; Malik, A.; Bhandari, P. Vital sign assessment. In: *StatPearls [internet].* : StatPearls Publishing, 2023.

Sax, D. R. et al. Evaluation of version 4 of the emergency severity index in us emergency departments for the rate of mistriage. *JAMA Network Open*, American Medical Association, v. 6, n. 3, p. e233404–e233404, 2023.

Schuller, B. W. et al. The interspeech 2021 computational paralinguistics challenge: Covid-19 cough, covid-19 speech, escalation & primates. *arXiv preprint arXiv:2102.13468*, 2021.

Seiger, N. et al. Improving the manchester triage system for pediatric emergency care: an international multicenter study. *PLoS One*, Public Library of Science San Francisco, USA, v. 9, n. 1, p. e83267, 2014.

Sengupta, N.; Sahidullah, M.; Saha, G. Lung sound classification using local binary pattern. *arXiv preprint arXiv:1710.01703*, 2017.

Shajari, S.; Kuruvinashetti, K.; Komeili, A.; Sundararaj, U. The emergence of ai-based wearable sensors for digital health technology: a review. *Sensors*, MDPI, v. 23, n. 23, p. 9498, 2023.

Sharma, G.; Prasad, D.; Umapathy, K.; Krishnan, S. Screening and analysis of specific language impairment in young children by analyzing the textures of speech signal. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Institute of Electrical and Electronics Engineers Inc., v. 2020-July, p. 964–967, 7 2020. ISSN 1557170X.

Sharma, G.; Umapathy, K.; Krishnan, S. Audio texture analysis of covid-19 cough, breath, and speech sounds. *Biomedical Signal Processing and Control*, v. 76, p. 103703, 7 2022. ISSN 17468094. Available on: <https://linkinghub.elsevier.com/retrieve/pii/S1746809422002257>.

Sharma, G.; Zhang, X. P.; Umapathy, K.; Krishnan, S. Audio texture and age-wise analysis of disordered speech in children having specific language impairment. *Biomedical Signal Processing and Control*, Elsevier, v. 66, p. 102471, 4 2021. ISSN 1746-8094.

Sharma, Y. P.; Agstam, S.; Yadav, A.; Gupta, A.; Gupta, A. Cardiovascular manifestations of covid-19: An evidence-based narrative review. *Indian journal of medical research*, Medknow, v. 153, n. 1-2, p. 7–16, 2021.

Shati, A.; Hassan, G. M.; Datta, A. Covid-19 detection system: a comparative analysis of system performance based on acoustic features of cough audio signals. In: IEEE. *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. 2023. p. 2706–2713.

Shen, Y. et al. Novel phenotypes of coronavirus disease: a temperature-based trajectory model. *Annals of Intensive Care*, Springer, v. 11, n. 1, p. 121, 2021.

Simonetti, J.; Lombardi, F.; Franciosa, C.; Viscuso, M.; Richeldi, L. Feasibility and safety of oxygen saturation remote monitoring in covid-19: A descriptive research. *Clinical Infection in Practice*, Elsevier, v. 20, p. 100240, 2023.

Smith, A. C. et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (covid-19). *Journal of telemedicine and telecare*, Sage Publications Sage UK: London, England, v. 26, n. 5, p. 309–313, 2020.

Spooner, A. et al. Benchmarking ensemble machine learning algorithms for multi-class, multi-omics data integration in clinical outcome prediction. *Briefings in Bioinformatics*, Oxford Academic, v. 26, n. 2, 2025.

Stasak, B.; Huang, Z.; Razavi, S.; Joachim, D.; Epps, J. Automatic detection of covid-19 based on short-duration acoustic smartphone speech analysis. *Journal of Healthcare Informatics Research*, v. 5, p. 201 – 217, 2021.

Swenson, K. E.; Hardin, C. C. Pathophysiology of hypoxemia in covid-19 lung disease. *Clinics in Chest Medicine*, v. 44, n. 2, p. 239, 2022.

Sönmez, Y. ; Varol, A. A speech emotion recognition model based on multi-level local binary and local ternary patterns. *IEEE Access*, v. 8, p. 190784–190796, 2020.

Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, v. 19, n. 6, p. 1635–1650, 2010.

Tan, X.; Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing*, IEEE, v. 19, n. 6, p. 1635–1650, 2010.

Tao, Q. et al. Clinical applications of smart wearable sensors. *Iscience*, Elsevier, v. 26, n. 9, 2023.

Tena, A.; Claria, F.; Solsona, F. Automated detection of covid-19 cough. *Biomedical Signal Processing and Control*, Elsevier, v. 71, p. 103175, 2022.

Tena, A.; Clarià, F.; Solsona, F. Automated detection of covid-19 cough. *Biomedical Signal Processing and Control*, v. 71, p. 103175 – 103175, 2021.

Teymouri, M. et al. Recent advances and challenges of rt-pcr tests for the diagnosis of covid-19. *Pathology-Research and Practice*, Elsevier, v. 221, p. 153443, 2021.

Tobin, M. J.; Laghi, F.; Jubran, A. Why covid-19 silent hypoxemia is baffling to physicians. *American journal of respiratory and critical care medicine*, American Thoracic Society, v. 202, n. 3, p. 356–360, 2020.

Topol, E. J. Is my cough covid-19? *The Lancet*, Elsevier, v. 396, n. 10266, p. 1874, 2020.

Torrente-Rodríguez, R. M. et al. Sars-cov-2 rapidplex: A graphene-based multiplexed telemedicine platform for rapid and low-cost covid-19 diagnosis and monitoring. *Matter*, v. 3, p. 1981 – 1998, 2020.

Townsend, B.; Plant, K.; Hodge, V. J.; Ashaolu, O.; Calinescu, R. Medical practitioner perspectives on ai in emergency triage. *Frontiers in Digital Health*, v. 5, 2023.

Triagenet.net. *Manchester Triage System*. 2022. <https://www.triagenet.net/classroom>. [Accessed: 2025-07-18].

Turan, C.; Lam, K.-M. Histogram-based local descriptors for facial expression recognition (fer): A comprehensive study. *Journal of visual communication and image representation*, Elsevier, v. 55, p. 331–341, 2018.

Uddin, S.; Haque, I.; Lu, H.; Moni, M. A.; Gide, E. Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction. *Scientific Reports*, Nature Publishing Group UK London, v. 12, n. 1, p. 6256, 2022.

University of Washington. *Environmental Health and Safety B . Ultraviolet Safety Sheet.* Seattle: University of Washington, 2017. 5 p.

Verde, L. et al. Exploring the use of artificial intelligence techniques to detect the presence of coronavirus covid-19 through speech and voice analysis. *IEEE Access*, Institute of Electrical and Electronics Engineers Inc., v. 9, p. 65750–65757, 2021. ISSN 21693536.

Villa-Parra, A. C. et al. Towards multimodal equipment to help in the diagnosis of covid-19 using machine learning algorithms. *Sensors*, MDPI, v. 22, n. 12, p. 4341, 2022.

Vitória, P. de. *Saúde de Vitória realiza quase 10 milhões de serviços em 2024*. 2025. Available on: <https://vitoria.es.gov.br/noticia/saude-de-vitoria-realiza-quase-10-milhoes-de-servicos-em-2024-52387>.

Waage, A. K. E.; Iwarsson, J. The effect of speaking rate on voice and breathing behavior. *Journal of Voice*, Elsevier, 2024.

Wallace, W. et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. *NPJ digital medicine*, Nature Publishing Group UK London, v. 5, n. 1, p. 118, 2022.

Wegen, M. E. van; Fransen, L. F.; Thijssen, W. A.; Alexandridis, G.; Groot, B. de. The association between urgency level and hospital admission, mortality and resource utilization in three emergency department triage systems: an observational multicenter study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, Springer, v. 33, n. 1, p. 72, 2025.

Wiersinga, W. J.; Rhodes, A.; Cheng, A. C.; Peacock, S. J.; Prescott, H. C. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (covid-19): a review. *Jama*, American Medical Association, v. 324, n. 8, p. 782–793, 2020.

Williams, B. A.; Jones, C. H.; Welch, V.; True, J. M. Outlook of pandemic preparedness in a post-covid-19 world. *npj Vaccines*, Nature Publishing Group UK London, v. 8, n. 1, p. 178, 2023.

World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. 2020. [Accessed August 01, 2025].

World Health Organization. *Coronavirus disease (COVID-19): Masks*. 2023. Accessed on May 19, 2024. Available on: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-masks>.

Worldometer. *Brazil COVID - Coronavirus Statistics*. 2024. [Accessed April 17, 2025].

Yazici, M. et al. Predictability of adult patient medical emergency condition from triage vital signs and comorbidities: a single-center, observational study. *BMC Emergency Medicine*, Springer, v. 24, n. 1, p. 185, 2024.

Yu, J.-H.; Weng, Y.-M.; Chen, K.-F.; Chen, S.-Y.; Lin, C.-C. Triage vital signs predict in-hospital mortality among emergency department patients with acute poisoning: a case control study. *BMC health services research*, Springer, v. 12, n. 1, p. 262, 2012.

Zaboli, A. et al. Comparing the national early warning score and the manchester triage system in emergency department triage: A multi-outcome performance evaluation. *Diagnostics*, MDPI, v. 15, n. 9, p. 1055, 2025.

Zachariasse, J. M. et al. Improving the prioritization of children at the emergency department: Updating the manchester triage system using vital signs. *PloS one*, Public Library of Science San Francisco, CA USA, v. 16, n. 2, p. e0246324, 2021.

Zhang, Z. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, v. 4, n. 11, p. 218, 2016.

Zhou, B.; Perel, P.; Mensah, G. A.; Ezzati, M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, Nature Publishing Group UK London, v. 18, n. 11, p. 785–802, 2021.

Zhou, Q. et al. Cough recognition based on mel-spectrogram and convolutional neural network. *Frontiers in Robotics and AI*, Frontiers Media S.A., v. 8, p. 112, 5 2021. ISSN 22969144.

Zhou, Z.-H. *Ensemble methods: foundations and algorithms*. : CRC press, 2025.

APPENDIX A – System Usability Scale

The System Usability Scale (SUS) questions are shown below (Brooke, 1995):

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

APPENDIX B – Post Study System Usability Questionnaire

The Post Study System Usability Questionnaire (PSSUQ) used in this article has all questions as presented in the questionnaire made by the original creator (Lewis, 1995):

1. Overall, I am satisfied with how easy it is to use this system.
2. It was simple to use this system.
3. I was able to complete the tasks and scenarios quickly using this system.
4. I felt comfortable using this system.
5. It was easy to learn to use this system.
6. I believe I could become productive quickly using this system.
7. The system gave error messages that clearly told me how to fix problems.
8. Whenever I made a mistake using the system, I could recover easily and quickly.
9. The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.
10. It was easy to find the information I needed.
11. The information was effective in helping me complete the tasks and scenarios.
12. The organization of information on the system screens was clear.
13. The interface of this system was pleasant.
14. I liked using the interface of this system.
15. This system has all the functions and capabilities I expect it to have.
16. Overall, I am satisfied with this system.

The questions are divided into three categories, described below:

- SYSUSE (System Utility): average score of questions 1 to 6
- INFOQUAL (Information Quality): average score of questions 7 to 12
- INTERQUAL (Interface Quality): average score of questions 13 to 16

APPENDIX C – Authorization Letter from SEMUS/PMV (COVID-19 Study)

Original version (Portuguese):

PREFEITURA DE VITÓRIA
Secretaria de Saúde

CARTA DE APRESENTAÇÃO

Origem	Destino	Data	Emitida por
PMV/SEMUS/ETSUS	PMV/SEMUS/ Todas UBS's	23/11/2022	Regina

Resumo do Assunto

ENCAMINHAMENTO DE PESQUISADOR

Sr(a). Diretor(a),

O projeto de pesquisa intitulado "Assistênte Médico Portátil Integrado para Auxílio ao Diagnóstico da COVID-19 e Outras Síndromes Respiratórias Através de Sinais Biomédicos e Técnicas de Inteligência Artificial", de autoria do pesquisador Teodiano Freire Bastos Filho, da Universidade Federal do Espírito Santo (UFES) foi aprovado pela Comissão Técnica de Pesquisa da PMV/SEMUS, instituída pela Portaria n.º 038/2021.

Esclarecemos que o presente tem como objetivo é utilizar o Assistente Médico Pessoal Integrado(AMPI) desenvolvido na UFES para fazer a coleta de dados biomédicos de voluntários em Postos de Saúde de Vitória - ES, e inferir possível contaminação pelo vírus SARS-CoV-2.

Ressaltamos que o pesquisador foi orientado que a liberação da pesquisa está condicionada à devolução dos resultados em forma digital e/ou apresentação oral para a Secretaria Municipal de Saúde (PMV/SEMUS) e que a não devolutiva dos resultados em até dois meses após o término desta referida pesquisa, implicará no indeferimento de novas solicitações do(s) pesquisador(es).

Solicitamos que a pesquisa seja viabilizada por este setor e informamos que esta autorização para realização da pesquisa tem validade por 1 ano.

Ressaltamos que cabe ao pesquisador o convite aos participantes, após acordo com o Diretor do Serviço.

Atenciosamente,

Josenan de Alcântara Almeida Costa
Diretora da Escola Técnica e Formação Profissional de Saúde

Translated version (English translation by the author):

To
The Directors of All Basic Health Unit (UBS)

Subject: Researcher Referral

Dear Director,

The research project entitled "Integrated Portable Medical Assistant for Supporting the Diagnosis of COVID-19 and Other Respiratory Syndromes Through Biomedical Signals and Artificial Intelligence Techniques", authored by researcher Teodiano Freire Bastos Filho from the Federal University of Espírito Santo (UFES), has been approved by the Technical Research Committee of PMV/SEMUS, established by Ordinance No. 038/2021.

The purpose of this project is to use the Integrated Personal Medical Assistant (AMPI) developed at UFES to collect biomedical data from volunteers at Health Centers in Vitória – ES and to infer possible contamination by the SARS-CoV-2 virus.

The researcher has been informed that authorization for the study is conditional upon returning the results in digital form and/or through an oral presentation to the Municipal Health Department (PMV/SEMUS). Failure to return the results within two months after the end of the study will result in the denial of future research requests by the researcher(s).

We request that this sector enable the execution of the research and inform that this authorization is valid for one (1) year.

We also emphasize that it is the researcher's responsibility to invite participants, after agreement with the Service Director.

Sincerely,

Josenan de Alcântara Almeida Costa
Director of the Technical School and Professional Health Training

APPENDIX D – Authorization Letter from SEMUS/PMV (MTS-Based Study)

Original version (Portuguese):

PREFEITURA DE VITÓRIA
www.vitoria.es.gov.br

SECRETARIA MUNICIPAL DE SAÚDE

rede bem estar

ETSUS ESCOLA TÉCNICA DE SAÚDE VITÓRIA

CARTA DE ANUÊNCIA

PREFEITURA DE VITÓRIA

Estado do Espírito Santo

DECLARAÇÃO DE ANUÊNCIA

Declaro, para fins de apresentação no Comitê de Ética, que a Secretaria Municipal de Saúde (PMV/SEMUS) está de acordo e possui infraestrutura adequada para a realização do projeto de pesquisa intitulado "USO DO ASSISTENTE MÉDICO PORTÁTIL INTEGRADO (AMPI) PARA MELHORA DA TRIAGEM EM CENTROS DE SAÚDE DE VITÓRIA-ES", da UFES, de autoria do(a) pesquisador(a) TEODIANO FREIRE BASTOS FILHO, que foi submetido à Comissão Técnica de Pesquisa da PMV/SEMUS, instituída pela Portaria n.º 038/2021.

Vitória, 21 de Dezembro de 2023.

Nº documento: 6391513/2023

Documento gerado eletronicamente através do sistema - REDE BEM ESTAR - em 21/12/2023 15:07:38
Em nome de JOSENAN DE ALCANTARA ALMEIDA COSTA - Diretor da Escola Técnica e Formação Profissional de Saúde - ETSUS

A integridade e autenticidade desse documento pode ser verificada através do link:
https://saude.vitoria.es.gov.br/Pesquisa/6391513/2023/43ced03c

Chave: 43ced03c-7458-4e94-b858-ed209850b75c
Data de Impressão: 21/12/2023 15:12:42 Página: 1/1

Translated version (English translation by the author):

LETTER OF AUTHORIZATION

I hereby declare, for submission to the Ethics Committee, that the Municipal Health Department (PMV/SEMUS) agrees with and has the appropriate infrastructure for conducting the research project entitled "Use of the Integrated Portable Medical Assistant (IPMA) to Improve Triage in Health Centers in Vitória-ES", from UFES, authored by researcher Teodiano Freire Bastos Filho, which was submitted to the Technical Research Committee of PMV/SEMUS, established by Ordinance No. 038/2021.

Document No.: 6391513/2023