

# Previsão de Sobrevida em Insuficiência Cardíaca: Pelos Olhos do Aprendizado de Máquina

**Roberto Almeida Burlamaque Catunda** – `rabc@cesar.school`,  
**Gabriel Reis de Melo Pires** – `grmp@cesar.school`,  
**Marina da Fonseca Frias de Siqueira Campos** – `mffcc@cesar.school`,  
**Joao Pedro Araújo** – `jpa@cesar.school`,  
**David Ian Pereira Paula** – `dipp@cesar.school`

<sup>1</sup>Centro de Estudos e Sistemas Avançados do Recife – C.E.S.A.R School, Recife, Brazil

**Abstract.** Mesmo nos dias atuais com os avanços medicinais e tecnológicos, é difícil prever com exatidão a predisposição de uma pessoa a ter insuficiências cardíacas. A complexidade do corpo humano por muitas vezes esconde os sintomas-chave que, quando agregados a outros, habilita a detecção de uma possível insuficiência cardíaca e a probabilidade de sobrevida do indivíduo. Esse artigo busca comparar métodos de aprendizado de máquina, levando em conta um conjunto de dados (features) já pre-determinados, observados em um paciente por equipes médicas, em prol de ajudar na detecção da probabilidade de sobrevida, caso essas features culminem em um ataque cardíaco dos mais diversos níveis. Foram testados 4 modelos de aprendizado de máquina, sendo eles: KNN Classifier, Decision Tree, Random Forest e Ensemble Learning que combina os 3 anteriores. Devido ao desbalanceamento dos dados, foram, também, aplicadas 4 técnicas distintas: Random Sampler, SMOTE, Borderline SMOTE e ADASYN, comparando: todos x todos x conjunto original, observando seus resultados e diferenças, por meio das métricas de Acurácia Precisão, Recall e F1-Score.

## 1. Introdução

A Insuficiência Cardíaca (IC) é definida clinicamente como uma síndrome complexa na qual o coração é incapaz de bombear sangue em uma taxa compatível com as necessidades metabólicas dos tecidos, ou só consegue fazê-lo à custa de elevadas pressões de enchimento. Tal condição está frequentemente associada à dilatação ou, por vezes, a hipertrofia do ventrículo esquerdo (Conforme a Figura 1, página 2). Esta condição não é uma doença isolada, mas sim o estágio final de diversas pa-

tologias cardiovasculares, caracterizando-se por anomalias estruturais ou funcionais que comprometem o enchimento ou a ejeção ventricular. Do ponto de vista epidemiológico, a IC representa um desafio global de saúde pública. Estima-se que a síndrome afete mais de 64 milhões de pessoas em todo o mundo, com uma prevalência que aumenta progressivamente com o envelhecimento populacional (SAVARESE et al., 2022). Dados apontam que a prevalência é de aproximadamente 1% a 2% na população adulta geral, su-

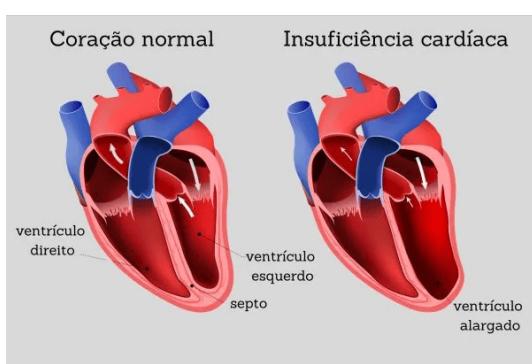
bindo para mais de 10% entre indivíduos com idade superior a 70 anos (Comitê Coordenador da Diretriz de Insuficiência Cardíaca, 2018).

Os pacientes com IC podem ser classificados de acordo com a fração de ejeção do ventrículo esquerdo (FEVE):

- IC com fração de ejeção preservada (IC-FEp): Pacientes com FEVE maior, ou igual a 50%
- IC de fração de ejeção levemente reduzida (CFEi): Pacientes com FEVE entre 41 e 49
- IC com fração de ejeção reduzida (IC-FEr): Pacientes com FEVE menor que 40%
- IC com FE melhorada (HFimpEF): Pacientes com uma FEVE basal menor, ou igual a 40%, um aumento maior, ou igual a 10 pontos da FEVE basal e uma segunda medição da FEVE maior que 40%

(Ministério da Saúde, gov.br, 2024)

**Figura 1: Comparação estrutural entre um coração saudável e um com IC.**



**Fonte:** Brasil Escola - UOL

O principal desafio é a dificuldade de prever o que acontecerá com cada indivíduo, a IC varia muito de pessoa para pessoa e geralmente vem acompanhada

de outros problemas de saúde, tornando-se complicada a identificação, principalmente no início, de quem corre maior risco de morte.

Conseguir estimar a capacidade de sobrevida com precisão é crucial para separarmos os pacientes entre os de alto e baixo risco, facilitando para a equipe médica poder tomar decisões melhores. No entanto, as ferramentas estatísticas tradicionais muitas vezes não dão conta de analisar tantas variáveis médicas ao mesmo tempo. É nesse ponto que o Aprendizado de Máquina se mostra essencial. Diferente da estatística clássica, esses algoritmos conseguem encontrar conexões complexas e padrões ocultos nos exames clínicos, recebendo dados brutos e transformando-os em previsões com alta taxa de acerto ajudando na análise e tomada de decisão do profissional.

Diante desse cenário, o objetivo principal deste trabalho é avaliar e comparar o desempenho de diferentes algoritmos de aprendizado de máquina na predição de sobrevida de pacientes com insuficiência cardíaca. Para lidar com o desafio comum de bases de dados médicos desbalanceadas, o estudo propõe ainda a aplicação e análise comparativa de diversas técnicas de reamostragem (*oversampling*), visando identificar qual combinação de modelo e tratamento de dados oferece a maior confiabilidade para os respectivos sintomas clínicos apresentados.

O restante deste artigo está organizado da seguinte maneira: A Seção 2 apresenta uma revisão dos trabalhos relacionados, situando o estudo no estado da arte da predição de risco em insuficiência cardíaca. A Seção 3 descreve o conjunto de dados utilizado, detalhando as etapas de pré-processamento e análise explo-

ratória. A Seção 4 delinea a metodologia, especificando as técnicas de reamortagem para balanceamento de classes e os algoritmos de aprendizado de máquina selecionados. A Seção 5 expõe os experimentos realizados e analisa os resultados obtidos. A Seção 6 discute as implicações clínicas desses achados, as limitações do estudo e aponta direções para trabalhos futuros. Por fim, a Seção 7 apresenta as referências bibliográficas utilizadas.

## 2. Trabalhos Relacionados

A aplicação de Aprendizado de Máquina na predição de sobrevivência em insuficiência cardíaca tem sido amplamente explorada. Tripoliti et al. (2017) destacam que métodos como Random Forest e Árvores de Decisão são frequentes na literatura devido à sua interpretabilidade. Chicco e Jurman (2020) demonstraram que apenas duas variáveis — creatinina sérica e fração de ejeção — podem ser suficientes para previsões robustas em determinados contextos.

Mais recentemente, Alex e Hasib (2025) apresentaram um estudo utilizando um classificador do tipo *Stacking Ensemble* (combinando CatBoost, LightGBM e XGBoost com um Meta-Learner MLP). Utilizando dados sintéticos gerados via Gretel AI baseados no dataset UCI, os autores alcançaram uma acurácia impressionante de 99.6%. Eles também empregaram a técnica SMOTE para平衡amento e Recursive Feature Elimination (RFE) para seleção de características.

Embora os resultados de Alex e Hasib (2025) sejam estado-da-arte, o uso de dados sintéticos pode simplificar fronteiras de decisão que, em dados clínicos reais e pequenos, são muito mais ruidosas. Diferentemente da abordagem deles, mas se baseando na mesma, este trabalho foca no **dataset original da UCI** (apenas 299

registros), testando se técnicas avançadas de oversampling (como ADASYN) conseguem extrair performance comparável usando modelos mais leves e explicáveis, sem recorrer à complexidade de caixas-pretas de *Deep Learning* ou geração de dados sintéticos massiva.

## 3. Descrição do Dataset e Preprocessamento

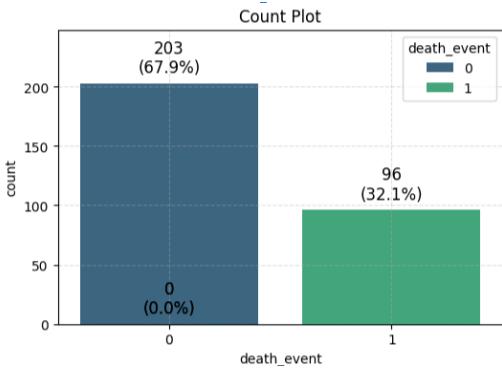
**Detalhes do Dataset.** O estudo utilizou o conjunto de dados Heart Failure Clinical Records, disponibilizado pelo UCI Machine Learning Repository, amplamente empregado em pesquisas voltadas à análise de risco clínico em pacientes com insuficiência cardíaca. Esse dataset reúne informações de 299 pacientes acompanhados ao longo de um período específico, apresentando doze características preditivas e uma variável alvo. As variáveis contemplam dimensões demográficas, clínicas e laboratoriais. Entre as características demográficas incluem-se a idade e o sexo do paciente; já as variáveis clínicas abrangem a presença de anemia, hipertensão arterial, diabetes, tabagismo e o tempo de acompanhamento médico. Os exames laboratoriais contemplam medidas fisiológicas e marcadores bioquímicos essenciais, como creatinina sérica, sódio sérico, fração de ejeção do ventrículo esquerdo, contagem de plaquetas e níveis de creatinina fosfoquinase. Por fim, a variável alvo (*"death\_event"*) indica se o paciente veio a óbito durante o acompanhamento, configurando o problema de classificação deste trabalho.

**Preprocessamento.** O processo de pré-processamento iniciou-se pela verificação de valores ausentes, confirmando que não há quaisquer lacunas nos dados, o que eliminou a necessidade de imputação e permitiu a continuação da análise sem

intervenções de reconstrução de valores. Em seguida, avaliou-se a natureza das variáveis. Todas as variáveis do conjunto de dados já se apresentavam em formato adequado e compatível com os modelos, dispensando a necessidade de técnicas de conversão de atributos, como de categóricos para numéricos, por exemplo. Posteriormente, a normalização das variáveis contínuas foi conduzida com o emprego do método StandardScaler, aplicado exclusivamente aos dados de treino após o train/test split, evitando-se qualquer forma de data leakage e garantindo coerência com as boas práticas descritas na literatura.

**Analise Exploratoria de Dados (EDA).** A análise exploratória permitiu uma compreensão inicial aprofundada do comportamento das variáveis. As estatísticas descritivas revelaram elevada variação em marcadores laboratoriais como CPK e contagem de plaquetas, além de uma faixa etária predominantemente avançada, típica de pacientes com insuficiência cardíaca. A variável alvo evidenciou um desbalanceamento significativo, com aproximadamente dois terços dos pacientes sobrevivendo e um terço indo a óbito. Essa assimetria foi confirmada por meio de um countplot (Confirmado pela figura 2, página 4), apontando a necessidade de técnicas específicas para o balanceamento das classes. Além disso, a geração de um pairplot possibilitou observar interações relevantes entre variáveis fisiológicas e clínicas, além de sugerir possíveis tendências de separação entre as classes, ainda que de natureza predominantemente não linear.

**Figura 2: Countplot ”death\_event”.**



Fonte: Autoria própria.

#### 4. Metodologia

**Estratégia Experimental de Balanceamento.** Uma das principais contribuições metodológicas deste estudo reside na investigação sobre qual técnica de geração de dados sintéticos oferece maior robustez para dados clínicos reais. Diferentemente do trabalho de referência (Alex e Hasib, 2025), que padronizou o uso do SMOTE, este projeto desenhou um cenário experimental comparativo envolvendo quatro abordagens distintas para tratar o desequilíbrio das classes:

1. *Random Oversampling* (Baseline);
2. SMOTE (Padrão da literatura);
3. *Borderline-SMOTE* (Foco na fronteira de decisão);
4. ADASYN (Foco na densidade de dificuldade).

Nenhuma técnica foi selecionada a priori; todas foram submetidas ao mesmo protocolo de validação cruzada para que os dados determinassem a abordagem superior.

**Modelos Selecionados.** Uma vez definido o método de balanceamento, procedeu-se à escolha dos modelos de *machine learning*. Foram adotados três algoritmos individuais: *K-Nearest Neighbors* (KNN), *Decision Tree* e *Random Forest*. O KNN foi selecionado por

sua simplicidade e boa resposta em *datasets* de baixa dimensionalidade (após padronização). A *Decision Tree* foi empregada devido à sua interpretabilidade e capacidade de capturar relações não lineares, configurada com profundidade limitada para evitar *overfitting*. O *Random Forest*, por sua vez, ofereceu maior robustez e menor variância.

Adicionalmente, foi construído um *Ensemble* baseado em *Soft Voting*, que agraga as probabilidades produzidas por cada modelo. A escolha pelo *Voting Classifier* — em detrimento do *Stacking* complexo utilizado por Alex e Hasib (2025) — justifica-se pela busca de um equilíbrio entre performance e interpretabilidade, evitando a opacidade de “caixas-pretas” profundas em um *dataset* de dimensão reduzida, mas mantendo a estratégia de combinação de modelos para compensar limitações individuais.

**Arquitetura de Deploy e Pipeline de Dados.** Para garantir que o modelo desenvolvido não permanecesse apenas em ambiente de prototipagem, foi construída uma arquitetura completa de MLOps para implantação e monitoramento. O sistema foi orquestrado utilizando **Docker Compose**, integrando múltiplos serviços essenciais:

- **FastAPI:** Atua como a interface de inferência, recebendo os dados do paciente em formato JSON vindos do ThingsBoard e retornando a predição de sobrevivência em tempo real.
- **MLflow:** Utilizado para o versionamento dos experimentos e armazenamento dos artefatos do modelo treinado, garantindo rastreabilidade.
- **MinIO & PostgreSQL:** Responsáveis pela persistência dos dados de entrada

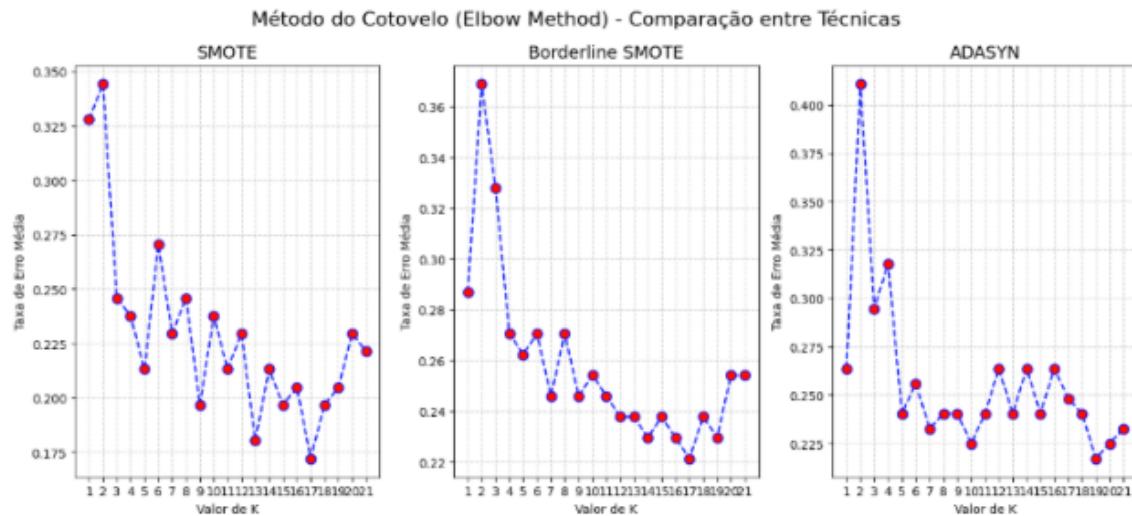
e dos resultados das predições, permitindo auditoria futura.

- **ThingsBoard:** Plataforma de IoT e MLOps utilizada para visualização dos dados via *dashboards*, simulando um ambiente de monitoramento clínico contínuo.

O fluxo de dados (*Dataflow*) foi desenhado da seguinte forma: os dados do paciente são enviados para a API, processados pelo modelo de *Random Forest* (treinado com SMOTE; Usamos o SMOTE para o workflow com intuito de facilitar a comparação direta com o artigo usado como base, mesmo que o ADASYN tenha se saído melhor nos nossos testes), e o resultado (Probabilidade de Sobrevida) é persistido no banco e enviado ao ThingsBoard para visualização em uma Dashboard. Este *pipeline* garante que o modelo esteja acessível e integrado a um fluxo de trabalho operacional real.

**Ajustes de Hiperparâmetros.** O ajuste de hiperparâmetros foi conduzido de forma controlada e coerente com o tamanho reduzido do dataset. Para o KNN, utilizou-se o método do cotovelo, avaliando-se diversas quantidades de vizinhos até identificar que três ofereciam a menor taxa de erro médio. Tanto a Decision Tree quanto a Random Forest foram configuradas com profundidade máxima igual a três, proporcionando regularização suficiente para evitar a segmentação excessiva de regiões do espaço de atributos, característica que levaria ao sobreajuste em conjuntos clínicos pequenos. O ensemble, por utilizar votação probabilística, não necessitou de ajustes adicionais.

**Figura 3: Método do cotovelo.**



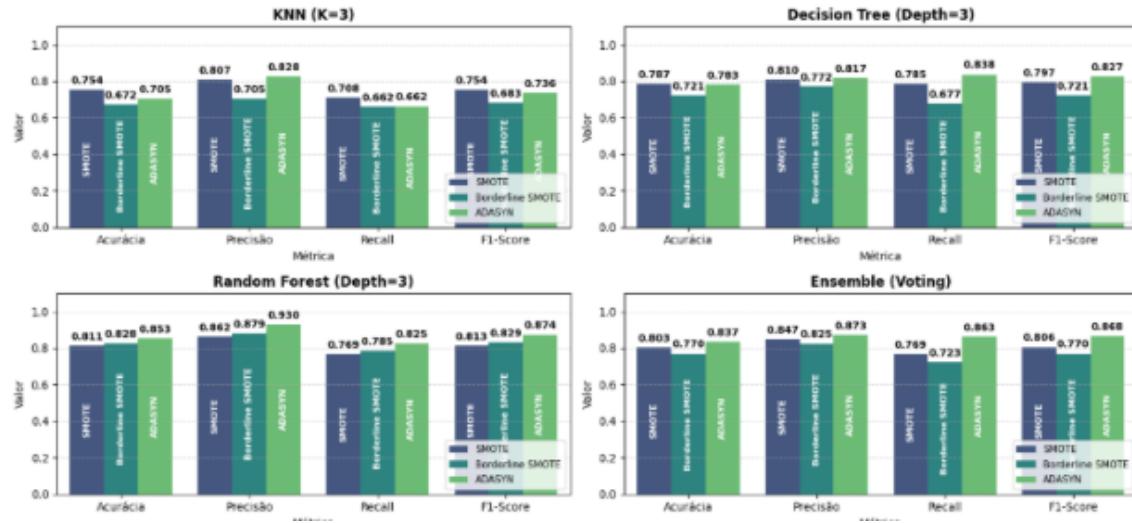
**Fonte:** Autoria própria.

**Estratégia de Validação.** A estratégia de validação adotada foi dividida em duas etapas complementares. A primeira consistiu na comparação entre as técnicas de oversampling, realizada por meio de validação cruzada estratificada em cinco dobras, garantindo que a distribuição da variável alvo fosse mantida em todos os subconjuntos e assegurando estimativas estáveis para os desempenhos de cada técnica. A segunda etapa correspondeu ao train/test split, aplicado após a definição do método de oversampling selecionado. A divisão em 70% para treino e 30% para teste seguiu a prática comum em estudos biomédicos de pequena escala e assegurou a avaliação final dos modelos em dados nunca utilizados durante o processo de treinamento.

**Métricas de Performance.** A avaliação da performance foi conduzida por meio de métricas consolidadas para problemas de classificação binária. A acurácia foi considerada apenas como indicador geral, já que tende a ser influenciada por desbalanceamento. Métricas mais sensíveis foram priorizadas. A precisão mensurou

a proporção de pacientes corretamente identificados como óbitos, enquanto o recall avaliou a capacidade do modelo em identificar todos os pacientes que de fato faleceram, dimensão crítica em estudos clínicos, onde falsos negativos representam risco significativo. O F1-score, utilizado tanto nas etapas de oversampling quanto na avaliação final dos modelos, sintetizou o equilíbrio entre precisão e recall, tornando-se a métrica central deste trabalho. Embora o projeto não tenha incluído a validação por meio da área sob a curva ROC (AUC-ROC), amplamente utilizada em contextos clínicos, as métricas empregadas permitiram uma caracterização robusta do comportamento preditivo dos modelos, em alinhamento com a metodologia científica adotada no artigo de referência.

**Figura 4: Comparação dos Métodos.**



Fonte: Autoria própria.

## 5. Experimentos e Resultados

**Configuração Experimental.** Os experimentos foram conduzidos em ambiente Python, usando o Google Colaboratory, foram importadas as biblioteca padrões, como: *Pandas* para fácil manipulação dos dados em forma de DataFrame, *Seaborn* e *Matplotlib* para plotagem dos dados em forma de gráfico, *scikit-learn* para modelagem e avaliação de métricas e *Imbalanced-Learn* para as técnicas de “Oversampling”, usado para balanceamento das features e target do Dataset. A estratégia experimental foi dividida em duas fases: (1) Seleção da técnica de balanceamento ideal via validação cruzada e (2) Avaliação final dos classificadores no conjunto de teste junto as técnicas escolhidas da 1<sup>a</sup> fase. Para garantir a reprodutibilidade, definiu-se uma “seed” aleatória (*random state* = 42) em todas as divisões de dados e inicializações de modelos. O conjunto de dados foi dividido na proporção de 70% para treinamento e 30% para teste, respeitando a estratificação (mesma quantidade percentual) da variável alvo para manter a proporção original de classes.

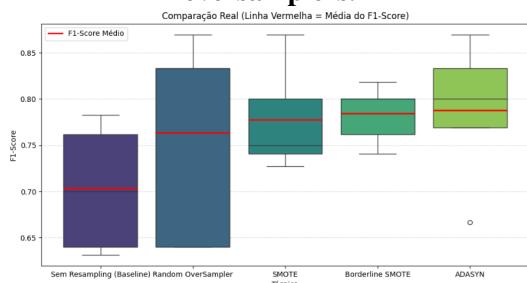
**1<sup>a</sup> fase: Análise das Técnicas de Oversampling.** Na primeira fase, avaliou-se o impacto de cada técnica de reamostragem no desempenho do classificador base (*Random Forest*). Utilizando da validacao cruzada ( $k = 3$ ), comparou-se o cenário (sem reamostragem) com quatro outros: Random Over-Sampler, SMOTE, Borderline-SMOTE e ADASYN. O objetivo foi identificar empiricamente qual método proporciona a melhor generalização. Conforme detalhado na Tabela 1, os resultados demonstraram uma evolução clara na qualidade da predição. O SMOTE tradicional obteve um F1-Score médio de 0.7775. O Borderline-SMOTE apresentou uma leve superioridade (0.7842). Contudo, foi o método **ADASYN** que emergiu com o melhor desempenho estatístico (0.7878), indicando que sua estratégia adaptativa é a mais eficaz para lidar com a complexidade deste conjunto de dados. Os resultados finais seguintes (Tabela 2) reportam o desempenho dos modelos utilizando o ADASYN, técnica que se manteve superior nas análises comparativas subsequentes.

**Tabela 1: Comparação das Técnicas de Oversampling (Validação Cruzada)**

Técnica	F1-Médio	Std
Baseline	0.7032	0.0615
Random Over.	0.7632	0.1015
SMOTE	0.7775	0.0522
Borderline	0.7842	0.0284
<b>ADASYN</b>	<b>0.7878</b>	0.0692

Fonte: Autoria própria.

**Figura 5: Comparação gráfica, oversamplers.**



Fonte: Autoria própria.

**2ª fase: Desempenho dos Classificadores.** Na segunda fase, os classificadores KNN, Decision Tree, Random Forest e o Ensemble (Soft Voting) foram treinados e avaliados no conjunto de teste, utilizando as técnicas de balanceamento que apresentaram melhor desempenho na fase anterior. A Tabela 2 sumariza os resultados obtidos com a aplicação do método ADASYN, que se mostrou consistentemente superior ou competitivo entre os modelos.

**Tabela 2: Resultados no Conjunto de Teste (Técnica ADASYN)**

Modelo	Acur.	Prec.	Rec.	F1
KNN (k=3)	0.705	0.828	0.663	0.736
Dec. Tree	0.783	0.817	0.838	0.827
<b>R. Forest</b>	<b>0.853</b>	<b>0.930</b>	0.825	<b>0.874</b>
Ensemble	0.837	0.873	<b>0.863</b>	0.868

Fonte: Autoria própria.

**Interpretação dos Resultados.** A análise comparativa permite inferir que a dificuldade primária deste conjunto de dados não reside apenas na escassez de exemplos da classe minoritária (óbitos), mas na complexidade da fronteira de decisão entre as classes. A superioridade do método ADASYN sobre o SMOTE convencional e o Random OverSampler sugere que a geração adaptativa de dados sintéticos — focada nas regiões onde o classificador encontra maior dificuldade — foi crucial para “desenhar” melhor essa fronteira, reduzindo a confusão entre pacientes estáveis e de alto risco.

Em relação aos modelos, o desempenho superior do *Random Forest* ( $F1 = 0.874$ ) demonstra a capacidade dos métodos de *ensemble* baseados em árvores de lidar com a não-linearidade das variáveis fisiológicas sem sofrer o *overfitting* observado na Árvore de Decisão única. O modelo KNN apresentou o desempenho inferior, indicando que a classificação baseada puramente em distância euclidiana pode não ser suficiente para capturar as nuances clínicas em um espaço de alta dimensionalidade.

Um ponto de destaque clínico é a métrica de *Recall* (Sensibilidade). Embora o Random Forest tenha o melhor equilíbrio geral (F1-Score), o modelo de *Ensemble* (Soft Voting) atingiu o maior *Recall* (0.863). No contexto médico, um alto *Recall* é frequentemente priorizado em detrimento da Precisão, pois o custo de um Falso Negativo (deixar de identificar um paciente que irá a óbito) é incomensuravelmente maior do que um Falso Positivo (alertar risco para um paciente que sobrevive). Portanto, a escolha final do modelo em um ambiente hospitalar real poderia pender para o Ensemble, dependendo do limiar de tolerância da equipe médica.

Por fim, os modelos baseados em árvores

trazem a vantagem de serem mais fáceis de interpretar. Diferente de modelos opacos (conhecidos como 'caixa-preta'), eles permitem visualizar como a decisão foi tomada, confirmando que indicadores médicos clássicos, como a fração de ejeção e a creatinina, são fundamentais para prever a sobrevivência (CHICCO; JURMAN, 2020).

## 6. Discussão

**Relevância Clínica e Insights.** Os resultados obtidos corroboram a hipótese de que o aprendizado de máquina pode atuar como uma ferramenta robusta de suporte à decisão clínica. A alta performance do modelo *Random Forest* (F1-Score de 0.87) em conjunto com o balanceamento ADASYN sugere que podemos ter uma maior probabilidade de achar pacientes de alto risco com base variáveis prontamente disponíveis em prontuários eletrônicos, sem a necessidade de exames invasivos ou excessivamente custosos, mas reiterando que esses podem ser fazer úteis, também, quando em conjunto com as técnicas apresentadas, o que geraria uma nova pesquisa por si só.

A confirmação de que variáveis como fração de ejeção e creatinina sérica são de grande importância para o prognóstico alinha-se com a fisiopatologia (estudo das doenças) conhecida da insuficiência cardíaca, onde a função de bomba comprometida e a disfunção renal formam um ciclo vicioso de agravamento do quadro clínico. Além disso, o desempenho competitivo do *Ensemble* no quesito sensibilidade (*Recall*) levanta uma discussão importante sobre o *trade-off* entre precisão e recuperação em ambientes de saúde: em triagens de emergência, pode ser preferível um modelo que "erre para mais" (falsos positivos) do que um que

deixe passar um caso grave (falsos negativos).

### Análise Crítica: SMOTE vs. ADASYN.

Nossos experimentos permitiram contestar a hegemonia do SMOTE em estudos similares. Enquanto o trabalho de referência (Alex e Hasib, 2025) adotou o SMOTE como padrão, nossa análise comparativa revelou que o ADASYN oferece um ganho de performance tangível em dados reais. Isso sugere que a simples interpolação linear (SMOTE) pode ser insuficiente em cenários clínicos ruidosos, e que métodos que priorizam exemplos difíceis (ADASYN) devem ser considerados como alternativas preferenciais em pesquisas futuras.

**Limitações do Estudo.** Apesar dos resultados promissores, este estudo possui limitações inerentes que devem ser consideradas. Primeiramente, a base de dados utilizada (299 registros) é relativamente pequena para os padrões de dados (como *Big Data*) em saúde, o que pode restringir a generalização do modelo para populações com características demográficas muito distintas da amostra original. Embora técnicas de *oversampling* como o ADASYN tenham mitigado o desbalanceamento, dados sintéticos são aproximações matemáticas que podem não capturar perfeitamente a complexidade biológica de casos reais raros. Outra limitação reside na natureza tabular e estática dos dados; a evolução temporal dos sintomas (séries temporais de exames) não foi considerada, o que poderia fornecer pistas adicionais sobre a deterioração do paciente.

**Trabalhos Futuros.** Para avançar nesta linha de pesquisa, sugere-se a validação dos modelos propostos em estruturas maiores e mais diversificadas, garantindo que o algoritmo não contenha vieses regionais ou étnicos. A incorporação de dados não

estruturados, como relatórios médicos em texto livre, exames de imagem (ecocardiogramas) e frames de vídeos, processados por redes neurais profundas (*Deep Learning*), somado as – citadas acima – series temporais, representa um caminho promissor para aumentar a acurácia preditiva. Por fim, recomenda-se a aplicação da análise da curva ROC (Receiver Operating Characteristic) e da área sob a curva (AUC) que são altamente usados em contextos clínicos e que nesse estudo foram esquecidos.

## Referências

- [1] ALEX, Merlin; HASIB, Maheen. Prediction Of Heart Failure Patient Survival Using Machine Learning. In: **AAAI Summer Symposium Series (SuSS-25)**. Dubai: Heriot-Watt University, 2025.
- [2] SAVARESE, G. et al. Global burden of heart failure: a comprehensive and updated review of epidemiology. **Cardiac Failure Review**, v. 8, 2022.
- [3] BRASIL. Ministério da Saúde. Insuficiência Cardíaca.
- [4] BRASIL ESCOLA. Insuficiência Cardíaca e Doenças Cardiovasculares. **Brasil Escola - UOL**, 2024. Disponível em: <<https://brasilescola.uol.com.br>>. Acesso em: 01 dez. 2024.
- [5] COMITÊ COORDENADOR DA DIRETRIZ DE INSUFICIÊNCIA CARDÍACA. Diretriz Brasileira de Insuficiência Cardíaca Crônica e Aguda. **Arquivos Brasileiros de Cardiologia**, v. 111, n. 3, 2018.
- [6] TRIPOLITI, E. E. et al. Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. **Computational and Structural Biotechnology Journal**, v. 15, p. 26-47, 2017.
- [7] CHICCO, Davide; JURMAN, Giuseppe. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. **BMC Medical Informatics and Decision Making**, v. 20, n. 1, p. 16, 2020.
- Gov.br**, 2024. Disponível em: <<https://www.gov.br/saude/pt-br>>. Acesso em: 01 dez. 2024.