

# INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL APLICADA NA ESTIMATIVA DE NÍVEIS DE OBESIDADE

## EXPLAINABLE ARTIFICIAL INTELLIGENCE APPLIED TO ESTIMATE OBESITY LEVELS

Willian de Oliveira Cardoso\*

Adonias Caetano de Oliveira\*\*

Rhyan Ximenes de Brito\*\*\*

### RESUMO

A obesidade representa um grave problema de saúde pública global, demandando ferramentas eficazes para sua avaliação e manejo. Este trabalho investiga a aplicação da Inteligência Artificial Explicável (XAI, do inglês *Explainable Artificial Intelligence*) na estimativa de níveis de obesidade, com o intuito de aprimorar a interpretabilidade e a confiabilidade dos modelos preditivos. Comparou-se o desempenho e a explicabilidade de modelos de Aprendizado de Máquina, como *Extra Trees Classifier* (ETC) e *Random Forest* (RF), com um modelo de Aprendizado Profundo (AP), utilizando um conjunto de dados sobre hábitos alimentares e estilo de vida. A metodologia incluiu o pré-processamento dos dados, o treinamento e a avaliação dos modelos por métricas como acurácia, precisão, *recall* e *F1-score*, além da aplicação da técnica *Local Interpretable Model-agnostic Explanations* (LIME), para explicar as previsões. Os resultados demonstram que os modelos ETC e RF apresentaram alta acurácia (superior a 91%), com baixo custo computacional e rápido tempo de inferência, sendo eficazes na classificação dos níveis de obesidade. O modelo de AP, embora com excelente desempenho de acurácia (92.5%) e AUC-ROC (97.5%), exigiu maior tempo de treinamento. A aplicação do LIME revelou a influência de variáveis como Índice de Massa Corporal, idade e nível de atividade física nas decisões dos modelos, contribuindo significativamente para a transparência e a confiança nos resultados. Conclui-se que a XAI é crucial para o entendimento e validação das predições em saúde. As variáveis identificadas como mais influentes pelos modelos, como IMC, idade e nível de atividade física, corroboram o conhecimento científico consolidado na literatura sobre os principais fatores associados à obesidade. Estes achados sugerem que os modelos desenvolvidos, especialmente o ETC e o RF pela sua eficiência e explicabilidade, possuem alto potencial para serem integrados em sistemas inteligentes de apoio clínico, fornecendo diagnósticos mais transparentes e auxiliando na tomada de decisões para intervenções personalizadas no combate à obesidade.

**Palavras-chave:** XAI. Obesidade. Aprendizado de Máquina. Aprendizado Profundo. LIME.

### ABSTRACT

---

\* Autor do artigo e Discente do Curso de Ciência da Computação, IFCE Campus Tianguá

\*\* Docente-orientador do Curso de Ciência da Computação, IFCE Campus Tianguá

\*\*\* Docente-coorientador do Curso de Ciência da Computação, IFCE Campus Tianguá

Obesity represents a severe global public health problem, demanding effective tools for its assessment and management. This work investigates the application of Explainable Artificial Intelligence (XAI) in estimating obesity levels, aiming to enhance the interpretability and reliability of predictive models. The performance and explainability of Machine Learning (ML) models, such as Extra Trees Classifier (ETC) and Random Forest (RF), were compared with a Deep Learning (DL) model, using a dataset on eating habits and lifestyle. The methodology included data preprocessing, model training, and evaluation using metrics such as accuracy, precision, recall, and F1-score, in addition to applying the Local Interpretable Model-agnostic Explanations (LIME) technique to explain predictions. The results demonstrate that the ETC and RF models achieved high accuracy (over 91%), with low computational cost and fast inference time, proving effective in classifying obesity levels. The DL model, although with excellent accuracy (92.5%) and AUC-ROC (97.5%) performance, required longer training time. The application of LIME revealed the influence of variables such as Body Mass Index (BMI), age, and physical activity level on the models' decisions, significantly contributing to the transparency and trustworthiness of the results. XAI is concluded to be crucial for understanding and validating health predictions. The variables identified as most influential by the models, such as BMI, age, and physical activity level, corroborate the consolidated scientific knowledge in the literature on the main factors associated with obesity. These findings suggest that the developed models, especially ETC and RF due to their efficiency and explainability, have high potential to be integrated into intelligent clinical support systems, providing more transparent diagnoses and assisting in decision-making for personalized interventions in combating obesity.

**Keywords:** XAI. Obesity. Machine Learning. Deep Learning. LIME.

## 1 INTRODUÇÃO

A obesidade é uma condição crônica caracterizada pelo acúmulo excessivo de gordura corporal, resultante de um desequilíbrio entre ingestão e gasto energético. Sua etiologia é multifatorial, envolvendo aspectos genéticos, metabólicos, emocionais e comportamentais, como hábitos alimentares inadequados e sedentarismo. Esse distúrbio nutricional tem se tornado um problema de saúde pública global, afetando diferentes faixas etárias e contextos socioeconômicos. Segundo a Organização Pan-Americana da Saúde (OPAS), a taxa de obesidade em crianças e adolescentes cresceu mais de dez vezes entre 1975 e 2016, evidenciando o impacto crescente dessa condição ao longo das décadas (NEVES et al., 2021).

A avaliação da obesidade no Brasil é amplamente realizada por métodos tradicionais, como o Índice de Massa Corporal (IMC), a circunferência da cintura e a relação cintura-quadril, de fácil aplicação e baixo custo (SOUZA et al., 2014; ABESO, 2016). Além desses, técnicas como a bioimpedância elétrica e as dobras cutâneas são utilizadas para estimar a composição corporal e o percentual de gordura (LINHARES et al., 2019). Métodos mais avançados, como a ultrassonografia, são estudados para avaliação da gordura corporal em pacientes obesos, demonstrando maior precisão em comparação às técnicas convencionais (SILVA et al., 2023).

No entanto, apesar de amplamente aplicados, esses métodos apresentam limitações, como a incapacidade de distinguir a distribuição da gordura corporal e possíveis imprecisões em indivíduos com obesidade severa, reforçando a necessidade de abordagens complementares e mais precisas.

A Inteligência Artificial (IA) têm se destacado como ferramentas inovadoras para superar as limitações das abordagens tradicionais no diagnóstico e monitoramento da obesidade. Modelos de IA baseados em imagens corporais e indicadores metabólicos, podem prever riscos de obesidade e suas comorbidades com uma acurácia superior à do IMC, oferecendo uma abordagem mais robusta e personalizada (ZHANG; LI; CHEN, 2021). Além disso, Smith, Brown e Johnson (2022) destacam que a combinação de IA com dados fenotípicos e genômicos permite uma estratificação mais detalhada dos pacientes obesos, facilitando intervenções clínicas mais eficazes e individualizadas. Essa integração não apenas melhora a eficiência diagnóstica, mas também possibilita a identificação de subtipos de obesidade, como a obesidade metabólica ou sarcopenica, que frequentemente passam despercebidos pelos métodos tradicionais.

Apesar do alto desempenho da IA na previsão de risco de obesidade, a adoção de modelos de IA na área da saúde exige transparência e interpretabilidade para garantir a confiança dos profissionais de saúde (KHATER; TAWFIK; SINGH, 2024). Nesse contexto, a Inteligência Artificial Explicável (XAI, do inglês *Explainable Artificial Intelligence*) surge como uma abordagem essencial. Ribeiro, Singh e Guestrin (2020) destacam que a falta de explicabilidade em modelos de IA pode limitar sua aplicação clínica, uma vez que os médicos precisam compreender as decisões tomadas pelos algoritmos para confiar em suas recomendações. Além disso, Chen, Li e Zhang (2021) reforçam que a XAI permite descrever como os modelos de IA chegam a determinadas conclusões, tornando os resultados mais transparentes e clinicamente interpretáveis. Isso é particularmente relevante no contexto da obesidade, onde a estratificação de pacientes e a personalização de tratamentos dependem de decisões precisas e bem fundamentadas. Estudos recentes, como o de Kumar, Singh e Patel (2023), argumentam que a XAI pode facilitar a integração de IA em protocolos clínicos, promovendo a aceitação e a adoção dessas tecnologias por profissionais de saúde.

Neste sentido de explicar os modelos de IA, este estudo busca responder à seguinte questão de pesquisa: “Como os modelos de IA identificam os níveis de obesidade?”. Para isso, o objetivo geral é explicar, por meio do método *Local Interpretable Model-agnostic Explanations* (LIME), o desempenho dos modelos de IA na identificação de níveis de obesidade. O LIME é uma técnica da XAI que permite interpretar as decisões de modelos de AM, tornando transparentes os critérios utilizados para classificar pacientes em diferentes níveis de obesidade (AZAD; KHAN; EL-GHANY, 2025). Ele funciona perturbando localmente os dados de entrada e observando como as previsões do modelo mudam, gerando explicações interpretáveis para cada predição individual. Essa abordagem é fundamental para aumentar a confiança dos profissionais de saúde na aplicação clínica dessas tecnologias, além de fornecer *insights* sobre como os modelos de

IA podem ser aprimorados para uma atuação mais precisa e personalizada (RIBEIRO; SINGH; GUESTRIN, 2016a).

Este trabalho está organizado da seguinte maneira: a Seção 2 aborda os conceitos fundamentais sobre obesidade e os classificadores de padrões utilizados na pesquisa; a Seção 3 revisa estudos relacionados que embasam esta proposta; a Seção 4 detalha os procedimentos metodológicos empregados, incluindo o tratamento dos dados e a condução dos experimentos computacionais; a Seção 5 apresenta e discute os resultados obtidos pelos modelos avaliados; e, por fim, a Seção 6 traz as principais considerações do estudo e sugestões de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Para a compreensão do leitor, esta seção apresenta os principais conceitos relacionados à obesidade (definição, classificação e impactos na saúde) e IA. Por fim, o trabalho discutirá as definições e tipos de métodos XAI, enfatizando sua importância para a transparência e confiança dos modelos.

### 2.1 Obesidade

O nível de obesidade pode ser determinado por meio do IMC, sendo calculado ao dividir o valor de massa corporal (em quilogramas) pela valor da altura (em metros) ao quadrado. O sobrepeso e a obesidade são caracterizados por um acúmulo excessivo de gordura corporal que pode comprometer a saúde (OMS, 2024a). Conforme as categorias estabelecidas pela Organização Mundial da Saúde (OMS), o IMC classifica os indivíduos em cinco categorias. Como mostrado no Quadro 1, a classificação normal refere-se ao IMC inferior a 25 kg/m<sup>2</sup>, sobrepeso com IMC a partir de 25 kg/m<sup>2</sup>, obesidade grau I com IMC a partir de 30 kg/m<sup>2</sup>, obesidade grau II com IMC a partir de 35 kg/m<sup>2</sup> e obesidade grau III quando o IMC é igual ou superior a 40 kg/m<sup>2</sup> (ABBADÉ, 2021).

Quadro 1 – Níveis de obesidade com base no IMC.

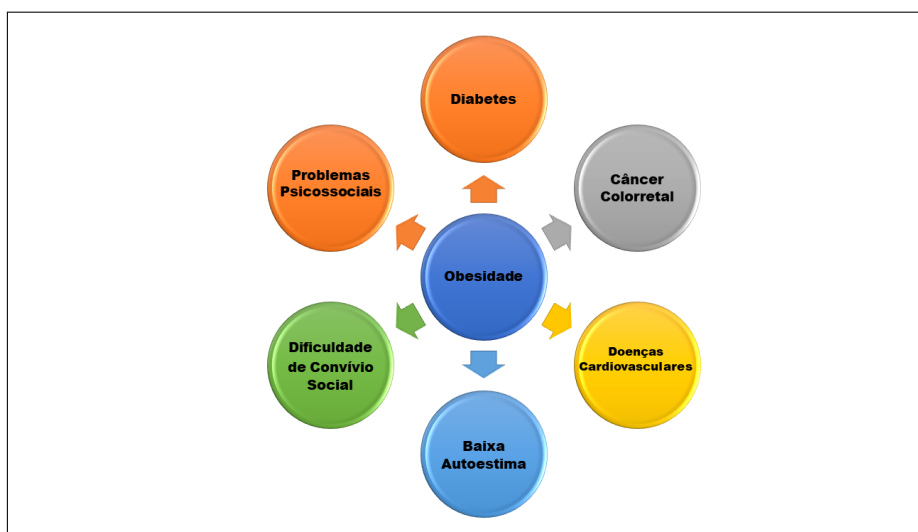
IMC (Kg/m <sup>2</sup> )	Classificação
< 25	Peso Normal
>= 25	Sobrepeso
>= 30	Obesidade I
>= 35	Obesidade II
>= 40	Obesidade III

Fonte: Adaptado de Abbade (2021)

A obesidade, tanto em crianças quanto em adultos, está associada a diversos riscos à saúde que podem levar a problemas imediatos e de longo prazo (ver Figura 1), como diabetes, câncer colorretal, doenças cardiovasculares, baixa autoestima, problemas psicossociais, além da

dificuldades relacionadas ao convívio social (SANTOS; ROCHA; DIAS, 2020). A obesidade gera um impacto significativo no sistema de saúde, refletido no aumento das internações hospitalares relacionadas a essa condição. Muitas dessas hospitalizações são decorrentes do agravamento de comorbidades associadas, enquanto outras ocorrem programadamente, como nos casos de cirurgias bariátricas (SAUT et al., 2024)

Figura 1 – Riscos a saúde causados pela obesidade.



Fonte: Adaptado de (SANTOS; ROCHA; DIAS, 2020)

A OMS (2022) prevê que, até 2025, cerca de 167 milhões de pessoas, incluindo adultos e crianças, terão sua saúde prejudicada devido ao sobrepeso ou à obesidade. De acordo com a OPAS (2024), em um novo estudo publicado pela *Lancet* com dados de 2022, mais de um bilhão de pessoas no mundo vivem com obesidade. Esse estudo mostra também que a obesidade desde 1990 está em uma crescente e que quadruplicou o número de crianças e adolescentes com obesidade na faixa de 5 a 19 anos. O excesso de peso na infância e adolescência afeta a saúde imediata de crianças e adolescentes e está associado a um maior risco e ao aparecimento mais precoce de algumas doenças não transmissíveis, tais como diabetes tipo 2 e doenças cardiovasculares (OMS, 2024b).

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), no Brasil, 41 milhões de indivíduos com 18 anos ou mais estavam obesos, representando uma em cada quatro pessoas nesta faixa etária. O excesso de peso afetava 60,3% da população com 18 anos ou mais, o que representa cerca de 96 milhões de pessoas. Dentre esses indivíduos, 62,6% eram mulheres e 57,5% eram homens (IBGE, 2023).

A Organização Mundial de Saúde (OMS) relata que a chave para prevenir a obesidade é agir precocemente, de preferência antes da concepção. Uma boa nutrição durante a gravidez, seguida de amamentação exclusiva até aos seis meses e continuada até aos dois anos ou mais, é melhor para todos os bebês e crianças pequenas (OMS, 2022).

O IMC é amplamente utilizado como indicador de obesidade, mas apresenta limitações significativas, como a incapacidade de distinguir entre massa gorda e massa magra, e a falta de sensibilidade para identificar variações na distribuição de gordura corporal. Essas limitações podem levar a diagnósticos imprecisos e à subestimação de riscos à saúde em indivíduos com composição corporal atípica, como atletas ou idosos ([BRAY et al., 2017](#)).

A medida da cintura é utilizada para avaliar a gordura abdominal e o risco de doenças cardiovasculares. No entanto, esse método não deve ser usado isoladamente para prever o excesso de gordura corporal, sendo recomendada sua associação com outros indicadores antropométricos para uma avaliação mais precisa. A bioimpedância elétrica, que estima a composição corporal com base na resistência elétrica dos tecidos, pode ser influenciada por fatores como o estado de hidratação, a temperatura corporal e a distribuição da gordura, o que pode afetar a precisão de sua estimativa ([SANT'ANNA et al., 2009](#)).

## 2.2 Inteligência Artificial

A IA refere-se à ciência e engenharia de criar máquinas capazes de simular a inteligência humana, realizando tarefas como tomada de decisão, compreensão da linguagem e resolução de problemas, para imitar o raciocínio humano e agir de maneira inteligente para resolver situações complexas ([HUNTS, 2023](#); [MONTEREI, 2023](#)).

A seguir são descritas os subcampos da IA, esta por sua vez compreende diversos campos de estudo, tais como Processamento de Linguagem Natural (NLP), Lógica *Fuzzy* e Visão Computacional. Dentro desse universo, destacam-se o Aprendizado de Máquina (AM), que permite aos sistemas aprenderem a partir de dados e experiências sem serem explicitamente programados, e o Aprendizado Profundo (AP), um subcampo que utiliza redes neurais profundas para resolver problemas complexos de forma autônoma ([RUSSELL; NORVIG, 2020](#)).

### 2.2.1 *Aprendizado de Máquina*

O AM é um subcampo da IA que envolve algoritmos capazes de aprender com experiências, extraíndo padrões de dados sem receber instruções explícitas, o que permite resolver tarefas realizadas por humanos de forma intuitiva e subjetiva. Um programa de computador está aprendendo se seu desempenho em uma classe de tarefas, medida por um critério de avaliação, melhora à medida que acumula experiência ([COZMAN; KAUFMAN, 2022](#)).

Existem três principais métodos de aprendizagem no campo do AM: Aprendizado Supervisionado, Aprendizado Não Supervisionado e Aprendizado por Reforço. A decisão sobre qual técnica utilizar depende do problema a ser resolvido e dos dados disponíveis. No Aprendizado Supervisionado, o modelo é treinado com dados que já estão rotulados, enquanto no Aprendizado Não Supervisionado, ele identificará padrões em dados que não possuem rótulos. Já

no Aprendizado por Reforço, um agente aprende a tomar decisões com base em recompensas e punições. Além disso, essas abordagens podem ser consideradas parte do aprendizado tradicional, que inclui tanto o supervisionado quanto o não supervisionado, ou ainda de métodos mais avançados como redes neurais e AP (COSTA; MOURA JÚNIOR, 2023).

No AM, existem muitos classificadores que desempenham papéis cruciais na tarefa de categorizar e rotular dados. Para Prati et al. (2008), os classificadores são algoritmos projetados para prever a classe ou rótulo de novos dados com base em padrões aprendidos a partir de um conjunto de exemplos de treinamento. Cada exemplo é descrito por um conjunto de atributos sendo associado a um rótulo denominado classe, que representa a classificação verdadeira do caso. O classificador aprende a partir desses exemplos para distinguir entre diferentes classes e fazer previsões sobre novos casos com base nos atributos fornecidos. O conjunto de exemplos de treinamento é a amostra utilizada para ensinar o classificador a identificar e categorizar os dados.

Os problemas de classificação, uma categoria fundamental em AM, são divididos em classificação de rótulo único (*Single-label*) e de rótulo múltiplo (*Multi-label*). Dentro da classificação de rótulo único, as subcategorias mais tradicionais são a classificação binária e a multiclasse. A classificação binária lida com a categorização de dados em apenas duas classes distintas, sendo valorizada por sua simplicidade e aplicabilidade em diversas situações. Já a classificação multiclasse envolve a atribuição de um dado a uma única categoria entre mais de duas opções, e, embora existam algoritmos dedicados, é frequentemente abordada através da decomposição em múltiplos subproblemas de classificação binária para maior eficiência (ER; VENKATESAN; WANG, 2016). Por exemplo, pode-se modelar o problema de estimação de nível de obesidade em classificação binária (obeso e não-obeso), como classificação multiclasse (peso insuficiente, normal, sobrepeso e obesidade) e como problema de classificação multirótulo (diabético e obeso).

Dentre os algoritmos de AM, este estudo avaliou o desempenho de *Extra Trees Classifier* (ETC), *Random Forest* (RF), *Gradient Boosting* (GB), *Multi-Layer Perceptron* (MLP) e *Support Vector Machine* (SVM). O ETC é um algoritmo de AM supervisionado que se fundamenta na utilização do método *ensemble* de *Decision Tree* (DT), podendo ser empregado tanto em tarefas de classificação quanto de regressão (MACEDO et al., 2023). Outro *ensemble* de DT é o RF. A partir de um conjunto de dados de treinamento, o RF combina os resultados de várias DT para melhorar o desempenho preditivo e a confiança na tomada de decisão, em comparação com o uso de um único classificador. Esse processo de combinação de vários modelos visa aumentar a precisão e a robustez das previsões (BARROS, 2019).

Já o GB é um algoritmo que eleva a precisão dos modelos ao combinar múltiplos modelos de baixa precisão, resultando em um modelo final altamente preciso (BAHAD; SAXENA, 2020). A MLP é uma rede neural *feedforward* composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio em uma camada está totalmente conectado a todos os neurônios da camada seguinte. A MLP propaga sinais de entrada para frente



através das camadas, aplicando funções de ativação para gerar a saída final. Essa estrutura permite a modelagem de relações complexas nos dados, tornando-o eficaz para tarefas de classificação (BRITO; OLIVEIRA, 2020).

Por fim, o SVM é um algoritmo de AM baseado nos princípios da minimização do risco estrutural da teoria do aprendizado estatístico, buscando equilibrar o erro de treinamento e a complexidade do modelo (SOUSA; BRITO; XIMENES, 2021). Dentro deste contexto, o *Support Vector Classification* (SVC) surge como uma aplicação específica da SVM voltada para problemas de classificação e identificação de *outliers*, mostrando-se particularmente eficaz em tarefas como a detecção de rostos em imagens e o reconhecimento de texto (AMARAL et al., 2021).

### 2.2.2 Aprendizado Profundo

O AP, também conhecido como *deep learning*, é uma subárea do AM que se destaca por sua capacidade de extrair e aprender automaticamente representações complexas a partir de representações mais simples dos dados. Em vez de depender do conhecimento prévio do desenvolvedor para pré-processar os dados, o AP permite que o algoritmo aprenda a mapear os dados de entrada e criar representações intermediárias para facilitar esse mapeamento. Isso é especialmente útil em tarefas complexas, onde aprender a representação dos dados é tão desafiador quanto aprender a própria função. As representações simples são organizadas em camadas, com diferentes níveis que se complementam para formar informações mais complexas (BOCHIE et al., 2020).

O AP diferencia-se do AM convencional por sua organização em camadas, permitindo que o modelo extraia representações sofisticadas de dados de forma autônoma, como a detecção de bordas e padrões complexos em imagens. Enquanto o AM tradicional depende de características definidas manualmente e é mais eficiente com recursos limitados, o AP se destaca por lidar com dados não estruturados e aprender atributos complexos automaticamente, embora exija maior poder computacional e muitos dados (WILLINGHOEFER, 2023).

As redes neurais *feedforward*, como as MLP, são formadas por camadas de nós interligados que processam informações sequencialmente, sem realimentação entre as camadas. Essas redes armazenam conhecimento baseado em dados experimentais, tornando-o acessível para uso em novas previsões. As MLP, em particular, se destacam por sua estrutura composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, onde cada nó da camada está conectado a todos os nós da próxima camada por meio de pesos (TISSOT; CAMARGO; POZO, 2012).

A criação do algoritmo de retropropagação foi um marco crucial para o avanço das redes neurais *feedforward*, pois permitiu que suas capacidades fossem amplamente reconhecidas, atraindo o interesse de pesquisadores de várias áreas do conhecimento. No entanto, uma desvan-



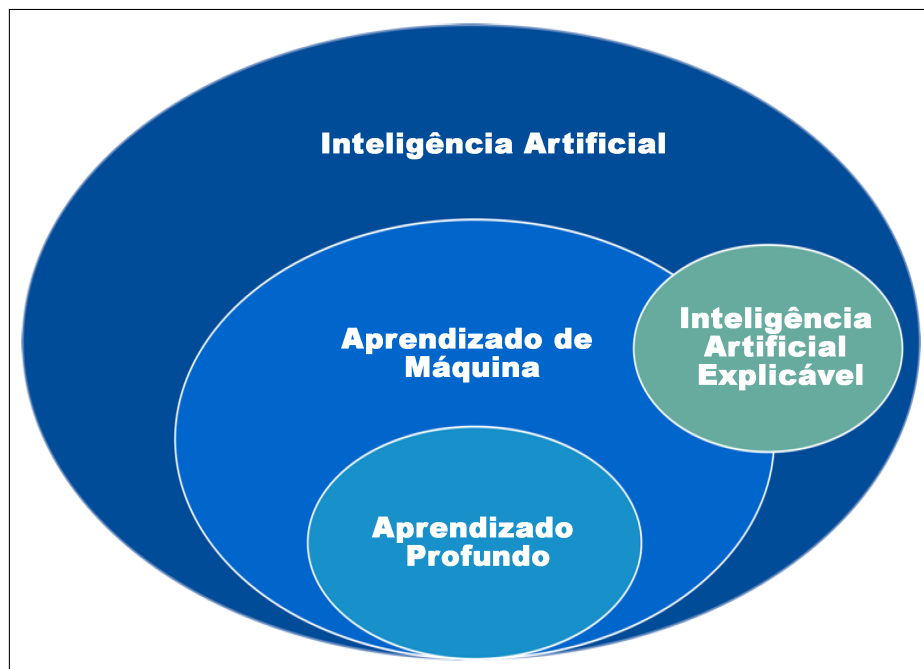
tagem desse algoritmo é sua tendência a se prender em mínimos locais ao lidar com problemas que envolvem funções não diferenciáveis ou multimodais, o que pode comprometer a eficácia da rede em determinadas situações (FILHO et al., 2011).

### 2.3 Inteligência Artificial Explicável

A XAI é uma área da IA que desenvolve ferramentas, técnicas e algoritmos capazes de produzir explicações claras, interpretáveis e intuitivas para humanos, permitindo que as previsões tomadas pela IA sejam mais compreensíveis (DAS; RAD, 2020).

Os métodos de XAI podem ser classificados em dois grupos principais: os intrínsecos, os quais são naturalmente compreensíveis, como a regressão linear e as árvores de decisão, e os *post hoc*, que oferecem explicações após a aplicação do modelo, valendo-se de técnicas como *SHapley Additive exPlanations* (SHAP) e *Grad-CAM*. A XAI é crucial para assegurar que diferentes profissionais, desde cientistas de dados até médicos, consigam entender e confiar nas previsões geradas pela IA (ZHANG; WENG; LUND, 2022). A Figura 2 ilustra a relação entre IA, AM, AP e XAI.

Figura 2 – Relação entre inteligência artificial, aprendizado de máquina, aprendizado profundo e inteligência artificial explicável.



Fonte: Adaptado de (ZHANG; WENG; LUND, 2022)

A explicabilidade é crucial em áreas como defesa, saúde, justiça e veículos autônomos, onde a confiança e a transparência são fundamentais. Além disso, a IA já vem sendo amplamente aplicada em diversas áreas, incluindo ciência, negócios, finanças e redes sociais, lidando com

diferentes tipos de dados, como texto, imagem, áudio e vídeo, e resolvendo problemas complexos de classificação, regressão e otimização (GOHEL; SINGH; MOHANTY, 2021).

Modelos de AM do tipo caixa preta ou caixa fechada estão sendo amplamente utilizados em domínios críticos, como justiça socioeconômica, ciência forense cibernética e justiça criminal. No entanto, esses modelos enfrentam desafios em ganhar a confiança do público devido à sua falta de transparência e responsabilidade, o que pode levar a sérias consequências, como decisões judiciais controversas em casos criminais (GOHEL; SINGH; MOHANTY, 2021). Esses modelos também, geralmente oferecem alto desempenho, mas sacrificam a explicabilidade, já que suas operações internas são invisíveis e difíceis de compreender, causando ambiguidade. Em contraste, os modelos de caixa branca ou caixa de vidro são transparentes, tornando todas as informações necessárias prontamente disponíveis, embora isso possa comprometer o desempenho. Esses modelos são frequentemente associados aos métodos de XAI, como o *Explainable Boosting Machine* e o *DeepLIFT*, que buscam oferecer soluções mais interpretáveis (DUELL, 2021).

A interpretabilidade diz respeito à habilidade de compreender as respostas de um modelo de AM de maneira que seja acessível a um ser humano, ajudando a identificar conexões de causa e efeito entre as entradas e saídas do sistema. Por outro lado, a explicabilidade envolve a compreensão dos mecanismos internos e dos processos que se desenrolam no interior do modelo durante as fases de treinamento e decisão. Apesar de ambos os conceitos estarem interligados, a explicabilidade oferece uma visão mais detalhada sobre as operações internas do modelo, enquanto a interpretabilidade se concentra na intuição e na clareza das saídas produzidas (LINARDATOS; PAPASTEFANOPOULOS; KOTSIANTIS, 2021). Enquanto a explicabilidade se refere à habilidade de elucidar as previsões produzidas por esses modelos, a interpretabilidade se concentra em como o modelo foi representado durante a fase de treinamento (ZHANG; WENG; LUND, 2022).

Sobre a importância das explicações, Das e Rad (2020) afirmam ser crucial que as explicações tornem o algoritmo de IA mais compreensível, pois isso aumenta a confiança dos usuários nas decisões tomadas e assegura que essas decisões sejam justas e imparciais. Portanto, para garantir transparência e confiança no processo de tomada de decisão em AM, é fundamental fornecer uma explicação clara ou uma solução que permita entender o funcionamento do modelo. Entretanto, embora as explicações no contexto de XAI possam melhorar a compreensão dos usuários sobre sistemas de IA, os benefícios em termos de confiança e aceitação dos usuários são variados, apontando possíveis discrepâncias entre as explicações algorítmicas e as necessidades dos usuários finais (LIAO; GRUEN; MILLER, 2020).

A interpretabilidade dos modelos de AM é essencial para garantir que decisões tomadas por esses sistemas não reforcem vieses discriminatórios presentes na sociedade. Ao tornar os modelos mais compreensíveis, é possível identificar deficiências e assegurar que requisitos éticos e legais sejam respeitados. Além disso, a interpretabilidade permite que humanos analisem as limitações dos modelos, promovendo maior confiabilidade nas decisões automatizadas (VIEIRA,

2023). A interpretabilidade, no contexto de sistemas de AM, pode ser entendida como a capacidade de explicar ou apresentar os resultados de maneira que sejam compreensíveis para um ser humano, sendo essencial para superar a incompletude na formalização do problema, dificultando a otimização e avaliação dos modelos (DOSHI-VELEZ; KIM, 2017).

A interpretabilidade dos modelos pode ser analisada de maneira local ou global. A abordagem local concentra-se no desempenho do modelo em relação a uma observação específica, simulando seu funcionamento em amostras similares. Por outro lado, a interpretabilidade global permite entender o comportamento geral do modelo, contemplando uma análise de todas as situações possíveis. A análise local tende a ser mais direta e viável, enquanto a abordagem global se mostra mais complexa, por requerer uma visão geral do modelo em sua totalidade (OLIVEIRA, 2020). Embora muitos estudos se dediquem a explicações locais, existe uma demanda crescente por métodos que proporcionem uma visão abrangente do modelo, pois a compreensão local pode não ser adequada para instaurar a confiança necessária antes de sua aplicação (VIEIRA, 2023).

Dentre os métodos XAI para interpretabilidade local, o LIME (RIBEIRO; SINGH; GUESTRIIN, 2016b), é uma técnica que utiliza modelos substitutos locais para explicar previsões individuais de modelos de AM caixa-preta. Em vez de treinar um modelo substituto global, o LIME foca em aproximar as previsões do modelo original para uma instância específica, gerando um novo conjunto de dados perturbado e ponderando as amostras pela proximidade à observação de interesse, permitindo maior compreensão local das decisões do modelo (MOLNAR, 2022).

O LIME se sobressai por sua abordagem agnóstica, que possibilita a explicação de qualquer modelo de aprendizado de máquina sem necessitar de acesso a seus mecanismos internos. Ele produz explicações que são facilmente interpretáveis, como a identificação de palavras ou elementos que influenciam as previsões, mesmo quando o modelo subjacente utiliza representações mais elaboradas e que não são intuitivas (MARY, 2019).

No contexto de dados tabulares, o LIME utiliza um *kernel* de suavização exponencial para definir a vizinhança ao redor de uma instância de interesse, ponderando as amostras conforme sua proximidade. No entanto, a escolha do *kernel* e sua largura pode ser problemática, especialmente em espaços de alta dimensão, pois a definição de uma vizinhança significativa e a comparação de distâncias entre diferentes recursos são desafios ainda não resolvidos de maneira ideal (MOLNAR, 2022).

*SHapley Additive exPlanations* (SHAP) (LUNDBERG; LEE, 2017) é um método de XAI baseado na teoria dos valores de *Shapley*, que fornece explicações locais para modelos de AM ao atribuir um valor de contribuição para cada característica, refletindo seu impacto na previsão de uma instância específica. Este método é amplamente reconhecido por sua fundamentação teórica robusta, embora seus resultados possam ser afetados por colinearidade entre características e dependência do modelo subjacente (SALIH et al., 2024).

### 3 TRABALHOS RELACIONADOS

Nesta seção, apresenta-se a seleção de trabalhos relacionados ao tema deste estudo. As pesquisas foram realizadas no Google Acadêmico em dezembro de 2024, utilizando *strings* de busca específicas e filtros temporais que consideraram publicações entre 2020 e 2024. Conforme descrito na subseção 4.1, o conjunto de dados utilizado neste trabalho foi o “*Estimation of Obesity Levels Based On Eating Habits and Physical Condition*” (PALECHOR; MANOTAS, 2019a). Por isso, foi adotada a seguinte *string* de busca: (“*Estimation of obesity levels UCI dataset*”OR “*Obesity Levels Dataset*”OR “*UCI Obesity Dataset*”) AND “*Kaggle*”AND (“*Machine Learning*”OR “*Deep Learning*”), voltada para encontrar trabalhos que utilizassem o mesmo conjunto de dados adotado neste estudo. Essa busca resultou em cinco trabalhos relevantes.

Quadro 2 – Seleção dos Trabalhos Relacionados

Tipo de busca	Trabalho selecionado	Mesmo Conjunto de Dados?	Algoritmos de IA utilizados	Método XAI
Google Acadêmico	Exame e avaliação de fatores de risco de obesidade com inteligência artificial explicável.	Sim	DT, RF, NB, KNN e XGBoost.	SHAP, InterpretML
Google Acadêmico	IA explicável para suporte à decisão no diagnóstico de comorbidades de obesidade.	Não	MLP, XGBoost, LR, KNN, RF, DT e Linear SVM.	SHAP
Google Acadêmico	Compreendendo o aprendizado de máquina e sua aplicação na estimativa de obesidade por IA explicável.	Sim	LP, LS, KNC, ETC, Linear SVC, RF, BC, DT, LDA, SVC, NuSVC, LR, XGBoost e LGBM.	SHAP
Google Acadêmico	Sistema de suporte à decisão clínica baseado em IA explicável para análise de comorbidade de obesidade.	Não	RF, DT, XGBoost, MLP, SVM, Linear SVM, KNN e LR.	SHAP
Google Acadêmico	Prevendo níveis de obesidade com alta precisão: insights de um modelo de aprendizado de máquina CatBoost.	Sim	LR, KNN, RF e NB.	-
Orientador	Software de estimativa do nível de obesidade baseado em árvores de decisão.	Não	DT, NB e LR.	-
Orientador	Estimativa dos Níveis de Obesidade com Base em Habitos Alimentares e Condição Física Atraves de Técnicas de Aprendizado de Máquina.	Sim	RF e SVM.	-
Orientador	Usando aprendizado de máquina explicável e dados do Fit-bit para investigar preditores de obesidade em adolescentes.	Não	EBM, LR e GBT.	-
Note: NB = Naive Bayes; KNN = K-Nearest Neighbors; XGBoost = Extreme Gradient Boosting; LR = Logistic Regression; Linear SVM = Linear Support Vector Machine; LP = Label Propagation; LS = Label Spreading; KNC = K-Nearest Neighbors Classifier; Linear SVC = Linear Support Vector Classifier; BC = Bagging Classifier; LDA = Linear Discriminant Analysis; NuSVC = Nu Support Vector Classifier; LGBM = Light Gradient Boosting Machine; EBM = Explainable Boosting Machine; GBT = Gradient Boosted Trees.				

Fonte: Elaborado pelo autor

Por fim, foram considerados trabalhos recomendados pelos orientadores, que não necessitaram de pesquisa direta, mas contribuíram significativamente para a fundamentação deste estudo. O Quadro 2 contém as informações principais desses trabalhos. A subseção 3.1 apresenta análise dos trabalhos selecionados e relacionados com este estudo, tanto aqueles que utilizaram o mesmo conjunto de dados deste estudo como aqueles principais estudos que se basearam em técnicas XAI para previsão de riscos ou níveis de obesidade além do conjunto de dados.

### 3.1 Análise dos trabalhos relacionados

Özkurt (2024) avaliou a eficácia de algoritmos de AM, incluindo classificação por DT, RF, *Naive Bayes (NB)*, *K-Nearest Neighbors (KNN)* e *Extreme Gradient Boosting (XGBoost)*. O algoritmo *XGBoost* exibiu a maior precisão 92,00%, *recall* e valores de *F1-score*. Além disso, técnicas de XAI, como SHAP e *InterpretML*, foram empregadas para compreender os efeitos dos parâmetros de obesidade e determinar quais parâmetros têm um impacto maior. Os resultados demonstraram que o algoritmo *XGBoost* supera outros algoritmos na detecção da obesidade, e as técnicas de XAI desempenham um papel crucial na compreensão dos parâmetros de obesidade, destacando que o histórico familiar de obesidade e fatores como consumo frequente de vegetais e o hábito de comer entre as refeições tem efeitos mais significativos.

Mahadi et al. (2024) tendo como objetivo prever a obesidade utilizando técnicas de AM e XAI, compara a performance de diversos algoritmos, incluindo DT, SVM, NB, KNN, RF e *XGBoost*. Os resultados indicam que o modelo *XGBoost* demonstrou o melhor desempenho, alcançando 95,98% de precisão, 95,98% de *recall* e 96,00% de *F1-score*. A pesquisa também empregou técnicas de XAI, como SHAP, para entender as contribuições dos parâmetros de entrada e identificar as características mais influentes na previsão da obesidade, como se o indivíduo monitora sua ingestão diária de calorias e a frequência do consumo de álcool, entre outros. As implicações do estudo apontam para o potencial de usar esses modelos preditivos e suas explicações para auxiliar na avaliação de risco e no desenvolvimento de estratégias de prevenção e gerenciamento da obesidade.

Maulana et al. (2024) em seu estudo tem o objetivo de prever a obesidade a partir de características demográficas, de estilo de vida e de saúde. O modelo foi comparado a outros algoritmos, como *Logistic Regression (LR)*, KNN, RF e Bayes Ingênuos, e demonstrou o melhor desempenho, com precisão de 95,98%, *recall* de 95,98% e pontuação F1 de 96,00%. A análise de importância de características destacou peso, altura e gênero como os fatores mais influentes na previsão de obesidade, seguidos por hábitos alimentares e atividade física. Com alta precisão e capacidade de lidar com variáveis categóricas, o modelo *CatBoost* oferece uma ferramenta valiosa para avaliação de risco de obesidade, suporte à tomada de decisões clínicas e desenvolvimento de estratégias de prevenção e gerenciamento da obesidade.

Lopes et al. (2021) em seu estudo, têm como objetivo principal avaliar o desempenho de classificadores de AM, como RF e SVM, na estimativa dos níveis de obesidade a partir de dados de hábitos alimentares e condição física. Os resultados preliminares, sob validação cruzada e *Hold-Out*, indicaram uma acurácia média de aproximadamente 87,84% para SVM e 95,18% para RF. Os autores observaram que sua abordagem se mostrou mais eficaz em reconhecer casos de sobrepeso e obesidade, um avanço em relação a esses casos mais críticos. As implicações do estudo sugerem que as técnicas de AM podem fornecer um suporte valioso na elaboração de planos nutricionais e no apoio a uma melhor classificação dos níveis de obesidade.

### 3.2 Trabalhos que não usam o mesmo Conjunto de Dados e não usam XAI

[Manotas et al. \(2019\)](#) tem como principal objetivo desenvolver um *software* para a estimativa do nível de obesidade, aplicando metodologias de *data mining* ao *dataset* "*Estimation of obesity levels based on eating habits and physical condition*". Os autores aplicaram a metodologia *SEMMA* para selecionar, explorar e modelar os dados, testando três algoritmos: DT, NB e LR. Os resultados confirmaram que a técnica de DT obteve a melhor precisão, alcançando 97,4%. As implicações do estudo apontam para a criação de um *software* capaz de auxiliar na previsão e classificação dos níveis de obesidade, oferecendo uma ferramenta prática para a análise e o combate a essa condição de saúde global.

[Kiss et al. \(2024\)](#) investiga os preditores da obesidade em adolescentes, utilizando dados de dispositivos Fitbit e modelos de AM explicáveis. Analisando dados de 2971 participantes, os autores empregaram modelos de AM do tipo "caixa de vidro", como Explainable Boosting Machines (EBM) para identificar os fatores mais influentes. Os resultados mostraram que identificar-se como parte de grupos raciais ou étnicos minoritários, baixa renda familiar, horário de dormir mais tardio, curta duração do sono, variabilidade no horário do sono, baixa contagem diária de passos e altas taxas de frequência cardíaca são preditores-chave da obesidade, com uma AUC média de 0,726. As implicações do estudo destacam a importância do sono inadequado, da inatividade física e das disparidades socioeconômicas no risco de obesidade, além de demonstrar a aplicabilidade clínica de tecnologia vestível (do inglês *wearable technology*), para monitoramento contínuo em adolescentes, informando intervenções e estratégias de tratamento.

### 3.3 Diferencial deste trabalho para os demais

Este trabalho aborda um tema ainda pouco explorado na literatura: a combinação de algoritmos de AM, técnicas XAI e a predição de obesidade em um único estudo. Enquanto a maioria das pesquisas se concentra em apenas um desses aspectos, este trabalho integra essas áreas para oferecer previsões precisas e explicáveis, atendendo à crescente demanda por transparência em sistemas de IA. A aplicação do LIME como técnica de XAI não somente amplia o entendimento sobre os modelos, mas também contribui significativamente para uma área acadêmica em rápida expansão.

Comparando algoritmos como *Extra Trees*, *Gradient Boosting*, *MLP*, *Random Forest* e *SVC* no contexto do conjunto de dados *Estimation of Obesity Levels UCI*, este estudo fornece análises robustas sobre o desempenho e a interpretabilidade de cada modelo. Ao integrar a explicação local do LIME às previsões, ele não apenas avança a aplicação prática de XAI no campo da obesidade, mas também oferece uma contribuição valiosa para a ciência de dados, ao facilitar o uso dessas tecnologias em problemas reais de saúde pública.



## 4 PROCEDIMENTOS METODOLÓGICOS

Em termos de objetivo, este estudo trata-se de uma pesquisa explicativa, pois não se limita à análise dos dados observados, isto é, o desempenho dos modelos de classificação de obesidade, mas também investiga suas por que e como esses modelos reconhecem os níveis de obesidade por meio do método XAI (PARANHOS; RODOLPHO, 2018; WAZLAWICK, 2020). A pesquisa combinou métodos qualitativos, avaliando a qualidade das predições dos modelos de IA por meio do método XAI, e métodos quantitativos, mensurando o desempenho dos modelos (ANDRADE, 2017; PARANHOS; RODOLPHO, 2018).

Este trabalho foi executado em quatro etapas metodológicas, a saber, (i) seleção do conjunto de dados, (ii) pré-processamento dos dados, (iii) avaliação dos modelos de classificação por métricas de desempenho e (iv) desenvolvimento de modelos explicáveis para interpretar as predições. As próximas subseções detalham a execução de cada etapa de pesquisa.

Todos os experimentos computacionais de cada etapa foram realizados utilizando a linguagem *Python*, empregando o ambiente *Google Colab*. As bibliotecas utilizadas incluíram *Pandas* e *NumPy* para manipulação de dados (WANG, 2019). E também *Scikit-learn* para construção e avaliação dos modelos e *Matplotlib/Seaborn* para visualização de resultados (KRAMER, 2016).

### 4.1 Seleção do Conjunto de Dados

O conjunto de dados utilizado neste trabalho é intitulado “*Estimation of Obesity Levels Based On Eating Habits and Physical Condition*” (PALECHOR; MANOTAS, 2019b), obtido a partir do repositório público *UCI Machine Learning*<sup>1</sup>. Ele contém informações sobre hábitos alimentares e condições físicas de indivíduos dos países do México, Peru e Colômbia, visando estimar os níveis de obesidade.

O conjunto de dados é composto por 2.111 registros e 17 atributos, descritos no Quadro 3, sendo 16 atributos de características e um atributo alvo (*NObesity*). Este conjunto de dados é problema de classificação multiclasse que reconhece sete níveis de massa corporal: Peso Insuficiente, Peso Normal, Sobrepeso Nível I, Sobrepeso Nível II, Obesidade Tipo I, Obesidade Tipo II e Obesidade Tipo III (PALECHOR; MANOTAS, 2019a).

Os atributos preditivos incluem variáveis relacionadas a hábitos alimentares, tais como consumo frequente de alimentos calóricos, vegetais e água, e condições física, por exemplo frequência de atividade física e tempo gasto em dispositivos eletrônicos. Todos os atributos são do tipo inteiro, e o conjunto de dados não contém valores ausentes. Desses dados 77% foram gerados sinteticamente utilizando a ferramenta *Weka* e a técnica *Synthetic Minority Over-sampling Technique* (SMOTE), enquanto 23% foram coletados diretamente de usuários

<sup>1</sup>Disponível em <https://archive.ics.uci.edu/>

Quadro 3 – Características do conjunto de dados e suas descrições.

Nº	Nome da Característica	Tipo de Dado	Descrição
1	Gender	Categórico	Gênero do indivíduo
2	Age	Contínuo	Idade do indivíduo
3	Height	Contínuo	Altura do indivíduo
4	Weight	Contínuo	Peso do indivíduo
5	family_history_with_overweight	Binário	Indica se o indivíduo tem um familiar que sofreu ou sofre de sobrepeso
6	FAVC	Binário	Indica se o indivíduo consome frequentemente alimentos com alto teor calórico
7	FCVC	Inteiro	Representa a frequência do consumo de vegetais nas refeições
8	NCP	Contínuo	Indica o número de refeições principais consumidas diariamente
9	CAEC	Categórico	Descreve o hábito do indivíduo de comer entre as refeições
10	SMOKE	Binário	Indica se o indivíduo fuma
11	CH2O	Contínuo	Representa a quantidade de água consumida diariamente
12	SCC	Binário	Indica se o indivíduo monitora sua ingestão diária de calorias
13	FAF	Contínuo	Descreve a frequência de atividade física
14	TUE	Inteiro	Tempo gasto com dispositivos tecnológicos, como celulares, videogames e televisão
15	CALC	Categórico	Indica a frequência do consumo de álcool
16	MTRANS	Categórico	Meio de transporte principal do indivíduo
17	NObeyesdad (Target)	Categórico	Nível de obesidade (Peso insuficiente, Peso normal, Sobrepeso Nível I, Sobrepeso Nível II, Obesidade Tipo I, II e III)

Fonte: Adaptado de (MAULANA et al., 2024)

por meio de uma plataforma web. Essa combinação de dados sintéticos e reais visa garantir a representatividade e a diversidade do conjunto de dados, tornando-o adequado para tarefas de classificação, regressão e *clustering* (PALECHOR; MANOTAS, 2019a).

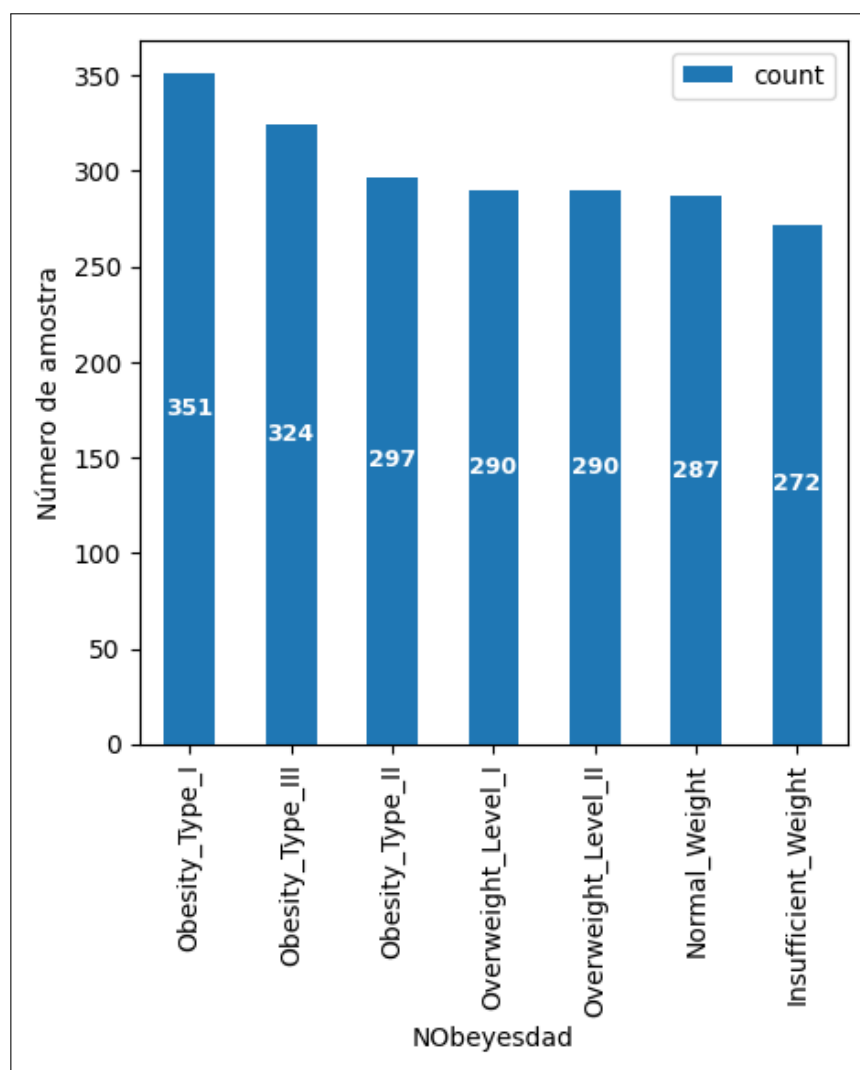
#### 4.2 Pré-processamento de dados

Conforme mostra a Figura 3, foi realizada uma análise da distribuição das amostras por classe no conjunto de dados. Esse procedimento permitiu visualizar a quantidade de amostras em cada categoria do atributo-alvo (*NObeyesdad*), garantindo um melhor entendimento sobre o balanceamento dos dados antes de aplicar qualquer transformação. Para isso, foi gerado um gráfico de barras ilustrando a distribuição das classes (ver Figura 3), auxiliando na identificação de possíveis desequilíbrios que poderiam influenciar o desempenho do modelo.

Embora a documentação do conjunto de dados indique a falta de valores ausentes, foi realizada uma etapa de imputação para garantir que não houvesse inconsistências. Essa abordagem assegura que, mesmo que eventuais falhas tenham ocorrido na importação dos dados ou em seu armazenamento, qualquer valor faltante seja devidamente tratado antes do treinamento do modelo. Para variáveis numéricas, a abordagem adotada foi a imputação da média dos valores

da respectiva coluna, garantindo que a ausência de dados não comprometesse a análise estatística. Já para variáveis categóricas, foi utilizada a estratégia de preenchimento com o valor mais frequente, assegurando que a informação inserida representasse o padrão mais comum daquela categoria. Essas transformações foram aplicadas por meio de uma funcionalidade de imputação da biblioteca *scikit-learn* (HOWLETT; JAIN, 2020).

Figura 3 – Balanceamento



Fonte: Elaborado pelo autor

Além disso, foi aplicada a normalização dos dados numéricos, uma técnica que ajusta as características para que tenham média zero e desvio padrão um. Esse processo ajusta os valores para uma escala padronizada, centralizando-os na média zero e ajustando o desvio padrão para um. Isso é essencial para melhorar a performance de algoritmos que são sensíveis a escalas diferentes, como modelos baseados em gradiente. Por exemplo, variáveis como *Age*, *Height* e *Weight* passaram por essa transformação, garantindo que diferenças nas ordens de grandeza não influenciassem o aprendizado do modelo (WANG, 2019).

Para variáveis categóricas, foi realizada a codificação *one-hot*, uma técnica que trans-

forma categorias em variáveis binárias, criando uma nova coluna para cada categoria. Essa abordagem permite que os algoritmos de aprendizado de máquina possam interpretar informações categóricas sem atribuir relações numéricas incorretas entre elas. Um exemplo claro desse processo ocorreu na variável *CALC*, que originalmente indicava a frequência do consumo de álcool como valores textuais (*Frequently*, *Sometimes*, *no*), mas após o pré-processamento foi transformada em três colunas (*CALC\_Frequently*, *CALC\_Sometimes* e *CALC\_no*), cada uma contendo valores binários (0 ou 1) para representar a presença ou ausência daquela categoria (ALEXANDER, 2022).

Por fim, todos os dados transformados foram organizados em um novo *DataFrame*, garantindo uma estrutura legível e pronta para ser utilizada no treinamento dos modelos. Isso facilita a interpretação dos dados e permite uma melhor visualização das transformações aplicadas, assegurando que todas as etapas do pré-processamento foram corretamente executadas e estruturadas.

#### 4.3 Avaliação dos modelos de classificação

Para garantir a robustez dos modelos de IA e uma melhor generalização para novos dados, o conjunto de dados foi dividido em subconjuntos de treino (80%) e teste (20%). Essa separação reduz o risco de sobreajuste (*overfitting*), permitindo uma avaliação mais precisa do desempenho do modelo. Além disso, foi aplicada a validação cruzada estratificada com 5 *folds* para garantir uma avaliação mais confiável do desempenho dos modelos. O ajuste fino (*fine-tuning*) foi realizado através do uso de um otimizador e de uma função de perda, visando melhorar a convergência e a acurácia do modelo (ALEXANDER, 2022).

Foram treinados modelos de IA para a estimativa dos níveis de obesidade seguindo a abordagem multiclasse, visando diferenciar os indivíduos em categorias específicas de obesidade (RUSSELL; NORVIG, 2010). Para essa abordagem, foram treinados os seguintes algoritmos: *Extra Trees Classifier*, *Random Forest*, *Gradient Boosting*, *SVC* e *MLP*.

O desempenho dos algoritmos foi avaliado por meio de métricas apropriadas para tarefas de classificação. As métricas de avaliação adotadas neste estudo foram selecionadas com o objetivo de proporcionar uma análise abrangente da *performance* dos modelos de classificação (SHAKHOVSKA, 2017). A acurácia foi utilizada como métrica global para quantificar a proporção de previsões corretas em relação ao total de amostras avaliadas (GOODFELLOW; BENGIO; COURVILLE, 2016). Contudo, considerando que a acurácia pode ser limitada em cenários com classes desbalanceadas, foram incorporadas métricas adicionais para uma avaliação mais robusta.

A precisão (*precision*) e a revocação (*recall*) foram calculadas para cada classe individualmente. A precisão indica a proporção de previsões corretas dentro de todas as amostras classificadas como pertencentes a uma determinada classe, enquanto a revocação mede a capacidade do modelo de identificar corretamente todas as ocorrências de uma classe específica

([MOHSENI, 2020](#)). A *F1-score*, que representa a média harmônica entre precisão e revocação, foi utilizada para fornecer um equilíbrio entre essas duas métricas ([SHAKHOVSKA, 2017](#)).

Para uma análise mais aprofundada do desempenho do modelo na diferenciação entre classes, foi utilizada a Curva Característica de Operação do Receptor (ROC, do inglês *Receiver Operating Characteristic*) e a métrica da área sob a curva (AUC, do inglês *Area Under the Curve*), que quantificam a capacidade do modelo de distinguir corretamente entre diferentes classes ([MACKIE et al., 2010](#)). Por fim, a matriz de confusão foi utilizada para visualizar a distribuição das previsões, facilitando a identificação de possíveis padrões de erro e relações entre classes ([GOODFELLOW; BENGIO; COURVILLE, 2016](#)).

#### 4.4 Criação do modelo de explicação

No contexto da estimativa dos níveis de obesidade, a explicabilidade dos modelos preditivos é essencial para que os profissionais de saúde compreendam as razões por trás das classificações realizadas. Para isso, utilizamos o método LIME, que permite explicar as decisões dos modelos de aprendizagem de máquina de forma local e intuitiva ([MOHSENI, 2020](#)).

A implementação dos modelos multiclasse foi realizada com a biblioteca LIME para Python. Nessa abordagem multiclasse, que classifica os indivíduos em categorias mais específicas de obesidade, o LIME foi empregado para analisar a contribuição de cada variável na diferenciação entre classes, tornando o processo de decisão mais transparente ([SHAKHOVSKA, 2017](#)).

A explicação dos modelos foi feita em dois níveis: global e local. A interpretação global avalia a importância geral de cada variável nos modelos, identificando quais fatores mais impactam a classificação dos níveis de obesidade. Já a interpretação local explicou a decisão para instâncias específicas, analisando quais atributos foram determinantes para cada previsão individual ([SAMEK et al., 2017](#)).

A análise de explicabilidade com LIME foi conduzida tanto para o modelo clássico de melhor desempenho, o ETC, quanto para o modelo de AP, levando em consideração os atributos de maior influência nas decisões mais difíceis de cada modelo. Para cada uma das sete classes de peso, cinco amostras foram inicialmente consideradas, e a mais problemática dentre elas foi selecionada para esta análise aprofundada, permitindo-nos investigar os cenários onde os modelos encontraram maior dificuldade. Essa dupla abordagem permitiu comparar não apenas a performance preditiva, mas também a interpretabilidade de diferentes paradigmas de modelagem, oferecendo percepções sobre como cada tipo de arquitetura utiliza as características para chegar a suas conclusões sobre os níveis de obesidade.

## 5 RESULTADOS E DISCUSSÕES

Esta seção está estruturada em duas subseções principais: a primeira apresenta a avaliação do desempenho dos modelos de classificação, com base em métricas quantitativas; a segunda aborda a explicação dos modelos, destacando os fatores que influenciaram as decisões preditivas.

### 5.1 Desempenho dos modelos de IA

A Tabela 1. apresenta o desempenho dos modelos clássico de AM avaliados conforme as métricas de Acurácia, Precisão, *Recall*, *F-Score* e AUC.

Tabela 1 – Desempenho dos classificadores e do modelo de AP nas métricas avaliadas.

Classificador	Acurácia	Precisão	Recall	F1-Score	AUC
Extra Trees Classifier	<b>92,4%</b>	<b>91,6%</b>	<b>93,0%</b>	<b>92,3%</b>	<b>97,0%</b>
Random Forest	91,5%	90,8%	92,8%	91,8%	96,5%
Gradient Boosting	90,6%	90,0%	91,0%	90,5%	96,3%
SVC	88,8%	88,0%	90,0%	89,0%	94,5%
MLP	89,0%	88,0%	89,0%	88,5%	95,0%
Rede Neural Feedforward	<b>95,51%</b>	<b>95,50%</b>	<b>95,39%</b>	<b>95,43%</b>	<b>99,79%</b>

Fonte: Elaborado pelo autor

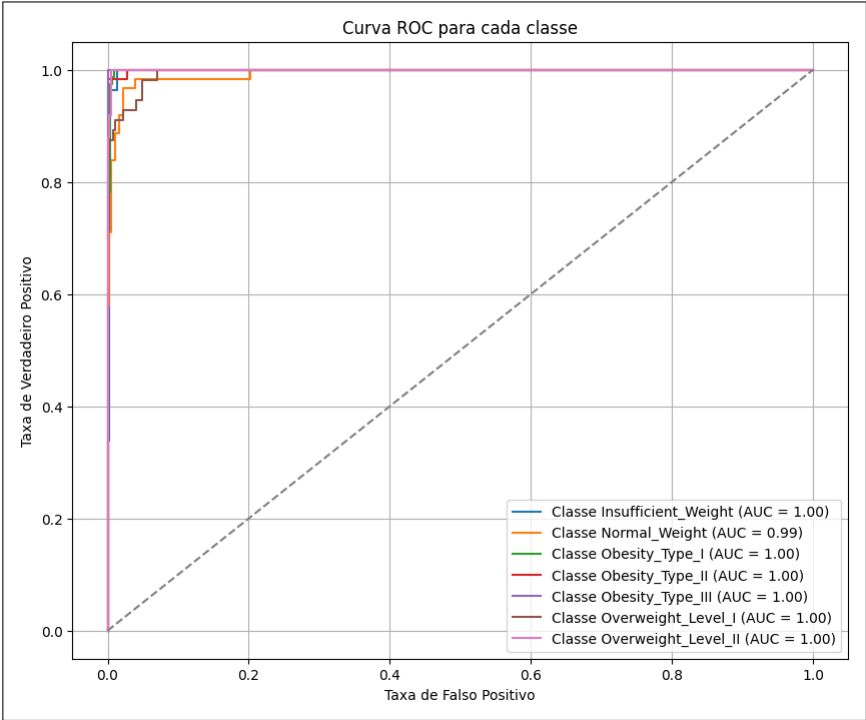
O ETC apresentou o melhor desempenho geral, com Acurácia de 92,4%, AUC de 97,0% e *Recall* de 93,0%, demonstrando alta robustez na distinção entre classes e eficácia na detecção de casos, aspecto crucial em diagnósticos médicos. O RF obteve resultados similares, com ligeira queda no *Recall* (92,8%), indicando possível limitação na identificação de casos menos frequentes. O *Gradient Boosting* também mostrou bom desempenho, embora com leve redução em Precisão (90%) e *Recall* (91%), sugerindo um viés na classificação de algumas classes. O SVC teve a menor Acurácia (88,8%), com desempenho aceitável no *Recall* (90%), porém com *F1-Score* (89%) inferior, refletindo menor equilíbrio entre precisão e sensibilidade. Já o MLP apresentou métricas intermediárias, com Acurácia de 89%, AUC de 95%, e *F1-Score* de 88,5%, evidenciando um desempenho equilibrado, porém ligeiramente inferior aos modelos baseados em árvores.

Os resultados do modelo AP indicaram um desempenho superior em acurácia e AUC quando comparados aos modelos de AM. A acurácia do modelo de AP ficou em torno de (95.51%), com um AUC de (99.79%), demonstrando a capacidade do modelo em aprender representações complexas dos dados. Ademais, pela Figura 4 observa-se que a curva ROC do modelo AP ficou acima aos demais modelos clássicos.

A matriz de confusão desse modelo demonstrou uma classificação bastante precisa, com a maioria das amostras corretamente identificadas em suas respectivas classes. Houve poucos erros de classificação, sendo que classes como *Obesity\_Type\_III* e *Obesity\_Type\_I* apresentaram

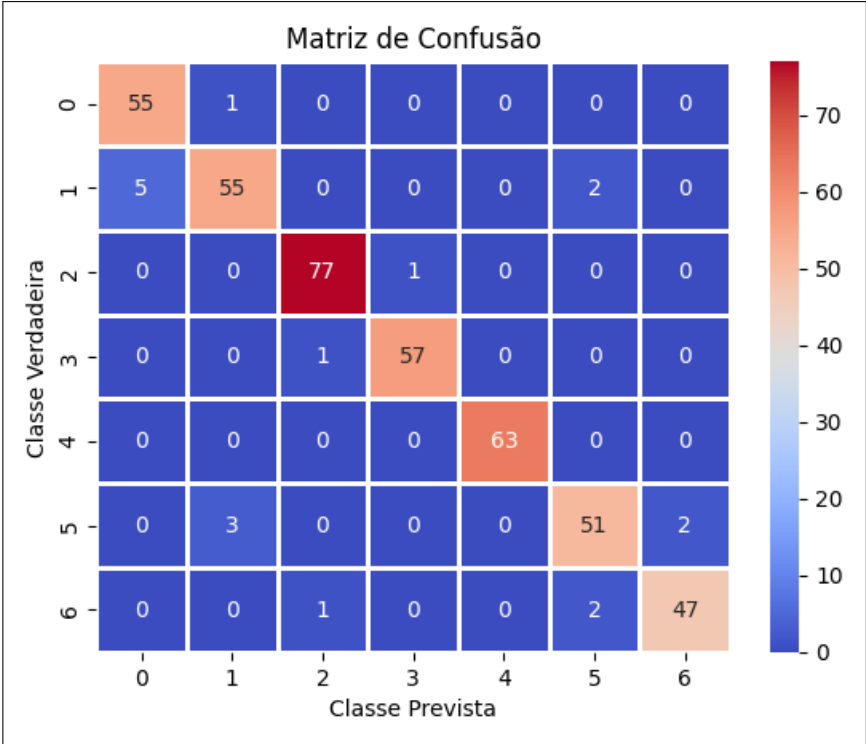
quase nenhum equívoco, evidenciando a capacidade do modelo em distinguir com clareza os diferentes níveis de obesidade, como pode ser vista na Figura 5.

Figura 4 – Curva ROC



Fonte: Elaborado pelo autor

Figura 5 – Matriz de Confusão



Fonte: Elaborado pelo autor



## 5.2 Explicação dos melhores modelos

### 5.2.1 Extra Trees Classifier

Ao aplicar o método LIME sobre as predições do ETC, escolhido por ter apresentado os melhores resultados de desempenho em comparação aos demais classificadores avaliados, e focar nas amostras mais desafiadoras em cada classe, foi possível identificar os atributos de maior influência nas decisões mais difíceis do modelo. Nestes cenários complexos, o peso permaneceu um fator de influência crucial; contudo, para classes como *Obesity\_Type\_I, II e III*, e *Normal\_Weight*, a probabilidade de acerto para a classe real foi notavelmente baixa, indicando que, apesar da correta classificação final, o modelo encontrou ambiguidades significativas, como podemos observar melhor na Tabela 2.

Tabela 2 – Informações das amostras com menor probabilidade de predição do classificador *Extra Trees* explicadas pelo LIME para cada classe.

I	P	A	FCVC	CAEC	G	Pr	MI	%
655	62,89	1,72	2,00	Sometimes	Fem	Insuff_Weight	Peso, CAEC às vezes, Alta MTRANS_Publ.	39%
478	70,00	1,70	1,83	Sometimes	Masc	Normal_Weight	Peso, FCVC Baixa, CAEC às vezes	2%
831	60,00	1,75	2,00	Always	Masc	Overweight_I	Peso Moderado, CAEC, SCC, CALC	0%
119	76,00	1,68	2,00	Sometimes	Fem	Normal_Weight (Erro) / Overweight_Level_II	Peso, Histórico Familiar, CAEC às vezes	–
1330	79,11	1,59	2,00	No	Fem	Obesity_I	Peso, Altura Baixa, CALC, NCP	0%
1586	84,85	1,74	2,00	Always	Masc	Obesity_II	Peso, CAEC Sempre, Gênero, MTRANS	1%
1889	150,00	1,80	2,00	Always	Masc	Obesity_III	Peso Muito Alto, CAEC, CALC	0%
Note: I = Número da Amostra; P = Peso (kg); A = Altura (m); FCVC = Frequência do consumo de vegetais nas refeições; CAEC = Hábito do indivíduo de comer entre as refeições; G = Gênero (Masc/Fem); Pr = Predição; MI = Atributos de maior influência; % = Probabilidade de acerto.								

Fonte: Elaborado pelo autor

A altura (*Height*), por sua vez, continuou a exercer influência moduladora, especialmente quando associada a valores mais baixos, potencializando o impacto do peso por meio do IMC e contribuindo para a dificuldade da distinção entre as categorias de peso. Além disso, a análise destas amostras de fronteira ou de menor confiança destaca a relevância de características comportamentais e de mobilidade na geração dessas complexidades, como evidenciado nos casos onde o modelo falhou ou hesitou, como na amostra 119 da classe *Overweight\_Level\_II*, a única predita incorretamente dentre todas as amostras consideradas.

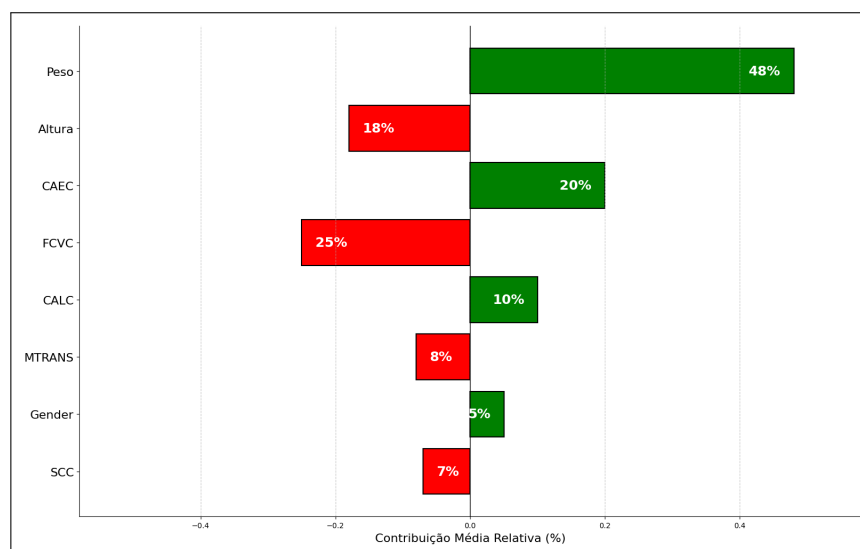
Mesmo nas amostras mais difíceis, algumas tendências nas características dos modelos analisados persistiram. Para as classes *Insufficient\_Weight* e *Normal\_Weight*, notamos que a presença de hábitos alimentares saudáveis, como a alta frequência de consumo de vegetais

(*FCVC=3*) e a ausência de consumo de alimentos fora de hora (*CAEC=No*), combinada a um peso dentro dos padrões adequados, foi crucial para a manutenção das predições corretas, mesmo quando a confiança do modelo era menor. Por outro lado, no contexto dos indivíduos classificados como *Overweight\_Level\_I* e *Overweight\_Level\_II*, a interação entre o peso e padrões de alimentação menos rigorosos (*CAEC=Sometimes* ou *CAEC=Frequently*) desempenhou um papel decisivo, direcionando as predições para essas categorias de sobrepeso, mesmo nos casos em que o classificador demonstrou maior dificuldade ou cometeu erros.

É crucial destacar um aspecto peculiar observado nas predições do modelo para as classes mais desafiadoras: em algumas amostras, como as de *Obesity\_Type\_I* (índice 1330), *Obesity\_Type\_II* (índice 1586) e *Obesity\_Type\_III* (índice 1889), a probabilidade direta de acerto para a classe real foi de 0.00 ou 0.01, mesmo com a predição final estando correta. Esse fenômeno indica que, nesses casos, a decisão do *ExtraTreesClassifier* foi baseada em uma votação unânime (ou quase unânime) das árvores internas, mesmo que a confiança numérica aparente fosse mínima.

O LIME, ao explicar a decisão do modelo em amostras difíceis, revelou que peso elevado (com 48% de contribuição) e hábitos alimentares como o *CAEC* (comer entre refeições, com 20% de contribuição), além do consumo de álcool (*CALC*, com 10%) e o gênero (*Gender*, com 5%), foram os principais fatores que positivamente direcionaram a predição para a classe correta. Por outro lado, variáveis como a frequência de consumo de vegetais (*FCVC*, com 25% de contribuição negativa), a altura (com 18% de contribuição negativa), o tipo de transporte (*MTRANS*, com 8%) e o monitoramento de calorias (*SCC*, com 7%) atuaram negativamente, afastando a amostra de outras classificações (especialmente de pesos mais baixos) e, assim, confirmando a predição da classe real. Na Figura 6, podemos observar as variáveis mais influentes, tanto positivas quanto negativas, confirmando que o *Extra Trees* não se baseia em um único fator, mas sim em uma combinação complexa e articulada de características antropométricas e comportamentais para determinar o risco de obesidade de cada indivíduo.

Figura 6 – Diagrama de força das variáveis mais influentes no modelo Extra Trees.



Fonte: Elaborado pelo autor

### 5.2.2 Aprendizado Profundo

Ao aplicar o método LIME sobre as amostras do classificador de AP, observamos que o peso permaneceu como um fator de influência crucial. Entretanto, ao contrário do ETC onde as probabilidades de acerto para a classe real foram notavelmente baixas em algumas classes, o modelo de AP demonstrou alta confiança em todas as amostras analisadas, mesmo as “menos confiantes” de seu próprio conjunto. Isso indica que, para estas amostras, o modelo de AP não encontrou ambiguidades significativas, validando a sua robustez. A altura (*Height*), por sua vez, continuou a exercer influência moduladora, especialmente quando associada a valores mais baixos, potencializando o impacto do peso por meio do IMC.

Tabela 3 – Informações das amostras com menor probabilidade de predição do modelo de AP explicadas pelo LIME para cada classe.

I	P	A	FCVC	CAEC	G	Pr	MI	%
443	53.00	1.72	2.00	Sometimes	Masc	Insufficient_Weight	Peso, CAEC, MTRANS	98.75
410	66.00	1.83	2.00	Sometimes	Masc	Normal_Weight	Peso, Altura, CAEC, FAF	89.72
1317	97.45	1.77	2.00	Sometimes	Masc	Obesity_Type_I	Peso, Altura, CAEC	99.81
1713	112.32	1.76	1.92	Sometimes	Masc	Obesity_Type_II	Peso, Gênero, CAEC, MTRANS	99.42
1924	137.86	1.79	3.00	Sometimes	Fem	Obesity_Type_III	Peso, Altura, Gênero, MTRANS	99.81
771	75.95	1.72	2.00	Frequently	Fem	Overweight_Level_I	Peso, MTRANS, CAEC, Gênero	93.89
1003	78.00	1.70	3.00	Sometimes	Fem	Overweight_Level_II	Peso, Gênero, CAEC, Idade	98.67

Note: I = Número da Amostra; P = Peso (kg); A = Altura (m); FCVC = Frequência do consumo de vegetais nas refeições; CAEC = Hábito do indivíduo de comer entre as refeições; G = Gênero (Masc/Fem); Pr = Predição; MI = Atributos de maior influência; % = Probabilidade de acerto.

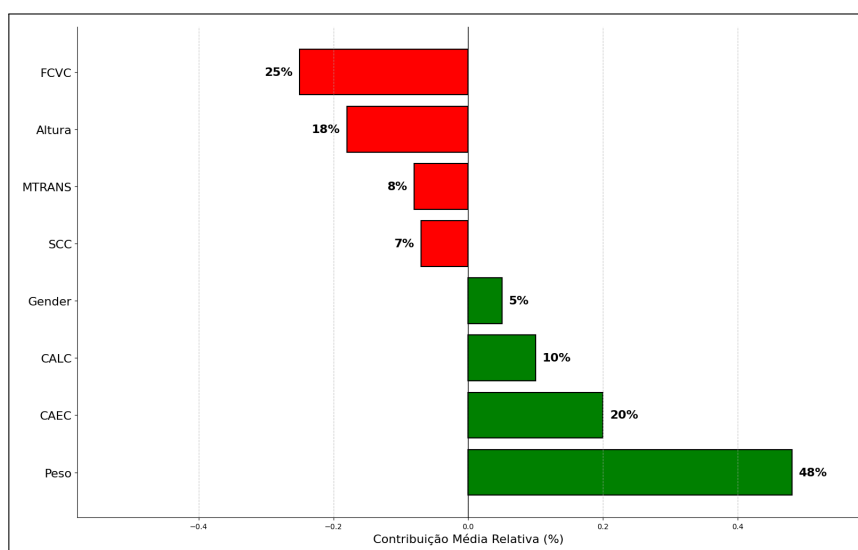
Fonte: Elaborado pelo autor

O LIME, ao explicar a decisão do modelo de AP nas amostras com menor probabilidade

de acerto, revelou um padrão complexo de influências. De forma consistente com o *Extra Trees Classifier*, o peso apresentou-se como o fator mais crucial, contribuindo com 48% para positivamente direcionar a predição para as classes de sobrepeso e obesidade. Hábitos alimentares como o *CAEC* (comer entre as refeições) também se mostraram altamente influentes, com 20% de contribuição positiva, seguidos pelo consumo de álcool (*CALC*), com 10%, e o Gênero (*Gender*), com 5%, que também direcionaram positivamente a predição para a classe correta.

Por outro lado, variáveis como a frequência de consumo de vegetais (*FCVC*), com 25% de contribuição negativa, a altura, com 18% de contribuição negativa, o tipo de transporte (*MTRANS*), com 8%, e o monitoramento de calorias (*SCC*), com 7%, atuaram negativamente, afastando a amostra de outras classificações (especialmente de pesos mais baixos) e, assim, confirmando a predição da classe real, na Figura 7, podemos observar essas variáveis mais influentes. O modelo de AP, portanto, não se baseia em um único fator, mas sim em uma combinação complexa e articulada de características antropométricas e comportamentais para determinar o risco de obesidade de cada indivíduo, de forma similar ao *Extra Trees*.

Figura 7 – Diagrama de força das variáveis mais influentes no modelo de AP.



Fonte: Elaborado pelo autor

Os modelos de IA identificam os níveis de obesidade aprendendo padrões em dados de estilo de vida e hábitos. Principalmente, o IMC, a idade e o nível de atividade física são as variáveis-chave reveladas pela XAI através do método LIME, que conferem transparência e confiança às predições.

## 6 CONCLUSÃO

Este estudo demonstrou a eficácia de modelos de IA na estimativa de níveis de obesidade e a importância da XAI para a interpretabilidade desses sistemas. Os modelos de AM, como ETC e RF, destacaram-se pela alta acurácia (superior a 91.0%), baixo custo computacional e rápido tempo de inferência, configurando-se como soluções robustas para classificação da obesidade. O modelo de AP (Rede Neural *Feedforward*), por sua vez, também exibiu excelente desempenho, com acurácia de 95.51% e AUC-ROC de 99.79%. A aplicação da técnica LIME foi crucial para revelar a influência de variáveis como IMC, idade e nível de atividade física nas predições, o que corrobora o conhecimento científico consolidado na literatura. Dessa forma, o objetivo de aprimorar a interpretabilidade e a confiabilidade dos modelos preditivos na estimativa de níveis de obesidade foi satisfatoriamente alcançado, oferecendo percepções valiosas para o diagnóstico e tratamento.

Para trabalhos futuros, sugere-se a exploração de outras técnicas de XAI, como SHAP, para complementar ou comparar com as análises do LIME, visando uma compreensão ainda mais aprofundada do comportamento dos modelos. Recomenda-se também a validação dos modelos em conjuntos de dados maiores e mais diversos, preferencialmente coletados em cenários clínicos reais, para testar sua generalização e robustez. Além disso, a implementação e avaliação desses modelos em sistemas inteligentes de apoio à decisão clínica é um passo natural, permitindo a validação da sua utilidade prática e o potencial para auxiliar profissionais de saúde na tomada de decisões e no desenvolvimento de intervenções personalizadas no combate à obesidade.

## REFERÊNCIAS

- ABBADE, E. B. Evolução da obesidade e doenças crônicas não transmissíveis nas populações das capitais do Brasil entre 2006 e 2018. *Medicina (Ribeirão Preto)*, v. 54, n. 1, p. e171413, ago. 2021. Disponível em: <https://www.revistas.usp.br/rmrp/article/view/171413>.
- ABESO. *Diretrizes Brasileiras de Obesidade 2016*. Associação Brasileira para o Estudo da Obesidade e da Síndrome Metabólica, 2016. Disponível em: <https://abeso.org.br/wp-content/uploads/2019/12/Diretrizes-Download-Diretrizes-Brasileiras-de-Obesidade-2016.pdf>.
- ALEXANDER, K. *Practical Explainable AI Using Python*. Cham: Springer, 2022. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-95088-1>. ISBN 978-3-030-95088-1.
- AMARAL, R. G. d. et al. Elaboração de algoritmo para aplicação de aprendizado de máquina a fim de reconhecer armadilhas geológicas do tipo domo salino em seções sísmicas. Brasil, 2021.
- ANDRADE, M. M. d. Introdução à metodologia do trabalho científico. 10. *São Paulo: Atlas*, 2017.
- AZAD, M.; KHAN, M. F. K.; EL-GHANY, S. A. Xai-enhanced machine learning for obesity risk classification: A stacking approach with lime explanations. *IEEE Access*, v. 13, p. 13847–13865, 2025.

BAHAD, P.; SAXENA, P. Study of AdaBoost and gradient boosting algorithms for predictive analytics. In: *International Conference on Intelligent Computing and Smart Communication 2019*. [S.l.]: Springer Singapore, 2020. p. 235–244.

BARROS, C. B. Classificadores de regressão logística, naive bayes e random forest na análise do tropismo do hiv-1 de subtipo b. Universidade Federal do Rio de Janeiro, 2019.

BOCHIE, K. et al. Aprendizado profundo em redes desafiadoras: Conceitos e aplicações. *Sociedade Brasileira de Computação*, 2020.

BRAY, G. A. et al. Obesity: a chronic relapsing progressive disease process. a position statement of the world obesity federation. *Obesity Reviews*, v. 18, n. 7, p. 715–723, 2017. Epub 2017 May 10. Disponível em: <https://doi.org/10.1111/obr.12551>.

BRITO, R. de; OLIVEIRA, A. de. Análise de características a partir do classificador mlp para auxílio no diagnóstico da covid-19. In: *Anais da VIII Escola Regional de Computação do Ceará, Maranhão e Piauí*. Porto Alegre, RS, Brasil: SBC, 2020. p. 89–95. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/ercemapi/article/view/11472>.

CHEN, Y.; LI, J.; ZHANG, H. Explainable machine learning for clinical decision support systems: A case study on obesity prediction. *IEEE Journal of Biomedical and Health Informatics*, IEEE, v. 25, n. 6, p. 2105–2114, 2021.

COSTA, E. L.; MOURA JÚNIOR, J. d. R. V. de. Aprendizado por reforço como ferramenta na tomada de decisão em investimentos em bolsa de valores. *OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA*, v. 21, n. 11, p. 22437–22457, Nov. 2023. Disponível em: <https://ojs.observatoriolatinoamericano.com/ojs/index.php/olel/article/view/2003>.

COZMAN, F. G.; KAUFMAN, D. Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação. *Revista USP*, n. 135, p. 195–210, 2022.

DAS, A.; RAD, P. *Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey*. 2020. Disponível em: <https://arxiv.org/abs/2006.11371>.

DOSHI-VELEZ, F.; KIM, B. *Towards A Rigorous Science of Interpretable Machine Learning*. 2017. Disponível em: <https://arxiv.org/abs/1702.08608>.

DUELL, J. A. *A Comparative Approach to Explainable Artificial Intelligence Methods in Application to High-Dimensional Electronic Health Records: Examining the Usability of XAI*. 2021. Disponível em: <https://arxiv.org/abs/2103.04951>.

ER, M. J.; VENKATESAN, R.; WANG, N. An online universal classifier for binary, multi-class and multi-label classification. In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.: s.n.], 2016. p. 003701–003706.

FILHO, E. L. d. S. et al. Previsão dos retornos do ibovespa utilizando redes neurais artificiais feedforward evolutivas. *Revista Produção Online*, v. 11, n. 4, p. 1114–1140, dez. 2011. Disponível em: <https://www.producaoonline.org.br/rpo/article/view/784>.

GOHEL, P.; SINGH, P.; MOHANTY, M. *Explainable AI: current status and future directions*. 2021. Disponível em: <https://arxiv.org/abs/2107.07045>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge: MIT Press, 2016. Disponível em: <https://www.deeplearningbook.org/>. ISBN 978-0-262-03561-3.

HOWLETT, R. J.; JAIN, L. C. (Ed.). *Smart Innovation, Systems and Technologies*. Cham: Springer, 2020. v. 93. (Smart Innovation, Systems and Technologies, v. 93). Disponível em: <https://link.springer.com/book/10.1007/978-3-030-30068-3>. ISBN 978-3-030-30068-3.

HUNTS, H. A. *Glossário de IA: Mais de 40 termos essenciais*. — *hunts.ai*. 2023. <https://hunts.ai/pt/glossario/>. [Accessed 06-08-2024].

IBGE, I. B. d. G. e. E. *Pesquisa do IBGE mostra aumento da obesidade entre adultos*. 2023. <https://www.gov.br/pt-br/noticias/saude-e-vigilancia-sanitaria/2020/10/pesquisa-do-ibge-mostra-aumento-da-obesidade-entre-adultos>. [Accessed 23-07-2024].

KHATER, T.; TAWFIK, H.; SINGH, B. Explainable artificial intelligence for investigating the effect of lifestyle factors on obesity. *Intelligent Systems with Applications*, v. 23, p. 200427, 2024. ISSN 2667-3053. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2667305324001017>.

KISS, O. et al. Using explainable machine learning and fitbit data to investigate predictors of adolescent obesity. *Sci. Rep.*, Springer Science and Business Media LLC, v. 14, n. 1, p. 12563, maio 2024.

KRAMER, O. Scikit-learn. In: *Machine Learning for Evolution Strategies*. Cham: Springer International Publishing, 2016. p. 45–53. ISBN 978-3-319-33383-0. Disponível em: [https://doi.org/10.1007/978-3-319-33383-0\\_5](https://doi.org/10.1007/978-3-319-33383-0_5).

KUMAR, S.; SINGH, R.; PATEL, A. Explainable ai in healthcare: Challenges and opportunities for clinical adoption. *npj Digital Medicine*, Nature Publishing Group, v. 6, n. 1, p. 45, 2023.

LIAO, Q. V.; GRUEN, D.; MILLER, S. Questioning the ai: Informing design practices for explainable ai user experiences. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020. (CHI '20). Disponível em: <http://dx.doi.org/10.1145/3313831.3376590>.

LINARDATOS, P.; PAPASTEFANOPOULOS, V.; KOTSIANTIS, S. Explainable ai: A review of machine learning interpretability methods. *Entropy*, v. 23, n. 1, 2021. ISSN 1099-4300. Disponível em: <https://www.mdpi.com/1099-4300/23/1/18>.

LINHARES, R. V. et al. Comparação entre bioimpedância elétrica e fórmulas antropométricas na estimativa da gordura corporal em mulheres obesas mórbidas. *Revista Brasileira de Obesidade, Nutrição e Emagrecimento*, v. 13, n. 79, p. 453–461, 2019. Disponível em: <https://www.rbone.com.br/index.php/rbone/article/download/1731/1124>.

LOPES, L. et al. Estimativa dos níveis de obesidade com base em hábitos alimentares e condição física através de técnicas de aprendizado de máquina. In: *Anais Estendidos da XXXIV Conference on Graphics, Patterns and Images*. Porto Alegre, RS, Brasil: SBC, 2021. p. 154–157. ISSN 0000-0000. Disponível em: [https://sol.sbc.org.br/index.php/sibgrapi\\_estendido/article/view/20029](https://sol.sbc.org.br/index.php/sibgrapi_estendido/article/view/20029).

LUNDBERG, S.; LEE, S.-I. *A Unified Approach to Interpreting Model Predictions*. 2017. Disponível em: <https://arxiv.org/abs/1705.07874>.



MACEDO, L. d. N. et al. *Comparação de algoritmos de aprendizado de máquina para prever futuras cepas do vírus da influenza*. Monografia de Graduação em Ciência da Computação. Universidade Federal de Uberlândia, 2023. Disponível em: <https://repositorio.ufu.br/handle/123456789/38329>.

MACKIE, I. et al. *Undergraduate Topics in Computer Science*. [S.l.]: Springer, 2010.

MAHADI, M. K. et al. Understanding machine learning its application in obesity estimation by explainable ai. In: *2024 International Conference on Inventive Computation Technologies (ICICT)*. [S.l.: s.n.], 2024. p. 112–117.

MANOTAS, A. De la H. et al. Obesity level estimation software based on decision trees. *Journal of Computer Science*, v. 15, p. 10, 01 2019.

MARY, S. Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review. In: \_\_\_\_\_. [S.l.: s.n.], 2019. p. 1269–1292. ISBN 978-3-030-22867-5.

MAULANA, A. et al. Predicting obesity levels with high accuracy: Insights from a catboost machine learning model. *Infolitika Journal of Data Science*, v. 2, n. 1, p. 17–27, May 2024. Disponível em: <https://heca-analitika.com/ijds/article/view/195>.

MOHSENI, M. *Explainable AI: Foundations, Developments, Prospects*. Cham: Springer, 2020. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-48191-4>. ISBN 978-3-030-48191-4.

MOLNAR, C. *Interpretable Machine Learning: A guide for making black box models explainable*. 2. ed. [s.n.], 2022. Disponível em: <https://christophm.github.io/interpretable-ml-book>.

MONTEREI, R. C. *Perspectivas do uso do aprendizado de máquina em bibliotecas: uma revisão sistemática de literatura*. 2023.

NEVES, S. C. et al. Os fatores de risco envolvidos na obesidade no adolescente: uma revisão integrativa. *Ciência Saúde Coletiva*, ABRASCO - Associação Brasileira de Saúde Coletiva, v. 26, p. 4871–4884, Oct 2021. ISSN 1413-8123. Disponível em: <https://doi.org/10.1590/1413-812320212611.3.30852019>.

OLIVEIRA, D. F. N. *Interpretabilidade de modelos de aprendizado profundo aplicados ao diagnóstico e prognóstico não supervisionado de falhas*. Tese (Doutorado) — Universidade de São Paulo, Agência USP de Gestão da Informação Acadêmica (AGUIA), 2020. Disponível em: <http://dx.doi.org/10.11606/D.3.2020.tde-08032021-102116>.

OMS, O. M. d. S. *Dia Mundial da Obesidade 2022: acelerar ação para acabar com a obesidade*. 2022. <https://www.paho.org/pt/noticias/4-3-2022-dia-mundial-da-obesidade-2022-acelerar-acao-para-acabar-com-obesidade>. [Accessed 23-07-2024].

OMS, O. M. d. S. *Obesity*. 2024. <https://www.who.int/health-topics/obesity>. [Accessed 23-07-2024].

OMS, O. M. d. S. *Obesity and overweight*. 2024. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>. [Accessed 23-07-2024].

- OPAS, O. P.-A. d. S. *Uma em cada oito pessoas, no mundo, vive com obesidade*. 2024. <https://www.paho.org/pt/noticias/1-3-2024-uma-em-cada-oito-pessoas-no-mundo-vive-com-obesidade>. [Accessed 23-07-2024].
- PALECHOR, F. M.; MANOTAS, A. de la H. Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from colombia, peru and mexico. *Data in Brief*, v. 25, p. 104344, 2019. ISSN 2352-3409. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352340919306985>.
- PALECHOR, F. M.; MANOTAS, A. de la H. *Estimation of Obesity Levels Based On Eating Habits and Physical Condition*. 2019. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5H31Z>.
- PARANHOS, L. R. L.; RODOLPHO, P. J. *Metodologia da Pesquisa Aplicada à Tecnologia*. SENAI-SP Editora, 2018. (Currículo Comum). ISBN 9788553400300. Disponível em: <https://books.google.com.br/books?id=ACBHDwAAQBAJ>.
- PRATI, R. C. et al. Curvas roc para avaliação de classificadores. *Revista IEEE América Latina*, v. 6, n. 2, p. 215–222, 2008.
- RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: ACM. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.], 2016. p. 1135–1144.
- RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. "why should i trust you?": Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 1135–1144. ISBN 9781450342322.
- RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. Explainable ai for healthcare: From black box to interpretable models. *Journal of Artificial Intelligence in Medicine*, Elsevier, v. 108, p. 101905, 2020.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3. ed. Upper Saddle River: Prentice Hall, 2010. ISBN 978-0136042594.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4. ed. Upper Saddle River, NJ: Pearson, 2020. ISBN 0-13-461099-7.
- SALIH, A. M. et al. A perspective on explainable artificial intelligence methods: Shap and lime. *Advanced Intelligent Systems*, Wiley Online Library, v. 7, n. 1, p. 2400304, 2024.
- SAMEK, W. et al. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham: Springer, 2017. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-58553-3>. ISBN 978-3-319-58553-3.
- SANT'ANNA, M. d. S. L. et al. Avaliação de gordura corporal pela bioimpedância elétrica e sua correlação com diferentes pontos anatômicos de medida da circunferência da cintura em crianças. *Jornal de Pediatria*, Sociedade Brasileira de Pediatria, v. 85, n. 1, p. 61–66, Jan 2009. ISSN 0021-7557. Disponível em: <https://doi.org/10.1590/S0021-75572009000100011>.

SANTOS, E. M. dos; ROCHA, M. M. S.; DIAS, T. de O. Obesidade infantil: uma revisão bibliográfica sobre fatores que contribuem para a obesidade na infância. *Revista Brasileira de Reabilitação e Atividade Física*, v. 9, n. 1, p. 57–62, 2020.

SAUT, M. G. S. et al. Internações por obesidade: Tendências epidemiológicas e impactos na saúde pública. *Journal of Medical and Biosciences Research*, v. 1, n. 4, p. 431–439, set. 2024. Disponível em: <https://journalmbr.com.br/index.php/jmbr/article/view/284>.

SHAKHOVSKA, N. *Advances in Intelligent Systems and Computing*. Cham: Springer, 2017. ISBN 978-3-319-69263-0.

SILVA, P. M. F. et al. MÉtodos de avaliaÇão da gordura corporal e anÁlise da ferramenta de ultrassonografia em pacientes obesos. *Arquivos de Ciências da Saúde da UNIPAR*, v. 27, n. 8, p. 4742–4753, ago. 2023. Disponível em: <https://revistas.unipar.br/index.php/saude/article/view/10418>.

SMITH, J.; BROWN, E.; JOHNSON, M. Integrating ai and phenotypic data for precision medicine in obesity. *Nature Communications*, Nature Publishing Group, v. 13, n. 1, p. 35285, 2022.

SOUSA, R.; BRITO, R.; XIMENES, J. Análise comportamental por meio de características com os classificadores svm e mlp para detecção de traços autísticos em adultos. In: *Anais da IX Escola Regional de Computação do Ceará, Maranhão e Piauí*. Porto Alegre, RS, Brasil: SBC, 2021. p. 9–16. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/ercemapi/article/view/17902>.

SOUZA, R. G. M. d. et al. Métodos de análise da composição corporal em adultos obesos. *Revista de Nutrição*, Pontifícia Universidade Católica de Campinas, v. 27, n. 5, p. 569–583, Sep 2014. ISSN 1415-5273. Disponível em: <https://doi.org/10.1590/1415-52732014000500006>.

TISSOT, H. C.; CAMARGO, L. C.; POZO, A. T. Treinamento de redes neurais feedforward: comparativo dos algoritmos backpropagation e differential evolution. In: SN. *Brazilian Conference on Intelligent Systems*. [S.l.], 2012.

VIEIRA, C. P. R. *Uso de agrupamento para alcançar explicabilidade global de modelos de aprendizado de máquina*. Tese (Doutorado) — Universidade de São Paulo. Agência de Bibliotecas e Coleções Digitais, 2023. Disponível em: <http://dx.doi.org/10.11606/D.100.2023.tde-22052023-205242>.

WANG, M. *An Introduction to Machine Learning Interpretability*. Cham: Springer, 2019. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-29229-2>. ISBN 978-3-030-29229-2.

WAZLAWICK, R. *Metodologia de Pesquisa para Ciência da Computação*. GEN LTC, 2020. ISBN 9788595151093. Disponível em: <https://books.google.com.br/books?id=K-PxzwEACAAJ>.

WILLINGHOEFER, E. Mapeamento sistemático da literatura em aprendizado de máquina para identificação de fibrilação atrial. Universidade Federal da Fronteira Sul, 2023.

ZHANG, X.; LI, W.; CHEN, H. Machine learning models for predicting obesity and its comorbidities: A comparative study. *Journal of Medical Internet Research*, JMIR Publications, v. 23, n. 6, p. e27218, 2021.

ZHANG, Y.; WENG, Y.; LUND, J. Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics*, v. 12, n. 2, 2022. ISSN 2075-4418. Disponível em: <https://www.mdpi.com/2075-4418/12/2/237>.

ÖZKURT, C. *Computers and Electronics in Medicine*, v. 1, n. 1, p. 12–17, Jul. 2024. Disponível em: <https://journals.adbascentific.com/cem/article/view/4>.