

Nº 179616

Inteligência artificial na saúde

Adriana Camargo Brito

*Palestra apresentada no Start
Bootcamp, 2025. .*

A série “Comunicação Técnica” compreende trabalhos elaborados por técnicos do IPT, apresentados em eventos, publicados em revistas especializadas ou quando seu conteúdo apresentar relevância pública.

PROIBIDO A REPRODUÇÃO, APENAS PARA CONSULTA.



Inteligência Artificial na saúde

Felipe Pacheco de Almeida Euphrásio
Adriana Camargo de Brito (colaboração)





Sobre mim

- Graduado em Engenharia de Controle e Automação
- Mestre em Ciências e Tecnologias Espaciais
- Doutorando em Engenharia Aeronáutica e Mecânica
- Pesquisador bolsista do IPT na Seção de Inteligência Artificial
- Mentor de IA de startups do Programa Start IA do Sebrae for Startups



Realização:



Execução:





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN



Realização:



Execução:





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN



Realização:



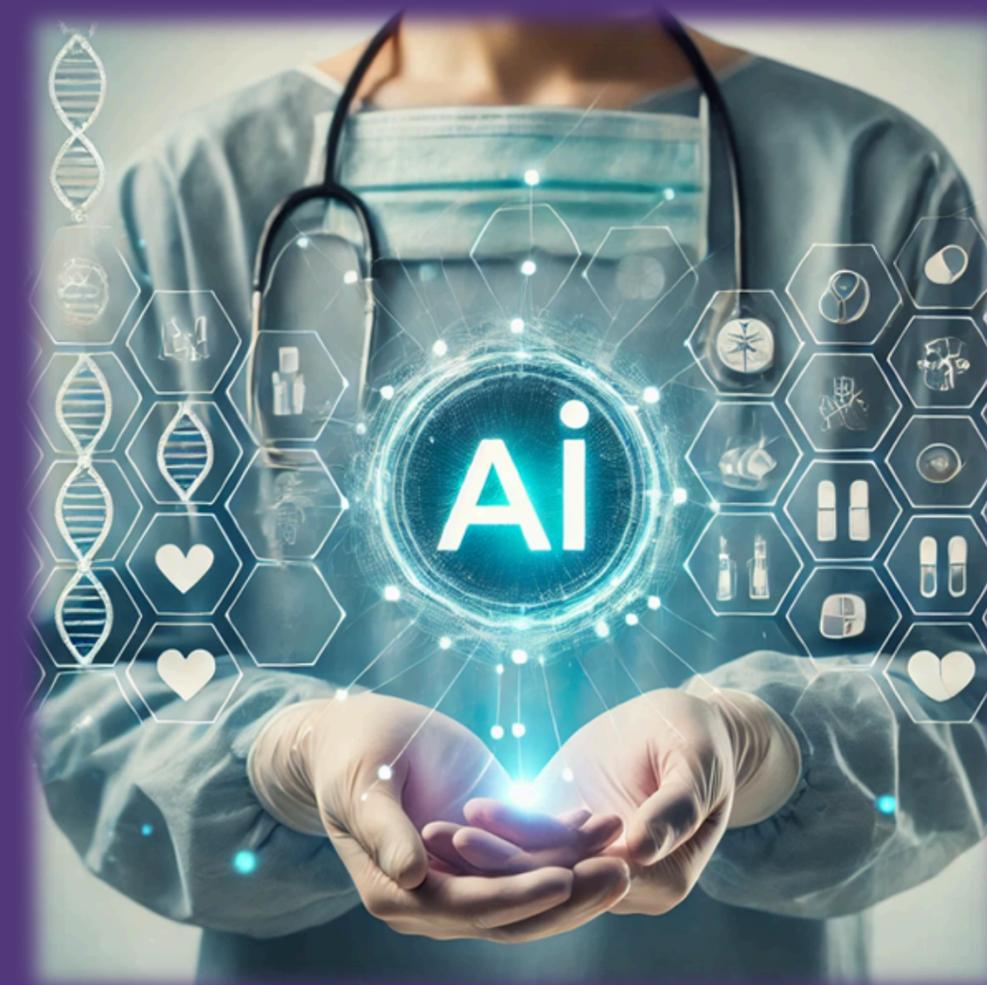
Execução:





IA na saúde

- Gerenciamento de registros e dados médicos
- Suporte ao diagnóstico e tratamento
- Sistemas de apoio à decisão clínica.
- Monitoramento de pacientes.
- Reconhecimento de imagem para diagnósticos.
- Escrita automatizada de prescrições.
- Cirurgia assistida por IA.



Realização:



Execução:





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN

Realização:



Execução:





Impactos de IA na saúde

- Diagnóstico Precoce
- Eficiência Operacional
- Monitoramento Contínuo e Cuidados Remotos
- Descoberta de Medicamentos
- Predição de Riscos
- Suporte à Decisão Clínica
- Terapia Digital
- Redução de Erros
- Gestão e Planejamento Hospitalar
- Integração de Dados e Interoperabilidade
- Assistência a Pacientes



Realização:



Execução:





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN



Realização:



Execução:





Desafios de IA na saúde



Limitações de algoritmo



Privacidade dos dados



Transparência e explicabilidade



Acessibilidade e equidade



Custo de implementação



Quantidade de dados

Realização:



Execução:





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN

Realização:



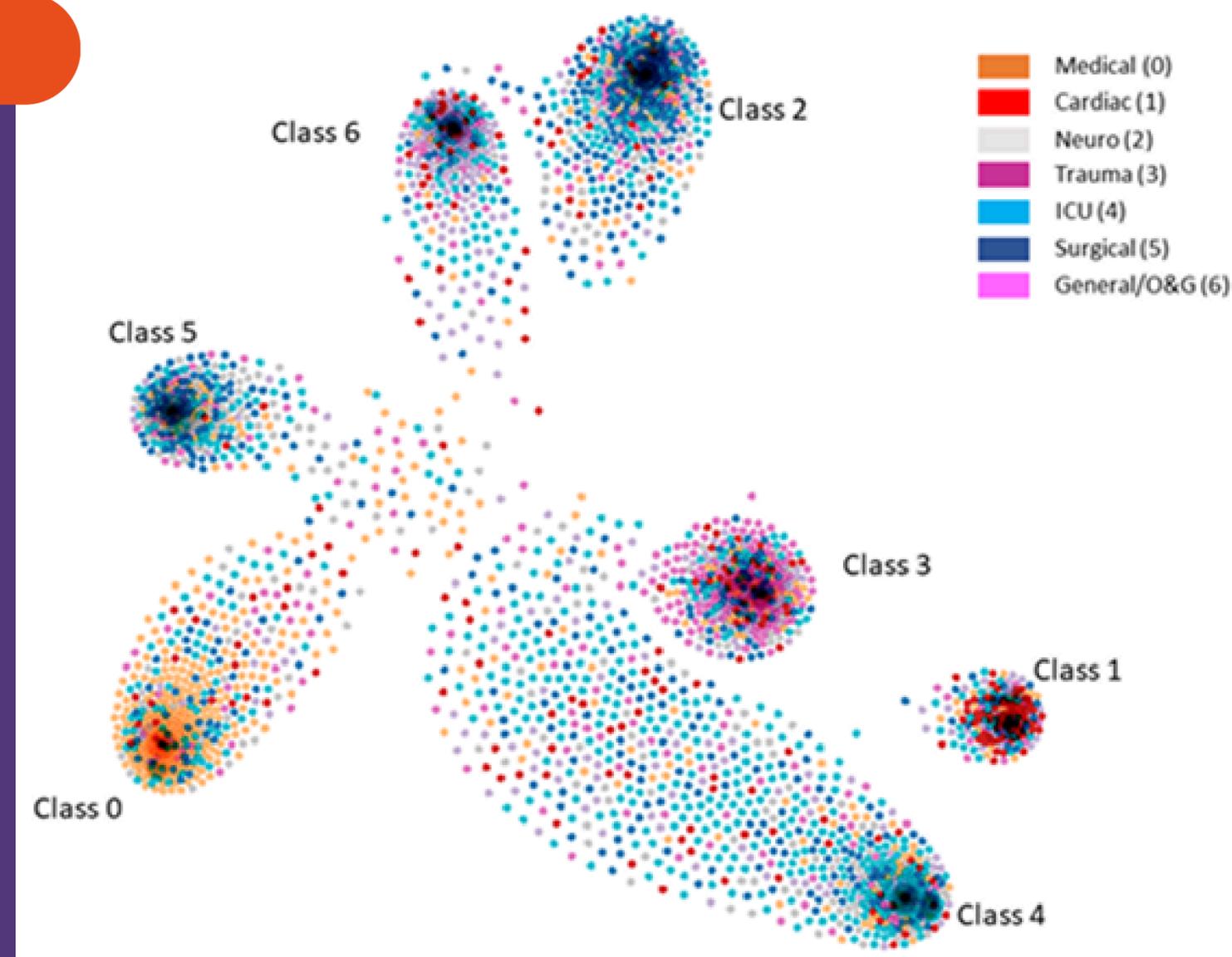
Execução:





Estudos de caso: artigos

- **Redes Profundas**
 - **Transparência:**
 - Raciocínio por trás das previsões
 - **Confiança**
 - Maior confiança na precisão das previsões e na tomada de decisões
 - **Melhoria Contínua**
 - A capacidade de interpretar os dados permite otimizar os modelos e aprimorar a precisão das previsões.



Fonte: El-Bouri et al., 2023

Realização:



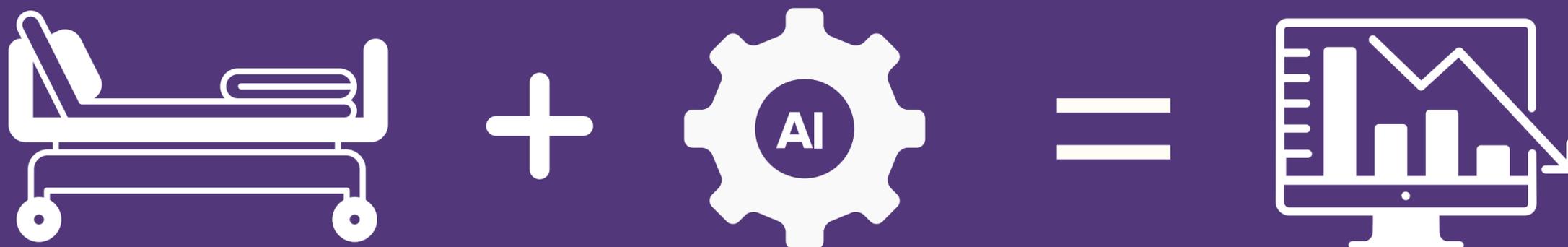
Execução:





Estudos de caso: artigos

- Previsão de localização de internações hospitalares



Fonte: KHANG et al., 2023

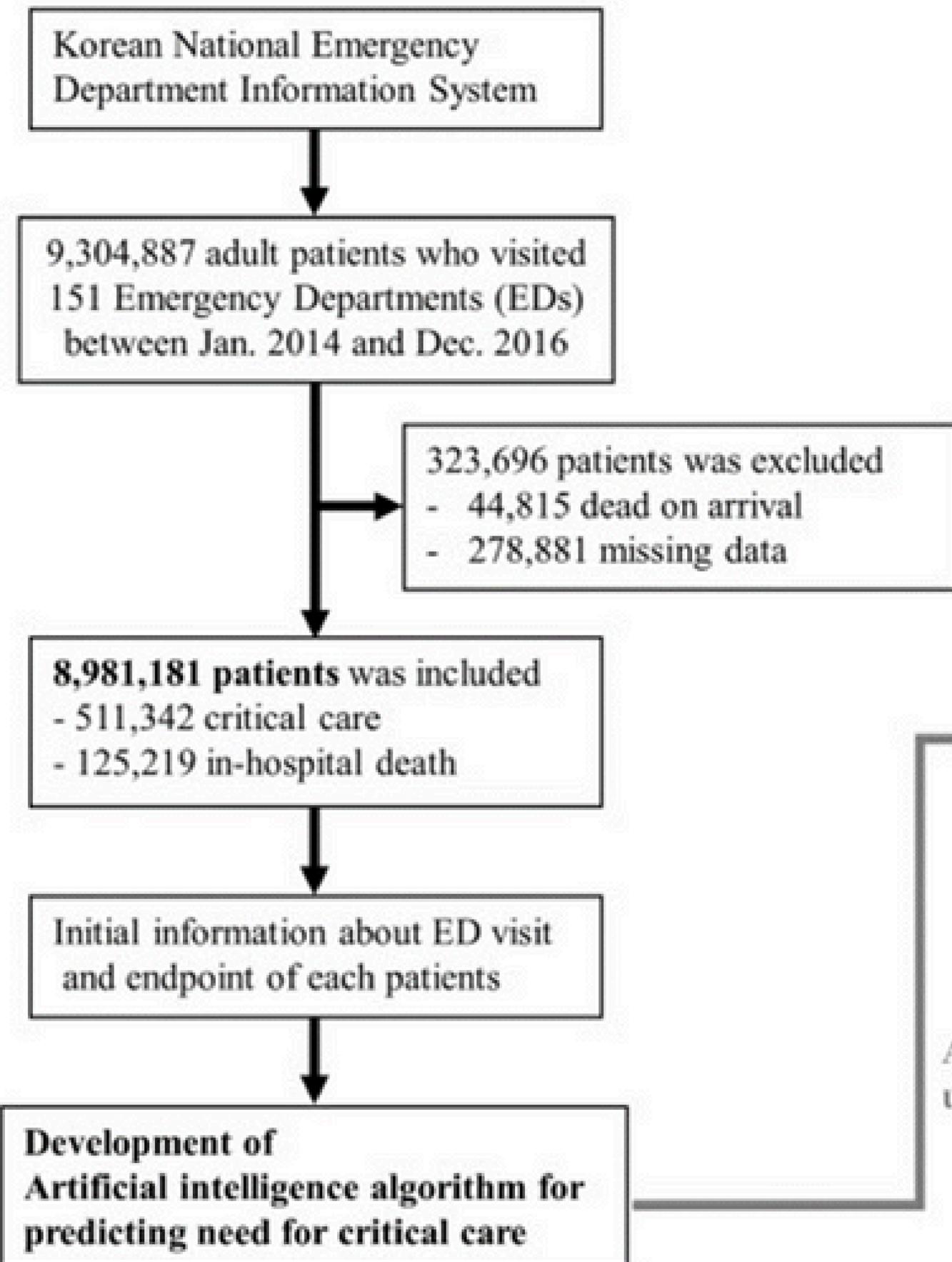
Realização:



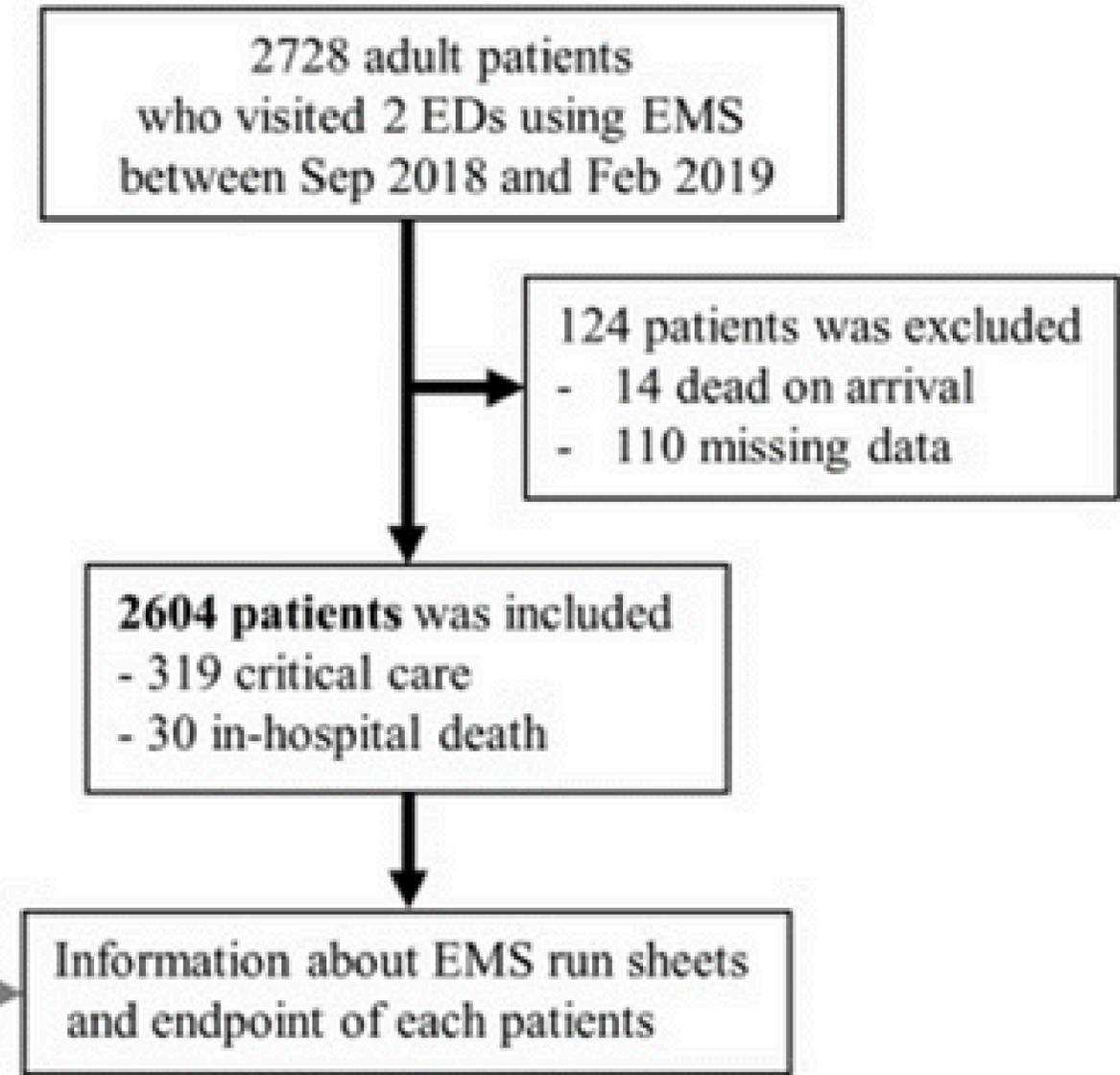
Execução:



Development data



Test data



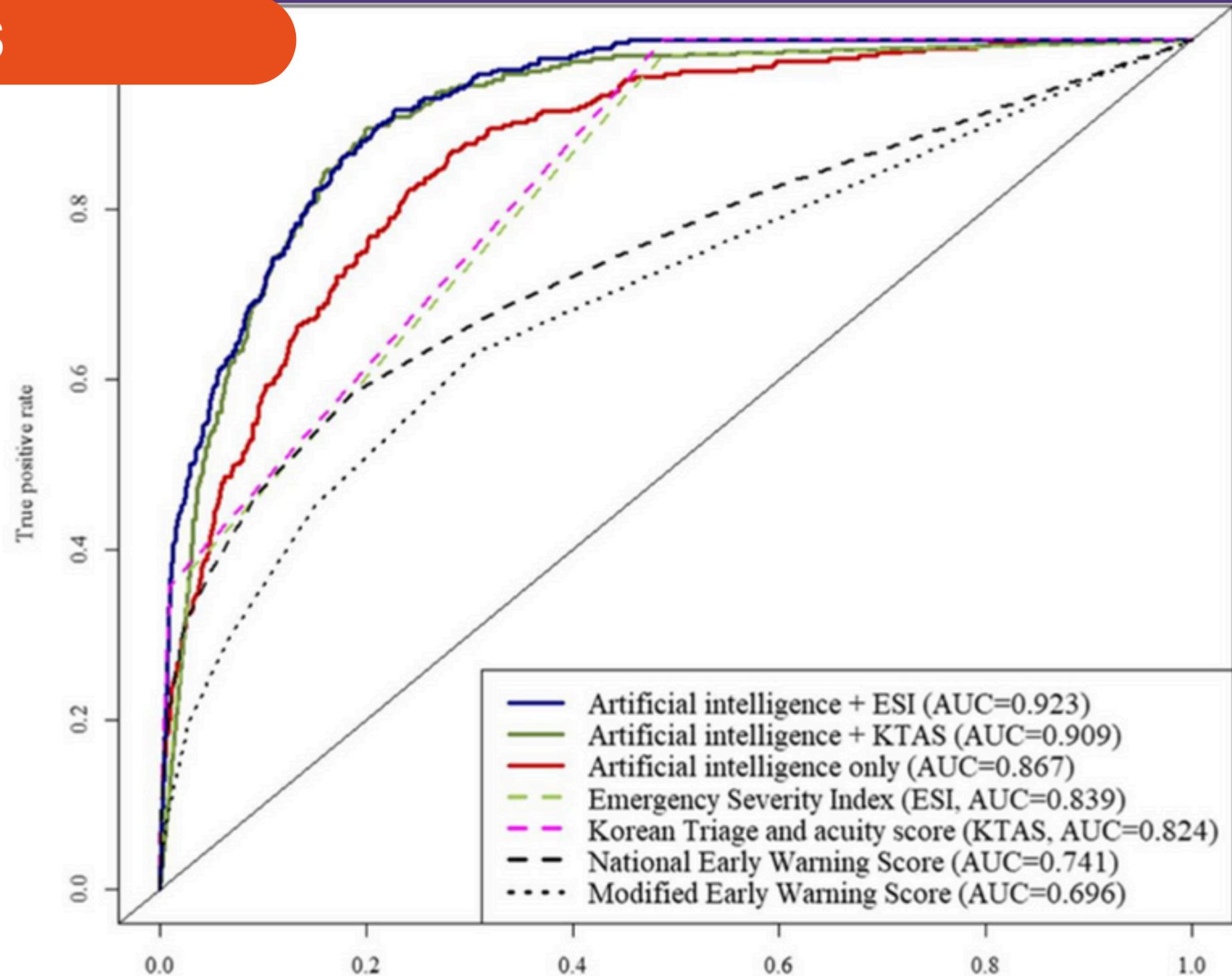
Accuracy test of developed algorithm using test data

Estudos de caso: artigos

- Previsão de localização de internações hospitalares



Fonte: KHANG et al., 2023





Estudos de caso: artigos

- Quais as vantagens de um sistema como esses citados nos dois artigos?



- Triagem Avançada
- Suporte à Decisão Clínica
- Gestão de Recursos
- Otimização Operacional
 - recursos materiais
 - recursos humanos



Realização:



Execução:





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN

Realização:



Execução:



Casos em Startups

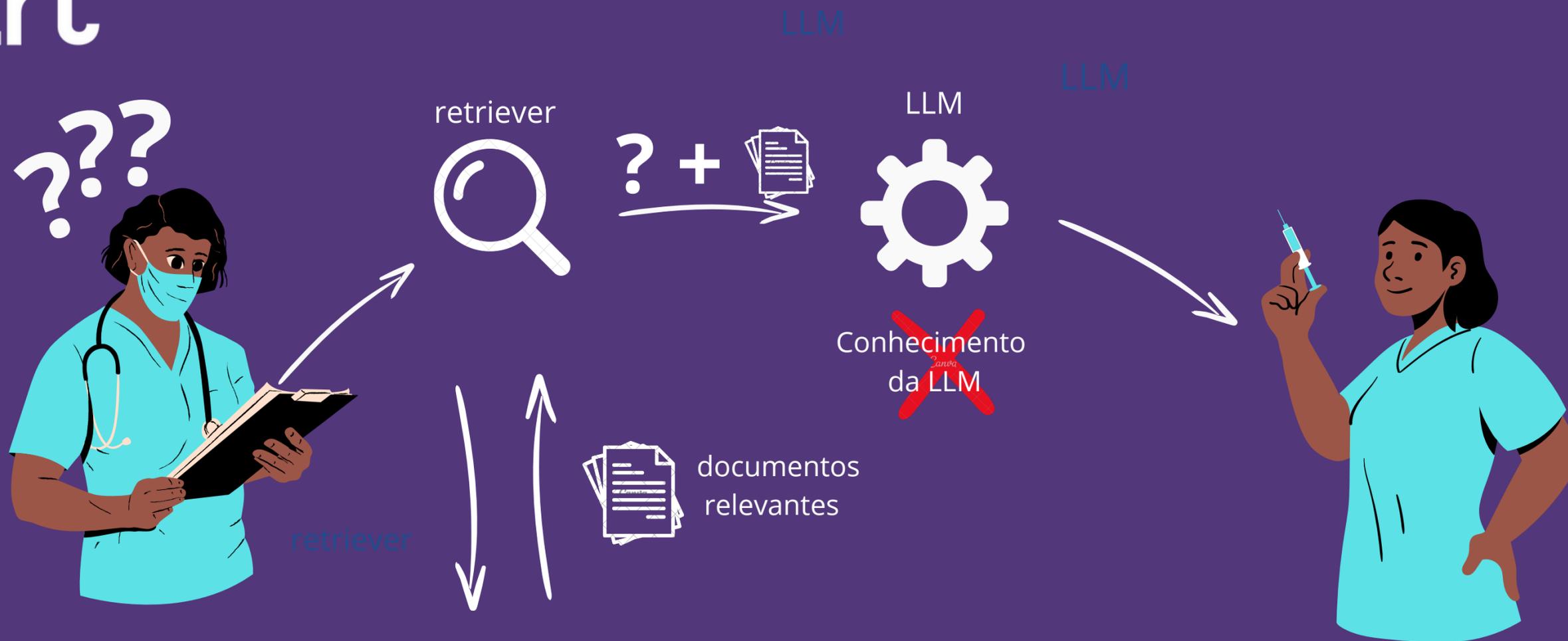
- **CHATBOT com LLM (Large Language Model) para auxiliar profissionais de imunização**
 - Objetivo: informações precisas e confiáveis sobre protocolos de imunização
 - Materiais: videoaulas, fotos, mapas mentais e e-books (guias) como base para a IA identificar as respostas corretas
 - IA: Geração Aumentada via Recuperação (RAG)

Realização:



Execução:





documentos relevantes



base de dados
vetorizada

Realização:



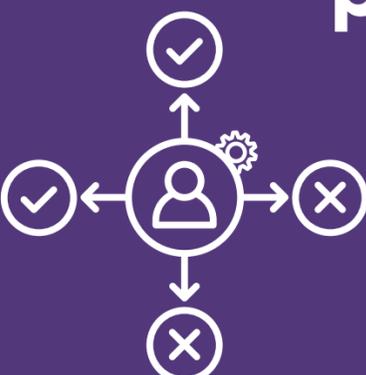
Execução:



Casos em Startups

• Classificação de mensagens de pacientes

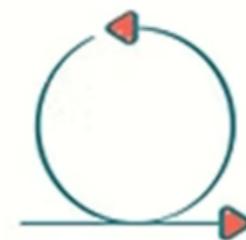
- Objetivo: identificar mensagens que precisam ser atendidas por equipe médica
- IA: Classificação usando IA
- 2 artigos internacionais
- 1 TCC MBA USP



Realização:



Execução:



Agile Health .tech



MACHINE LEARNING MODEL FOR CLASSIFICATION OF HEALTH SYSTEM USER MESSAGES REGARDING THE PRESENCE OF HEALTH NEEDS

ABSTRACT

The electronic transmission of messages represents a mode of communication employed across various services, including those in the healthcare sector. This study introduces a system for the entire world of patients, wherein messages are received and categorized either for automatic responses or for redirection to a human responder, depending on health-related necessities. The database employed in this study is one with two imbalanced classes. To address this issue, **SMOTE** 3.0 was utilized to generate additional examples for the minority class in the database. Three scenarios were compared, the first using the original imbalanced database, the second with a database of synthetic messages generated by **CLUSTOC**, and the third with **CLUSTOC**-generated messages, albeit containing colloquial language and errors in Portuguese. In all models, the **SVM** algorithm was employed. The scenario in which **SVM** demonstrated the best discrimination capacity between messages was the one trained on the database of synthetic messages without Portuguese errors.

Keywords: Natural Language Processing; text classification; SVM

MODELO DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE MENSAGENS DE USUÁRIOS DO SISTEMA DE SAÚDE QUANTO À PRESENÇA DE NECESSIDADES DE SAÚDE

RESUMO

O envio de mensagens eletrônicas é uma forma de comunicação usada em diversos serviços, incluindo os da área de saúde. Neste trabalho, é proposto um sistema de triagem para os pacientes, em que mensagens são recebidas e classificadas em mensagens que são respondidas automaticamente ou mensagens que devem ser direcionadas para seres humanos, em função de necessidades de saúde. A base de dados utilizada neste trabalho é uma base com as duas classes desbalanceadas. Para lidar com esse problema, foi utilizado o **SMOTE** 3.0 para gerar mais exemplos da classe minoritária na base de dados. Foram comparados 3 cenários, o primeiro com dados na base original desbalanceada, o segundo com a base de dados de mensagens sintéticas geradas pelo **CLUSTOC** e o terceiro com mensagens do **CLUSTOC**, mas contendo linguagem coloquial e erros de português. Em todos os modelos foi utilizado o algoritmo **SVM**. O cenário onde o **SVM** apresentou melhor capacidade de discriminação entre as mensagens foi aquele treinado na base de dados com mensagens sintéticas sem erros de português.

Palavras-chave: Processamento de Linguagem Natural; classificação de texto; SVM

MODELO DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE MENSAGENS DE USUÁRIOS DO SISTEMA DE SAÚDE QUANTO À PRESENÇA DE NECESSIDADES DE SAÚDE

Adriana Camargo de Brito
Instituto de Pesquisas Tecnológicas do Estado de São Paulo

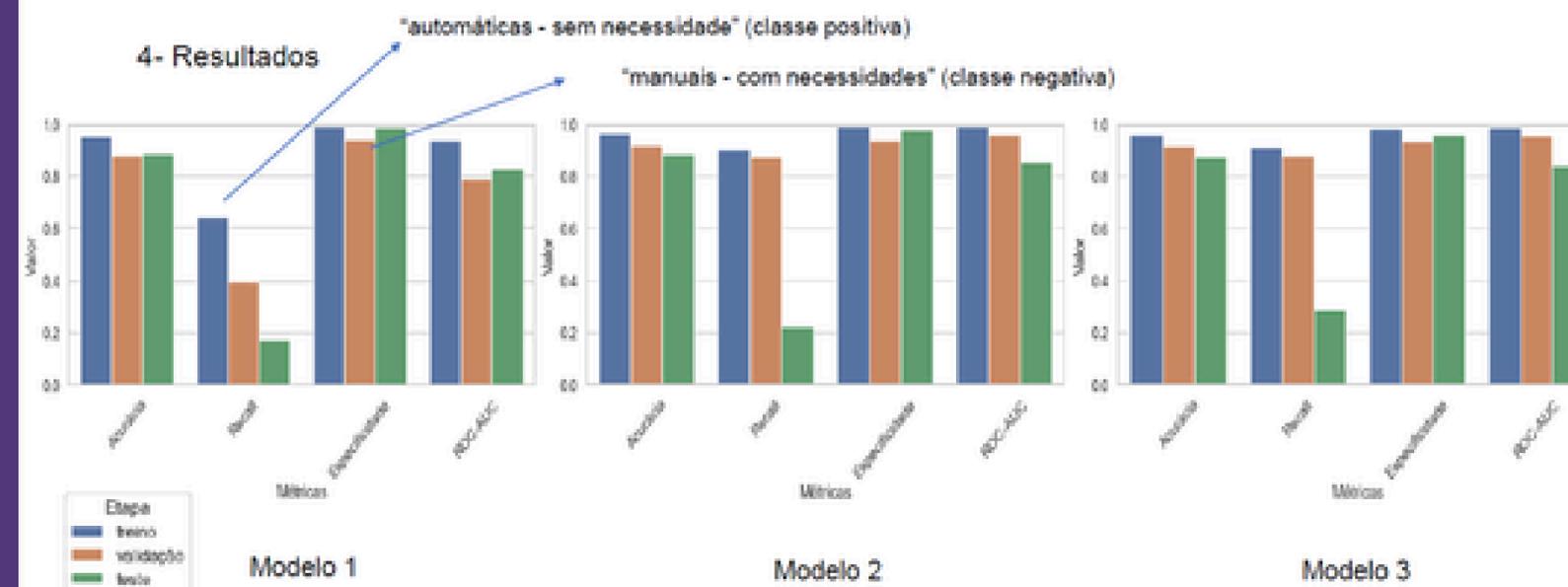
Gustavo Torres Custodio
Instituto de Pesquisas Tecnológicas do Estado de São Paulo

Rogério Silicani Ribeiro
Agile HealthTech

Rubens Carvalho Silveira
Agile HealthTech

Renata Luciria Monteiro
Agile HealthTech

Rodrigo Cabrera Castaldoni
Agile HealthTech



Enhancing Proactive Patient Outreach with Medical Message Classification Using Synthetic Data from ChatGPT

Adriana Camargo de Brito
Digital Transformation
Institute for Technological Research
São Paulo, Brazil
ORCID: 0000-0001-4019-6061

Adriano Galvão Lual
Digital Transformation
Institute for Technological Research
São Paulo, Brazil
ORCID: 0000-0001-7114-6830

Gustavo Torres Custódio
Digital Transformation
Institute for Technological Research
São Paulo, Brazil
ORCID: 0000-0002-3215-1691

Edilson José Rodrigues
Digital Transformation
Institute for Technological Research
São Paulo, Brazil
ORCID: 0000-0001-3505-7202

Michelle Marcia Viana Martins
Department of Economics
Federal University of Minas
Gerais, Brazil
ORCID: 0000-0002-0790-5625

Vinícius Monteiro de Paula Guirado
Department of Neurology
University of São Paulo Medical School
São Paulo, Brazil
ORCID: 0000-0002-7712-5609

Abstract—The increasing demand for proactive patient outreach and follow-up within health insurance providers and large hospital systems has led to the adoption of automated message classification systems. Although state-of-the-art tools like Azure and OpenAI all-urges are available, their costs are prohibitive for small companies. Therefore, using cost-effective traditional tools for text classification is still a valuable alternative. This study investigates how OpenAI's ChatGPT can generate synthetic data to balance class distributions in medical message datasets and improve the performance of binary classification models. The primary aim is to classify messages that indicate medical needs, aiding in patient outreach and efficient resource allocation. We applied Natural Language Processing (NLP) techniques, including Logistic Regression, Naive Bayes, Random Forest, XGBoost, and BERT-based, a Portuguese transformer-based model. Models were trained on both original and augmented datasets, and results showed a significant increase in classification accuracy, particularly with BERT-based, whose performance improved from 84% to 99%. Synthetic data also reduced overfitting by narrowing the gap between training and validation accuracies while improving model robustness on unseen data. Our findings demonstrate that synthetic data from ChatGPT effectively addresses class imbalance, improving the robustness and accuracy of medical message classification models. However, challenges remain in correctly classifying non-medical messages, indicating that further refinement in synthetic data generation is necessary for optimal performance in real-world scenarios. These results suggest that such approaches can significantly enhance automated patient engagement efforts.

Index Terms—Machine learning, Natural language processing, Text classification, Synthetic data, Transformer models, Healthcare communication, Medical applications, Clinical decision support systems, Data augmentation, and Imbalanced datasets.

We thank our financial funds to support from the Institute for Technological Research and São Paulo Research Foundation (FAPESP) grant number 2023/01014-0, 2023/01014-0 and 2023/01014-0.

I. INTRODUCTION

In the healthcare sector, research has shown that text classifiers can effectively reduce human moderator involvement with messages on digital platforms [1], [2], [3], [4], [5], [6], and [7]. Particularly in the medical field [4], the sequence of words and their context is decisive. Consequently, researchers employed more robust Natural Language Processing (NLP) pre-processing techniques that utilize embeddings [2], [3], [4], [5], [6]. These studies used classifiers based on Convolutional Neural Networks (CNN) [3] or combinations of neural networks (CNNs, Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM) [6]. Other researchers applied more complex network architectures with transformers, like Bidirectional Encoder Representations from Transformers (BERT) [7], or combining neural networks with Generative Pre-trained Transformers (GPT) [3], [4].

Recent advancements in generative models, such as GPTs, have demonstrated their ability to address imbalanced datasets by generating synthetic data that maintains contextual integrity and linguistic nuance [8]. While achieving satisfactory classification performance necessitates sufficient and balanced training data, challenges remain in ensuring that augmented samples are not only numerous but also contextually meaningful [10]–[12].

This study applies ChatGPT-generated synthetic data to address class imbalances in medical text classification, demonstrating its potential to improve generalization across diverse datasets. While GPT-based augmentation reduces noise and enhances robustness [13]–[15], it also inherits biases and risks (semantic drifts from training data [9], [16].

PAPER 34 ENHANCING PROACTIVE PATIENT OUTREACH WITH MEDICAL MESSAGE CLASSIFICATION USING SYNTHETIC DATA FROM CHATGPT

AUTHORS

Adriana Camargo de Brito (1)
Adriano Galvão Lual (2)
Gustavo Torres Custódio (3)

Edilson José Rodrigues (1)
Michelle Marcia Viana Martins (2)
Vinícius Monteiro de Paula Guirado (3)

AFFILIATIONS

(1) Institute for Technological Research, São Paulo, Brazil
(2) Federal University of Minas Gerais, Minas Gerais, Brazil
(3) University of São Paulo Medical School, São Paulo, Brazil

INTRODUCTION

The growing need for proactive patient outreach has driven the use of automated message classification. While advanced tools exist, their high costs limit small companies. Traditional, cost-effective methods remain a viable alternative.

OBJECTIVE

This article aims to identify the effect of data augmentation with synthetic data generated by ChatGPT 3.5 on the performance of traditional and Transformer models for the binary classification of health insurance user messages regarding the presence or absence of medical needs.

METHODOLOGY

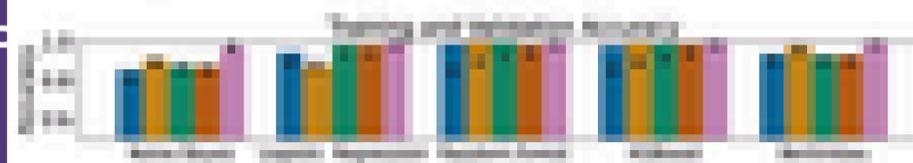
The primary aim is to classify messages that indicate medical needs, aiding in patient outreach and efficient resource allocation. We applied Natural Language Processing (NLP) techniques, including Logistic Regression, Naive Bayes, Random Forest, XGBoost, and BERT-based, a Portuguese transformer-based model. Models were trained on both original and augmented datasets.



DATABASE

A. Camargo de Brito, A. Galvão Lual, G. Torres Custódio, E. José Rodrigues, M. Marcia Viana Martins, and V. Monteiro de Paula Guirado, "Dataset for the IEEE/ACM/CIHM paper 'Enhancing Proactive Patient Outreach with Medical Message Classification Using Synthetic Data from ChatGPT,'" Dec. 2024. [Online]. Available: [https://github.com/PT-ITP-ITP/IEEE/CIHM/2024/gt](https://github.com/PT-ITP/IEEE/CIHM/2024/gt)

Training and Validation Accuracy



Specificity and Recall



Realização:



Execução:



MBA USP ESALO

Data augmentation com dados sintéticos em modelos tradicionais e Transformers para classificação de mensagens médicas

Adriana Camargo de Brito
Michelle Marcia Viana Martins

Realização:



Execução:



Premio Melhor TCC Data Science e Analytics 2024

Casos em Startups

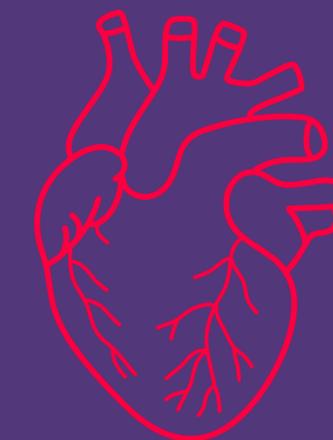
- **Previsão de eventos cardíacos maiores - MACE**

- Objetivo:

- Análise Preditiva: A IA prevê resultados e riscos, auxiliando na tomada de decisões estratégicas e personalizadas.

- Desafios:

- Entender e tratar 20 planilhas com informações de tratamento, internações, variados CIDs, medicamentos e etc.
- Separação de dados por raça, idade, histórico, pessoas que possuem e não possuem MACE
- Treinamento com histórico: variação em tamanhos de históricos, em períodos de ocorrência



Realização:



Execução:



Casos em Startups

- **Estimador de doença cardíaca**
 - Auxílio a decisões clínicas
 - Importância da informação precoce
 - Informação relevante ao paciente

Idade (entre 28 e 77 anos):

Sexo:

Masculino

Tipo de Dor no Peito:

Angina Típica

Pressão Sanguínea em Repouso (entre 0 e 200 mm Hg):

Colesterol (entre 0 e 603 mg/dl):

Açúcar no Sangue em Jejum:

Menos de 120 mg/dl

Eletrocardiograma em Repouso:

Normal

Frequência Cardíaca Máxima Alcançada (entre 60 e 202):

Angina Induzida por Exercício:

Não

Depressão do Segmento ST Induzida por Exercício (entre -2.6 e 6.2):

Inclinação do Segmento ST:

Ascendente

Enviar

Realização:



Execução:



Casos em Startups



- **Análise de Recidiva em Câncer a partir de Histórico Clínico e Exames Médicos**
- **Porque personalizar tratamento?**
 - 28% das mortes: pacientes com câncer de mama e pulmão - INCA
 - 40% dos pacientes com câncer de mama e pulmão podem ter recidiva
 - 5x mais custos quando há recidiva



OncoAI

Realização:



Execução:



Casos em Startups



- **Análise de Recidiva em Câncer a partir de Histórico Clínico e Exames Médicos**
 - Permitir uma visualização mais acessível de dados de pacientes
 - Tipo de dados a serem coletados:
 - Dados clínicos dos pacientes
 - Dados de pacientes diagnosticados
 - LGPD

Realização:



Execução:





Casos em Startups



- **Análise de Recidiva em Câncer a partir de Histórico Clínico e Exames Médicos**
 - Etapas do aprendizado de máquina



Análise do
Prontuário



Análise de
Imagem



Algoritmo



Predição da
Recidiva

Realização:

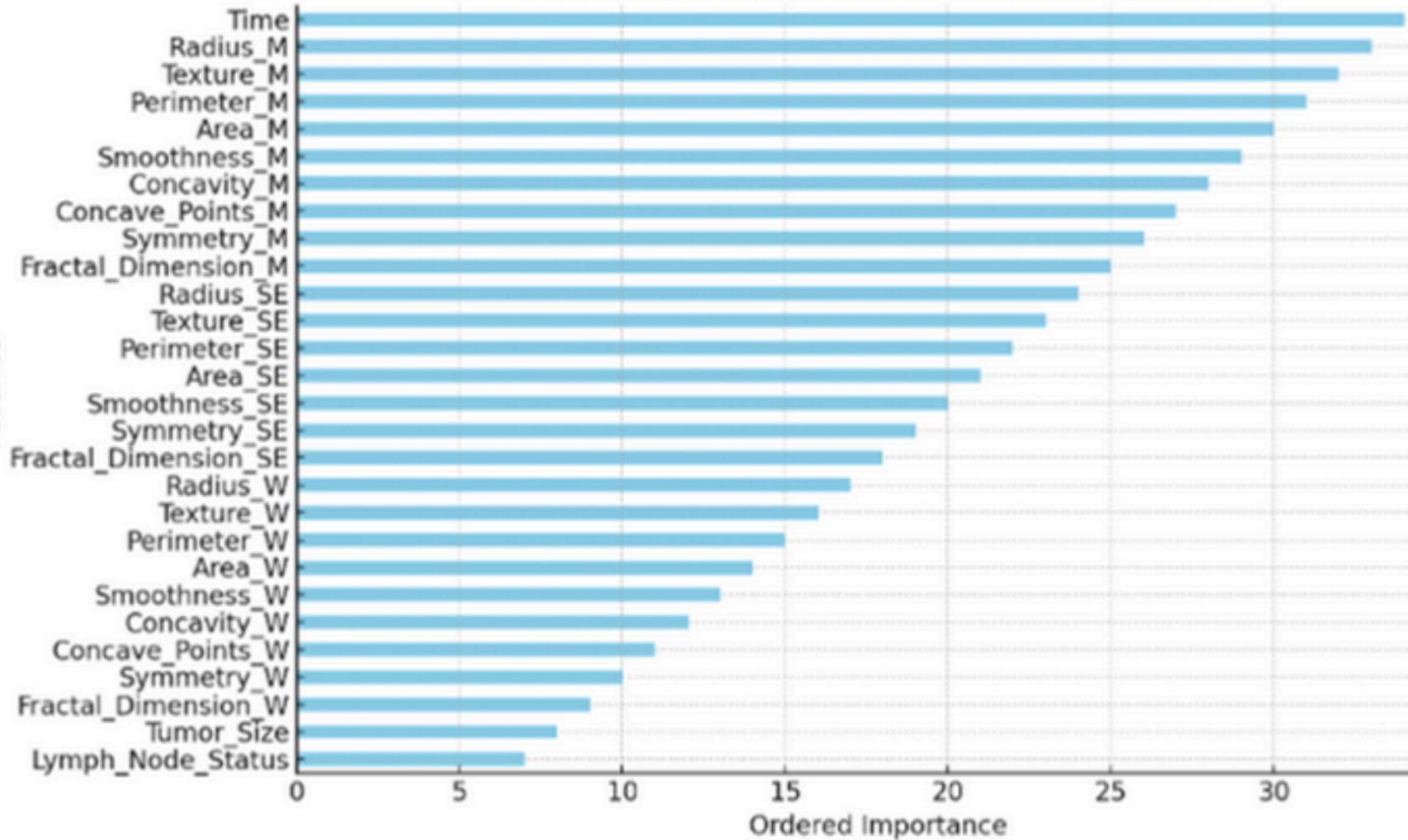


Execução:





Feature Selection Based on Kbest Method



	Accuracy	Precision	F-1 Score	Recall	ROC/AUC
AdaBoost					
Kbest	0.6667	0.7244	0.6414	0.6667	0.6615
Linear	0.6471	0.6872	0.6244	0.6471	0.6423
Complete	0.6471	0.6872	0.6314	0.6471	0.6431
kNN					
Kbest	0.6078	0.6079	0.6069	0.6078	0.6069
Linear	0.6078	0.6095	0.6048	0.6078	0.6062
Complete	0.5882	0.6052	0.5661	0.5882	0.5838
Naive Bayes					
Kbest	0.6667	0.6893	0.6547	0.6667	0.6631
Linear	0.6275	0.6438	0.6140	0.6275	0.6238
Complete	0.5882	0.5886	0.5863	0.5882	0.5869
SVM					
Kbest	0.5882	0.6154	0.5571	0.5882	0.5831
Linear	0.6275	0.6699	0.5993	0.6275	0.6223
Complete	0.6275	0.6544	0.6074	0.6275	0.6231
Random Forest					
Kbest	0.5490	0.5527	0.5328	0.5490	0.5454
Linear	0.5294	0.5287	0.5283	0.5294	0.5285
Complete	0.6471	0.6618	0.6370	0.6471	0.6438



Casos em Startups



- **Aplicação de Inteligência Artificial em Análise de Procedimento Odontológico**
 - Correlação entre o antes e depois do paciente
 - Aplicativo de gerenciamento de dados e controle de pacientes
 - Uso de Inteligência Artificial para Detecção de Fraudes

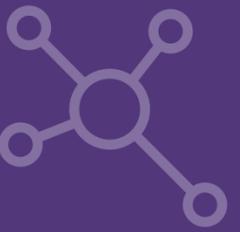


Realização:

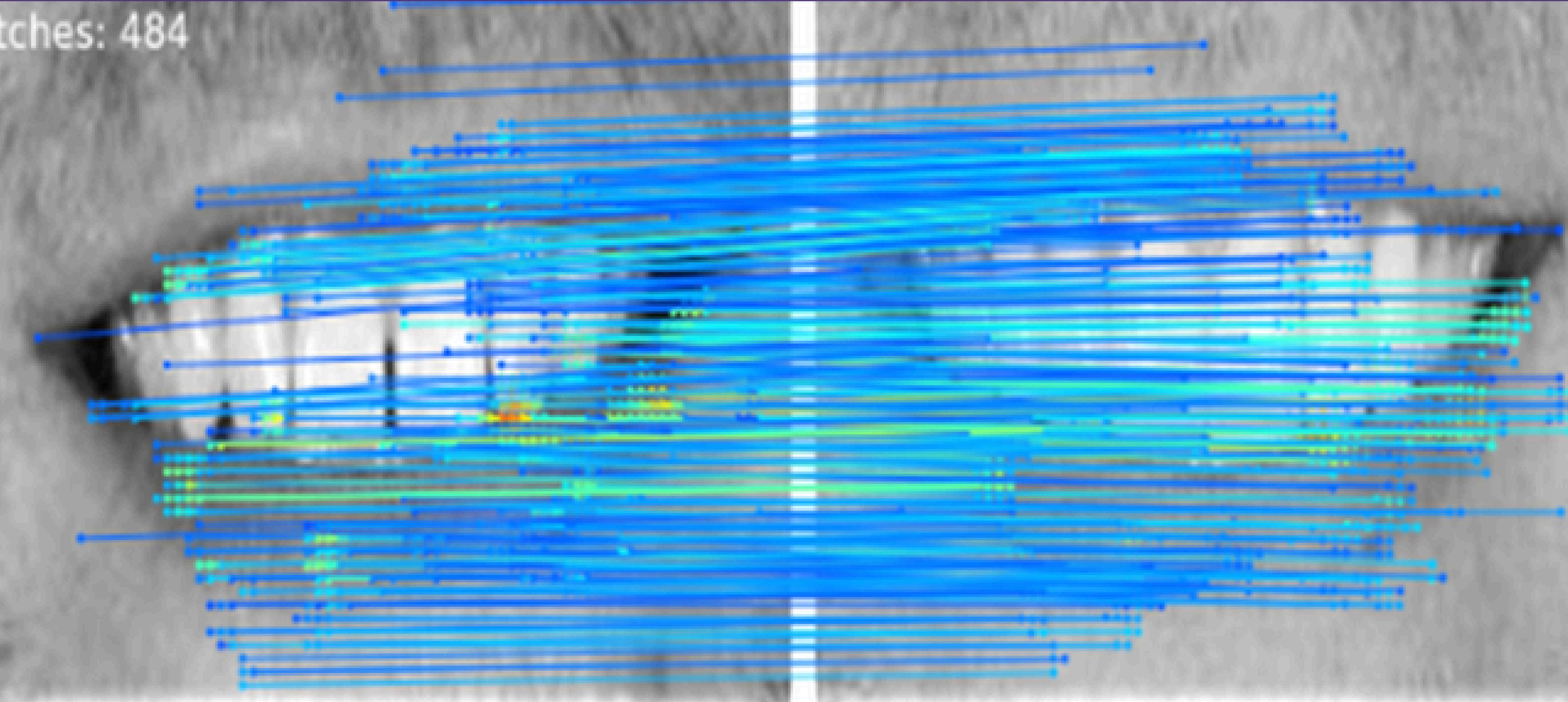


Execução:





atches: 484





Roteiro

- Aplicações de Inteligência Artificial na Saúde
- Impactos da Inteligência Artificial na Saúde
- Desafios da Inteligência Artificial na Saúde
- Estudos de Caso
 - Casos Publicados em Artigos;
 - Casos em Startups: Programa IPT–Sebrae (IA Factory)
 - Casos Desenvolvidos por Mim: IA Factory
- Plataforma IASMIN

Realização:



Execução:





Plataforma IASMIN



Plataforma IAsmin

A Evolução da Indústria Brasileira para o Padrão 4.0

Realização:



Execução:





Plataforma IASMIN

Patrocinadores fundadores



Instituição sede



Realização:



Execução:





Plataforma IASMIN

Instituições participantes

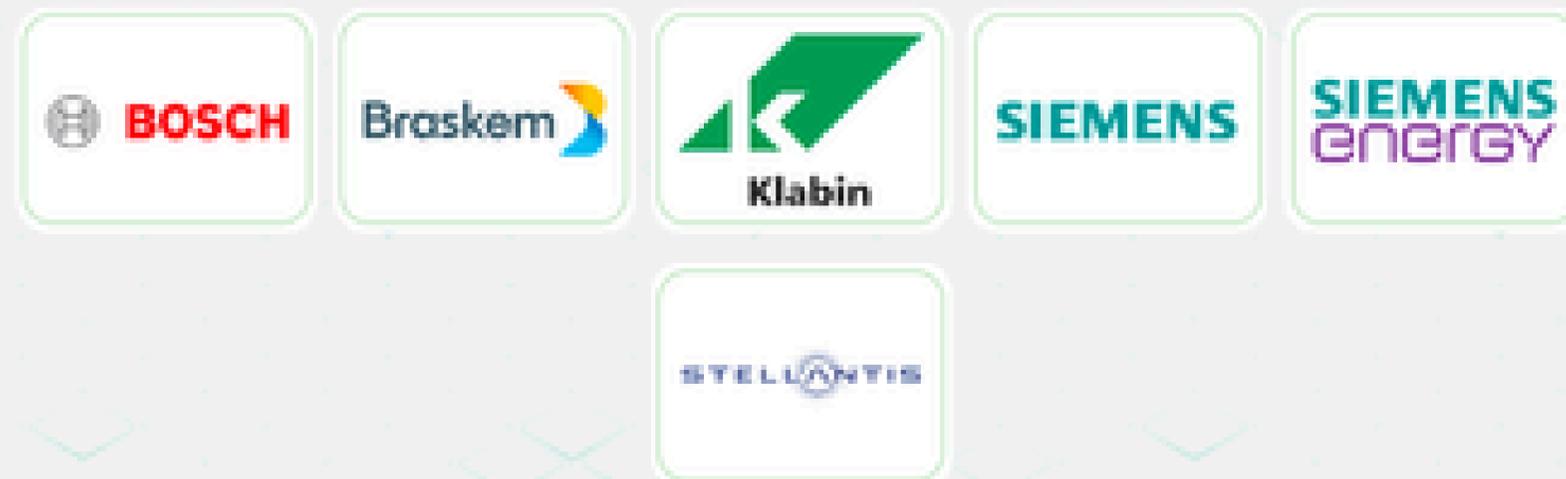
Nacionais



Internacionais



Empresas



Realização:



Execução:





Referências

R. El-Bouri, D. W. Eyre, P. Watkinson, T. Zhu and D. A. Clifton, "Hospital Admission Location Prediction via Deep Interpretable Networks for the Year-Round Improvement of Emergency Patient Care," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 1, pp. 289-300, Jan. 2021, doi: 10.1109/JBHI.2020.2990309.

Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019 Jun;6(2):94-98. doi: 10.7861/futurehosp.6-2-94. PMID: 31363513; PMCID: PMC6616181.

Pradeep, R., Praveen kumar, S., Sasikumar, S., Valarmathie, P., & Gopirajan, P.V. (2021). Artificial Intelligence-Based Automation System for Health Care Applications: Medbot. Advances in Intelligent Systems and Computing.

GHOURI, Arsalan Mujahid et al. An Artificial-Intelligence-Based omnichannel blood supply chain: A pathway for sustainable development. Journal of Business Research, v. 164, 2023. ISSN 0148-2963. Disponível em: <https://doi.org/10.1016/j.jbusres.2023.113980>. Acesso em: 08 nov. 2023.

“LOGÍSTICA TOTAL. Logística Hospitalar. Disponível em: <https://logisticatotal.com.br/logistica-hospitalar/>. Acesso em: 05 de Novembro de 2023.”

KANG, Da-Young; CHO, Kyung-Jae; KWON, Oyeon; KWON, Joon-myung; JEON, Ki-Hyun; PARK, Hyunho; LEE, Yeha; PARK, Jinsik; OH, Byung-Hee. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. ResearchGate, 2020. Disponível em: [Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services - PMC]. Acesso em: 05 de Novembro de 2023.

Realização:



Execução:



start



Obrigado!



person-2



felipephelp1234@gmail.com
fpacheco@ipt.br
felipe.euphrasio@ga.ita.br



**Agradecemos a
sua participação**

Um excelente
evento!



Realização:



Execução:

