

João Victor Sousa

Título do seu TCCe

Ilhéus - Bahia

2025

João Victor Sousa

Título do seu TCCe

Trabalho apresentado ao Curso de Y da Universidade X como requisito parcial para obtenção do título de Bacharel.

Universidade Estadual de Santa Cruz
Curso de Graduação em Ciência da Computação

Orientador: Otacílio José Pereira

Ilhéus - Bahia
2025

Resumo

Escreva aqui o resumo do seu trabalho. Inclua objetivos, metodologia, resultados e conclusões.

Palavras-chave: palavra1. palavra2. palavra3.

Sumário

Sumário	3	
Lista de tabelas	4	
Lista de ilustrações	5	
1	INTRODUÇÃO	7
2	FUNDAMENTAÇÃO TEÓRICA	8
2.1	ECG	8
2.1.1	O funcionamento do coração	8
2.1.2	Tipos de arritmia	8
2.1.3	O batimento cardíaco	8
2.1.4	O que é ECG?	9
3	METODOLOGIA	10
3.1	Particionamento dos dados e classes	10
3.2	Métricas	11
3.3	Tipos de redes usadas	13
3.4	Pré-processamento e entrada para os modelos	14
3.5	Arquiteturas	14
4	RESULTADOS E DISCUSSÕES	17
4.1	Resultados do modelo GRU	17
4.2	Resultados do modelo híbrido GRU e CNN	23
5	ANÁLISE DE ERROS NO PIOR <i>FOLD</i>	30
5.1	Análise de erros do modelo híbrido CNN com GRU	30
6	CONCLUSÃO	31
	REFERÊNCIAS	32

Lista de tabelas

Tabela 1	–	Particionamento inter-paciente proposto por Chazel et al.	10
Tabela 2	–	Mapeamento das anotações originais do MIT-BIH para as classes AAMI.	11
Tabela 3	–	Exemplo de matriz de confusão binária	13
Tabela 4	–	Resultados do GRU (N vs. V) na validação	17
Tabela 5	–	Resultados do GRU (N vs. V) no treino	17
Tabela 6	–	Resultados do modelo híbrido CNN e GRU (N vs. V) na validação	24
Tabela 7	–	Resultados do modelo híbrido CNN e GRU (N vs. V) no treino	24
Tabela 8	–	Total dos erros e acertos por paciente no <i>fold</i> de validação	30

Lista de ilustrações

Figura 1 – Curva ROC de um Classificador Perfeito: Comparação com Modelo Aleatório.	12
Figura 2 – Curva Precisão–Recall de um Classificador Perfeito: Comparação com Baseline.	13
Figura 3 – Arquitetura GRU pura.	15
Figura 4 – Arquitetura híbrida CNN e GRU.	16
Figura 5 – Métricas do modelo 3 por fold <i>fold</i>	18
Figura 6 – Matriz de confusão do modelo 3 em seu melhor <i>fold</i>	19
Figura 7 – Matriz de confusão do modelo 3 em seu pior <i>fold</i>	20
Figura 8 – Curva <i>ROC</i> modelo 3 em seu melhor <i>fold</i>	20
Figura 9 – Curva <i>ROC</i> do modelo 3 em seu melhor <i>fold</i>	21
Figura 10 – Curva precisão vs <i>recall</i> do modelo 3 em seu pior <i>fold</i>	22
Figura 11 – Curva precisão vs <i>recall</i> do modelo 3 em seu melhor <i>fold</i>	23
Figura 12 – Métricas do modelo híbrido CNN com GRU por <i>fold</i>	24
Figura 13 – Matriz de confusão modelo 4 em seu melhor <i>fold</i>	25
Figura 14 – Matriz de confusão modelo 4 em seu melhor <i>fold</i>	26
Figura 15 – Curva <i>ROC</i> modelo 4 em seu melhor <i>fold</i>	26
Figura 16 – Curva precisão vs <i>recall</i> do modelo 4 em seu melhor <i>fold</i>	27
Figura 17 – Curva <i>ROC</i> modelo 4 em seu pior <i>fold</i>	28
Figura 18 – Curva precisão vs <i>recall</i> do modelo 4 em seu melhor <i>fold</i>	28

Lista de Siglas

AUC	Area Under Curve
AP	Average Precision
RNN	Redes Neurais Recorrentes
CNN	Redes Neurais Convolucionais
LSTM	Long Short-Term Memory
AAMI	Association for the Advancement of Medical Instrumentation
GRU	Gated Unit Recurrent
ECG	Eletrocardiograma
TP	Verdadeiro Positivo
FP	Falso Positivo
TN	Verdadeiro Negativo
FN	Falso Negativo

1 Introdução

Texto da introdução.

2 Fundamentação Teórica

2.1 ECG

2.1.1 O funcionamento do coração

O coração é um órgão muscular composto por quatro câmaras que se contraem em sequência regular, bombeando o sangue de forma eficiente (CASCINO; SHEA, 2025). As contrações são controladas por correntes elétricas que percorrem o coração com precisão e velocidade regulada.

O processo inicia no nódulo sinoatrial, localizado no átrio direito, que funciona como marcapasso natural do coração. A frequência cardíaca é determinada pela frequência de disparos desse nódulo, sendo modulada pelo sistema nervoso autônomo e por hormônios presentes na corrente sanguínea (CASCINO; SHEA, 2025). O sistema nervoso simpático aumenta a frequência cardíaca, enquanto o parassimpático, por meio do nervo vago, a reduz. Hormônios como a epinefrina (adrenalina) e norepinefrina (noradrenalina), produzidos pelo sistema simpático, também elevam a frequência cardíaca. Além disso, o hormônio tireoidiano liberado pela glândula da tireoide exerce influência na frequência cardíaca.

Em repouso, a frequência cardíaca em adultos varia entre 60 e 100 batimentos por minuto, sendo geralmente mais baixa em indivíduos jovens e em bom condicionamento físico (CASCINO; SHEA, 2025).

2.1.2 Tipos de arritmia

As arritmias podem ser classificadas de forma simplificada em três tipos principais:

- Taquicardia: frequência excessivamente rápida;
- Bradicardia: frequência excessivamente lenta;
- Irregular: quando os impulsos percorrem o coração por vias irregulares.

2.1.3 O batimento cardíaco

O batimento cardíaco inicia-se no nódulo sinoatrial, cuja corrente elétrica atravessa o átrio direito e, em seguida, o átrio esquerdo, promovendo sua contração. O sangue é então impulsionado dos átrios para os ventrículos. A corrente elétrica passa pelo nódulo atrioventricular, único ponto de conexão entre átrios e ventrículos, que retarda o impulso, garantindo enchimento completo dos ventrículos.

Em seguida, o impulso segue pelo feixe de His, que se divide em ramos para conduzir a corrente a cada ventrículo, permitindo sua ativação uniforme e subsequente contração, bombeando o sangue para o corpo (CASCINO; SHEA, 2025).

2.1.4 O que é ECG?

Segundo Cascino e Shea (CASCINO; SHEA, 2025), o eletrocardiograma (ECG) é um exame não invasivo que registra a atividade elétrica do coração. Ele é realizado pela colocação de eletrodos na pele do paciente, geralmente 12, chamados de derivações. Esses eletrodos registram tanto a direção quanto a magnitude da corrente elétrica.

O registro resultante gera uma onda que reflete a atividade elétrica do coração. Cada etapa do ciclo cardíaco é representada na morfologia do traçado: a onda **P** corresponde à ativação dos átrios, o complexo **QRS** à ativação dos ventrículos e a onda **T** ao processo de repolarização ventricular.

O ECG é uma ferramenta fundamental no diagnóstico de problemas cardíacos, permitindo identificar, por exemplo, episódios de infarto do miocárdio, oferta insuficiente de sangue e oxigênio ao coração (isquemia), hipertrofia das paredes cardíacas e diferentes tipos de arritmias.

3 Metodologia

Primeiramente, foi necessário definir qual banco de dados seria utilizado para o treinamento e a validação. Optou-se pelo *MIT-BIH Arrhythmia Database* (MOODY; MARK, 2005), recomendado pela AAMI. O banco é composto por 58 registros de eletrocardiograma (ECG), cada um com 30 minutos de duração. Os 23 primeiros registros foram selecionados aleatoriamente a partir de um conjunto de 4000 gravações de 24 horas realizadas em pacientes ambulatoriais do Beth Israel Deaconess Medical Center. Os 25 registros restantes foram escolhidos de modo a incluir arritmias raras, mas clinicamente significativas.

Uma característica importante desse banco é a anotação de cada batimento cardíaco em torno do pico R, realizada por três cardiologistas independentes.

3.1 Particionamento dos dados e classes

Os dados foram particionados seguindo a estratégia inter-paciente proposta por Chazel et al. (apud Silva et al. (2025)), na qual batimentos de um mesmo paciente não podem aparecer simultaneamente nos conjuntos de treinamento e validação. O objetivo é garantir a capacidade de generalização do modelo para diferentes pacientes. Além disso, conforme recomendado pela AAMI, registros de pacientes com marcapasso foram excluídos.

Os registros 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 e 230 são normalmente chamados de DS1. Os demais (100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 e 234) de DS2.

De acordo com a AAMI (apud Silva et al. (2025)), são definidas cinco classes de arritmia: N, SVEB, VEB, F e Q, correspondentes a batimento normal, batimento supraventricular ectópico, batimento ventricular ectópico, fusão de batimento ventricular e normal e batimento não classificado, respectivamente. A Tabela 1 apresenta a distribuição dessas classes no conjunto de dados.

Tabela 1 – Particionamento inter-paciente proposto por Chazel et al.

Conjunto	N	SVEB	VEB	F	Q	Total
DS1	45 866	944	3 788	415	8	51 021
DS2	44 259	1 837	3 221	388	7	49 712
Total	90 125	2 781	7 009	803	15	100 733

Fonte: Adaptado de Silva et al. (2025).

O conjunto DS1 foi então subdividido em treinamento e validação por meio de

validação cruzada, inicialmente com duas partições (dois *folds*) e, posteriormente, com cinco partições (cinco *folds*) nos modelos finais.

É importante ressaltar que o MIT-BIH detalha muito outros tipos de batimentos. Essas cinco classes são uma proposta da AAMI, feita a partir do agrupamento desses tipos. Uma lista das anotações pode ser encontrada em (GOLDBERGER et al., n.d.). O mapeamento entre os tipos descritos originalmente e as cinco classes foi feito da seguinte maneira:

Tabela 2 – Mapeamento das anotações originais do MIT-BIH para as classes AAMI.

Anotação Original	Classe AAMI
N, e, j, L, R	N (Normal)
A, a, J, S	S (Supraventricular)
V, E	V (Ventricular)
F, f	F (Fusão)
Q, ?, /	Q (Desconhecida)

Fonte: Elaborado pelo autor.

Considerando o desbalanceamento dos dados e visando maior simplicidade, adotou-se a classificação binária; onde a classe positiva corresponde a arritmia ventricular e a negativa ao batimento normal.

A arritmia ventricular é o tipo arritmico mais prevalente no MIT-BIH e apresenta uma morfologia marcante. Ela é composta dos tipos arritmicos: contração prematura contração ventricular prematura (classe V) e batimento de escape ventricular (classe E).

3.2 Métricas

As métricas utilizadas para avaliar o desempenho dos modelos foram: sensibilidade, precisão, acurácia, *F1-score*, AUC (*Area Under the Curve*) e AP (textitAverage Precision). Esses dois últimos são exibidos juntos aos gráficos *ROC* e *PR*, respectivamente.

A sensibilidade representa a capacidade do modelo em identificar corretamente as classes positivas, isto é, os batimentos arritmicos. Sua equação é dada por:

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (3.1)$$

em que TP são os verdadeiros positivos e FN os falsos negativos.

A precisão, por sua vez, indica a proporção de batimentos classificados como arritmicos que realmente pertencem a essa classe:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (3.2)$$

onde FP representa os falsos positivos. Precisão e sensibilidade estão relacionadas por um *trade-off*. No contexto médico, prioriza-se elevada sensibilidade, ainda que à custa de menor precisão, uma vez que falsos negativos são mais prejudiciais que falsos positivos.

O $F1-score$ é a média harmônica entre precisão e sensibilidade, buscando um equilíbrio entre ambas:

$$F1-score = \frac{2 \cdot \text{Precisão} \cdot \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (3.3)$$

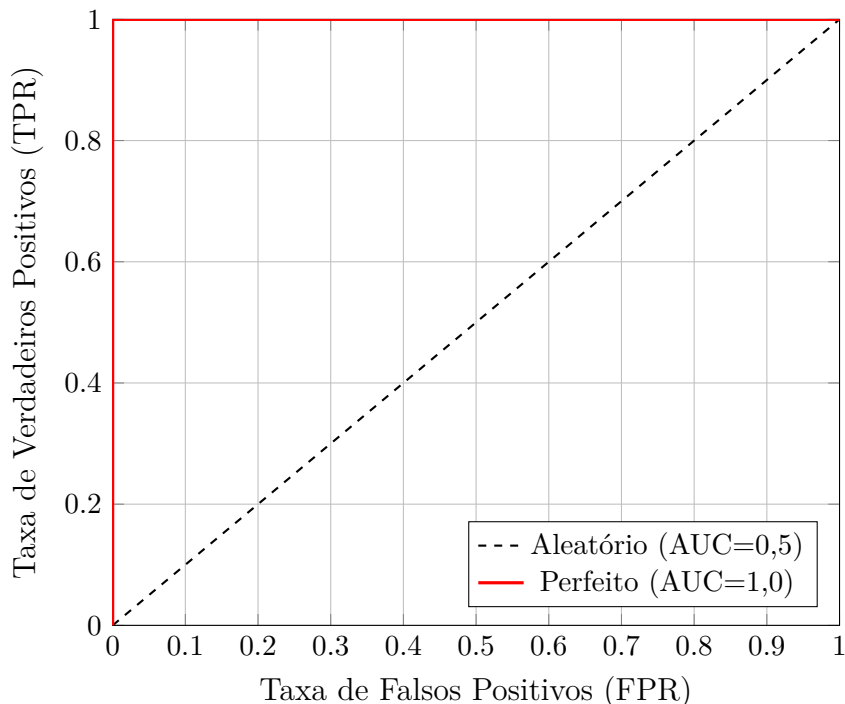
A acurácia corresponde ao acerto global do modelo, considerando tanto as classes positivas quanto as negativas:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

A AUC mede a capacidade do modelo em separar as classes positivas das negativas, variando entre 0 e 1. Valores próximos de 1 indicam separação perfeita, enquanto 0,5 corresponde a um modelo com desempenho equivalente ao acaso, o *baseline*.

Essa métrica é calculada a partir da área sob a curva ROC. Na Figura 1, é ilustrado a curva ROC de um classificador perfeito.

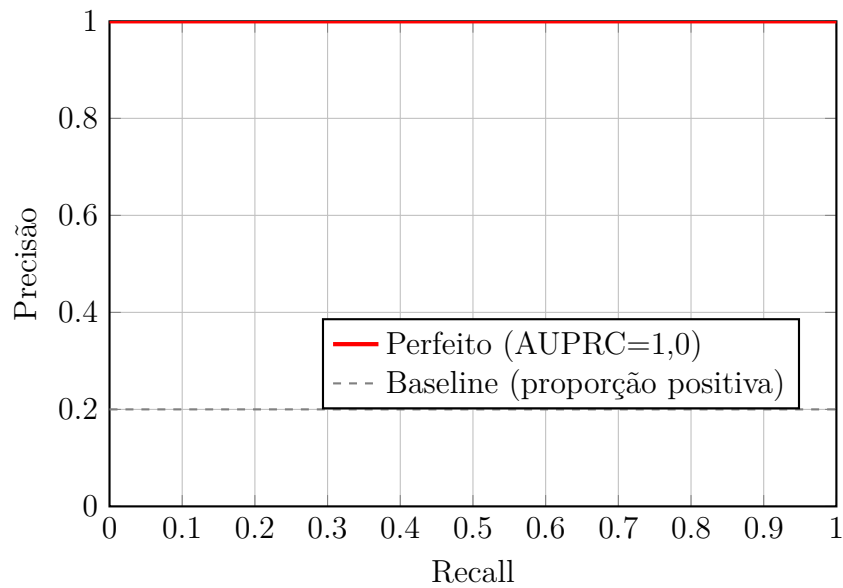
Figura 1 – Curva ROC de um Classificador Perfeito: Comparação com Modelo Aleatório.



Fonte: Elaborado pelo autor.

Já a curva PR, Precisão vs *Recall* representa a precisão em função do *recall*. Na figura Figure 1, é ilustrada a curva PR de um classificador perfeito.

Figura 2 – Curva Precisão–Recall de um Classificador Perfeito: Comparação com Baseline.



Fonte: Elaborado pelo autor.

Na figura ?? representa uma curva PR genérica, a curva mostra a precisão em função do *recall*. Como ilustrado, conforme aumenta-se o *recall*, há uma diminuição da precisão e vice-versa; representado o *trade-off*. O AP é a área do gráfico; quanto maior, melhor. Um modelo aleatório teria um AP igual a frequência da classe positiva.

A matriz de confusão, por fim, fornece uma representação tabular dos acertos e erros do modelo, como exemplificado na Tabela 3.

Tabela 3 – Exemplo de matriz de confusão binária

Classe Verdadeira	Classe Preditada	
	Positiva	Negativa
Positiva	TP	FN
Negativa	FP	TN

Fonte: Elaborado pelo autor.

Essas métricas em conjunto permitem avaliar não apenas a proporção global de acertos, mas também a capacidade do modelo em detectar corretamente arritmias, aspecto essencial em aplicações médicas.

3.3 Tipos de redes usadas

Inicialmente, foram escolhidas redes neurais recorrentes (RNNs) e seus subtipos, como LSTM e GRU. Segundo James et al. (2023), esse tipo de rede apresenta grande potencial para lidar com dados sequenciais, como no processamento de linguagem natural,

previsão de preços e outros tipos de séries temporais. Como o componente temporal é relevante para o diagnóstico das arritmias, optou-se por esse tipo de modelo.

Além das RNNs, foram utilizadas redes neurais convolucionais (CNNs), conhecidas por sua habilidade em reconhecer padrões em diferentes domínios James et al. (2023). Em particular, CNNs unidimensionais (1D-CNNs) têm se mostrado eficazes na análise de sinais fisiológicos, sendo amplamente aplicadas à classificação de ECG (NAROTAMO et al., 2024).

A motivação para essa combinação está na complementaridade entre os modelos: enquanto as RNNs são eficazes na captura de dependências temporais, as CNNs se destacam na identificação de características morfológicas do sinal.

3.4 Pré-processamento e entrada para os modelos

Antes de usar o sinal do ECG como entrada, ele precisou passar por uma etapa de pré-processamento que consistiu em uma limpeza de ruídos e segmentação. Para a diminuição do ruído, foi utilizando um filtro passa-alta de ordem 5 de 0,5 hz, seguido por uma filtragem de linha de energia de 60hz. Isso foi importante para remover ruídos musculares e ruídos oriundos da alimentação dos aparelhos.

Em seguida, os batimentos foram segmentados em batimentos individuais. Nas duas etapas foram utilizadas a biblioteca NeuroKit2 (MAKOWSKI et al., 2021)

O objetivo do trabalho é a classificação de batimentos cardíacos em duas classes: normais e arritmia ventricular. Para isso, os modelos recebem uma sequência de 16 batimentos e realizam a classificação do último batimento da sequência. Cada sequência é composta exclusivamente por batimentos de um único paciente

Tanto as CNNs quanto as RNNs recebem como entrada uma matriz tridimensional com a seguinte estrutura: (tamanho do batch, tamanho da sequência, número de features).

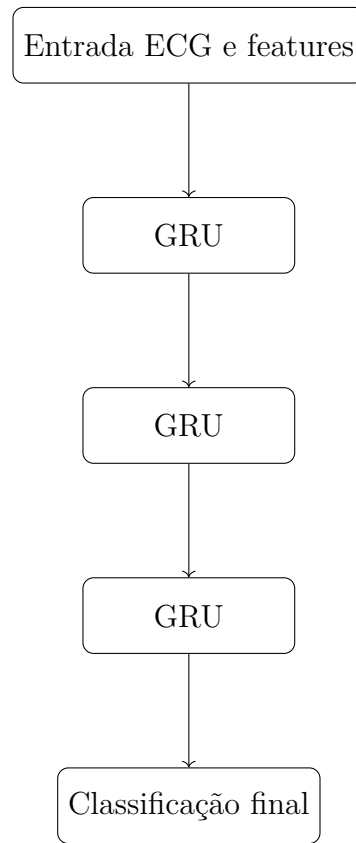
Para otimização do processo de treinamento, foram utilizados os mecanismos de *early stopping* e *reduce on plateau*, responsáveis por limitar o número de épocas e ajustar dinamicamente a taxa de aprendizagem, respectivamente.

3.5 Arquiteturas

Foram testadas dois tipos de arquiteturas, uma é o uso de RNNs puras e a outra é uma arquitetura híbrida com CNNs.

A primeira arquitetura de pura é composta por três camadas de GRUs com 256 unidades ocultas. Essa arquitetura foi utilizada em Narotamo et al. (2024), onde obteve o melhor desempenho. A diferença é que nesse trabalho, além da rede receber o sinal do ECG, ela também recebeu os intervalos RRs pré e pós:

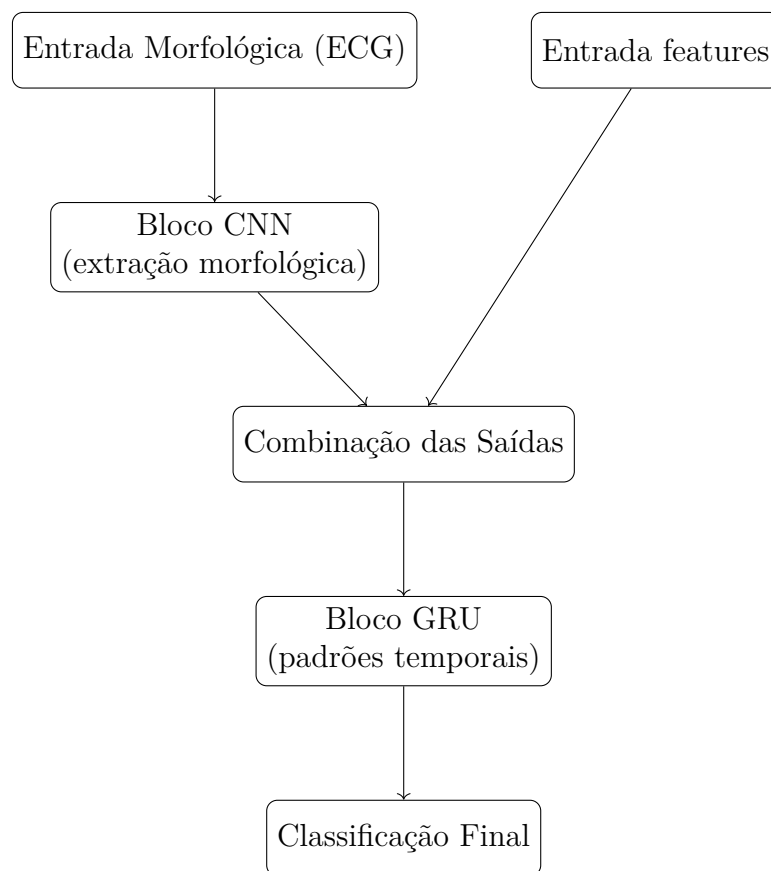
Figura 3 – Arquitetura GRU pura.



Fonte: Elaborado pelo autor.

A segunda rede é uma híbrida de CNN com GRU:

Figura 4 – Arquitetura híbrida CNN e GRU.



Fonte: Elaborado pelo autor.

O bloco de CNN precisou ser aplicado em cada batimento dentro da sequência. Trata-se de duas camadas de CNN com 32 e 64 filtros respectivamente e cada uma seguida por uma camada de *batch normalization* e *global max pooling* para evitar sobre ajuste e reduzir as *features* respectivamente.

Enquanto que a rede da figura 3 recebeu o ECG concatenado com as *features*, a rede híbrida as recebeu separadas, sendo conectadas após o processamento das CNNs.

4 Resultados e discussões

4.1 Resultados do modelo GRU

A seguir, os resultados alcançados pelo modelo descrita em 3:

Tabela 4 – Resultados do GRU (N vs. V) na validação

Métrica	Média	Desvio Padrão
Precisão	0.8515	0.1825
<i>Recall</i>	0.8039	0.0795
<i>F1-Score</i>	0.8060	0.0760
Acurácia	0.9640	0.0278

Fonte: Elaborado pelo autor.

Na tabela 4, tem-se as métricas médias com seus respectivos desvio padrão na *cross-validação* de cinco *folds*. Os resultados indicam que o modelo achou aproximadamente 80% dos casos positivos, com um desvio padrão relativamente baixo, indicando boa estabilidade. Além disso, a precisão do modelo foi maior que seu *recall*, indicando um perfil mais conservador na classificação.

A seguir os resultados no treino:

Tabela 5 – Resultados do GRU (N vs. V) no treino

Métrica	Média	Desvio Padrão
Precisão	0.9872	0.0121
<i>Recall</i>	0.9782	0.0150
F1-Score	0.9827	0.0134
Acurácia	0.9969	0.0024

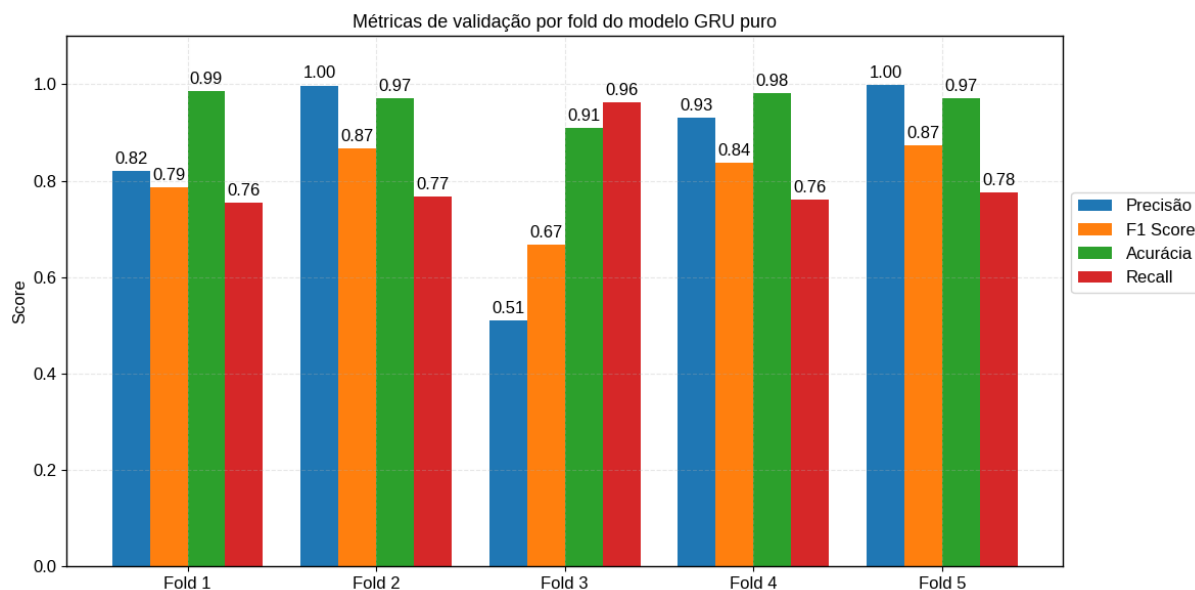
Fonte: Elaborado pelo autor.

Comparando os resultados da tabela 5 com o da 4, há uma evidência de sobreajuste; isto é, o modelo apresenta uma falha em sua capacidade de generalização. Durante o treino de um modelo de aprendizado de máquina, busca-se a partir de uma amostra da população estimar uma curva que melhor se encaixa na população. Modelos flexíveis como uma rede neural tem uma grande capacidade de se ajustar a essa amostra e, caso ela seja pequena, por exemplo, o modelo pode acabar aprendendo as particularidades da amostra ao invés de padrões generalizáveis.

No caso do MIT-BIH; o desbalanceamento junto com as diferenças entre os batimentos de cada paciente pode ter causado esse sobre-ajuste.

Na figura 5, está os resultados alcançado pelo modelo em cada *fold* na validação:

Figura 5 – Métricas do modelo 3 por fold *fold*



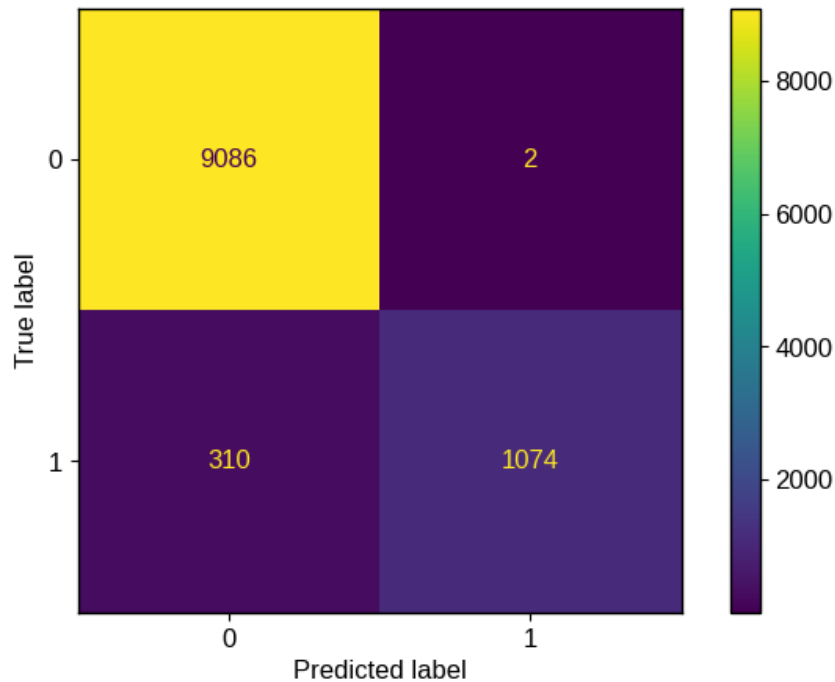
Fonte: Elaborado pelo autor.

No *fold* três, o modelo obteve sua menor precisão, aproximadamente 0,51 porém obteve um alto *recall*, aproximadamente 0,96. Essa discrepância sugere que neste *fold*, havia batimentos normais que fugiam do padrão aprendido no treino, fazendo com que o modelo confundisse eles com batimentos da classe V. Nos demais *folds*, a precisão foi maior que o *recall*, sugerindo a presença de arritmias com características mais sutis, que fizeram com que o modelo as confundisse com batimentos normais.

Considerando o *f1-score*, o terceiro *fold* foi eleito o pior. Como o *fold* cinco empatou com o segundo por esse mesmo critério, como desempate, aquele com o maior *recall*, o quinto, foi considerado o melhor.

Na figura 6, está a matriz de confusão do modelo em seu melhor *fold*:

Figura 6 – Matriz de confusão do modelo 3 em seu melhor *fold*



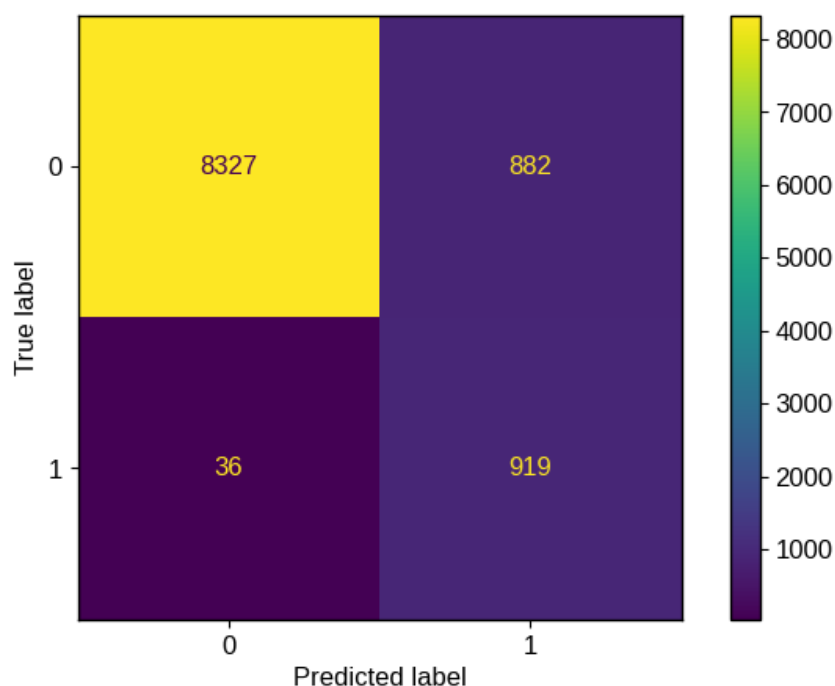
Fonte: Elaborado pelo autor.

Na matriz, é possível ver o desbalanceamento das classes. Neste *fold*, o número de sequencias pertencentes a classe negativa é 9.088, enquanto que 1.384 pertencem a positiva; ou seja, aproximadamente, 13,21% de todas as sequencias são da classe positiva.

A maioria dos erros cometidos são de falsos negativos; o modelo classificou 310 sequencias arrítmicas como normais e apenas duas normais como arrítmicas. Algo que já era evidenciado no gráfico 5, pois sua precisão foi maior que seu *recall*.

Na figura 7, é dada a matriz de confusão no pior *fold*

Figura 7 – Matriz de confusão do modelo 3 em seu pior *fold*

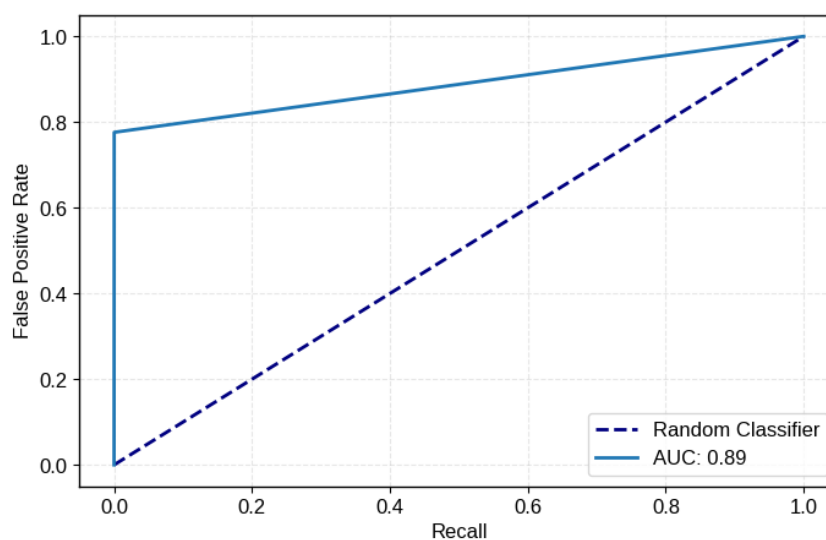


Fonte: Elaborado pelo autor.

Aqui o desbalanceamento foi mais severo; havia 9.209 classes negativas e 955 classes positivas; 9,39% aproximadamente. Neste *fold*, a situação se inverte: a maioria dos erros foram de falsos positivos, confirmando o que foi visto no gráfico 5.

A seguir, na figura 15, a curva ROC no melhor *fold*:

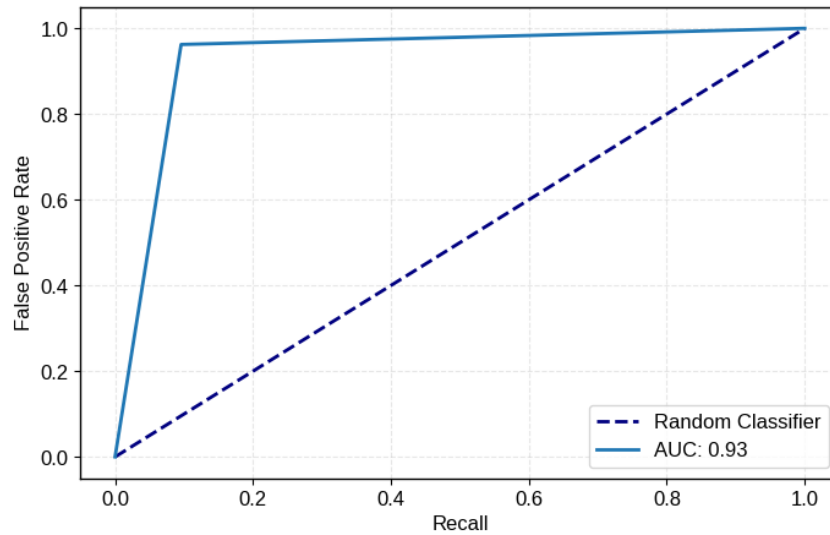
Figura 8 – Curva ROC modelo 3 em seu melhor *fold*



Fonte: Elaborado pelo autor.

Considerando que o *baseline*, um classificador aleatório, tem um *ROC* de 0,5, o melhor foi significativamente melhor.

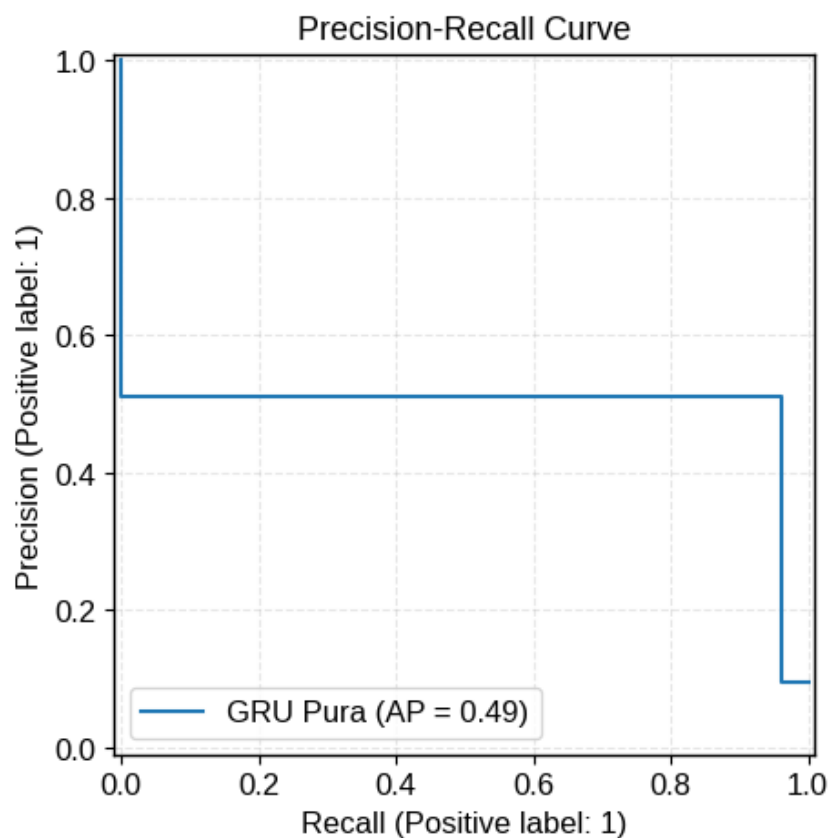
Figura 9 – Curva *ROC* do modelo 3 em seu melhor *fold*



Fonte: Elaborado pelo autor.

No pior fold, 9, o modelo ainda conseguiu manter uma performance satisfatória, com um *ROC* de 0,87. Entretanto, devido ao desbalanceamento dos conjuntos, o desempenho pode ser melhor analisado com a curva PR:

Figura 10 – Curva precisão vs *recall* do modelo 3 em seu pior *fold*

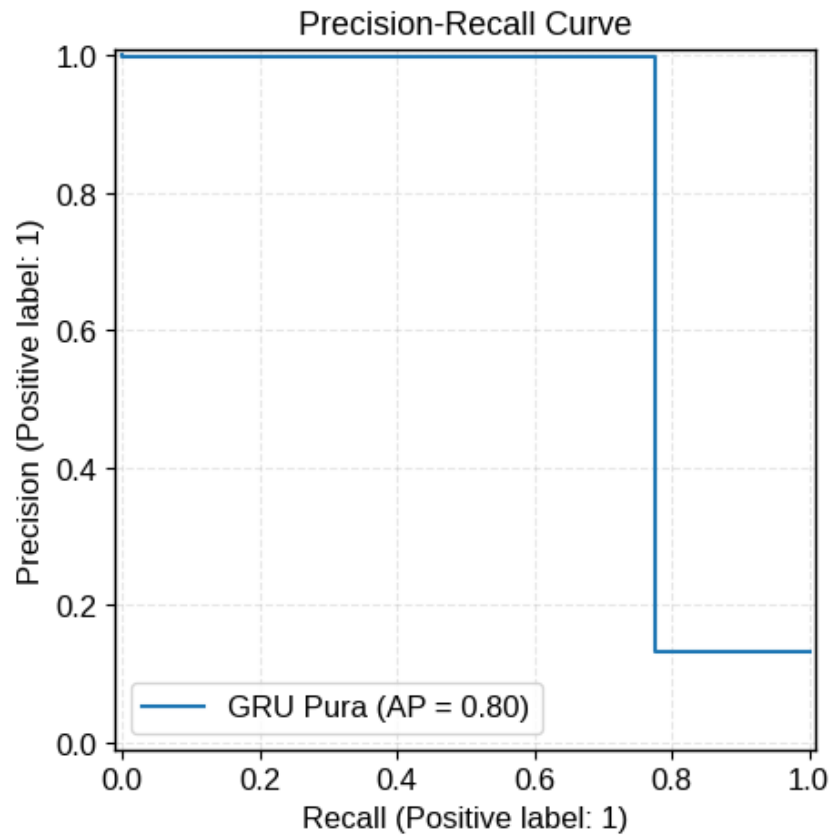


Fonte: Elaborado pelo autor.

Nesse gráfico, o *baseline* não é fixo, mas igual a prevalência da classe positivos. No pior *fold*, a proporção foi de aproximadamente 9,39%, contrastando com o 49% alcançado pelo modelo. Entretanto, a precisão foi baixa. Pelo gráfico, é possível ver que, por exemplo, seria possível ter um recall de 80% porém com uma precisão menor que 60%.

No melhor caso:

Figura 11 – Curva precisão vs *recall* do modelo 3 em seu melhor *fold*



Fonte: Elaborado pelo autor.

Nesse *fold*, o modelo alcançou um *AP* de 80%, enquanto que a proporção de casos positivos foi de 13,21%. No melhor caso, entretanto, o modelo para ter 80% de *recall*, teria que baixar sua precisão para menos de 20%.

Apesar do desbalanceamento, o modelo alcançou resultados satisfatórios, considerando o extremo desbalanceamento do conjunto.

4.2 Resultados do modelo híbrido GRU e CNN

O modelo híbrido apresentou resultado superior em relação ao modelo descrito em 3. Na tabela 6 abaixo, é possível ver que o modelo obteve média maior em todas as métricas. Apesar disso, o modelo obteve um desvio padrão maior no *recall* e *F1-score*.

Tabela 6 – Resultados do modelo híbrido CNN e GRU (N vs. V) na validação

Métrica	Média	Desvio Padrão
Precisão	0.8800	0.1684
<i>Recall</i>	0.8726	0.0857
<i>F1-Score</i>	0.8593	0.0866
Acurácia	0.9730	0.0258

Fonte: Elaborado pelo autor.

Na tabela 7, é possível observar que ainda há *overfit* porém a diferença entre os resultados do treino e validação do modelo 4 é menor quando comparado com o modelo 3. Por exemplo, a diferença percentual entre o *f1-score* do primeiro modelo foi de aproximadamente 9,56% enquanto que para o segundo, foi de, aproximadamente, 17,98%

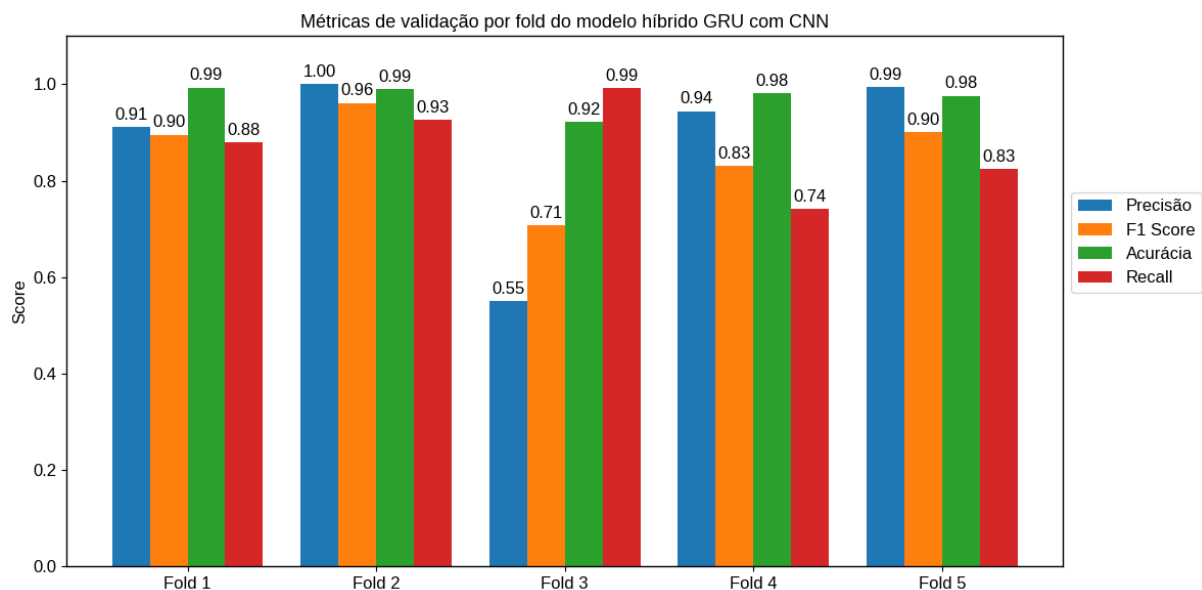
Tabela 7 – Resultados do modelo híbrido CNN e GRU (N vs. V) no treino

Métrica	Média	Desvio Padrão
Precisão	0.9698	0.0180
<i>Recall</i>	0.9313	0.0268
<i>F1-Score</i>	0.9502	0.0222
Acurácia	0.9915	0.0035

Fonte: Elaborado pelo autor.

Na figura 12, estão os resultados obtidos pelo modelo híbrido em cada *fold*.

Figura 12 – Métricas do modelo híbrido CNN com GRU por *fold*



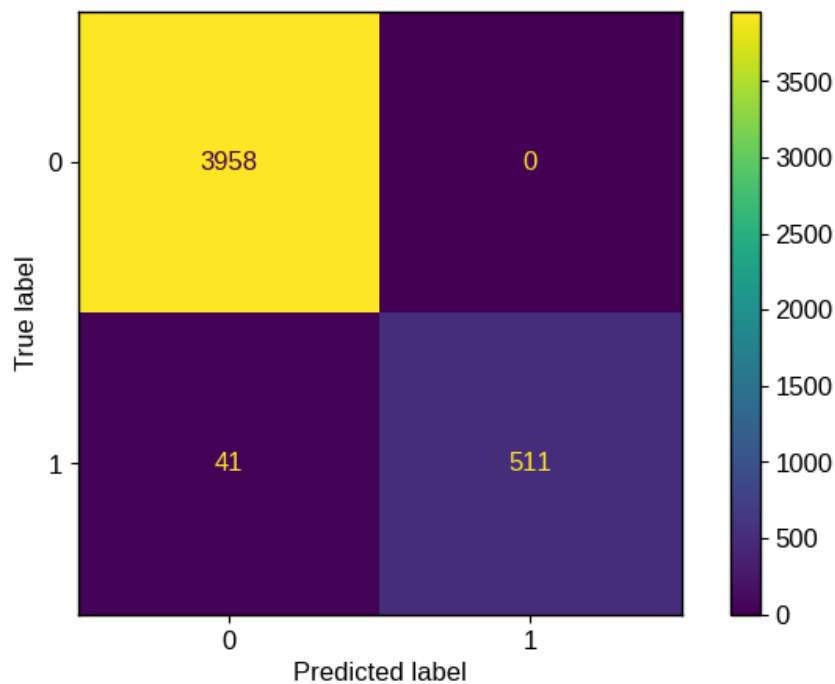
Fonte: Elaborado pelo autor.

O modelo manteve a tendência de ter uma precisão acima da acurácia na maioria dos *folds*. É possível observar também que o modelo obteve um *recall* maior que o modelo GRU puro em todos os *folds* e uma precisão, no geral, maior ou igual. Sendo as exceções, os *folds* quatro e cinco, porém a diferença foi de, aproximadamente, 0,01 pontos percentuais.

A seguir, o desempenho do modelo em seu melhor e pior *fold*. Repetindo os critérios descritos na seção 4.1, o melhor *fold* foi o segundo e o pior foi o terceiro.

Na figura 13, é ilustrada a matriz de confusão do modelo em seu melhor *fold*:

Figura 13 – Matriz de confusão modelo 4 em seu melhor *fold*

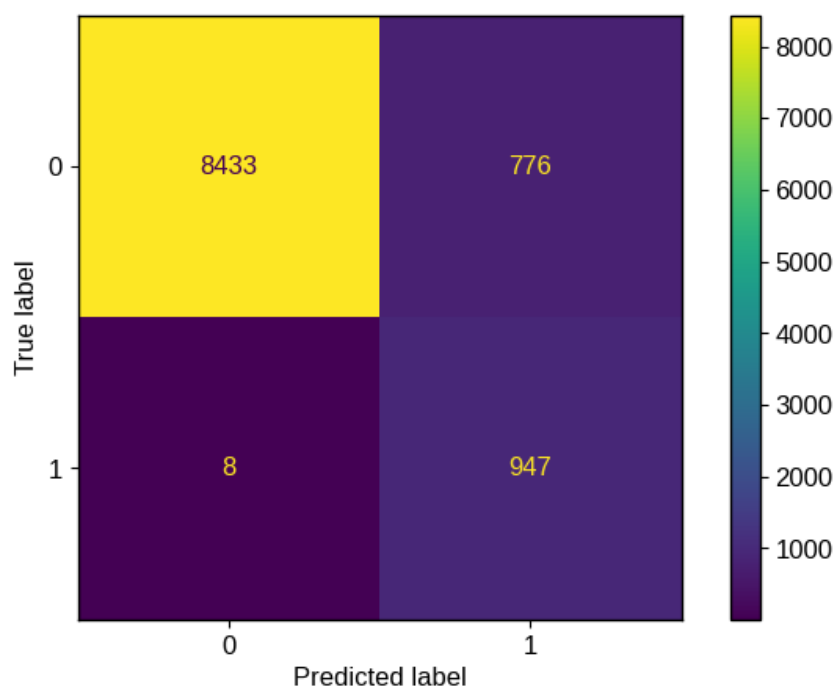


Fonte: Elaborado pelo autor.

Como pode ser visto no gráfico; o modelo não cometeu nenhum erro de falso positivo e errou 41 arritmias, classificando-as como batimentos normais.

Na matriz de confusão do pior *fold*, ilustrado na figura 14, novamente, a situação se inverte; a quantidade de erros de falsos negativos foi menor que as de falsos positivos, refletindo em um *recall* maior que a precisão; como pode ser observado no gráfico 12.

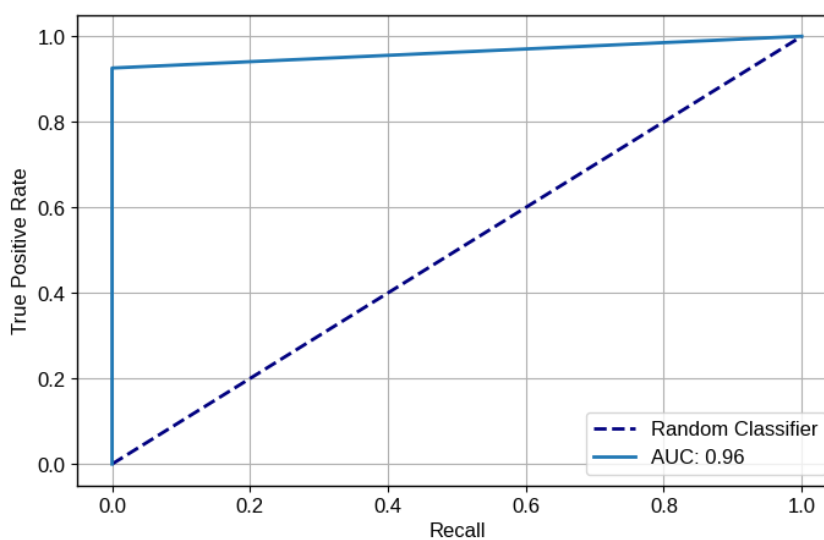
Figura 14 – Matriz de confusão modelo 4 em seu melhor *fold*



Fonte: Elaborado pelo autor.

No melhor, o modelo híbrido obteve um AP de 0,96, como pode ser visto na figura 15:

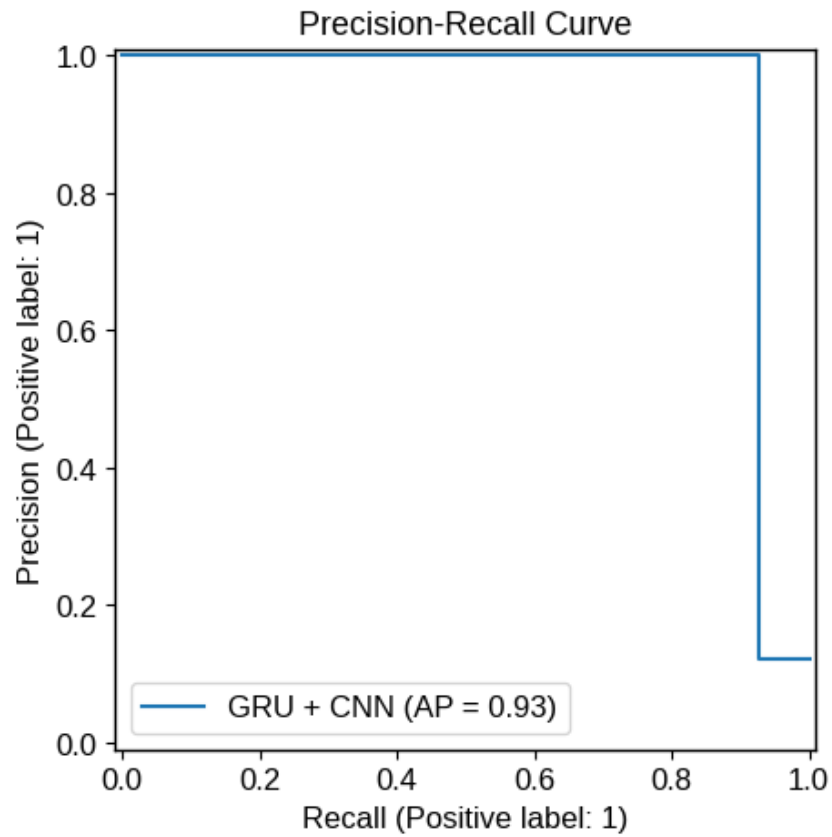
Figura 15 – Curva *ROC* modelo 4 em seu melhor *fold*



Fonte: Elaborado pelo autor.

Considerando a curva PR, o AP também foi maior:

Figura 16 – Curva precisão vs *recall* do modelo 4 em seu melhor *fold*

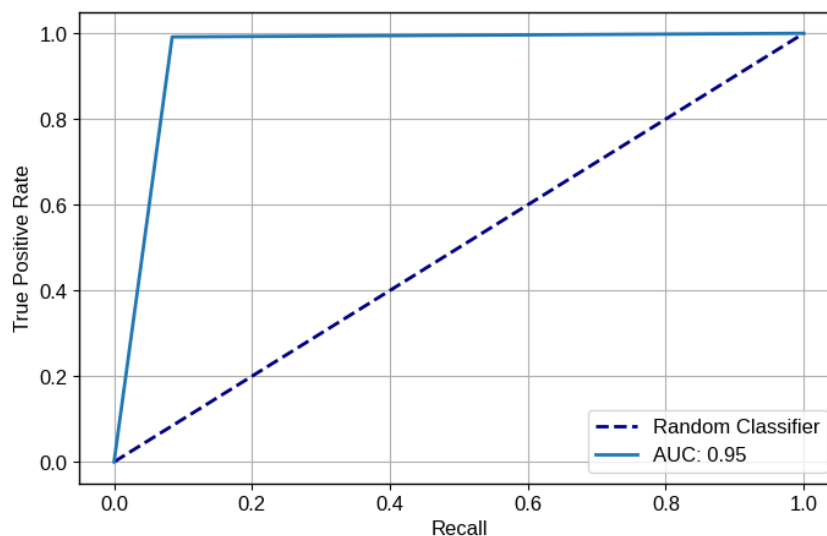


Fonte: Elaborado pelo autor.

Pelo gráfico 16, nesse *fold*, o modelo conseguiria manter um *recall* de até 80% enquanto sua precisão permanece em torno de quase 100%, desempenho superior ao modelo GRU puro.

A curva *ROC* desse modelo no pior *fold* é descrita a seguir:

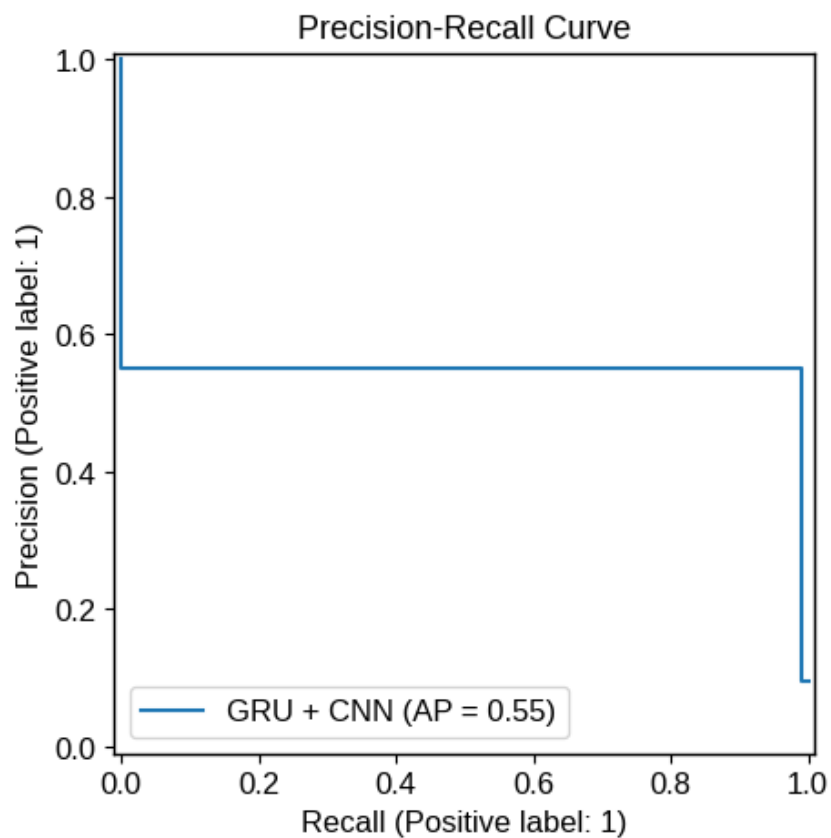
Figura 17 – Curva *ROC* modelo 4 em seu pior *fold*



Fonte: Elaborado pelo autor.

O AP foi de 0,87, igual ao obtido no modelo GRU puro. Já a curva PR descrita na figura

Figura 18 – Curva precisão vs *recall* do modelo 4 em seu melhor *fold*



Fonte: Elaborado pelo autor.

Nesse cenário, o modelo conseguiria manter um *recall* de até 80% com um pouco menos de 60% de precisão.

De modo geral, os modelos exibiram um perfil semelhante em seu pior e melhor caso. No pior, a sensibilidade foi maior, resultados em maiores erros de falso positivo, como resultado, o *recall* foi alto e a precisão foi baixa. No melhor caso, houve um equilíbrio maior e, apesar do *recall* mais baixo, a alta precisão aumentou o *f1-score*.

Em contextos médicos, é preferível um *recall* maior, pois falsos negativos são mais danosos que um falso positivo; isto é, é melhor dizer que um batimento normal é arritmico do que o contrário. Entretanto, uma precisão muito baixa pode indicar que o modelo é tão bom quanto um modelo aleatório; o que o tornaria inútil.

Como evidenciado pelo AP, ambos os modelos se saíram bem melhor do que esse *baseline*.

5 Análise de erros no pior *fold*

Pelos critérios adotados, o pior *fold* de ambos os modelos foi o terceiro. Devido ao *recall* maior que a precisão, supõe-se que a razão era por haver batimentos normais com características mais heterodoxas o que pode ter confundido os modelos. Para obter um melhor entendimento dessa questão, foi feita uma breve análise dos batimentos e pacientes desse *fold*.

5.1 Análise de erros do modelo híbrido CNN com GRU

Na tabela 8 a seguir, é possível ver que a maioria dos erros foi oriunda de um paciente, o 203.

Tabela 8 – Total dos erros e acertos por paciente no *fold* de validação

Pacientes	Erros	Acertos
119	0	1972
203	772	2186
205	11	2616
209	1	2606

Fonte: Elaborado pelo autor.

Aproximadamente, 98,46% de todos os erros foram desse paciente. O modelo errou em torno de 35,31% de seus batimentos. Conforme visto na figura 14, a maioria desses erros são de falsos positivos.

Segundo as anotações do MIT-BIH, disponíveis em (GOLDBERGER et al., n.d.), o paciente 203 é considerado como muito difícil. As anotações ainda citam a presença de mudança de morfologia no complexo QRS e contrações ventriculares prematuras (PVC) de múltiplas formas.

6 Conclusão

Texto da conclusão.

Referências

CASCINO, T.; SHEA, M. J. *Eletrocardiograma (ECG)*. 2025. MSD Manual de Medicina. Revisado por Jonathan G. Howlett. Disponível em: <<https://www.msdmanuals.com/pt/casa/dist%C3%BArbios-do-cora%C3%A7%C3%A3o-e-dos-vasos-sangu%C3%ADneos/diagn%C3%B3stico-de-dist%C3%BArbios-do-cora%C3%A7%C3%A3o-e-dos-vasos-sangu%C3%ADneos/eletrocardiograma>>.

GOLDBERGER, A. L. et al. *MIT-BIH Arrhythmia Database Directory*. n.d. Acesso em: 15 out. 2025. Disponível em: <<https://archive.physionet.org/physiobank/database/html/mitdbdir/intro.htm>>.

JAMES, G. et al. *An Introduction to Statistical Learning: with applications in Python*. Cham: Springer, 2023. Disponível em: <<https://www.statlearning.com/>>. Acesso em: 28 ago. 2025.

MAKOWSKI, D. et al. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, Springer Science and Business Media LLC, v. 53, n. 4, p. 1689–1696, feb 2021. Disponível em: <<https://doi.org/10.3758%2Fs13428-020-01516-y>>.

MOODY, G. B.; MARK, R. G. *MIT-BIH Arrhythmia Database*. 2005. Disponível em: <<https://www.physionet.org/physiobank/database/mitdb/>>. Acesso em: 28 set. 2025.

NAROTAMO, H. et al. Deep learning for ecg classification: A comparative study of 1d and 2d representations and multimodal fusion approaches. *Biomedical Signal Processing and Control*, Elsevier, 2024.

SILVA, G. A. L. et al. *A Systematic Review of ECG Arrhythmia Classification: Adherence to Standards, Fair Evaluation, and Embedded Feasibility*. 2025. ArXiv:2503.07276v1. Acesso em: 28 set. 2025.