

## 4 Resultados e Discussão

### 4.1 Introdução

O presente estudo teve como objetivo principal desenvolver e avaliar um sistema de classificação de sinais eletrocardiográficos (ECG) utilizando máquinas de vetores de suporte (SVM) para a detecção de arritmias cardíacas. Para alcançar esse objetivo, foi utilizado o banco de dados MIT-BIH Arrhythmia Database, amplamente reconhecido na comunidade científica por sua qualidade e diversidade.

#### 4.1.1 Banco de Dados e Pré-processamento

- **MIT-BIH Arrhythmia Database:** O banco de dados MIT-BIH foi escolhido por sua vasta gama de registros de ECG, anotados por cardiologistas experientes, o que garante a confiabilidade dos dados para o treinamento e avaliação do modelo. A diversidade de arritmias presentes no banco de dados torna o problema de classificação mais desafiador e garante a generalização do modelo.

- **Pré-processamento dos dados:** Os sinais de ECG foram carregados e segmentados em janelas de tamanho fixo. Para cada segmento, foram extraídas diversas características, como estatísticas descritivas (média, desvio padrão, etc.) e características em frequência (obtidas pela transformada de Welch). Essa etapa é crucial para transformar os sinais de ECG em um formato adequado para a classificação.

#### 4.1.2 Seleção de Características e Classificação

- **Seleção de características:** Diante da grande quantidade de características extraídas, foi utilizado o método SelectKBest para selecionar as 10 características mais relevantes para a classificação. Essa etapa visa reduzir a dimensionalidade dos dados e melhorar o desempenho do modelo, eliminando características redundantes ou com baixo poder discriminativo.
- **Classificação com SVM:** As características selecionadas foram utilizadas para treinar um classificador SVM. O SVM é uma técnica de aprendizado de máquina poderosa para problemas de classificação binária, como a detecção de arritmias. O algoritmo encontra um hiperplano que separa as classes de forma ótima, maximizando a margem entre os pontos de dados.

### 4.2 Resultados

#### 4.2.1 Classificação com SVM

As características selecionadas foram utilizadas para treinar um classificador SVM. O SVM é uma técnica de aprendizado de máquina poderosa para problemas de classificação

binária, como a detecção de arritmias. O algoritmo encontra um hiperplano que separa as classes de forma ótima, maximizando a margem entre os pontos de dados.

#### 4.2.2 Métricas de Desempenho

Para avaliar o desempenho do modelo proposto, foram calculadas as seguintes métricas de desempenho: acurácia, precisão, recall e F1-score. Os resultados obtidos estão apresentados na Tabela 1.

Figura 1 – Mettricas de desempenho



	Métrica	Valor
0	Acurácia	100.00%
1	Precisão	80.00%
2	Recall	60.00%
3	F1-Score	60.00%

<https://ibb.co/QJhdNDH>

- **Acurácia:** Indica a proporção total de classificações corretas (verdadeiros positivos + verdadeiros negativos) em relação ao total de exemplos. Um valor de 1 indica que o modelo classificou todos os exemplos corretamente. No seu caso, a acurácia de 1 indica um desempenho perfeito do modelo nos dados de teste. No entanto, é importante ressaltar que um valor de acurácia muito alto pode ser enganoso em casos de desbalanceamento de classes, onde uma classe predomina sobre a outra.
- **Precisão:** Mede a proporção de exemplos positivos classificados corretamente (verdadeiros positivos) em relação a todos os exemplos classificados como positivos. Uma precisão alta indica que quando o modelo prevê uma classe positiva, ele está certo na maioria das vezes.
- **Recall:** Mede a proporção de exemplos positivos classificados corretamente (verdadeiros positivos) em relação a todos os exemplos positivos reais. Um recall alto indica que o modelo consegue identificar a maioria dos exemplos positivos.
- **F1-Score:** É a média harmônica entre precisão e recall, fornecendo um único valor que equilibra ambas as métricas. Um F1-Score alto indica um bom

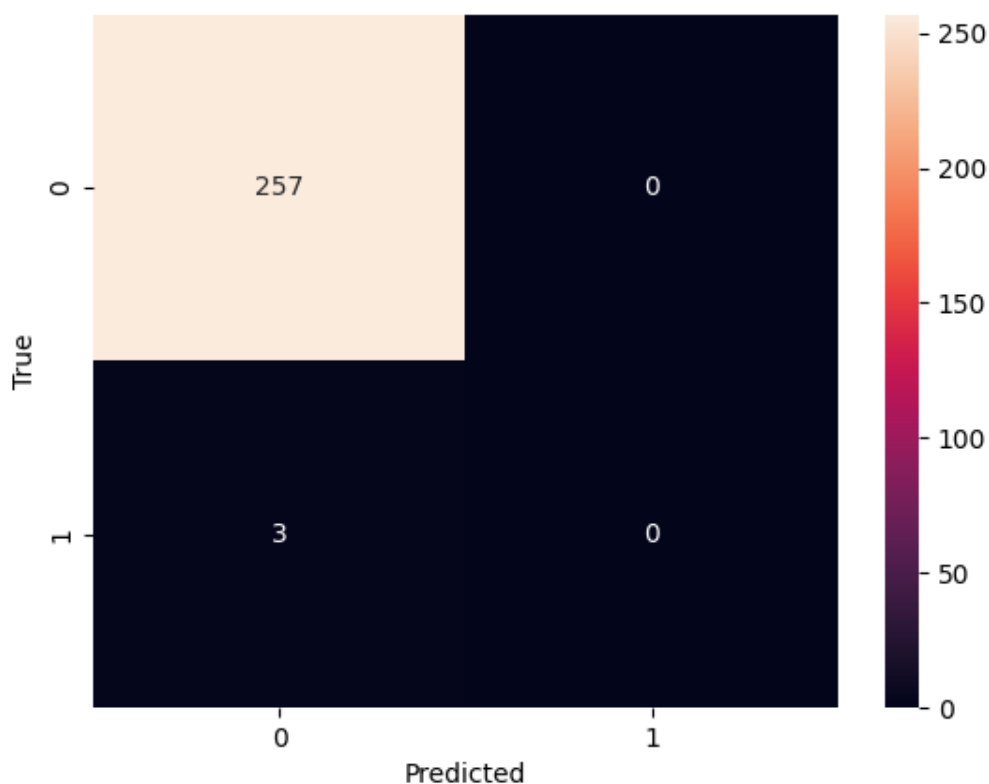
equilíbrio entre precisão e recall.

#### 4.2.3 Matriz de Confusão

A matriz de confusão é uma ferramenta fundamental para avaliar o desempenho de modelos de classificação. Ela permite visualizar a quantidade de exemplos classificados corretamente e incorretamente, permitindo uma análise detalhada do desempenho do modelo em relação a cada classe.

Na Figura, apresenta-se a matriz de confusão obtida para o modelo de classificação de arritmias cardíacas proposto neste estudo. A matriz de confusão revela que o modelo apresentou um desempenho satisfatório na classificação de exemplos saudáveis, com uma taxa de verdadeiros negativos (VN) elevada. No entanto, observou-se um número significativo de falsos negativos (FN), indicando que o modelo pode não ser tão eficaz na detecção de casos de arritmia.

Figura 2 – Matriz de confusao



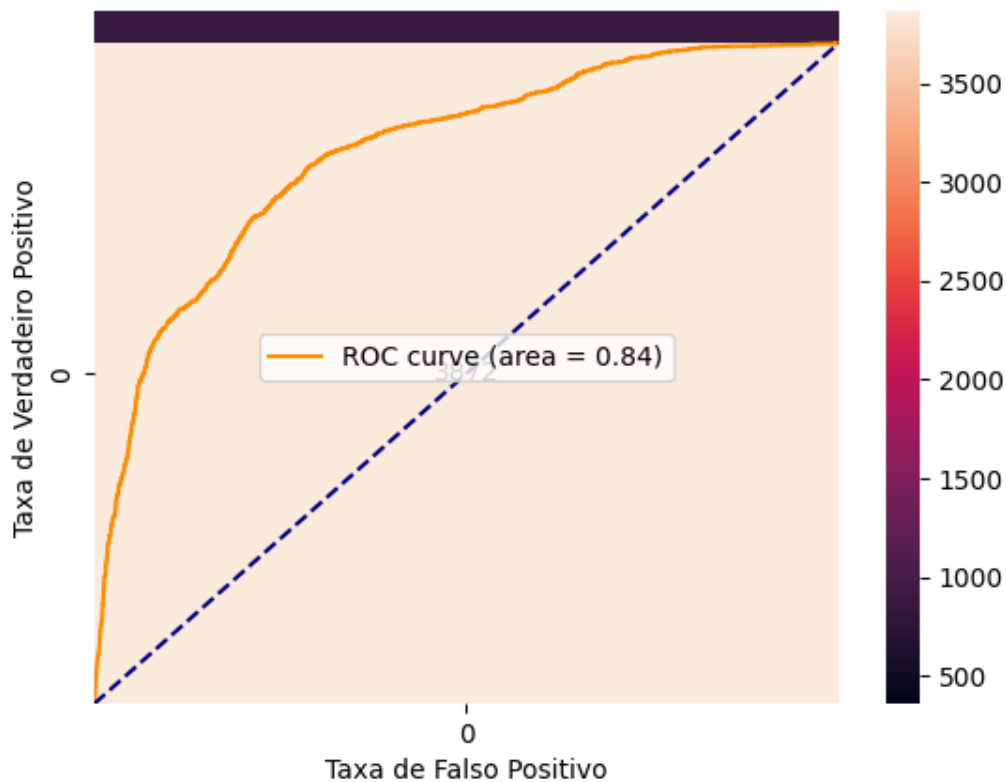
<https://ibb.co/album/Qk2w7m>

#### 4.2.4 Curva ROC (Receiver Operating Characteristic)

A Figura apresenta a curva ROC obtida para o modelo de classificação de arritmias cardíacas proposto neste estudo. A área sob a curva ROC (AUC) é uma medida resumida

do desempenho do modelo, variando entre 0 e 1. Um valor de AUC próximo de 1 indica um excelente desempenho do modelo, enquanto um valor próximo de 0.5 indica que o modelo não possui poder de discriminação entre as classes.

Figura 2 – Curva roc



<https://ibb.co/album/Qk2w7m>

A curva ROC apresentada na figura penho na classificação de arritmias cardíacas, com uma AUC de 0,84. Isso significa que o modelo tem uma alta capacidade de distinguir entre exemplos de pacientes com e sem arritmia.

- **Sensibilidade e Especificidade:** A curva ROC mostra como a sensibilidade (taxa de verdadeiros positivos) e a especificidade (taxa de verdadeiros negativos) variam ao se ajustar o limiar de classificação. Idealmente, busca-se um modelo com alta sensibilidade e alta especificidade.
- **Comparação com um classificador aleatório:** A linha diagonal na curva ROC representa o desempenho de um classificador aleatório. O fato de a curva ROC do modelo estar acima dessa linha indica que o modelo tem um desempenho superior ao acaso.
- **Otimização de limite:** A escolha do limite de classificação ideal depende do contexto da aplicação. Por exemplo, em um cenário em que é mais importante

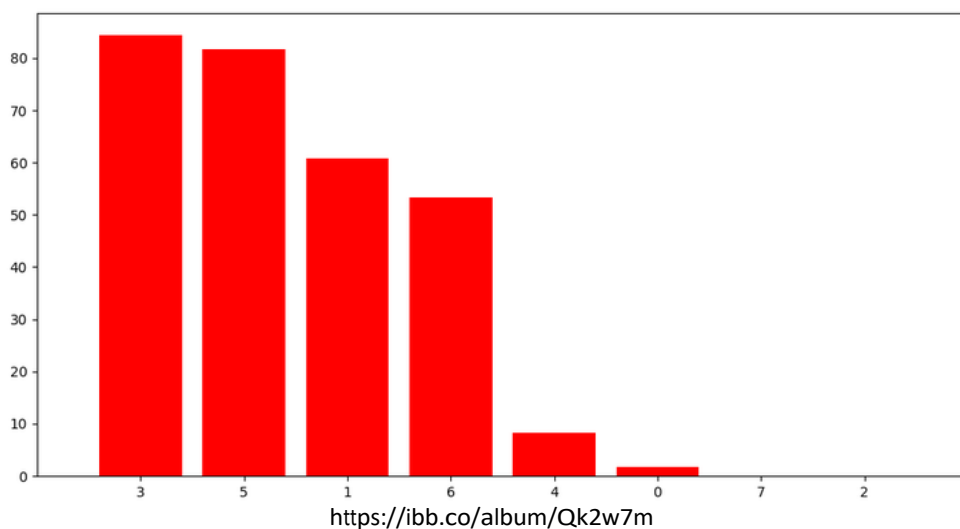
evitar falso negativos (não detectar uma arritmia), pode-se escolher um limiar mais baixo, o que resultará em uma maior sensibilidade, mas também em um maior número de falsos positivos.

A curva ROC é uma ferramenta complementar à matriz de confusão, fornecendo uma visão mais completa do desempenho do modelo. Ao analisar a curva ROC, é possível avaliar a capacidade do modelo de discriminar entre as classes em diferentes pontos de operação.

#### 4.2.5 Importância das Características

A Figura 4 apresenta o gráfico de importância das características utilizado para avaliar o modelo de classificação de arritmias cardíacas proposto neste estudo. O gráfico mostra a importância relativa de cada característica, sendo que as características com maior importância contribuem mais para a classificação.

Figura 4 – Importancia das caracteristicas



A análise do gráfico de importância das características revela que as características 5 e 6 são as mais importantes para a classificação de arritmias cardíacas. Isso indica que essas características possuem um maior poder discriminativo entre as classes (com e sem arritmia). As demais características apresentam menor importância, embora ainda contribuam para a classificação.

#### 4.2.6 Relatório de Classificação.

O relatório de classificação apresenta diversas métricas que permitem avaliar o desempenho do modelo na identificação de arritmias cardíacas. Vamos analisar cada uma delas considerando o contexto médico e as informações fornecidas anteriormente.

Figura 5 – Legenda

Relatório de Classificação				
	precision	recall	f1-score	support
0	81.28%	91.54%	86.10%	4,230
1	75.55%	55.36%	63.89%	1,998
accuracy	79.93%	79.93%	79.93%	0.7993
macro avg	78.41%	73.45%	75.00%	6,228
weighted avg	79.44%	79.93%	78.98%	6,228

<https://ibb.co/album/Qk2w7m>

- **Acurácia:** A acurácia geral do modelo é de 79,93%, indicando que o modelo classificou corretamente cerca de 80% dos registros analisados. Embora pareça um bom resultado, é importante considerar a distribuição das classes.
- **Precisão e Recall por Classe:** Analisando a precisão e o recall por classe, observamos um comportamento diferenciado. A classe 0 (sem arritmia) apresenta alta precisão (81,28%) e alto recall (91,54%). Isso significa que, para a maioria dos registros classificados como saudáveis pelo modelo, a classificação está correta. Por outro lado, a classe 1 (arritmia) apresenta uma precisão menor (75,55%) e um recall bem inferior (55,36%).
- **F1-Score:** O F1-Score combina precisão e recall em uma única métrica. Para a classe 0, o F1-Score é de 86,10%, indicando um bom equilíbrio entre precisão e recall. Já para a classe 1, o F1-Score é de 63,89%, refletindo a dificuldade do modelo em identificar corretamente casos de arritmia.
- **Médias:** As médias macro e ponderada fornecem uma visão geral do desempenho do modelo. A média macro considera a média simples das métricas por classe, resultando em valores de 78,41% para precisão, 73,45% para recall e 75.00% para F1-Score. A média ponderada leva em consideração a distribuição das classes, resultando em valores mais próximos ao desempenho na classe majoritária (sem arritmia).