



## Mayara Martins Perroni

Endereço para acessar este CV: <http://lattes.cnpq.br/5104624049915172>

ID Lattes: **5104624049915172**

Última atualização do currículo em 04/02/2025

Graduanda em Informática Biomédica pela Universidade de São Paulo (USP), com foco em Machine Learning aplicado à saúde. Concluiu, em 2022, um Treinamento Técnico II FAPESP, no qual trabalhou na segmentação de imagens de tomografia computadorizada para o desenvolvimento de um software de realidade aumentada voltado a procedimentos cirúrgicos. De 2023 a 2024, participou do projeto PRECARE-ML (Prevenção de Eventos Cardiovasculares Adversos Maiores) como bolsista de iniciação científica PUB, aplicando o modelo OMOP-CDM da OHDSI para a organização de coortes clínicas e o desenvolvimento de modelos de aprendizado de máquina. Durante o projeto, contribuiu para a padronização de vocabulários médicos e desenvolveu soluções para evitar o overfitting local, colaborando com a generalização dos modelos para diferentes populações. No Trabalho de Conclusão de Curso (TCC), aplicou técnicas de aprendizado de máquina para a estruturação de campos de texto livre em prontuários eletrônicos de saúde. O projeto envolveu o uso de modelos de processamento de linguagem natural (NLP) para a extração e padronização de informações clínicas relevantes, como fatores de risco (tabagismo, etilismo e obesidade) e sintomas descritos de forma não estruturada nos prontuários. Foram testadas abordagens como Modelos de Linguagem de Grande Escala (LLMs) para capturar nuances e contextos complexos dos textos, além de técnicas clássicas como o TF-IDF para tokenização e vetorização dos textos. Essa combinação permitiu a criação de um pipeline automatizado para processar esses campos e incorporá-los em modelos preditivos de aprendizado de máquina, aumentando a precisão das previsões de MACE em diferentes contextos clínicos.  
(Texto informado pelo autor)

## Identificação

### Nome

Mayara Martins Perroni 

### Nome em citações bibliográficas

PERRONI, M. M.

### Lattes iD



<http://lattes.cnpq.br/5104624049915172>

### País de Nacionalidade

Brasil

## Formação acadêmica/titulação

## 2021 - 2025

Graduação em Informática Biomédica.  
Universidade de São Paulo, USP, Brasil.  
Título: Aprendizado de Máquina para Estruturação de Texto Livre e Predição de Eventos Cardiovasculares em Dados Clínicos Multicêntricos.  
Orientador: Paulo Mazzoncini de Azevedo Marques.

## 2017 - 2019

Ensino Médio (2º grau).  
Instituto Federal - campus São Carlos, IFSP, Brasil.

## 2012 - 2016

Ensino Fundamental (1º grau).  
E.E.Antônio Militão de Lima - São Carlos, MILITÃO, Brasil.

## Atuação Profissional

---

Instituto Federal de São Paulo, IFSP, Brasil.

### Vínculo institucional

## 2017 - 2018

Vínculo: Bolsista, Enquadramento  
Funcional: Monitoria, Carga horária: 40

### Outras informações

Monitoria de Algoritmos e de Biologia.

StepWise Inteligência e Serviços de Apoio à Decisão, STEPWISE, Brasil.

### Vínculo institucional

## 2021 - 2021

Vínculo: Celetista, Enquadramento  
Funcional: Data Annotator, Carga horária: 40

### Outras informações

Trabalho temporário (3 meses) para atuar como Data Annotator na empresa Serasa Experian, sendo a Stepwise a empresa terceirizada mediadora do contrato.

Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, Brasil.

#### **Vínculo institucional**

**2021 - 2023**

Vínculo: Bolsista, Enquadramento  
Funcional: Treinamento Técnico, Carga  
horária: 40, Regime: Dedicação exclusiva.

#### **Outras informações**

A criação de novas formas de treinamentos médicos especializados e interativos se torna importante frente ao cenário relativo a mortalidade ligada a eventos preveníveis. Segundo o II Anuário da Segurança Assistencial Hospitalar no Brasil, em 2017, aproximadamente 98 pacientes morreram a cada dia devido a eventos graves e preveníveis. Durante a Fase I do PIPE constatamos a demanda de mercado por ferramentas que utilizem a realidade aumentada para ensino de práticas médicas. Essa constatação foi obtida a partir da opinião de especialistas em relação a um protótipo com funcionalidades básicas, desenvolvido ao longo do projeto. A avaliação feita a partir de questionários com os usuários do protótipo mostra que a simulação médica pode contribuir para melhor formação dos alunos de áreas médicas, com potenciais benefícios para os pacientes. A Fase II deste projeto visa aprimorar o protótipo apresentado na Fase I, implementando técnicas de visualização digital e manufatura aditiva (MA) para criar uma plataforma de treinamento de cirurgias complexas. Para tanto, a anatomia de pacientes reais serão reproduzidas simultaneamente em um protótipo 3D e numa visualização digital interativa. Por um lado, o uso da MA permite a mimetização de casos individuais no treinamento, incluindo aspectos de uma anatomia única de determinado paciente a ser tratado. De outro, a realidade aumentada (RA) fornece indicadores para que o cirurgião visualize a anatomia interna para agir externamente. Para o desenvolvimento tecnológico e comercial na Fase II serão realizadas as seguintes etapas: 1. Desenvolvimento de uma plataforma digital, nomeada SOMA, para hospedagem de cursos médicos que utilizam simuladores desenvolvidos pela TagmaVisions. Essa plataforma será usada como captação e fidelização de clientes, além de ser um canal de vendas; 2. Desenvolvimento de um produto físico, impresso em 3D, que permita o

Linhas de pesquisa

---

1.

Analise Preditiva
2.

Processamento de Linguagem Natural
3.

Dados Clínicos Multicêntricos
4.

Estruturação de Dados Clínicos
5.

Large Language Models (LLM)
6.

Term Frequency-Inverse Document Frequency (TF-IDF)

Projetos de pesquisa

---

2023 - 2024

Estudo de organização de coorte para aprendizado de máquina utilizando o modelo OMOP-CDM da OHDSI

Descrição: O crescimento da tecnologia da informação em saúde abriu as portas para as organizações deste setor manterem seus dados em várias formas e em vários volumes. Tais dados têm sido usados para gerar conhecimento e apoiar a solução de vários problemas. A heterogeneidade da organização dos dados nas diferentes organizações tende a levar os modelos de aprendizado de máquina (AM) a overfitting local, ajustando-os somente aos detalhes específicos do conjunto de dados de treinamento, sem capacidade de generalização de seus resultados. Para dar suporte à análise de dados multicêntricos e heterogêneos, foram implementadas técnicas de Extraction-Transformation>Loading (ETL). Neste projeto, dados extraídos de registros eletrônicos de saúde (RES) foram padronizados utilizando-se o

Observational Medical Outcomes Partnership (OMOP) - Common Data Model (CDM), que é um padrão de dados da comunidade aberta Observational Health Data Sciences and Informatics (OHDSI)..

Situação: Concluído; Natureza: Pesquisa.  
Alunos envolvidos: Graduação: (1) .

Integrantes: Mayara Martins Perroni -  
Integrante / Paulo Mazzoncini de Azevedo Marques - Coordenador.

## 2017 - 2018

SAMBA E AMOR NA CIDADE DO ?  
POGRÉSSIO?: OS POBRES URBANOS NA  
LÍRICA AMOROSA DE ADONIRAN  
BARBOSA

Descrição: O presente trabalho apresenta os resultados finais do projeto de pesquisa Samba e amor na cidade do ? pogrêssio?: os pobres urbanos na lírica amorosa de Adoniran Barbosa, cujo objeto de análise é a representação da experiência das camadas pobres urbanas no samba lírico-amoroso do compositor paulista Adoniran Barbosa. Aliando a pesquisa bibliográfica à pesquisa e análise poética, busca-se descrever as transformações do cenário urbano paulistano e identificar seu impacto na conformação da temática amorosa nas canções do sambista. Propõe-se como objetivo analisar o papel que a representação poética das vivências populares, ligadas à expansão urbana e industrial da cidade de São Paulo, desempenha na estruturação dos personagens, das narrativas e diálogos amorosos elaborados nas canções do compositor..

Situação: Concluído; Natureza: Pesquisa.

Integrantes: Mayara Martins Perroni -  
Integrante / Giordano Barbin Bertelli -  
Coordenador.

## Projetos de desenvolvimento

---

### 2024 - Atual

Aprendizado de Máquina para  
Estruturação de Texto Livre e Predição de  
Eventos Cardiovasculares em Dados  
Clínicos Multicêntricos

Descrição: O avanço da tecnologia da informação em saúde possibilitou a coleta e armazenamento de grandes volumes de dados em formatos variados, essenciais para apoiar a tomada de decisões clínicas. No entanto, a heterogeneidade desses dados pode comprometer o desempenho de modelos de aprendizado de máquina

(AM), causando sobreajuste (overfitting) e limitando a generalização dos resultados. Para lidar com esse desafio, foi implementado um processo de Extraction, Transformation and Loading (ETL), padronizando dados de registros eletrônicos de saúde (RES) utilizando o Observational Medical Outcomes Partnership (OMOP) - Common Data Model (CDM), da comunidade Observational Health Data Sciences and Informatics (OHDSI). Esse processo garantiu a uniformidade dos dados, permitindo sua integração em diferentes instituições de saúde. O objetivo foi desenvolver e testar um classificador para identificar fumantes a partir de textos livres extraídos de prontuários eletrônicos. Para isso, foram aplicadas duas abordagens de aprendizado de máquina: Large Language Models (LLM) e Term Frequency-Inverse Document Frequency (TF-IDF). Os textos, padronizados no OMOP-CDM, passaram por etapas de pré-processamento, categorização e análise comparativa. O LLM focou na interpretação semântica e contextual dos textos, enquanto o TF-IDF identificou padrões de palavras-chave associados ao tabagismo. Após os testes, ambos os modelos foram avaliados em termos de precisão, recall e capacidade de generalização, considerando a variabilidade dos dados. Os resultados indicam que o LLM oferece uma abordagem mais eficaz para categorizar fatores de risco em dados não estruturados, proporcionando uma análise mais profunda dos textos livres. Por outro lado, o TF-IDF se mostrou vantajoso em contextos com padrões de texto mais controlados e repetitivos. Este estudo contribui para o desenvolvimento de classificadores automáticos que podem ser integrados a sistemas de prontuários eletrônicos, aprimorando a categorização de informações críticas, como o histórico de tabagismo, e servindo de base para futuras análises de fatores de risco em registros padronizados..

Situação: Em andamento; Natureza: Desenvolvimento.

Alunos envolvidos: Graduação: (1) .

Integrantes: Mayara Martins Perroni - Integrante / Paulo Mazzoncini de Azevedo Marques - Coordenador.

## **2021 - 2023**

Desenvolvimento de uma plataforma para treinamento em cirurgias de pacientes específicos usando realidade aumentada e manufatura aditiva

Descrição: A criação de novas formas de treinamentos médicos especializados e interativos se torna importante frente ao cenário relativo a mortalidade ligada a eventos preveníveis. Segundo o II Anuário da Segurança Assistencial Hospitalar no

Brasil, em 2017, aproximadamente 98 pacientes morreram a cada dia devido a eventos graves e preveníveis. Durante a Fase I do PIPE constatamos a demanda de mercado por ferramentas que utilizem a realidade aumentada para ensino de práticas médicas. Essa constatação foi obtida a partir da opinião de especialistas em relação a um protótipo com funcionalidades básicas, desenvolvido ao longo do projeto. A avaliação feita a partir de questionários com os usuários do protótipo mostra que a simulação médica pode contribuir para melhor formação dos alunos de áreas médicas, com potenciais benefícios para os pacientes. A Fase II deste projeto visa aprimorar o protótipo apresentado na Fase I, implementando técnicas de visualização digital e manufatura aditiva (MA) para criar uma plataforma de treinamento de cirurgias complexas. Para tanto, a anatomia de pacientes reais serão reproduzidas simultaneamente em um protótipo 3D e numa visualização digital interativa. Por um lado, o uso da MA permite a mimetização de casos individuais no treinamento, incluindo aspectos de uma anatomia única de determinado paciente a ser tratado. De outro, a realidade aumentada (RA) fornece indicadores para que o cirurgião visualize a anatomia interna para agir externamente. Para o desenvolvimento tecnológico e comercial na Fase II serão realizadas as seguintes etapas: 1. Desenvolvimento de uma plataforma digital, nomeada SOMA, para hospedagem de cursos médicos que utilizam simuladores desenvolvidos pela TagmaVisions. Essa plataforma será usada como captação e fidelização de clientes, além de ser um canal de vendas; 2. Desenvolvimento de um produto físico, impresso em 3D, que permita o treinamento de práticas médicas com interações virtuais por meio da tecnologia de RA, utilizando dispositivos móveis como smartphones e tablets. Este produto será ser disponibilizado na plataforma digital em parceria com centros de ensino; 3. Uma plataforma digital que utiliza o design instrucional para integrar conhecimento à plataforma; e 4. Uma proposta de modelo de negócios capaz de gerir o conhecimento em nível de complexidade suficientes para atender esta demanda com alto grau de especialização. O projeto de produto comercial deverá considerar a melhor forma de aliar a reprodução da anatomia da coluna vertebral nos dois mundo, tanto físico quanto virtual. Durante o projeto buscaremos elaborar um pipeline inteligente que acelere o entendimento da jornada do cliente, incremente as funcionalidades da interface com o usuário e execute o processamento das imagens e dos modelos anatômicos através da impressão 3D. O próximo passo será a integração desta segmentação com o aplicativo de RA, que receberá as malhas tridimensionais e as integrará com o modelo impresso. Os

produtos de cada fase iterativa do desenvolvimento serão testados pela sua apresentação como recursos didáticos em cada um dos 10 módulos do Curso de Aperfeiçoamento na USP de Ribeirão Preto. A cada módulo do curso ciclos consecutivos de construir - medir - aprender serão efetivados com os menores esforços e tempo de desenvolvimento possíveis. Ao final do período deste projeto teremos o produto pronto para comercialização..

Situação: Concluído; Natureza: Desenvolvimento.

Alunos envolvidos: Graduação: (1) / Doutorado: (1) .

Integrantes: Mayara Martins Perroni - Integrante / Carlo Rondinoni - Coordenador.

## Idiomas

---

### Inglês

Compreende Bem, Fala Bem, Lê Bem, Escreve Bem.

### Espanhol

Compreende Razoavelmente, Fala Razoavelmente, Lê Razoavelmente, Escreve Razoavelmente.

### Português

Compreende Bem, Fala Bem, Lê Bem, Escreve Bem.

## Produções

---

### Produção bibliográfica

### Apresentações de Trabalho

1.

★ **PERRONI, M. M.**; Paulo Mazzoncini de Azevedo Marques . Estudo de organização de coorte para aprendizado de máquina utilizando o modelo OMOP-CDM da OHDSI. 2024. (Apresentação de Trabalho/Congresso).

2.

Paulo Mazzoncini de Azevedo Marques ; **PERRONI, M. M.** . Estudo de organização de coorte para aprendizado de máquina utilizando o modelo OMOP-CDM da OHDSI.. 2024. (Apresentação de Trabalho/Simpósio).



## Demais tipos de produção técnica

1.

★ **PERRONI, M. M.**. Estudo de organização de coorte para aprendizado de máquina utilizando o modelo OMOP-CDM da OHDSI. 2024. (Relatório de pesquisa).

## Eventos

---

### Participação em eventos, congressos, exposições e feiras

1.

3º Workshop de Inovação, Pesquisa, Ensino e Extensão. SAMBA E AMOR, NA CIDADE DO ?POGRESSION?: OS POBRES URBANOS NA LÍRICA AMOROSA DE ADONIRAN BARBOSA. 2018. (Congresso).

## Inovação

---

### Projeto de desenvolvimento tecnológico

#### 2024 - Atual

Aprendizado de Máquina para Estruturação de Texto Livre e Predição de Eventos Cardiovasculares em Dados Clínicos Multicêntricos

Descrição: O avanço da tecnologia da informação em saúde possibilitou a coleta e armazenamento de grandes volumes de dados em formatos variados, essenciais para apoiar a tomada de decisões clínicas. No entanto, a heterogeneidade desses dados pode comprometer o desempenho de modelos de aprendizado de máquina (AM), causando sobreajuste (overfitting) e limitando a generalização dos resultados. Para lidar com esse desafio, foi implementado um processo de Extraction, Transformation and Loading (ETL), padronizando dados de registros eletrônicos de saúde (RES) utilizando o Observational Medical Outcomes Partnership (OMOP) - Common Data

Model (CDM), da comunidade Observational Health Data Sciences and Informatics (OHDSI). Esse processo garantiu a uniformidade dos dados, permitindo sua integração em diferentes instituições de saúde. O objetivo foi desenvolver e testar um classificador para identificar fumantes a partir de textos livres extraídos de prontuários eletrônicos. Para isso, foram aplicadas duas abordagens de aprendizado de máquina: Large Language Models (LLM) e Term Frequency-Inverse Document Frequency (TF-IDF). Os textos, padronizados no OMOP-CDM, passaram por etapas de pré-processamento, categorização e análise comparativa. O LLM focou na interpretação semântica e contextual dos textos, enquanto o TF-IDF identificou padrões de palavras-chave associados ao tabagismo. Após os testes, ambos os modelos foram avaliados em termos de precisão, recall e capacidade de generalização, considerando a variabilidade dos dados. Os resultados indicam que o LLM oferece uma abordagem mais eficaz para categorizar fatores de risco em dados não estruturados, proporcionando uma análise mais profunda dos textos livres. Por outro lado, o TF-IDF se mostrou vantajoso em contextos com padrões de texto mais controlados e repetitivos. Este estudo contribui para o desenvolvimento de classificadores automáticos que podem ser integrados a sistemas de prontuários eletrônicos, aprimorando a categorização de informações críticas, como o histórico de tabagismo, e servindo de base para futuras análises de fatores de risco em registros padronizados..  
Situação: Em andamento; Natureza: Desenvolvimento.  
Alunos envolvidos: Graduação: (1) .

Integrantes: Mayara Martins Perroni - Integrante / Paulo Mazzoncini de Azevedo Marques - Coordenador.

## 2021 - 2023

Desenvolvimento de uma plataforma para treinamento em cirurgias de pacientes específicos usando realidade aumentada e manufatura aditiva

Descrição: A criação de novas formas de treinamentos médicos especializados e interativos se torna importante frente ao cenário relativo a mortalidade ligada a eventos preveníveis. Segundo o II Anuário da Segurança Assistencial Hospitalar no Brasil, em 2017, aproximadamente 98 pacientes morreram a cada dia devido a eventos graves e preveníveis. Durante a Fase I do PIPE constatamos a demanda de mercado por ferramentas que utilizem a realidade aumentada para ensino de práticas médicas. Essa constatação foi obtida a partir da opinião de especialistas em relação a um protótipo com

funcionalidades básicas, desenvolvido ao longo do projeto. A avaliação feita a partir de questionários com os usuários do protótipo mostra que a simulação médica pode contribuir para melhor formação dos alunos de áreas médicas, com potenciais benefícios para os pacientes. A Fase II deste projeto visa aprimorar o protótipo apresentado na Fase I, implementando técnicas de visualização digital e manufatura aditiva (MA) para criar uma plataforma de treinamento de cirurgias complexas. Para tanto, a anatomia de pacientes reais serão reproduzidas simultaneamente em um protótipo 3D e numa visualização digital interativa. Por um lado, o uso da MA permite a mimetização de casos individuais no treinamento, incluindo aspectos de uma anatomia única de determinado paciente a ser tratado. De outro, a realidade aumentada (RA) fornece indicadores para que o cirurgião visualize a anatomia interna para agir externamente. Para o desenvolvimento tecnológico e comercial na Fase II serão realizadas as seguintes etapas: 1. Desenvolvimento de uma plataforma digital, nomeada SOMA, para hospedagem de cursos médicos que utilizam simuladores desenvolvidos pela TagmaVisions. Essa plataforma será usada como captação e fidelização de clientes, além de ser um canal de vendas; 2. Desenvolvimento de um produto físico, impresso em 3D, que permita o treinamento de práticas médicas com interações virtuais por meio da tecnologia de RA, utilizando dispositivos móveis como smartphones e tablets. Este produto será disponibilizado na plataforma digital em parceria com centros de ensino; 3. Uma plataforma digital que utiliza o design instrucional para integrar conhecimento à plataforma; e 4. Uma proposta de modelo de negócios capaz de gerir o conhecimento em nível de complexidade suficientes para atender esta demanda com alto grau de especialização. O projeto de produto comercial deverá considerar a melhor forma de aliar a reprodução da anatomia da coluna vertebral nos dois mundos, tanto físico quanto virtual. Durante o projeto buscaremos elaborar um pipeline inteligente que acelere o entendimento da jornada do cliente, incremente as funcionalidades da interface com o usuário e execute o processamento das imagens e dos modelos anatômicos através da impressão 3D. O próximo passo será a integração desta segmentação com o aplicativo de RA, que receberá as malhas tridimensionais e as integrará com o modelo impresso. Os produtos de cada fase iterativa do desenvolvimento serão testados pela sua apresentação como recursos didáticos em cada um dos 10 módulos do Curso de Aperfeiçoamento na USP de Ribeirão Preto. A cada módulo do curso ciclos consecutivos de construir - medir - aprender serão efetivados com os menores esforços e tempo de

desenvolvimento possíveis. Ao final do período deste projeto teremos o produto pronto para comercialização..

Situação: Concluído; Natureza: Desenvolvimento.

Alunos envolvidos: Graduação: (1) / Doutorado: (1) .

Integrantes: Mayara Martins Perroni -  
Integrante / Carlo Rondinoni -  
Coordenador.

Página gerada pelo Sistema Currículo Lattes em 14/03/2025 às 17:55:49

Somente os dados identificados como públicos pelo autor são apresentados na consulta do seu Currículo Lattes.

[Configuração de privacidade na Plataforma Lattes](#)