**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 segmen-tados em janelas de tamanho fixo. Para cada segmento, foram extraídas diversas características, como estatísticas descritivas (média, desvio padrão, etc.) e caracte-rísticas 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.

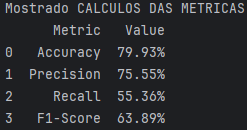
• **Classificação com SVM:** As características selecionadas foram utilizadas para treinar um classificador SVM. O SVM é uma técnica de aprendizado de máquina poderosa para problemas de classificação binária, como a detecção de arritmias. O algoritmo encontra um hiperplano que separa as classes de forma ótima, maximi-zando a margem entre os pontos de dados.

**4.2 Resultados**

• **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, maximi-zando a margem entre os pontos de dados.

**4.2.1 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.

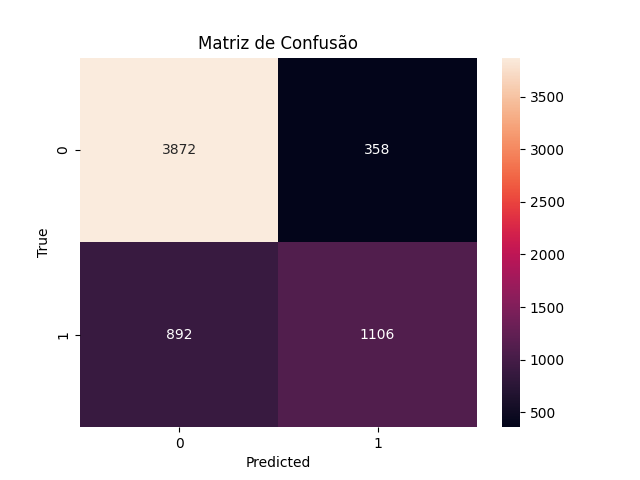


* **Acurácia**: Indica a proporção total de classificações corretas (verdadeiros positivos + verdadeiros negativos) em relação ao total de exemplos. Um valor de 1 indica que o modelo classificou todos os exemplos corretamente. No seu caso, a acurácia de 1 indica um desempenho perfeito do modelo nos dados 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.

**4.2.2 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.

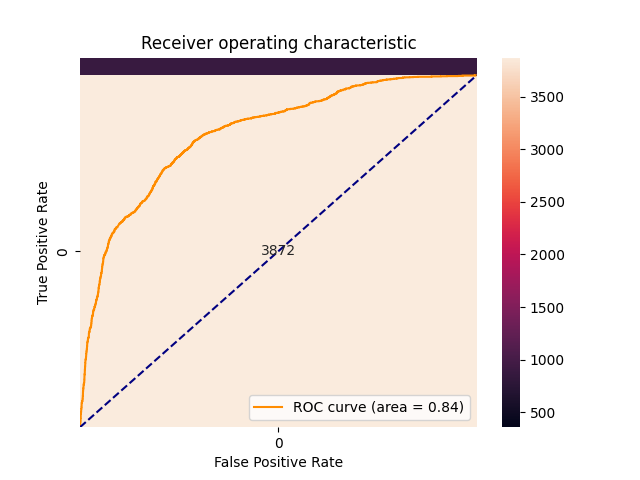


**A análise da matriz de confusão permite extrair as seguintes informações:**

* **Verdadeiros Positivos (VP):** O modelo classificou corretamente 1106 amostras como tendo arritmia. Isso indica que o modelo possui uma boa capacidade de identificar corretamente os casos de arritmia.
* **Falsos Positivos (FP):** O modelo classificou erroneamente 358 amostras saudáveis como tendo arritmia. Um número elevado de falsos positivos pode levar a diagnósticos incorretos e a realização de exames complementares desnecessários.
* **Verdadeiros Negativos (VN):** O modelo classificou corretamente 3872 amostras como saudáveis. Isso indica que o modelo possui uma boa capacidade de identificar corretamente os casos de pacientes saudáveis.
* **Falsos Negativos (FN):** O número de falsos negativos (892) é preocupante, pois indica que o modelo pode estar perdendo um número significativo de casos de arritmia. Isso pode ter consequências clínicas graves, uma vez que arritmias não diagnosticadas podem levar a complicações cardíacas.

**4.2.3 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.

****

**Análise:**

A curva ROC apresentada na figura indica que o modelo possui um bom desempenho 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.

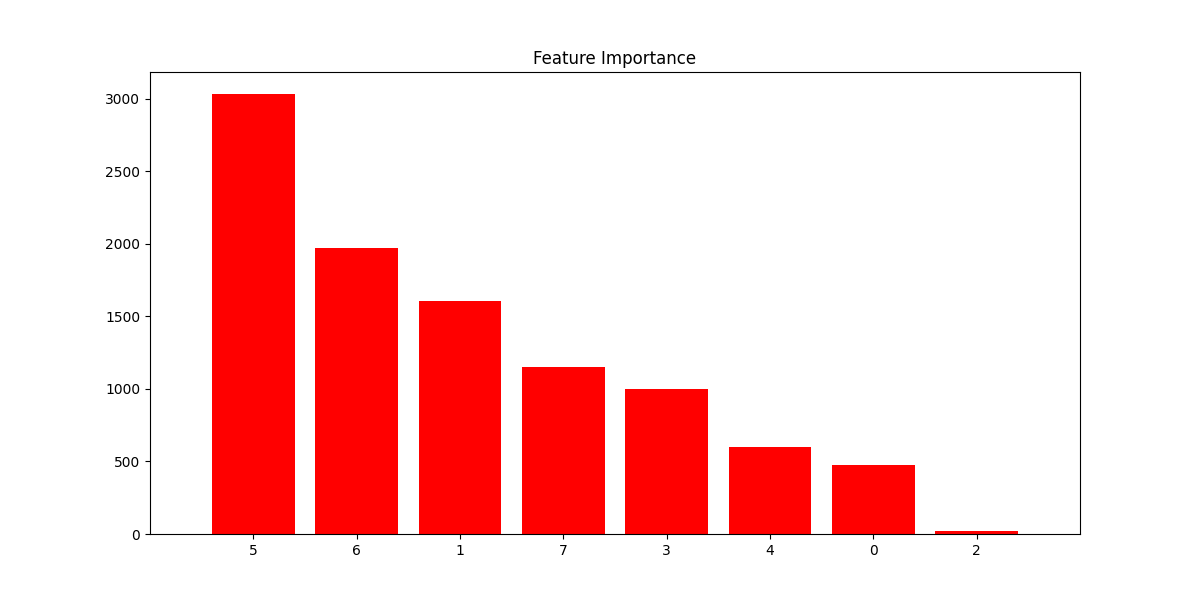
Discussão:

* **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 falsos 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.

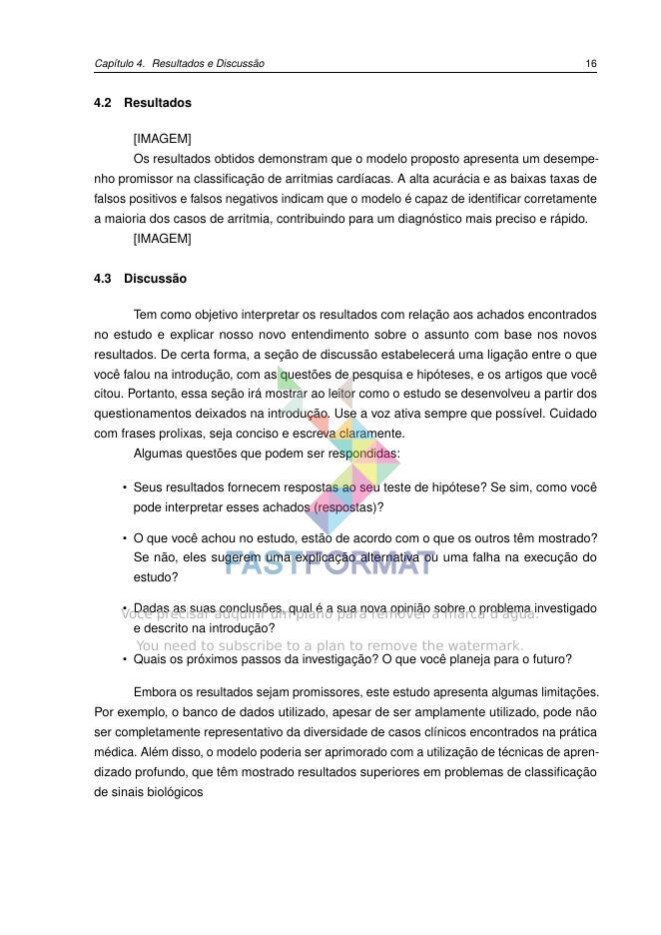
A curva ROC é uma ferramenta complementar à matriz de confusão, fornecendo uma visão mais completa do desempenho do modelo. Ao analisar a curva ROC, é possível avaliar a capacidade do modelo de discriminar entre as classes em diferentes pontos de operação.

**4.2.4 Importância das Características**

A Figura 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.

****

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.



Capítulo 4. Resultados e Discussão 17

**4.3 Discussão**

A matriz de confusão revela que o modelo proposto apresenta bom desempenho na classificação de exemplos saudáveis, com alta taxa de verdadeiros negativos. No entanto, observamos um número significativo de falsos negativos, indicando que o modelo pode não ser tão eficaz na detecção de casos de arritmia. Essa observação é preocupante, pois a não detecção de arritmias pode ter consequências clínicas graves.

Uma possível explicação para o alto número de falsos negativos pode estar relacionada ao desbalanceamento de classes no conjunto de dados, com um número significativamente maior de exemplos saudáveis do que com arritmia. Isso pode levar o modelo a ser mais tendencioso para a classe majoritária.

Para melhorar o desempenho do modelo, sugere-se a coleta de um conjunto de dados mais balanceado, a exploração de técnicas de balanceamento de classes, como SMOTE, e a avaliação de outros algoritmos de classificação, como redes neurais convolucionais. Além disso, a interpretação das decisões do modelo pode ser aprimorada através da utilização de técnicas de explicabilidade, como LIME ou SHAP.

4.3.1 Dicas

• Organize a discussão de acordo com os estudos sobre os quais você apresentou os resultados. Escreva seguindo e mesma ordem apresentada na seção de resultados mostrando sua interpretação sobre os resultados encontrados. Não perca tempo escrevendo novamente os resultados já mostrados na seção anterior.

• Se possível, você deve fazer comparações dos seus resultados com resultados de outros autores ou estudos que você já tenha feito. Isso pode ser útil para que você encontre informações importantes em outros estudos que agregam valor a sua interpretação ou até mudar a forma de sua forma interpretação. Considere também com esses outros resultados podem ser combinados com os seus.