

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Deep Learning no Delineamento de ECG: Revisão e Comparação entre CNNs e Transformers

Mariana Trevizani Nocelli

JUIZ DE FORA
JANEIRO, 2026

Deep Learning no Delineamento de ECG: Revisão e Comparação entre CNNs e Transformers

MARIANA TREVIZANI NOCELLI

Universidade Federeal de Juiz de Fora

Instituto de Ciências Exatas

Departamento da Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Luciana Conceição Dias Campos

JUIZ DE FORA

JANEIRO, 2026

DEEP LEARNING NO DELINEAMENTO DE ECG: REVISÃO E COMPARAÇÃO ENTRE CNNS E TRANSFORMERS

Mariana Trevizani Nocelli

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE IN-
TEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Luciana Conceição Dias Campos
Doutora em Engenharia Elétrica

Bernardo Martins Rocha
Doutor em Modelagem Computacional

Priscila Vanessa Zabala Capriles Goliatt
Doutora em Modelagem Computacional

JUIZ DE FORA
20 DE JANEIRO, 2026

Resumo

O eletrocardiograma (ECG) é essencial para o diagnóstico cardiovascular, mas sua interpretação manual é demorada e sujeita à variabilidade. Embora técnicas de aprendizado profundo (*deep learning*) prometam automatizar a segmentação de ondas, persistem desafios relacionados à escassez de dados anotados e à reproduzibilidade de arquiteturas complexas.

Este trabalho combina uma revisão sistemática da literatura (2015–2025) com um estudo experimental de replicação. A revisão, composta por 42 estudos, mapeou a transição tecnológica de CNNs simples para modelos profundos e híbridos com mecanismos de atenção. A análise teórica identificou a predominância de bases públicas (QTDB, LUDB) e evidenciou que o *Data Augmentation* consolidou-se como um componente crítico — e não apenas acessório — para mitigar o severo desbalanceamento de classes em cenários de bases pequenas.

Complementarmente, o estudo prático avaliou a reproduzibilidade de duas arquiteturas representativas utilizando a base LUDB com validação estatística robusta. Os resultados demonstraram a superioridade da abordagem puramente convolucional (Residual U-Net), que atingiu um F1-Score de $0,9544 (\pm 0,002)$, superando o modelo híbrido SEResUTer (F1-Score: $0,8373 \pm 0,007$), mesmo quando este último foi otimizado com aumento de dados. Conclui-se que, embora *Transformers* representem a tendência teórica, modelos convolucionais robustos permanecem como a escolha mais eficiente e estável para aplicações clínicas com restrição de dados rotulados.

Palavras-chave: Eletrocardiograma (ECG), aprendizado profundo, segmentação semântica, redes neurais convolucionais, reproduzibilidade.

Resumo

The electrocardiogram (ECG) is essential for cardiovascular diagnosis, but its manual interpretation is time-consuming and subject to inter-observer variability. Although deep learning techniques promise to automate wave segmentation, challenges persist regarding the scarcity of annotated data and the reproducibility of complex architectures.

This work combines a systematic literature review (2015–2025) with an experimental replication study. The review, comprising 42 studies, mapped the technological transition from simple 1D CNNs to deep and hybrid models with attention mechanisms. Theoretical analysis identified the predominance of public databases (QTDB, LUDB) and highlighted that Data Augmentation has consolidated itself as a critical component—rather than just an accessory—to mitigate severe class imbalance in Small Data scenarios.

Complementarily, the practical study evaluated the reproducibility of two representative architectures using the LUDB database with robust statistical validation. Results demonstrated the superiority of the purely convolutional approach (Residual U-Net), which achieved an F1-Score of 0.9544 (± 0.002), outperforming the hybrid SEResUTer model (F1-Score: 0.8373 ± 0.007), even when the latter was optimized with data augmentation. It is concluded that, while Transformers represent the theoretical trend, robust convolutional models remain the most efficient and stable choice for clinical applications with limited labeled data.

Keywords: Electrocardiogram (ECG), deep learning, semantic segmentation, convolutional neural networks, reproducibility.

Agradecimentos

Primeiramente, agradeço a Deus, pela vida e por me conceder a força e o equilíbrio mental necessários para não desistir diante dos desafios desta jornada.

Aos meus pais, Mario e Ana, pelo apoio incondicional, pelos sacrifícios feitos para que eu pudesse estudar e por serem meu porto seguro.

À minha orientadora, Professora Dra. Luciana, agradeço imensamente pela paciência, pelas correções e por todo o conhecimento compartilhado. Obrigada por guiar meus passos acadêmicos.

Aos professores do curso, que contribuíram para minha formação profissional dividindo suas experiências e saberes.

Aos meus amigos, tanto os de longa data quanto os que a faculdade me presenteou. Obrigada pelas conversas e por tornarem o fardo mais leve e a caminhada mais divertida.

Conteúdo

Lista de Figuras	4
Lista de Tabelas	6
Lista de Abreviações	7
1 Introdução	9
1.1 Objetivos	12
1.2 Organização do Trabalho	13
2 Fundamentação Teórica	14
2.1 Perceptron	14
2.2 Redes Neuronais Multicamadas (MLP)	16
2.3 <i>Deep Learning</i>	17
2.3.1 Redes Neurais Convolucionais (CNN)	19
2.3.2 Redes Neurais Recorrentes (RNN) e Long Short-Term Memory (LSTM)	21
2.4 Redes Neurais Residuais (ResNet)	22
2.4.1 O Bloco Residual e o Fluxo do Gradiente	23
2.4.2 Blocos Squeeze-and-Excitation (SE)	25
2.4.3 Residual U-Net: A Integração de Conexões Residuais na Segmentação	26
2.4.4 Mecanismos de Atenção e a Arquitetura Transformer	27
2.4.5 Transformers e Modelos Híbridos no ECG	28
2.5 Funções de Perda Ponderadas	30
2.6 Métricas de Avaliação de Desempenho	31
2.6.1 Métricas baseadas na Matriz de Confusão	32
2.6.2 Métricas de Sobreposição e Similaridade	32
2.6.3 Métricas de Erro Temporal	33
3 Trabalhos Relacionados	34
3.1 Panorama da Literatura e Revisões Sistemáticas	34
3.2 Evolução das Arquiteturas: De CNNs a Transformers	35
3.2.1 Primeira Geração: Adaptação de CNNs (2015–2019)	35
3.2.2 Segunda Geração: U-Net e Modelos Residuais (2020–2022)	35
3.2.3 Terceira Geração: Transformers e Arquiteturas Híbridas (2023–Presente)	36
4 Revisão Sistemática da Literatura	37
4.1 Questões de Pesquisa Norteadoras	37
4.2 Metodologia da Revisão	38
4.2.1 Fonte de dados e estratégia de busca	38
4.2.2 Critérios de elegibilidade e exclusão	39
4.2.3 Seleção de estudos	40
4.2.4 Extração de dados	40
4.3 Resultados e Discussão da Revisão	42
4.3.1 Arquiteturas de <i>Deep Learning</i> (P1)	42

4.3.2	Bases de Dados e Data Augmentation (P2)	44
4.3.3	Métricas de Avaliação (P3)	46
4.4	Síntese da Revisão	48
5	Delineamento Experimental e Arquiteturas	50
5.1	Fluxo do Experimento	50
5.2	Definição dos Estudos de Caso	50
5.3	Base de Dados Utilizada (LUDB)	52
5.3.1	Caracterização dos Dados	52
5.4	Protocolo de Avaliação e Ferramentas	53
5.5	Fundamentação Matemática da Implementação	55
5.5.1	Função de Perda: Entropia Cruzada Categórica	55
5.5.2	Estratégia de Ponderação de Classes (Class Weights)	56
5.6	Arquiteturas Implementadas	56
5.6.1	Caso 1: Residual U-Net (Abordagem Convolucional)	56
5.6.2	Caso 2: SEResUTer (Abordagem Híbrida)	57
5.6.3	Estratégias de Estabilização para o Modelo Híbrido	58
6	Resultados Experimentais e Discussão	60
6.1	Resultados do Caso 1: Residual U-Net	60
6.2	Resultados do Caso 2: SEResUTer	61
6.3	Análise Comparativa e Conclusão dos Experimentos	61
7	Conclusão e Trabalhos Futuros	63
7.1	Síntese dos Resultados e Objetivos	63
7.2	Limitações do Estudo	64
7.3	Trabalhos Futuros	64
	Bibliografia	66

Lista de Figuras

1.1	Representação esquemática das ondas do ECG e sua correlação com o ciclo cardíaco.	10
1.2	Exemplos de patologias que dificultam a segmentação automática. Em (a), a Fibrilação Atrial elimina a Onda P e insere ruído na linha de base. Em (b), o Infarto eleva o segmento ST, alterando drasticamente a morfologia da repolarização ventricular.	11
2.1	Representação esquemática do Perceptron de Rosenblatt.	15
2.2	Estrutura de uma Rede Neural Multicamadas (MLP). À direita, o detalhamento de um neurônio oculto, evidenciando o viés (b) e o processamento interno de soma e ativação.	16
2.3	Comparação do fluxo de trabalho. Em cima, a abordagem tradicional exige intervenção humana na engenharia de características. Em baixo, o <i>Deep Learning</i> unifica extração e classificação em uma única estrutura treinável (end-to-end).	17
2.4	Diferença estrutural entre a Convolução 2D e 1D. Em (a), o filtro varre a imagem em dois eixos. Em (b), o filtro desloca-se apenas no tempo, reduzindo drasticamente a complexidade computacional.	21
2.5	Estrutura interna de uma célula LSTM. Os retângulos representam as funções de ativação (sigmoide e tangente hiperbólica) que atuam como portões de controle. Os círculos indicam as operações pontuais (multiplicação e adição) que atualizam o estado da célula, permitindo o fluxo seletivo de informação.	22
2.6	Diagrama de um bloco residual. O caminho da esquerda (seta vermelha) contorna as camadas de peso, permitindo que o gradiente flua livremente durante a retropropagação.	24
2.7	Diagrama esquemático do bloco Squeeze-and-Excitation (SE). O mecanismo comprime a informação espacial (Squeeze) para calcular a importância de cada canal (Excitation), utilizando esses pesos para recalibrar a entrada original.	25
2.8	Arquitetura U-Net 1D. O caminho de contração (esquerda) extrai o contexto, enquanto o caminho de expansão (direita) recupera a localização precisa. As setas tracejadas (<i>skip Connections</i>) transferem detalhes de alta resolução do encoder para o decoder, essenciais para a segmentação precisa das ondas.	27
2.9	Comparação do mecanismo de captura de contexto. Em (a), a CNN observa apenas uma vizinhança local. Em (b), o mecanismo de atenção conecta x_6 a todos os pontos da sequência (dependências de longo alcance).	29
2.10	Representação esquemática do mecanismo de Atenção (<i>Scaled Dot-Product</i>). O modelo utiliza a similaridade entre Q e K para calcular pesos que filtram a informação relevante vinda de V	30
2.11	Matriz de confusão para segmentação de ECG.	31

4.1	Fluxograma PRISMA 2020 detalhando o processo de seleção dos estudos. Dos 175 registros iniciais, 42 atenderam a todos os critérios de inclusão e compõem o corpus desta revisão.	41
4.2	Evolução das arquiteturas de <i>deep learning</i> aplicadas ao ECG (2018–2025).	44
4.3	Evolução temporal da estratégia de validação (2018–2024). O gráfico evi- dencia a inversão do paradigma: o uso de bases únicas (cinza) cede lugar à integração de múltiplas bases (azul), impulsionada pela necessidade de generalização dos modelos.	46
4.4	Distribuição das categorias de <i>data augmentation</i> utilizadas nos estudos de ECG analisados.	47
5.1	Fluxo metodológico da etapa de replicação experimental. Ambos os mode- los foram submetidos a pipelines similares de pré-processamento e avaliação para garantir a comparabilidade dos resultados.	51
5.2	Exemplo de um ciclo cardíaco segmentado na base LUDB. As áreas colori- das representam as anotações de especialistas para Onda P, QRS e Onda T.	52
5.3	Distribuição percentual das classes na base de treinamento, evidenciando o predomínio da classe Fundo.	54
5.4	Diagrama da arquitetura Residual U-Net com blocos SE utilizada no Es- tudo de Caso 1.	57
5.5	Diagrama da arquitetura híbrida SEResUTer, integrando U-Net com Trans- former Encoder no gargalo.	58

Lista de Tabelas

4.1	Resumo das estratégias de <i>data augmentation</i> empregadas nos estudos selecionados.	47
4.2	Principais métricas de avaliação identificadas na revisão sistemática.	48
4.3	Síntese das Questões de Pesquisa e Respostas Obtidas	49
5.1	Distribuição dos diagnósticos de ritmo cardíaco presentes nos registros da base LUDB.	53
5.2	Hiperparâmetros e especificidades das arquiteturas comparadas.	55
6.1	Desempenho Estatístico do Modelo Convolucional (Média de 5 Rodadas). .	60
6.2	Impacto do Data Augmentation no Modelo Híbrido (Média de 5 Rodadas). .	61
6.3	Comparativo final: Residual U-Net vs. SEResUTer (Melhores Configurações). .	62

Lista de Abreviações

ABC	Área Abaixo da Curva (<i>Area Under the Curve</i>)
ANN	Rede Neural Artificial (<i>Artificial Neural Network</i>)
AVC	Acidente Vascular Cerebral
BPM	Batimentos Por Minuto
CCE	<i>Categorical Cross-Entropy</i> (Entropia Cruzada Categórica)
CNN	Rede Neural Convolucional (<i>Convolutional Neural Network</i>)
CPSC	<i>China Physiological Signal Challenge</i>
DCC	Departamento de Ciência da Computação
DCV	Doenças Cardiovasculares
DL	<i>Deep Learning</i> (Aprendizado Profundo)
DSC	<i>Dice Similarity Coefficient</i>
ECG	Eletrocardiograma
FN	Falso Negativo
FP	Falso Positivo
GBD	<i>Global Burden of Disease</i>
GPU	Unidade de Processamento Gráfico (<i>Graphics Processing Unit</i>)
IoU	Índice de Jaccard (<i>Intersection over Union</i>)
LSTM	<i>Long Short-Term Memory</i>
LUDB	<i>Lobachevsky University Database</i>
MAD	Desvio Absoluto Médio (<i>Mean Absolute Deviation</i>)
MIT-BIH	<i>Massachusetts Institute of Technology - Beth Israel Hospital</i>
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
NSTDB	<i>Noise Stress Test Database</i>
PPV	Valor Preditivo Positivo (<i>Positive Predictive Value</i> ou Precisão)
PRISMA	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
QTDB	<i>QT Database</i>

ReLU	<i>Rectified Linear Unit</i>
RNN	Rede Neural Recorrente (<i>Recurrent Neural Network</i>)
ROC	<i>Receiver Operating Characteristic</i>
SE	<i>Squeeze-and-Excitation</i>
SGD	<i>Stochastic Gradient Descent</i>
TCC	Trabalho de Conclusão de Curso
TP	Verdadeiro Positivo
UFJF	Universidade Federal de Juiz de Fora
VN	Verdadeiro Negativo
WHO	Organização Mundial da Saúde (<i>World Health Organization</i>)

1 Introdução

As doenças cardiovasculares (DCV) representam, atualmente, o maior desafio para a saúde pública global. Dados recentes do estudo *Global Burden of Disease (GBD)*(Instituto de Métrica e Avaliação em Saúde (IHME), 2025) de 2023 indicam que a doença isquêmica do coração permanece como a principal causa de morte no mundo, seguida pelo Acidente Vascular Cerebral (AVC), que ocupa a terceira posição no ranking global de mortalidade. O estudo destaca ainda que, embora haja avanços, a mortalidade por doenças não transmissíveis continua persistentemente elevada, evidenciando desigualdades críticas no acesso a diagnóstico e tratamento, especialmente em regiões com menores recursos. Esse cenário global reflete-se diretamente na realidade brasileira: segundo o Ministério da Saúde, cerca de 400 mil pessoas faleceram no país em 2022 em decorrência de problemas cardiovasculares (Brasil. Ministério da Saúde, 2022).

Diante desse cenário crítico, a rapidez e a precisão no diagnóstico tornam-se fatores determinantes para a sobrevivência do paciente. Nesse contexto, o eletrocardiograma (ECG) consolida-se como a ferramenta padrão-ouro para a avaliação inicial da atividade cardíaca, por ser um exame não invasivo, de baixo custo e alta disponibilidade clínica (HALL; GUYTON, 2017).

A interpretação do exame baseia-se na análise de padrões morfológicos específicos que refletem o ciclo cardíaco. O traçado eletrocardiográfico é composto por três deflexões principais, cada uma correspondendo a um evento fisiológico distinto, como pode ser visto na Figura 1.1. Segundo Hall e Guyton (2017), a primeira onda, denominada Onda P, representa a despolarização atrial, momento em que o impulso elétrico se propaga pelos átrios gerando sua contração. Na sequência, identifica-se o Complexo QRS, uma estrutura de maior amplitude que sinaliza a despolarização ventricular. Devido à grande massa muscular dos ventrículos responsável por bombear sangue para todo o corpo, este evento elétrico é o mais proeminente do registro (DUBIN, 2000). O ciclo encerra-se com a Onda T, que indica a repolarização ventricular, fase crítica em que o miocárdio recupera seu potencial elétrico preparando-se para o batimento seguinte.

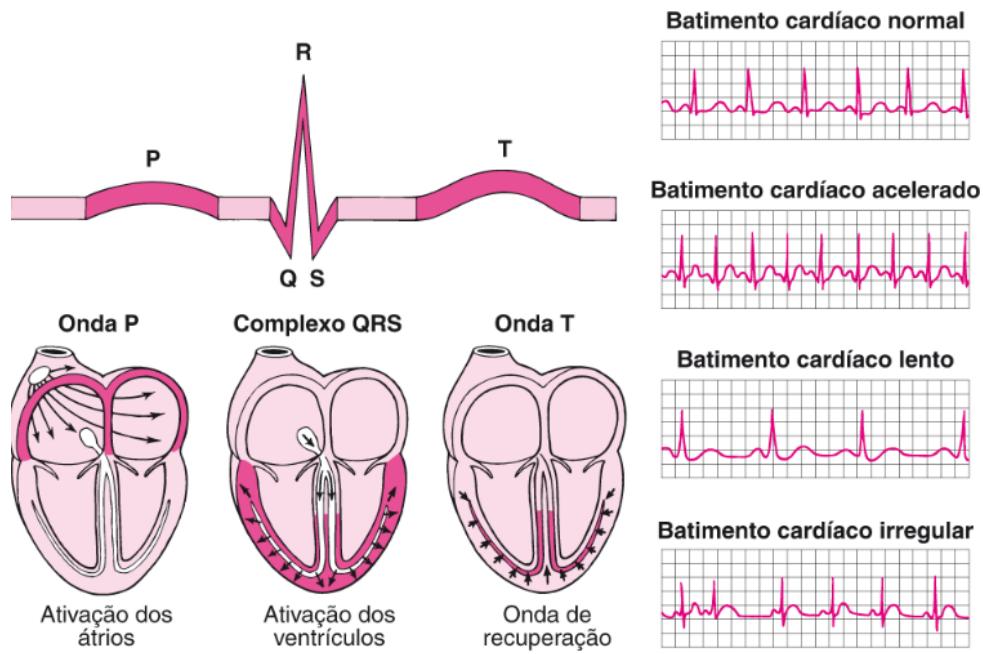


Figura 1.1: Representação esquemática das ondas do ECG e sua correlação com o ciclo cardíaco.

Fonte: Adaptado de Manuais MSD (2024).

Tradicionalmente, a identificação dessas ondas e de seus intervalos é realizada por especialistas de forma manual, um processo trabalhoso, demorado e sujeito à variabilidade inter e intraobservador (ALAM; AGUIRRE; STULTZ, 2024). Essa tarefa torna-se ainda mais desafiadora na presença de patologias, como ilustrado na Figura 1.2, onde arritmias ou isquemias alteram drasticamente a morfologia padrão, dificultando a delimitação precisa das ondas. Essa limitação é especialmente crítica em cenários de alta demanda, como unidades de terapia intensiva e monitoramento contínuo, nos quais a velocidade e a precisão do diagnóstico são cruciais. A automatização dessa análise, por meio de algoritmos e técnicas de aprendizado de máquina, surge como uma alternativa promissora para mitigar tais dificuldades (GOOVAERTS et al., 2018).

Nos últimos anos, o avanço das técnicas de aprendizado profundo (*deep learning*) transformou a análise de ECG. Diferentemente dos métodos tradicionais, que dependiam de extração manual de características, as redes neurais profundas são capazes de aprender representações complexas diretamente a partir do sinal bruto (FAUST et al., 2018). Essa evolução viabilizou modelos capazes de realizar a segmentação automática das ondas P, QRS e T, ampliando o potencial para aplicações clínicas em tempo real e para o

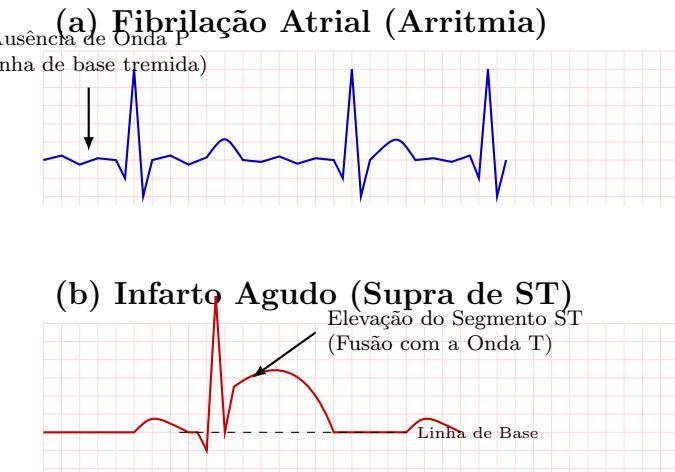


Figura 1.2: Exemplos de patologias que dificultam a segmentação automática. Em (a), a Fibrilação Atrial elimina a Onda P e insere ruído na linha de base. Em (b), o Infarto eleva o segmento ST, alterando drasticamente a morfologia da repolarização ventricular.

Fonte: Elaborada pelo autor (2025).

desenvolvimento de sistemas de apoio à decisão médica.

Apesar desses avanços, a transição desses modelos para a prática clínica enfrenta barreiras significativas. A principal delas reside na generalização: modelos treinados em ambientes controlados frequentemente falham ao processar sinais do mundo real. Isso ocorre, em grande parte, porque o desenvolvimento de algoritmos modernos ainda depende substancialmente de bases de dados ‘clássicas’, adquiridas com tecnologias de décadas passadas. Exemplos notáveis são a *Massachusetts Institute of Technology - Beth Israel Hospital (MIT-BIH) Arrhythmia Database*, lançada em 1980 (MOODY; MARK, 2001), e a QT Database, de 1997 (LAGUNA et al., 1997). Embora pioneiras, essas coleções apresentam limitações frente às demandas atuais de *Deep Learning*: possuem um número reduzido de pacientes e sinais que não refletem integralmente a diversidade demográfica nem os artefatos dos dispositivos modernos (HONG et al., 2020). Somado a isso, há uma notável escassez de anotações detalhadas validadas por múltiplos especialistas, o que restringe o treinamento supervisionado, além do fato de muitos modelos operarem como “caixas-pretas”, dificultando a interpretabilidade clínica.

1.1 Objetivos

O objetivo principal deste trabalho é investigar o estado da arte em delimitação temporal de início e fim de ondas do eletrocardiograma (ECG) com técnicas de *deep learning*, aliando uma revisão sistemática da literatura recente (2015–2025) a uma análise experimental. Dessa forma, o estudo engloba simultaneamente uma perspectiva teórica e uma perspectiva prática, visando oferecer uma síntese crítica da literatura e avaliar a reproduzibilidade e o desempenho de arquiteturas representativas.

Para cumprir esse propósito geral, o trabalho foi estruturado em dois eixos complementares:

- (i) **Revisão Sistemática**: mapeamento das arquiteturas de *deep learning* predominantes na literatura, das bases públicas utilizadas e das métricas de avaliação aplicadas à segmentação das ondas P, QRS e T.
- (ii) **Replicação Experimental**: implementação e análise comparativa de duas arquiteturas distintas, sendo uma convolucional e outra híbrida com mecanismos de atenção, com o objetivo de avaliar sua viabilidade prática, estabilidade e desempenho em condições reais de treinamento.

Para operacionalizar esses eixos, foram definidos os seguintes objetivos específicos:

- **Identificar e categorizar** as arquiteturas de *deep learning* mais recorrentes na literatura recente para a segmentação de ondas do ECG;
- **Mapear as bases de dados** públicas predominantes e as estratégias de *Data Augmentation* empregadas para mitigar a escassez de dados anotados;
- **Levantar as métricas de avaliação** padronizadas utilizadas para aferir o desempenho dos modelos de segmentação;
- **Realizar um estudo experimental de replicação** para avaliar a viabilidade prática, a estabilidade e o desempenho comparativo entre modelos puramente convolucionais e modelos híbridos com mecanismos de atenção.

1.2 Organização do Trabalho

Além desta introdução, a monografia está estruturada em mais seis capítulos, organizados da seguinte forma:

O Capítulo 2 estabelece a base teórica necessária para o entendimento dos modelos de inteligência artificial. Inicia-se com os fundamentos de redes neurais (*Perceptron e MLP*) e avança para arquiteturas de *Deep Learning*, detalhando o funcionamento das Redes Neurais Convolucionais (CNNs), Redes Recorrentes (LSTMs) e os mecanismos de atenção (*Transformers*). Também são apresentados os conceitos de Redes Residuais, funções de perda ponderadas e as métricas de avaliação de desempenho utilizadas.

O Capítulo 3 apresenta os trabalhos relacionados, oferecendo um panorama da evolução das arquiteturas aplicadas ao ECG. O capítulo descreve a transição histórica das abordagens, dividindo-as em gerações: desde as primeiras adaptações de CNNs, passando pela consolidação da U-Net e modelos residuais, até a emergência recente dos *Transformers* e arquiteturas híbridas.

O Capítulo 4 descreve a Revisão Sistemática da Literatura. Nele, são definidas as questões de pesquisa norteadoras e detalhada a metodologia de revisão (fontes de dados e critérios de seleção). Além disso, apresenta-se a discussão dos resultados da revisão, sintetizando as arquiteturas de *Deep Learning* predominantes no estado da arte.

O Capítulo 5 dedica-se ao delineamento experimental e à implementação das arquiteturas. Esta seção define o fluxo do experimento, os protocolos de avaliação e a fundamentação matemática da implementação. São detalhados os dois estudos de caso realizados: a replicação da Residual U-Net (abordagem convolucional) e do modelo SE-ResUTer (abordagem híbrida).

O Capítulo 6 expõe os resultados experimentais e a discussão. São analisados os dados obtidos em cada estudo de caso e realizada uma análise comparativa entre os modelos, avaliando a viabilidade de replicação e o desempenho de segmentação.

Por fim, o Capítulo 7 apresenta as conclusões do trabalho, integrando os achados teóricos com as evidências práticas. O capítulo oferece uma síntese das respostas às questões de pesquisa, discute as limitações do estudo e aponta direções para trabalhos futuros.

2 Fundamentação Teórica

Este capítulo estabelece as bases necessárias para a compreensão dos modelos desenvolvidos neste trabalho. A discussão parte das unidades fundamentais de processamento (Perceptrons) e avança cronologicamente para arquiteturas de aprendizado profundo, abordando as Redes Neurais Convolucionais (CNNs), Redes Recorrentes (LSTMs) e arquiteturas de segmentação semântica (U-Net e Transformers). Além disso, são definidos os conceitos matemáticos subjacentes, como transformações lineares e funções de otimização, essenciais para o entendimento do treinamento *end-to-end* (ponta a ponta) aplicado ao processamento de sinais de ECG.

2.1 Perceptron

O perceptron foi proposto originalmente por Rosenblatt (1958) como um modelo probabilístico e representa um dos primeiros esforços formais para modelar computacionalmente o comportamento de um neurônio biológico. Trata-se do modelo mais elementar das redes neurais artificiais, capaz de realizar classificação supervisionada.

Matematicamente, o funcionamento do perceptron baseia-se em uma transformação linear dos dados de entrada, cuja resultante é submetida a uma não-linearidade. Uma transformação linear é uma operação que mapeia vetores de um espaço para outro preservando a aditividade e a homogeneidade, a saída final y é expressa pela aplicação da função de ativação (φ) sobre a soma ponderada das entradas (x) pelos pesos sinápticos (w) adicionada a um viés (b):

$$y = \varphi \left(\sum_{i=1}^n w_i \cdot x_i + b \right) = \varphi(\mathbf{w}^T \mathbf{x} + b) \quad (2.1)$$

A função de ativação (φ) é responsável por determinar a saída. Embora em redes profundas se utilizem funções não-lineares complexas, o conceito mais simples seria a função identidade ($f(x) = x$), que transmitiria o sinal sem alteração, preservando a

linearidade do sistema. Contudo, para tarefas de classificação, utiliza-se geralmente uma função degrau ou sigmoide.

A estrutura completa pode ser visualizada na Figura 2.1. Nesse esquema, as entradas x_n são ponderadas pelos pesos w_n , somadas ao viés b e submetidas à função de ativação (originalmente uma função degrau) para gerar a saída y .

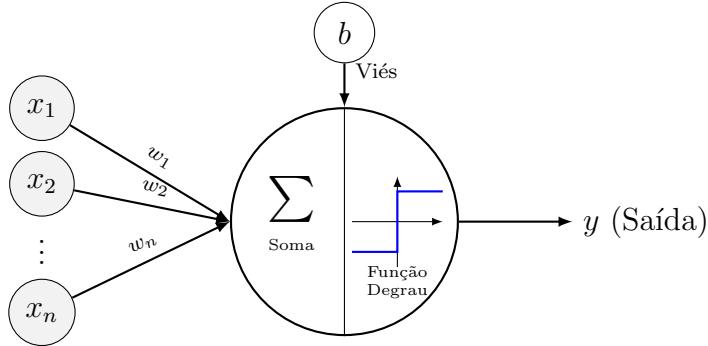


Figura 2.1: Representação esquemática do Perceptron de Rosenblatt.

Fonte: Baseado em Rosenblatt (1958)

Do ponto de vista conceitual, essa estrutura define o perceptron como um classificador linear, sendo capaz de separar dados apenas quando estes são linearmente separáveis no espaço de características. Conforme discutido por Goodfellow, Bengio e Courville (2016), essa limitação estrutural restringe sua aplicabilidade a problemas do mundo real, nos quais os padrões de interesse apresentam relações não lineares complexas.

No contexto de sinais biomédicos, como o eletrocardiograma, essa limitação torna-se ainda mais evidente, uma vez que a morfologia das ondas P, QRS e T é influenciada por múltiplos fatores fisiológicos e patológicos que não podem ser adequadamente modelados por fronteiras de decisão lineares (SHARMA; JOSHI, 2025; CHOU et al., 2024). Por exemplo, a distinção entre uma onda T normal e uma onda P sobreposta a ruído frequentemente requer a análise de dependências temporais e não lineares, algo inviável para um perceptron simples.

Apesar de suas limitações práticas, o perceptron possui relevância histórica e conceitual fundamental, pois estabelece os princípios básicos do aprendizado supervisionado, do ajuste iterativo de pesos a partir do erro e da construção de modelos neurais, servindo de base para o desenvolvimento de arquiteturas mais avançadas, como os perceptrons multicamadas e as redes neurais profundas.

2.2 Redes Neuronais Multicamadas (MLP)

Com o objetivo de superar as restrições impostas pelo perceptron simples, surgiram as redes neuronais multicamadas, comumente denominadas *Multilayer Perceptrons* (MLP) (RUMELHART; HINTON; WILLIAMS, 1986; HAYKIN, 2009). Essas arquiteturas estendem o modelo original ao introduzir uma ou mais camadas ocultas entre a camada de entrada e a camada de saída, permitindo a composição de múltiplas transformações não lineares sucessivas. A estrutura geral de uma MLP e o detalhamento do funcionamento de um de seus neurônios ocultos podem ser visualizados na Figura 2.2.

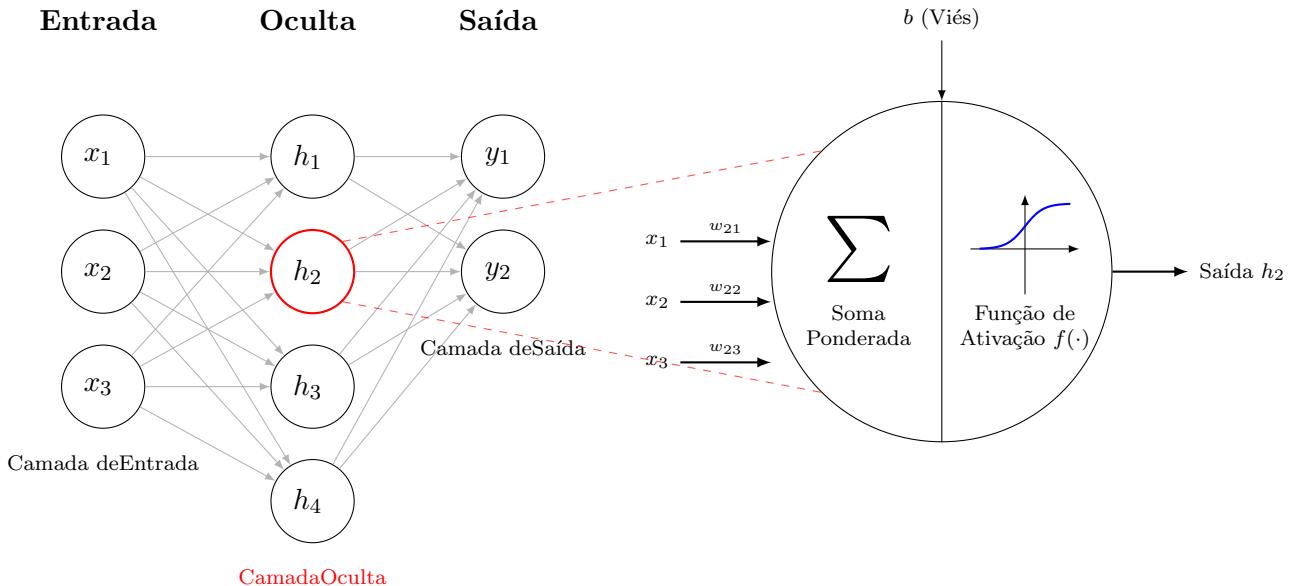


Figura 2.2: Estrutura de uma Rede Neural Multicamadas (MLP). À direita, o detalhamento de um neurônio oculto, evidenciando o viés (b) e o processamento interno de soma e ativação.

Fonte: Baseado em Haykin (2009)

A principal inovação das MLPs reside na capacidade de modelar relações não lineares complexas por meio da utilização de funções de ativação não lineares nas camadas ocultas, como ReLU, sigmoide ou tangente hiperbólica. Esse avanço teórico está associado ao conhecido Teorema da Aproximação Universal, segundo o qual uma rede neural com ao menos uma camada oculta e número suficiente de neurônios é capaz de aproximar qualquer função contínua em domínios compactos, sob condições apropriadas (GOODFELLOW; BENGIO; COURVILLE, 2016).

O treinamento dessas redes é viabilizado pelo algoritmo de retropropagação do

erro (*backpropagation*), que permite o ajuste eficiente dos pesos por meio de métodos baseados em gradiente (RUMELHART; HINTON; WILLIAMS, 1986). Embora as MLPs apresentem maior poder de representação quando comparadas ao perceptron simples, elas ainda enfrentam limitações relevantes ao processar dados estruturados no tempo ou no espaço. No caso de sinais de ECG, as MLPs tendem a processar a entrada de forma global, sem a hierarquia local inerente às convoluções. Consequentemente, essas redes são menos eficazes em explorar a correlação local entre amostras vizinhas ou em garantir a invariância à translação, ou seja, a capacidade de reconhecer um mesmo padrão morfológico (como uma onda QRS) independentemente de sua posição temporal dentro da janela analisada (SHARMA; JOSHI, 2025). Essa característica limita o desempenho em tarefas de segmentação precisa, onde a forma e a posição relativa das ondas são mais determinantes que seus valores absolutos de amplitude.

2.3 Deep Learning

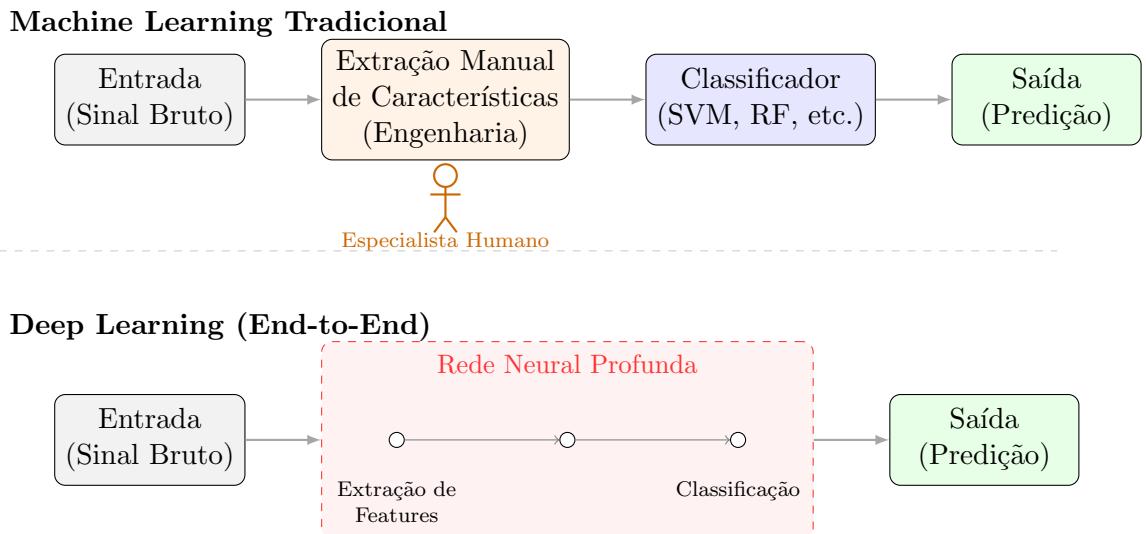


Figura 2.3: Comparação do fluxo de trabalho. Em cima, a abordagem tradicional exige intervenção humana na engenharia de características. Em baixo, o *Deep Learning* unifica extração e classificação em uma única estrutura treinável (end-to-end).

Fonte: Baseado em Goodfellow, Bengio e Courville (2016).

O termo *Deep Learning* refere-se a um subconjunto do aprendizado de máquina baseado em redes neurais artificiais profundas, caracterizadas pela presença de múltiplas camadas ocultas hierarquicamente organizadas. Nesse paradigma, as redes são capazes de

aprender representações distribuídas em diferentes níveis de abstração, nas quais camadas mais próximas da entrada capturam padrões simples, enquanto camadas mais profundas modelam estruturas progressivamente mais complexas (GOODFELLOW; BENGIO; COURVILLE, 2016). Embora arquiteturas profundas possam ser vistas como uma generalização das MLPs tradicionais, elas incorporam mecanismos estruturais específicos que favorecem a aprendizagem eficiente de padrões complexos em grandes volumes de dados.

Essa abordagem representa uma mudança significativa em relação aos métodos clássicos de análise de sinais, que dependiam fortemente da extração manual de características. No contexto do ECG, abordagens tradicionais baseavam-se em regras heurísticas e descritores projetados por especialistas, como intervalos temporais, amplitudes e relações morfológicas pré-definidas (PAN; TOMPKINS, 1985; MARTÍNEZ et al., 2004). Em contraste, modelos de *deep learning* reduzem a necessidade de engenharia manual de características, aprendendo representações discriminativas diretamente a partir do sinal bruto. Entretanto, em cenários supervisionados, como o presente trabalho, esse processo depende de anotações especializadas previamente definidas, as quais incorporam conhecimento fisiológico fundamental para o treinamento e a avaliação dos modelos. A Figura 2.3 ilustra essa mudança de paradigma, comparando o fluxo de trabalho clássico, dependente de engenharia manual, com a abordagem *end-to-end* do *Deep Learning*.

A eficácia do *deep learning* em aplicações biomédicas está intimamente relacionada à sua habilidade de lidar com sinais ruidosos, variabilidade interpaciente e padrões sutis associados a condições patológicas (MEHRI et al., 2023; PEIMANKAR; PUTHUSSERYPADY, 2021). Essas características tornam as redes profundas particularmente adequadas para tarefas de segmentação de sinais fisiológicos, como a identificação precisa dos instantes de início e fim das ondas P, QRS e T (PEIMANKAR; PUTHUSSERYPADY, 2021). No entanto, a adoção bem-sucedida dessas técnicas depende do uso de arquiteturas especializadas capazes de explorar adequadamente a estrutura local e temporal dos dados, o que motiva o emprego de redes convolucionais, redes recorrentes e modelos híbridos (NURMAINI et al., 2021), discutidos nas seções subsequentes.

2.3.1 Redes Neurais Convolucionais (CNN)

As Redes Neurais Convolucionais (CNNs) constituem uma das principais classes de arquiteturas profundas utilizadas na análise de sinais biomédicos. Elas representam um avanço significativo sobre as redes densas tradicionais (MLP) por serem capazes de extrair características hierárquicas diretamente dos dados brutos. Essa capacidade viabiliza o aprendizado ***end-to-end*** (de ponta a ponta), um paradigma onde todo o processo, desde a extração de atributos até a classificação final, é realizado por uma única rede neural otimizada conjuntamente, dispensando a engenharia manual de características (*feature engineering*) necessária em métodos clássicos (LECUN; BENGIO; HINTON, 2015).

Embora tenham sido originalmente desenvolvidas para o processamento de imagens, as CNNs podem ser adaptadas para sinais unidimensionais por meio da convolução 1D, tornando-se particularmente adequadas para o tratamento de séries temporais como o eletrocardiograma (GOODFELLOW; BENGIO; COURVILLE, 2016). Diferentemente das CNNs 2D, que deslocam filtros em dois eixos espaciais (altura e largura), as CNNs 1D deslocam seus kernels¹ apenas ao longo do eixo temporal. Isso reduz a complexidade computacional e adequa-se à natureza sequencial do vetor de ECG (KIRANYAZ; INCE; GABBOUJ, 2016).

A Figura 2.4 demonstra visualmente essa diferença estrutural: enquanto a CNN 2D exige varredura espacial, a CNN 1D opera exclusivamente ao longo do tempo, capturando padrões morfológicos locais da onda cardíaca.

Para evidenciar a eficiência computacional, convém comparar as operações. Na CNN 2D, utilizada em imagens, o valor de cada pixel de saída depende de um somatório duplo (varredura em altura e largura):

$$y[i, j] = \sum_m \sum_n x[i + m, j + n] \cdot w[m, n] \quad (2.2)$$

Onde:

- $y[i, j]$ representa o valor do pixel de saída na posição espacial (i, j) ;
- x é a matriz de entrada (imagem 2D);

¹No contexto de redes neurais, um *kernel* (ou filtro) é uma matriz de pesos treináveis que desliza sobre os dados de entrada para extrair características relevantes, como bordas ou padrões específicos.

- w é a matriz do kernel (filtro) com dimensões $m \times n$;
- Os somatórios percorrem as duas dimensões do filtro (altura e largura).

Já na CNN 1D, a operação é simplificada para um somatório simples ao longo do eixo temporal:

$$y[t] = \sum_k x[t+k] \cdot w[k] \quad (2.3)$$

Onde:

- $y[t]$ representa o valor de saída na posição temporal t ;
- x é o vetor de entrada (sinal de ECG);
- w é o kernel (filtro) de tamanho k ;
- A operação de soma percorre o tamanho do filtro, multiplicando os pesos pelos valores do sinal localmente.

Essa redução de dimensionalidade (de $O(N^2)$ para $O(N)$ no contexto do filtro) justifica por que as CNNs 1D são mais rápidas e exigem menos memória, permitindo o processamento de longos registros de ECG em tempo real (KIRANYAZ; INCE; GABBOUJ, 2016).

No contexto da análise de ECG, os filtros convolucionais atuam como detectores de características morfológicas locais, sendo capazes de identificar padrões associados às ondas P, QRS e T diretamente a partir do sinal bruto. A operação de convolução permite que a mesma estrutura de filtro seja aplicada ao longo de todo o sinal, conferindo à rede um grau de invariância a pequenas variações temporais, comuns em registros reais devido a ruídos, variações fisiológicas e diferenças interpaciente (KIRANYAZ; INCE; GABBOUJ, 2016).

Além disso, a hierarquização das camadas convolucionais possibilita o aprendizado progressivo de representações, nas quais camadas iniciais capturam padrões simples, como inclinações e picos, enquanto camadas mais profundas modelam estruturas mais complexas do traçado eletrocardiográfico. Essa característica torna as CNNs especialmente eficazes em tarefas de segmentação, nas quais é necessário identificar com precisão

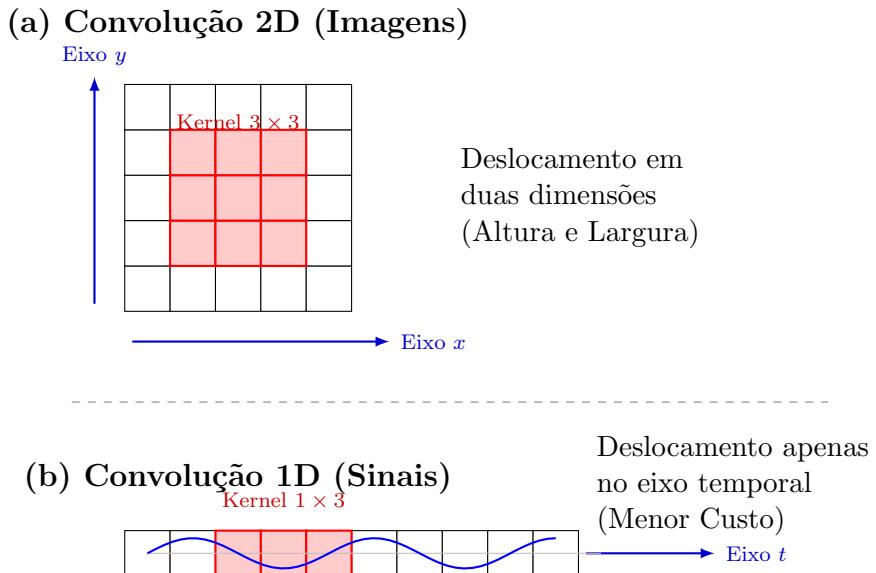


Figura 2.4: Diferença estrutural entre a Convolução 2D e 1D. Em (a), o filtro varre a imagem em dois eixos. Em (b), o filtro desloca-se apenas no tempo, reduzindo drasticamente a complexidade computacional.

Fonte: Baseado em Kiranyaz, Ince e Gabbouj (2016).

os limites temporais das ondas cardíacas. Diversos estudos demonstram o sucesso de arquiteturas convolucionais na segmentação automática de ECG, consolidando-as como uma abordagem de referência na área (ACHARYA et al., 2017; DURAJ et al., 2022).

2.3.2 Redes Neurais Recorrentes (RNN) e Long Short-Term Memory (LSTM)

Embora as CNNs sejam eficazes na captura de padrões locais, elas apresentam limitações na modelagem explícita de dependências temporais de longo alcance. Para lidar com essa característica intrínseca dos sinais fisiológicos, foram propostas as redes neurais recorrentes (Recurrent Neural Networks – RNN), que incorporam conexões cíclicas capazes de manter informações de estados anteriores ao longo da sequência (GOODFELLOW; BENGIO; COURVILLE, 2016).

Entretanto, as RNNs tradicionais sofrem com o problema do desvanecimento do gradiente². Como solução, foram introduzidas as redes Long Short-Term Memory (LSTM), que utilizam mecanismos de portas (*gates*) para controlar o fluxo de informações

²O desvanecimento do gradiente (*Vanishing Gradient*) ocorre quando os gradientes utilizados para atualizar os pesos da rede tornam-se extremamente pequenos durante a retropropagação. Isso impede que as camadas iniciais de redes profundas aprendam corretamente, pois seus pesos param de ser ajustados significativamente.

relevantes ao longo do tempo, permitindo a preservação de dependências de maior duração (HOCHREITER; SCHMIDHUBER, 1997). Conforme detalhado na Figura 2.5, essas portas (esquecimento, entrada e saída) regulam matematicamente o que deve ser preservado ou descartado em cada passo da sequência.

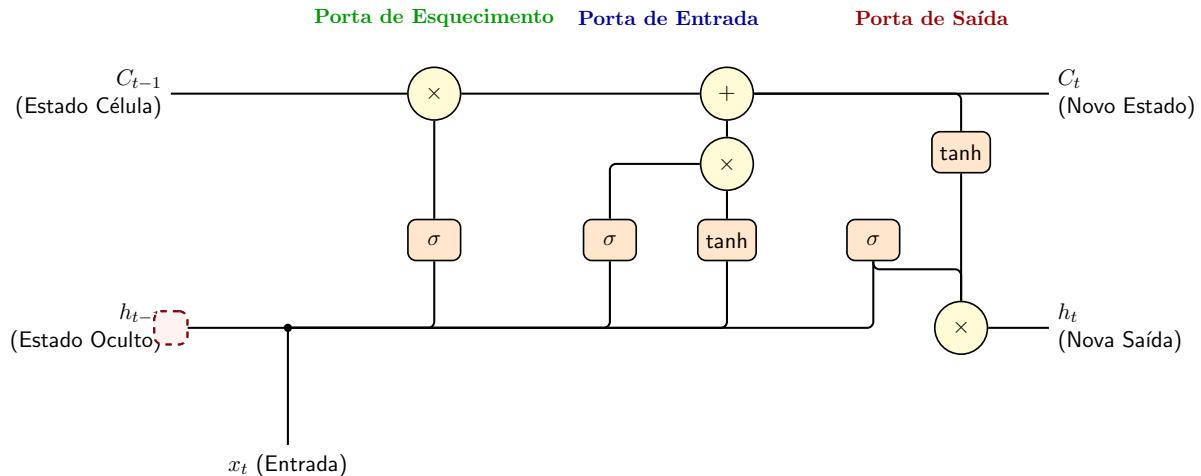


Figura 2.5: Estrutura interna de uma célula LSTM. Os **retângulos** representam as funções de ativação (sigmoide e tangente hiperbólica) que atuam como portões de controle. Os **círculos** indicam as operações pontuais (multiplicação e adição) que atualizam o estado da célula, permitindo o fluxo seletivo de informação.

Fonte: Baseado em Hochreiter e Schmidhuber (1997) e Olah (2015).

Na análise de ECG, as LSTM têm sido empregadas para capturar relações temporais entre diferentes batimentos e para modelar a dinâmica global do sinal. Contudo, o custo computacional elevado e a dificuldade de paralelização limitam sua aplicação isolada em tarefas de segmentação densa. Por essa razão, abordagens híbridas que combinam CNNs para extração local de características e LSTM para modelagem temporal têm sido amplamente exploradas, apresentando ganhos de desempenho em cenários específicos (YILDIRIM, 2018).

2.4 Redes Neurais Residuais (ResNet)

Em teoria, adicionar mais camadas a uma rede neural deveria aumentar sua capacidade de modelar funções complexas. No entanto, na prática, observa-se que o aumento excessivo da profundidade leva a um problema crítico conhecido como desvanecimento do gradiente (*vanishing gradient*). Durante o treinamento via retropropagação (*backpropagation*), o sinal de erro é multiplicado sucessivamente pelas derivadas das camadas conforme retorna

da saída para a entrada. Em redes muito profundas, essas multiplicações sucessivas de valores pequenos fazem com que o gradiente tenda a zero nas primeiras camadas, impedindo que elas aprendam ou atualizem seus pesos (HE et al., 2016).

Para solucionar esse problema de fluxo do gradiente e permitir o treinamento de redes com centenas de camadas, He et al. (2016) introduziram as Redes Neurais Residuais (ResNets). A inovação central é a introdução de conexões de salto (*skip connections* ou *shortcut connections*).

2.4.1 O Bloco Residual e o Fluxo do Gradiente

Diferente das redes tradicionais, as ResNets utilizam “blocos residuais” onde a entrada x é somada diretamente à saída das camadas de processamento. A operação realizada por um bloco residual é descrita matematicamente pela Equação 2.4, e sua estrutura gráfica pode ser visualizada na Figura 2.6.

$$y = \mathcal{F}(x, \{W_i\}) + x \quad (2.4)$$

Onde:

- x é o vetor de entrada;
- $\mathcal{F}(x)$ é a função residual aprendida pelas camadas (convoluções);
- O termo $+x$ é a conexão de identidade que preserva a informação original.

A intuição por trás dessa abordagem é que é mais fácil para a rede aprender que a função residual deve ser zero (ou seja, $\mathcal{F}(x) \approx 0$, resultando em $y \approx x$) do que aprender uma função identidade complexa do zero usando várias camadas não-lineares.

No contexto da análise de eletrocardiogramas, essa propriedade é particularmente valiosa. Como a morfologia do sinal cardíaco possui uma estrutura base bem definida e repetitiva (a sequência P-QRS-T), a conexão de identidade permite que a rede preserve essa integridade estrutural ao longo das camadas profundas. Em vez de tentar reconstruir as características da onda do zero a cada convolução, o que poderia degradar detalhes finos como entalhes no QRS ou a baixa amplitude da onda P, a rede aprende apenas os

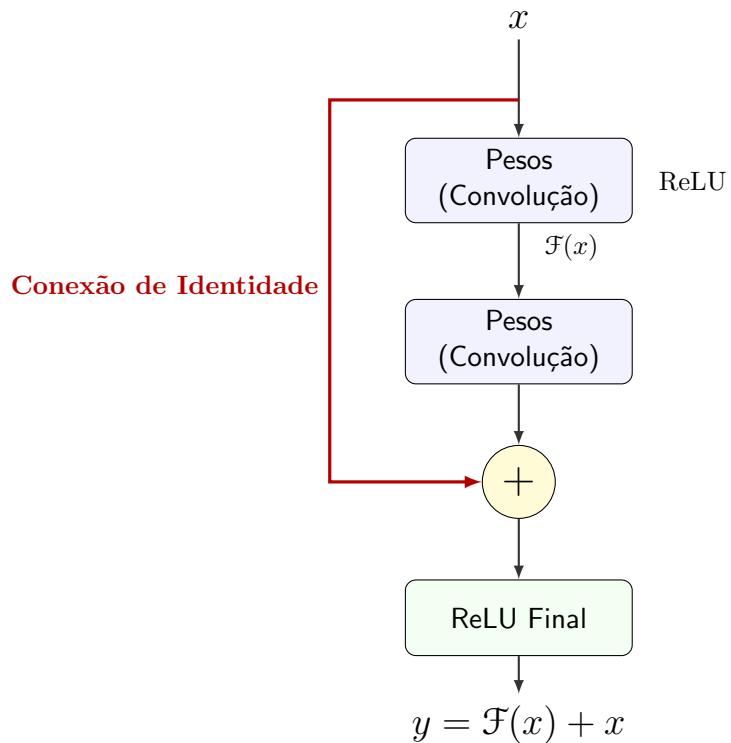


Figura 2.6: Diagrama de um bloco residual. O caminho da esquerda (seta vermelha) contorna as camadas de peso, permitindo que o gradiente fluia livremente durante a retropropagação.

Fonte: Adaptado de He et al. (2016).

“resíduos” necessários para refinar a segmentação. Isso significa que o modelo foca seus recursos computacionais em aprender as variações sutis e as bordas exatas das ondas, mantendo o fluxo do sinal original preservado através do atalho, estratégia que tem se mostrado eficaz na detecção de padrões complexos em séries temporais biomédicas (HANNUN et al., 2019).

Além disso, essa estrutura resolve o problema do fluxo do gradiente. Durante a retropropagação, o gradiente do erro em relação à entrada x passa a ter um termo aditivo constante de 1 proveniente da conexão de identidade ($\frac{\partial y}{\partial x} = \frac{\partial \mathcal{F}}{\partial x} + 1$). Esse termo “1” atua como uma “rodovia” (*highway*) para o gradiente, garantindo que o sinal de erro fluia diretamente para as camadas iniciais sem desaparecer, mesmo que os pesos \mathcal{F} sejam muito pequenos (HE et al., 2016). Isso viabilizou o treinamento das arquiteturas U-Net profundas e dos modelos híbridos utilizados neste trabalho.

Essa característica foi determinante para viabilizar o uso de arquiteturas profundas na segmentação de ECGs. Modelos mais profundos conseguem capturar dependências temporais de longo alcance e contextos globais do batimento sem sofrer com a degradação

da precisão na localização das ondas. Esse avanço permitiu a construção de arquiteturas híbridas robustas, como as baseadas em Residual U-Net, fundamentais para o desempenho superior relatado em estudos recentes da literatura (DURAJ et al., 2022; ZHANG; LIU; WANG, 2018).

2.4.2 Blocos Squeeze-and-Excitation (SE)

Enquanto as convoluções tradicionais operam misturando informações espaciais e de canais simultaneamente, os blocos *Squeeze-and-Excitation* (SE), propostos por Hu, Shen e Sun (2018), visam recalibrar explicitamente a interdependência entre os canais da rede.

O funcionamento do bloco ocorre em duas etapas principais:

1. **Squeeze (Compressão):** realiza-se uma operação de *Global Average Pooling* para agregar a informação espacial de cada canal em um único descritor numérico global;
2. **Excitation (Excitação):** utiliza-se uma rede neural *fully connected* leve para aprender pesos de importância para cada canal (entre 0 e 1).

O resultado é uma multiplicação elemento a elemento que enfatiza características informativas e suprime as irrelevantes, permitindo que a rede foque nos mapas de características mais importantes para a segmentação do ECG.

A Figura 2.7 ilustra esse fluxo de informação, destacando como o vetor de pesos recalibra os canais de entrada.

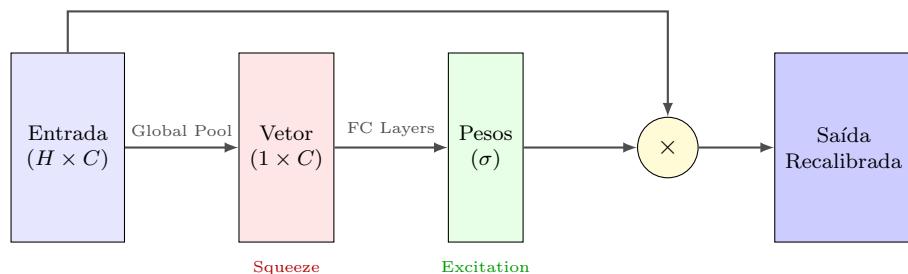


Figura 2.7: Diagrama esquemático do bloco Squeeze-and-Excitation (SE). O mecanismo comprime a informação espacial (Squeeze) para calcular a importância de cada canal (Excitation), utilizando esses pesos para recalibrar a entrada original.

Fonte: Adaptado de Hu, Shen e Sun (2018).

2.4.3 Residual U-Net: A Integração de Conexões Residuais na Segmentação

A arquitetura U-Net, base para este trabalho, é estruturalmente derivada dos ***Autoencoders***. Conforme definido por Goodfellow, Bengio e Courville (2016), um autoencoder é uma rede neural composta por duas etapas simétricas: um **codificador** (*encoder*), que comprime progressivamente a entrada para extrair apenas as características mais relevantes (representação latente), e um **decodificador** (*decoder*), que tenta reconstruir o sinal original a partir dessa versão comprimida.

Embora eficazes para redução de ruído, os autoencoders tradicionais sofrem com a perda de detalhes espaciais durante o processo de compressão, o que é crítico em segmentação médica, contexto no qual a localização temporal precisa dos eventos é determinante. A U-Net resolve esse problema introduzindo as **conexões de salto** (*skip connections*), visualizadas pelas setas tracejadas na Figura 2.8. Essas conexões transferem as características de alta resolução diretamente do encoder para o decoder, permitindo uma reconstrução precisa das fronteiras das ondas (RONNEBERGER; FISCHER; BROX, 2015).

Apesar da eficiência da U-Net clássica, o aumento de sua profundidade para capturar padrões de ECG muito complexos pode levar a dificuldades de convergência no treinamento. Para superar essa barreira, a literatura propôs a *Residual U-Net* (ResUNet), uma arquitetura híbrida que integra os blocos residuais (discutidos na Seção 2.4) diretamente na estrutura da U-Net, facilitando o fluxo do gradiente e permitindo o treinamento de redes mais profundas sem degradação do desempenho (??DURAJ et al., 2022).

Conforme ilustrado na Equação 2.5, a saída y de um bloco residual não é apenas o resultado das convoluções $\mathcal{F}(x)$, mas a soma deste resultado com a entrada original x :

$$y_{bloco} = \mathcal{F}_{conv}(x) + x \quad (2.5)$$

Essa alteração arquitetural traz duas vantagens cruciais para a segmentação de ECG:

1. **Fluxo de Informação Facilitado:** O atalho residual ($+x$) funciona como uma via

expressa para os gradientes, permitindo que a rede aprenda funções de identidade e evitando a degradação do desempenho em redes profundas (GOODFELLOW; BENGIO; COURVILLE, 2016);

2. Preservação de Características: A conexão residual ajuda a preservar as informações originais do sinal dentro do bloco, vital para não perder detalhes de baixa amplitude, como a onda P.

No trabalho de (DURAJ et al., 2022), replicado nesta pesquisa, essa estrutura é aprimorada com a inserção de blocos *Squeeze-and-Excitation* (SE) dentro das unidades residuais. Essa adição permite que a rede facilite o treinamento (via resíduos) e também aprenda a ponderar a importância de cada canal de filtragem (via SE), sem utilizar mecanismos de atenção global complexos, mantendo um custo computacional eficiente para processamento de ECG.

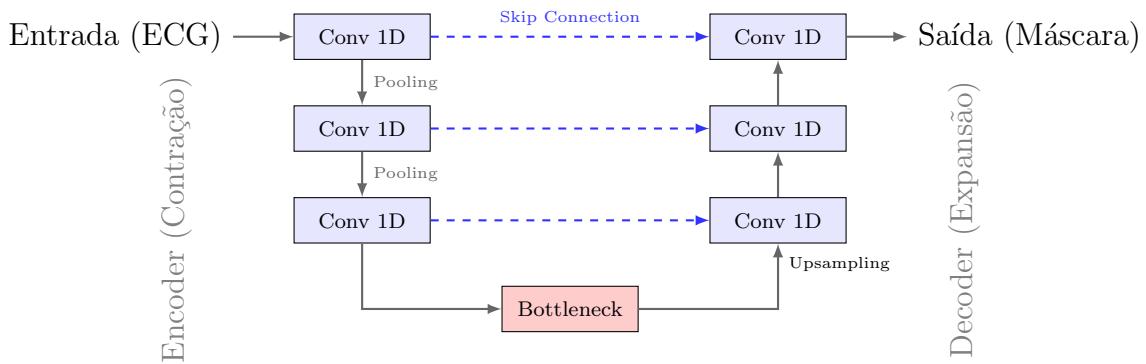


Figura 2.8: Arquitetura U-Net 1D. O caminho de contração (esquerda) extrai o contexto, enquanto o caminho de expansão (direita) recupera a localização precisa. As setas tracejadas (*skip Connections*) transferem detalhes de alta resolução do encoder para o decoder, essenciais para a segmentação precisa das ondas.

Fonte: Baseado em Ronneberger, Fischer e Brox (2015) e Duraj et al. (2022).

2.4.4 Mecanismos de Atenção e a Arquitetura Transformer

Para superar as limitações de memória das redes recorrentes e o campo receptivo local das convoluções, Bahdanau, Cho e Bengio (2014) introduziram o conceito de **Mecanismo de Atenção**. Originalmente desenhado para tradução automática, esse mecanismo permite que a rede “foque” dinamicamente em diferentes partes da sequência de entrada a cada

passo do processamento, superando o gargalo de tentar comprimir toda a informação em um único vetor de contexto.

Evoluindo esse conceito, Vaswani et al. (2017) propuseram o **Transformer**, uma arquitetura que abandona totalmente a recorrência sequencial em favor do processamento paralelo global. O núcleo dessa arquitetura é a **Autoatenção** (*Self-Attention*), que relaciona diferentes posições de uma única sequência para computar uma representação mais rica da mesma, capturando dependências de longo alcance instantaneamente.

Matematicamente, a autoatenção opera sobre três vetores projetados a partir da entrada: *Query* (Q), *Key* (K) e *Value* (V). A saída é uma soma ponderada dos valores, onde os pesos são calculados através da similaridade entre Q e K , conforme a Equação 2.6:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (2.6)$$

onde $\sqrt{d_k}$ é um fator de escala para estabilizar os gradientes. No contexto do eletrocardiograma, essa operação permite que o modelo aprenda que a morfologia de um complexo QRS atual (Q) deve ser interpretada à luz de batimentos anteriores (K), recuperando informações contextuais (V) essenciais para distinguir ritmos normais de arritmias complexas, mesmo que distantes temporalmente.

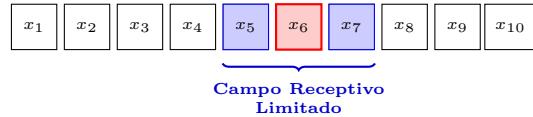
2.4.5 Transformers e Modelos Híbridos no ECG

A literatura sobre a análise automática de ECG tem evoluído progressivamente de arquiteturas puramente convolucionais para modelos capazes de capturar dependências temporais de longo alcance. Enquanto abordagens baseadas em CNNs, como a Residual U-Net discutida anteriormente (DURAJ et al., 2022), aprimoram a segmentação por meio de conexões residuais e mecanismos de recalibração local (SE), elas permanecem fundamentalmente limitadas pelo campo receptivo das convoluções.

Nesse contexto, os *Transformers* emergem como uma alternativa promissora. A sua capacidade de modelar relações globais torna-os particularmente adequados para a análise de ECG, onde arritmias e padrões patológicos podem depender de correlações temporais distantes, que escapam à visão local de uma CNN.

A Figura 2.9 ilustra essa distinção fundamental: enquanto a convolução observa apenas uma vizinhança restrita de pontos vizinhos, o mecanismo de atenção (base do *Transformer*) conecta cada ponto do sinal a todos os outros pontos da sequência, permitindo uma compreensão abrangente do contexto rítmico global.

(a) CNN: Contexto Local



(b) Transformer: Contexto Global

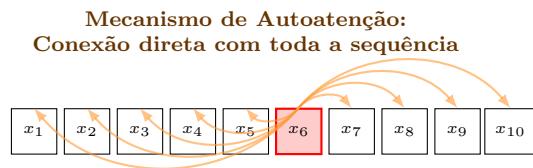


Figura 2.9: Comparaçāo do mecanismo de captura de contexto. Em (a), a CNN observa apenas uma vizinhança local. Em (b), o mecanismo de atenção conecta x_6 a todos os pontos da sequência (dependências de longo alcance).

Fonte: Baseado em Vaswani et al. (2017).

O modelo SEResUTer, proposto por Li et al. (2023) e também explorado neste trabalho, representa essa evolução híbrida. Diferentemente da abordagem puramente convolucional de Duraj et al. (2022), o SEResUTer introduz um módulo *Transformer* especificamente no gargalo (*bottleneck*) da rede. Sua composição integra:

1. A estrutura *U-Net* para a localização precisa;
2. Blocos Residuais e *Squeeze-and-Excitation* (similares aos de Duraj) no encoder/decoder para extração de features locais;
3. Mecanismos de Atenção (*Transformer*) na camada mais profunda para capturar o contexto global da série temporal, conforme detalhado na Figura 2.10.

Essa abordagem híbrida visa unir o melhor dos dois mundos: a eficiência local das CNNs (validada por Duraj et al. (2022)) com a capacidade de modelagem global dos Transformers (proposta por Li et al. (2023)).

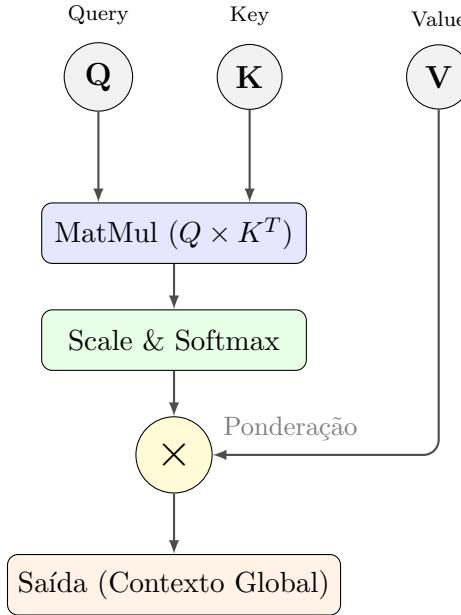


Figura 2.10: Representação esquemática do mecanismo de Atenção (*Scaled Dot-Product*). O modelo utiliza a similaridade entre Q e K para calcular pesos que filtram a informação relevante vinda de V .

Fonte: Adaptado de Vaswani et al. (2017).

2.5 Funções de Perda Ponderadas

Para o treinamento de classificação multiclasse, utiliza-se comumente a Entropia Cruzada Categórica (*Categorical Cross-Entropy*). Conforme discutido por Goodfellow, Bengio e Courville (2016), essa função mede a divergência entre a distribuição de probabilidade predita e a real.

Em cenários de segmentação de ECG, onde a classe “fundo” (linha isoelétrica) é majoritária, o desbalanceamento pode enviesar o modelo. Para corrigir isso, aplica-se uma ponderação na função de perda (*Weighted Cross-Entropy*), onde atribui-se um peso maior (w_c) aos erros cometidos nas classes minoritárias (Ondas P, QRS, T), penalizando mais severamente o modelo quando ele falha em detectar uma onda do que quando ele classifica incorretamente o fundo.

Matematicamente, a função de perda ponderada (L_{WCCE}) introduz um termo escalar w_c para cada classe c , que multiplica o valor logarítmico da perda. A equação formal é dada por:

$$L_{WCCE} = - \sum_{c=1}^C w_c \cdot y_c \cdot \log(\hat{y}_c) \quad (2.7)$$

Onde:

- C é o número de classes (no caso deste trabalho: Fundo, P, QRS, T);
- y_c indica se a classe c é a correta (1 ou 0);
- \hat{y}_c é a probabilidade predita pelo modelo;
- w_c é o peso calculado inversamente à frequência da classe.

Dessa forma, se a classe c for uma onda rara (peso w_c alto), o erro resultante será amplificado, forçando o gradiente a ser mais agressivo na correção dos pesos da rede para aquela classe específica.

2.6 Métricas de Avaliação de Desempenho

A avaliação quantitativa de modelos de segmentação e classificação de ECG baseia-se na comparação entre as previsões geradas pela rede neural e as anotações manuais realizadas por especialistas (*ground truth*). Essas métricas são derivadas da matriz de confusão, que contabiliza os Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), Falsos Positivos (FP) e Falsos Negativos (FN), como exemplo na figura Figura 2.11.

		Classe Real (Especialista)	
		Positivo	Negativo
Classe Predita (Modelo)	Positivo	Verdadeiro Positivo (TP) <i>Acerto: Onda Detectada</i>	Falso Positivo (FP) <i>Erro: Alarme Falso</i>
	Negativo	Falso Negativo (FN) <i>Erro: Onda Perdida</i>	Verdadeiro Negativo (TN) <i>Acerto: Fundo Ignorado</i>

Figura 2.11: Matriz de confusão para segmentação de ECG.
 Fonte: Elaborada pelo autor (2025).

2.6.1 Métricas baseadas na Matriz de Confusão

A Acurácia mede a proporção global de acertos do modelo. Embora intuitiva, ela pode ser enganosa em cenários de alto desbalanceamento de classes, como no ECG, onde o fundo (linha isoelétrica) predomina sobre as ondas (DURAJ et al., 2022). É definida por:

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

Para mitigar o viés da acurácia, utilizam-se métricas focadas na classe positiva. A Sensibilidade (ou *Recall*) avalia a capacidade do modelo de detectar corretamente os eventos de interesse (ondas), sendo crítica para evitar falsos negativos em diagnósticos clínicos (DURAJ et al., 2022):

$$Sensibilidade = \frac{TP}{TP + FN} \quad (2.9)$$

Simultaneamente, a Precisão avalia a confiabilidade das detecções positivas, indicando o quanto o modelo é robusto a falsos alarmes (ruído classificado como onda) (DURAJ et al., 2022):

$$Precisão = \frac{TP}{TP + FP} \quad (2.10)$$

2.6.2 Métricas de Sobreposição e Similaridade

Para tarefas de segmentação, onde o objetivo é delimitar uma região temporal, utiliza-se frequentemente o *F1-Score* (também conhecido como coeficiente Dice). Sendo a média harmônica entre precisão e sensibilidade, esta métrica penaliza tanto falsos positivos quanto falsos negativos, sendo mais robusta ao desbalanceamento de classes (DURAJ et al., 2022):

$$F1-Score = 2 \cdot \frac{Precisão \cdot Sensibilidade}{Precisão + Sensibilidade} \quad (2.11)$$

Outra métrica comum em segmentação semântica é o índice de Jaccard, ou *Intersection over Union (IoU)*, que mede a razão entre a interseção e a união das máscaras

preditas e reais. É uma métrica mais rígida que o F1-Score, penalizando severamente desalinhamentos (DURAJ et al., 2022):

$$IoU = \frac{TP}{TP + FP + FN} \quad (2.12)$$

2.6.3 Métricas de Erro Temporal

Além das métricas de classificação, a precisão da localização dos pontos fiduciais (início e fim das ondas) é avaliada pelo Erro Médio Absoluto (MAD). Medido em milissegundos (ms), o MAD calcula a média das distâncias absolutas entre o ponto predito pelo modelo e a anotação do especialista, sendo fundamental para validar a aplicabilidade clínica da segmentação em medições de intervalos como QT e QRS.

3 Trabalhos Relacionados

A aplicação de aprendizado profundo na análise de eletrocardiogramas (ECG) passou por uma evolução acelerada na última década. O campo transicionou da simples classificação de arritmias para tarefas complexas de segmentação semântica, buscando delimitar com precisão os pontos fiduciais (ondas P, QRS e T). Esta seção apresenta essa evolução cronológica, partindo de revisões sistemáticas que mapeiam o cenário geral até as arquiteturas híbridas mais recentes que fundamentam a parte prática deste trabalho.

3.1 Panorama da Literatura e Revisões Sistemáticas

Para compreender o estado da arte, revisões recentes oferecem um diagnóstico das tendências predominantes. Wasimuddin et al. (2020), em um survey abrangente, estruturaram o pipeline de análise de ECG em etapas de aquisição, pré-processamento e classificação, destacando a integração incipiente entre técnicas clássicas e *deep learning*. Mais recentemente, Tihak, Konjicija e Boskovic (2022) analisaram 32 artigos focados na detecção de fibrilação atrial, constatando a hegemonia das Redes Neurais Convolucionais (CNN) e o surgimento de modelos híbridos (CNN-LSTM).

Contudo, a análise crítica desses trabalhos revela uma lacuna importante: a literatura existente concentra-se majoritariamente na classificação global de arritmias (diagnóstico “doente” vs “saudável”), oferecendo pouca profundidade sobre as arquiteturas específicas para a segmentação semântica das ondas morfológicas. Essa constatação, de que as revisões atuais não respondem integralmente às especificidades da segmentação de ondas, justifica a condução de uma nova Revisão Sistemática da Literatura neste trabalho. O protocolo e a execução desta nova revisão, desenhada especificamente para responder às questões de pesquisa propostas (P1-P3), são detalhados a seguir no Capítulo 4.

3.2 Evolução das Arquiteturas: De CNNs a Transformers

A evolução dos modelos para segmentação de ECG pode ser dividida cronologicamente em três gerações principais, conforme detalhado a seguir.

3.2.1 Primeira Geração: Adaptação de CNNs (2015–2019)

Os primeiros esforços significativos concentraram-se na adaptação de Redes Neurais Convolucionais (CNNs), originalmente desenvolvidas para imagens bidimensionais, para o processamento de sinais unidimensionais. Trabalhos pioneiros demonstraram que CNNs 1D eram capazes de extrair características automaticamente, superando abordagens baseadas em engenharia manual de atributos.

Embora estudos preliminares, como o de Vijaya, Kumar e Verma (1998), já tivessem explorado o uso de redes neurais para anotação automática de sinais de ECG, essas abordagens eram fortemente limitadas pelo reduzido poder computacional disponível à época e pela escassez de bases de dados extensas e adequadamente anotadas, o que restringiu sua adoção em larga escala.

Foi apenas a partir do final da década de 2010 que os avanços em hardware e a maior disponibilidade de bases públicas permitiram o treinamento eficaz de redes convolucionais profundas para esta tarefa. Nesta fase, o foco da maioria dos trabalhos permaneceu centrado na detecção de picos R e do complexo QRS, com atenção limitada à segmentação precisa das ondas P e T.

3.2.2 Segunda Geração: U-Net e Modelos Residuais (2020–2022)

A virada de chave para a segmentação completa ocorreu com a adaptação da arquitetura U-Net para sinais 1D. Originalmente proposta para imagens biomédicas, a estrutura encoder-decoder provou-se ideal para gerar máscaras de segmentação temporal. Neste contexto, Duraj et al. (2022) representam um marco importante ao proporem a *1D Residual U-Net* com mecanismos de *Squeeze-and-Excitation* (SE). Utilizando a base LUDB (KALYAKULINA et al., 2020), os autores demonstraram que a inclusão de conexões resi-

duais e blocos SE permitia à rede focar em canais informativos e ignorar ruídos, alcançando alta precisão na delimitação de ondas P e T, historicamente mais difíceis de segmentar que o complexo QRS. Este trabalho consolidou a U-Net como o *baseline* robusto para a área.

3.2.3 Terceira Geração: Transformers e Arquiteturas Híbridas (2023–Presente)

A fase atual é marcada pela introdução de mecanismos de atenção global para superar as limitações das CNNs em capturar dependências de longo alcance. Li et al. (2023) introduziram o modelo SEResUTer, uma arquitetura híbrida que insere blocos Transformer no “gargalo” (*bottleneck*) da U-Net. A premissa é que, enquanto as convoluções tratam das características locais (morfologia da onda), o Transformer analisa o contexto global (ritmo e relação entre batimentos). Esta abordagem reflete a fronteira do conhecimento, onde a interpretabilidade e a capacidade de generalização para múltiplas bases de dados tornam-se os novos desafios a serem superados.

4 Revisão Sistemática da Literatura

Conforme discutido no capítulo 3, embora a literatura apresente revisões abrangentes sobre o uso de inteligência artificial em eletrocardiografia, como os trabalhos de Wasimuddin et al. (2020) e Tihak, Konjicija e Boskovic (2022), observa-se uma lacuna específica. As revisões existentes concentram-se majoritariamente em tarefas de classificação diagnóstica ou em panoramas generalistas do processamento de sinais.

Não foi identificada, até o momento, uma revisão sistemática recente dedicada exclusivamente a mapear e comparar as arquiteturas de *Deep Learning* voltadas para a **segmentação semântica** das ondas P, QRS e T. A distinção é crucial: enquanto a classificação busca apenas rotular um registro, a segmentação exige a delimitação temporal precisa de cada componente morfológico, tarefa que demanda arquiteturas e métricas de avaliação distintas.

Diante dessa necessidade, este capítulo apresenta uma Revisão Sistemática da Literatura conduzida especificamente para preencher essa lacuna. O estudo segue o protocolo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (PAGE et al., 2021), com o objetivo de responder às questões de pesquisa P1, P2 e P3.

4.1 Questões de Pesquisa Norteadoras

Para guiar ambas as etapas metodológicas, foram formuladas três questões de pesquisa (P) fundamentais, focadas na compreensão do cenário atual da literatura:

- **P1:** Quais são as arquiteturas de *deep learning* mais utilizadas para a segmentação de ondas do ECG na literatura recente (2015–2025)?
- **P2:** Quais bases de dados públicas são predominantes no treinamento desses modelos e quais estratégias de enriquecimento de dados (*Data Augmentation*) são empregadas?
- **P3:** Quais são as métricas de avaliação mais adotadas para aferir o desempenho

dos modelos de segmentação de ECG?

A seguir, são detalhados os métodos específicos para cada eixo.

4.2 Metodologia da Revisão

Nesta seção são apresentados os procedimentos adotados para a condução da revisão sistemática, com o objetivo específico de responder às perguntas **P1**, **P2** e **P3**. O protocolo seguiu as recomendações do PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (PAGE et al., 2021). A aplicação das diretrizes PRISMA contribui para garantir transparência, reproduzibilidade e rigor metodológico, orientando desde a busca até a seleção dos estudos.

4.2.1 Fonte de dados e estratégia de busca

As buscas foram conduzidas em três bases de dados reconhecidas internacionalmente pela relevância na área de ciência e engenharia biomédica: IEEEExplore (IEEE, 2025), PubMed (National Center for Biotechnology Information, 2025) e Web of Science (CLARIVATE, 2025). A estratégia de busca foi elaborada com o uso de operadores booleanos, expressões truncadas e termos específicos relacionados ao objetivo da revisão. A seguinte expressão foi utilizada:

(“ ECG ”OR “ electrocardiogram ”) AND (“ wave delineation ”OR “ P wave detect ”OR “ T wave detect ”OR “ QRS detect* ”OR “ QRS complex* ”OR “ automated annotation ”) AND (“ deep learning ”OR “ neural network* ”) AND (“ semantic segmentation ”OR “ time-series segmentation ”OR “ signal processing ”)

As buscas foram realizadas no mês de junho de 2025, considerando o intervalo de publicação entre 2015 e 2025. Foram incluídos apenas artigos publicados em inglês ou português brasileiro, disponíveis em texto completo e que abordassem a identificação das ondas do ECG por meio de redes neurais profundas.

Para organização, triagem e seleção dos registros, foi utilizado o software Rayyan.ai (OUZZANI et al., 2016), uma plataforma gratuita para revisões sistemáticas. A ferramenta permitiu:

- Importar os resultados das buscas diretamente das bases;
- Eliminar automaticamente artigos duplicados;
- Facilitar a triagem por meio da leitura rápida de títulos e resumos;
- Aplicar etiquetas de inclusão/exclusão com base nos critérios definidos;
- Adicionar anotações colaborativas durante a avaliação dos estudos.

4.2.2 Critérios de elegibilidade e exclusão

Os critérios de inclusão e exclusão foram definidos previamente à leitura dos artigos, com base no escopo da revisão:

Critérios de Inclusão

- Estudos que utilizaram redes neurais profundas (ex.: CNN, LSTM, U-Net, Transformers) para detecção, predição ou delineamento de ondas ou pontos-chave do ECG (P, QRS, T, R-peak, etc.);
- Trabalhos com aplicação prática ou validação em bases públicas de ECG, como *Lobachevsky University Database* (LUDB) (KALYAKULINA et al., 2020), MIT-BIH (MOODY; MARK, 2001) e *QT Database* (QTDB) (LAGUNA et al., 1997);
- Artigos revisados por pares;
- Publicados entre 2015 e 2025.

Critérios de Exclusão

- Artigos de revisão, editoriais, capítulos de livro ou anais de evento sem acesso ao texto completo;

- Estudos que não utilizam *deep learning* para detecção, predição ou delineamento de ondas/pontos do ECG;
- Trabalhos que têm como objetivo exclusivamente a classificação de arritmias ou doenças, sem qualquer foco na detecção de ondas/pontos do ECG (mesmo que usem *deep learning*);
- Trabalhos que não apresentavam detalhes suficientes sobre a arquitetura do modelo, dados utilizados ou métricas de avaliação.

4.2.3 Seleção de estudos

A seleção dos artigos seguiu as três etapas recomendadas para revisões sistemáticas:

1. Identificação e remoção de duplicatas: utilizando a função automática do Rayyan;
2. Triagem por títulos e resumos: eliminação de registros irrelevantes ou fora dos critérios definidos;
3. Leitura completa dos artigos selecionados: análise criteriosa para verificação final de elegibilidade.

A busca inicial resultou em 175 registros. Após a identificação de 62 duplicatas e a resolução de conflitos, foram excluídos 33 trabalhos nesta fase inicial. Assim, 142 artigos permaneceram para a triagem por título e resumo. Desses, 47 estudos foram selecionados para leitura na íntegra. Com base na análise completa, 42 artigos atenderam aos critérios e foram incluídos na análise final. O processo completo de seleção está representado na Figura 4.1.

4.2.4 Extração de dados

Para cada um dos estudos incluídos, foi realizada a extração de dados por meio de uma planilha padronizada. Os itens extraídos foram selecionados estratégicamente para responder às perguntas de pesquisa (Ps) norteadoras deste trabalho:

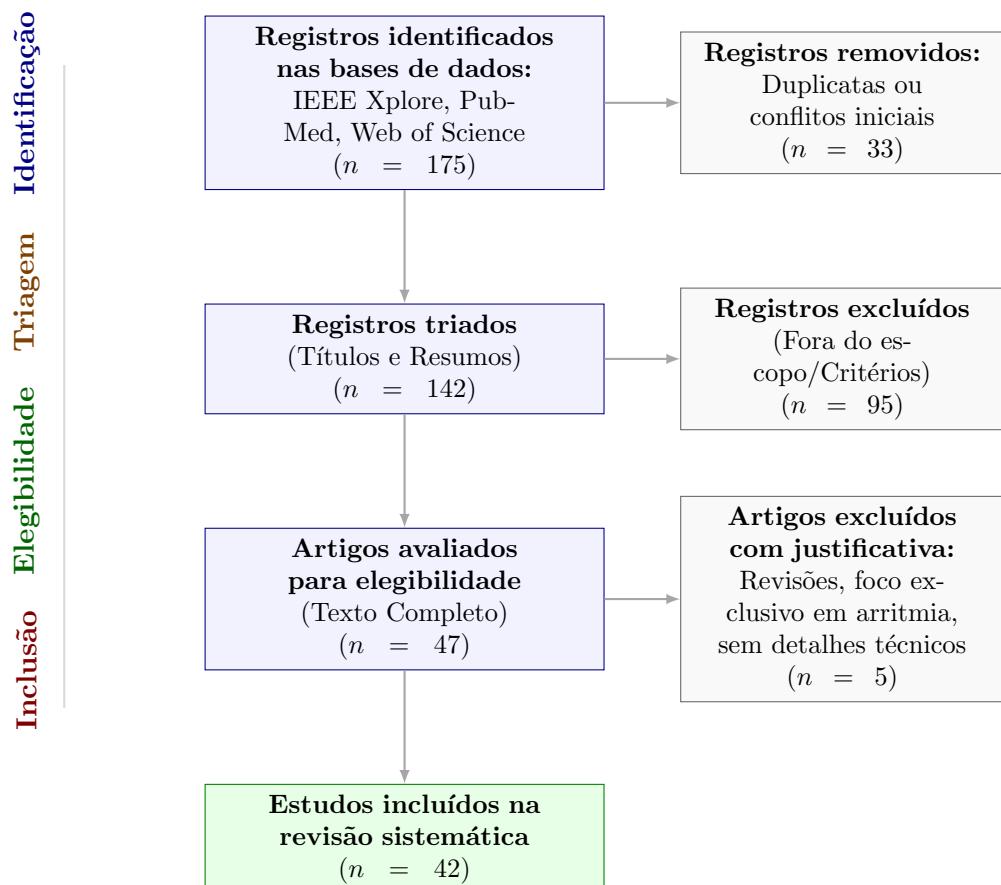


Figura 4.1: Fluxograma PRISMA 2020 detalhando o processo de seleção dos estudos. Dos 175 registros iniciais, 42 atenderam a todos os critérios de inclusão e compõem o corpus desta revisão.

Fonte: Elaborada pelo autor (2025).

- **Para responder à P1 (Arquiteturas):** Foram extraídos dados sobre a arquitetura de rede neural aplicada (ex.: CNN, LSTM, U-Net, Transformers) e suas configurações específicas;
- **Para responder à P2 (Bases de Dados):** Levantou-se a base de dados de ECG utilizada (pública ou privada), bem como as técnicas de pré-processamento e estratégias para lidar com limitações (como *data augmentation*, filtragem e segmentação por janelas);
- **Para responder à P3 (Métricas):** Foram registradas as métricas de avaliação de desempenho utilizadas (ex.: F1-score, IoU, acurácia, erro médio de tempo) para aferir a qualidade da segmentação;
- **Informações Complementares:** Principais contribuições relatadas e limitações metodológicas mencionadas pelos autores.

A avaliação da qualidade metodológica dos estudos foi feita de maneira qualitativa, considerando a clareza na descrição dos métodos, a completude dos resultados apresentados e a solidez da validação dos modelos. Foram valorizados estudos com validação cruzada, comparação com métodos clássicos e uso de múltiplas bases de dados. Os procedimentos descritos nesta seção forneceram a base metodológica necessária para a análise dos resultados apresentada a seguir.

4.3 Resultados e Discussão da Revisão

Com base nos 42 estudos selecionados através do protocolo descrito acima, apresenta-se a seguir a análise detalhada para responder às questões norteadoras.

4.3.1 Arquiteturas de *Deep Learning* (P1)

A análise cronológica dos 42 trabalhos revisados permite compreender a evolução do uso de redes neurais profundas (*Deep Learning*) na segmentação e classificação de sinais de ECG. Observou-se uma transição clara entre três fases distintas: início (2018-2019), consolidação (2020-2021) e sofisticação e integração de técnicas avançadas (2022 em diante).

Na fase inicial (2018-2019), os trabalhos concentram-se em arquiteturas convolucionais básicas (CNN 1D) e *Fully Convolutional Networks* (FCNs), com foco em tarefas fundamentais, como a detecção de picos R e complexos QRS. Embora algumas abordagens tenham explorado redes recorrentes (LSTMs e BLSTMs) para capturar dependências temporais, a aplicação em tempo real era limitada. Métricas como *Sensibilidade* (Se), *Valor Preditivo Positivo* (PPV) e *F1-score* já alcançavam valores próximos a 99% em bases clássicas como MIT-BIH, indicando o potencial dessas redes para análises de alta precisão.

Com a maturação das pesquisas (2020-2021), verifica-se a popularização da U-Net e suas variantes 1D, adaptadas para a segmentação de ondas P, QRS e T. Modelos híbridos CNN-LSTM ou CNN-BiLSTM tornam-se frequentes, combinando a extração espacial de características das CNNs com a memória temporal das redes recorrentes. O desempenho reportado nesse período é consistentemente elevado, com F1-scores acima de 97% em delineação completa e acurácia próxima de 99% para detecção de QRS, como observado em métodos baseados em U-Net e arquiteturas dilatadas (DCNN-LSTM). Este avanço indica o início da consolidação de soluções capazes de lidar com maior variabilidade de sinais.

A partir de 2022, observa-se um movimento em direção à sofisticação arquitetural e à integração de mecanismos de atenção e Transformers. Modelos como SEResUTer e ECG_DEEPNet ilustram essa tendência, incorporando blocos residuais (ResNet), *Squeeze-and-Excitation*, autoatenção e codificadores Transformer a estruturas U-Net ou redes híbridas. Estas abordagens buscam maior robustez a ruídos generalização para múltiplas derivações e até mesmo capacidade de análise de ECG em dispositivos vestíveis. Resultados recentes indicam F1-scores superiores a 99% para detecção de QRS e desempenho consistente na identificação de ondas P e T, ainda que estas últimas permaneçam mais desafiadoras devido a menor amplitude e à variabilidade morfológica.

Como pode ser observado na Figura 4.2, o levantamento realizado evidencia um amadurecimento das técnicas de *Deep Learning* em ECG, que passaram de soluções baseadas em CNNs simples para modelos híbridos e de múltiplos componentes, capazes de oferecer altíssima precisão e maior potencial de aplicação clínica. Ainda assim, desafios persistem, como a generalização para dados externos, a interpretabilidade dos modelos

e a escassez de bases anotadas com alta qualidade, que continuam a motivar pesquisas recentes.

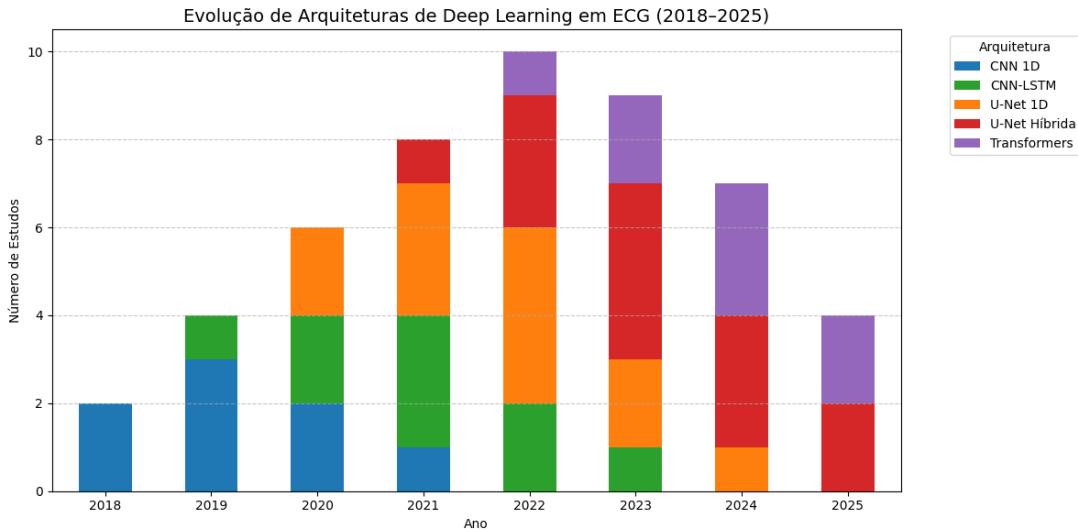


Figura 4.2: Evolução das arquiteturas de *deep learning* aplicadas ao ECG (2018–2025).

Fonte: Elaborada pelo autor(a).

Em síntese, a evolução arquitetural observada nos últimos anos acompanha a crescente necessidade de modelos capazes de lidar com variações morfológicas, ruídos e requisitos de aplicabilidade clínica. Embora CNNs continuem sendo referência em desempenho, arquiteturas híbridas vêm ganhando destaque, apesar de maior custo computacional.

4.3.2 Bases de Dados e Data Augmentation (P2)

A análise dos 42 estudos revisados evidencia uma clara transição no uso de bases de dados públicas e privadas ao longo do período de 2018 a 2024, refletindo mudanças tanto na complexidade dos modelos quanto nas exigências de generalização clínica.

Período Inicial (2018–2019): Predominância de Bases Clássicas. Nos anos iniciais, observou-se forte concentração no uso de bases tradicionais, especialmente o MIT-BIH Arrhythmia Database (MITDB), utilizado por 75% dos estudos (ex: (SODMANN et al., 2018); Camps et al., 2018). Esse domínio deve-se à sua ampla disponibilidade, anotação manual confiável e foco em arritmias. Paralelamente, a QT Database (QTDB) foi empregada em 50% dos trabalhos. A respeito de sua acessibilidade, essas bases apresentavam limitações quanto à diversidade patológica e representatividade de cenários reais, como ambientes ambulatoriais ou sinais com ruído originário de dispositivos vestíveis.

Transição (2020–2021): Expansão para Bases Especializadas e Ambulatoriais. A partir de 2020, identificou-se uma expansão para bases mais especializadas. A LUDB (Lobachevsky University Database) passou a representar 37,5% da amostra (ex: (XU et al., 2023)), oferecendo dados multiderivação (12 derivações) com anotações detalhadas para tarefas de segmentação semântica. Simultaneamente, conjuntos como a CPSC 2019/2021 e a Noise Stress Test Database (NSTDB) introduziram registros mais ruidosos e realistas, refletindo condições clínicas mais desafiadoras. Aproximadamente 68,75% dos estudos nessa fase passaram a adotar múltiplas bases em combinação, sinalizando uma busca por maior robustez na validação cruzada dos modelos.

Consolidação (2022–2024): Diversidade e Personalização. Entre 2022 e 2024, consolidou-se uma tendência de uso de bases mais diversificadas e híbridas, tanto públicas quanto privadas. O PTB-XL, atualmente o maior banco público com ECGs de 12 derivações e metadados clínicos, foi destaque em estudos recentes (ex: (SILVA et al., 2024)). Simultaneamente, a incorporação de bases privadas, como a do Chang Gung Memorial Hospital (2023) e CardioCloud Medical (2023), passou a representar uma estratégia de personalização para contextos clínicos específicos, embora isso levante questões sobre reproduzibilidade e acesso (Zhu et al., 2024). Houve ainda a adoção inicial de bases sintéticas, especialmente em estudos com modelos de difusão, sinalizando uma nova fronteira para treinamento de redes profundas com dados simulados.

A análise temporal entre 2018 e 2024 revela um crescimento expressivo no número de estudos que adotam múltiplas bases de dados para validação de modelos de ECG. Entre 2018 e 2019, apenas 31,25% dos trabalhos incluíam duas ou mais bases, com foco majoritário em conjuntos tradicionais como a MITDB. A partir de 2020, essa proporção salta para 68,75%, acompanhando o surgimento de modelos mais complexos e a necessidade de validar desempenho em cenários clínicos variados. Já no triênio 2022–2024, aproximadamente 75% dos estudos passaram a combinar três ou mais bases, incluindo públicas, privadas e sintéticas.

O aumento na média de bases por estudo, de 1,3 em 2018 para 3,2 em 2024, marca uma mudança de paradigma na pesquisa com ECG baseada em *deep learning*, conforme ilustrado na Figura 4.3. A integração de múltiplas fontes tornou-se uma exigência

metodológica para garantir desempenho clínico confiável.

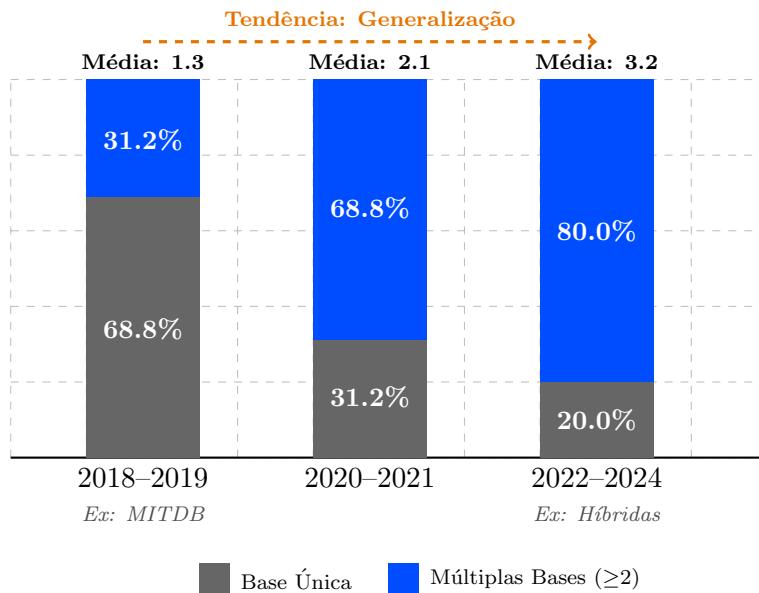


Figura 4.3: Evolução temporal da estratégia de validação (2018–2024). O gráfico evidencia a inversão do paradigma: o uso de bases únicas (cinza) cede lugar à integração de múltiplas bases (azul), impulsionada pela necessidade de generalização dos modelos.

Fonte: Elaborada pelo autor (2025).

Como parte da análise sobre as estratégias utilizadas para mitigar as limitações das bases de dados (P2), investigaram-se também as técnicas de *data augmentation* relatadas nos estudos. Verificou-se que aproximadamente 37% empregaram tais técnicas com o objetivo de mitigar limitações recorrentes, como número reduzido de amostras e desbalanceamento. A Figura 4.4 apresenta a distribuição das categorias de *data augmentation* entre os estudos analisados.

As estratégias mais comuns (Tabela 4.1) incluíram janelamento com sobreposição, transformações do sinal (inversão, escalonamento) e injeção de ruído realista. De forma geral, os trabalhos que aplicaram *data augmentation* apresentaram modelos com maior capacidade de generalização e tolerância a sinais de baixa qualidade.

4.3.3 Métricas de Avaliação (P3)

A terceira pergunta de pesquisa (P3) buscou identificar as métricas mais adotadas na literatura para aferir o desempenho dos modelos. A análise revelou que a escolha das métricas reflete a natureza híbrida do problema de segmentação de ECG, exigindo tanto

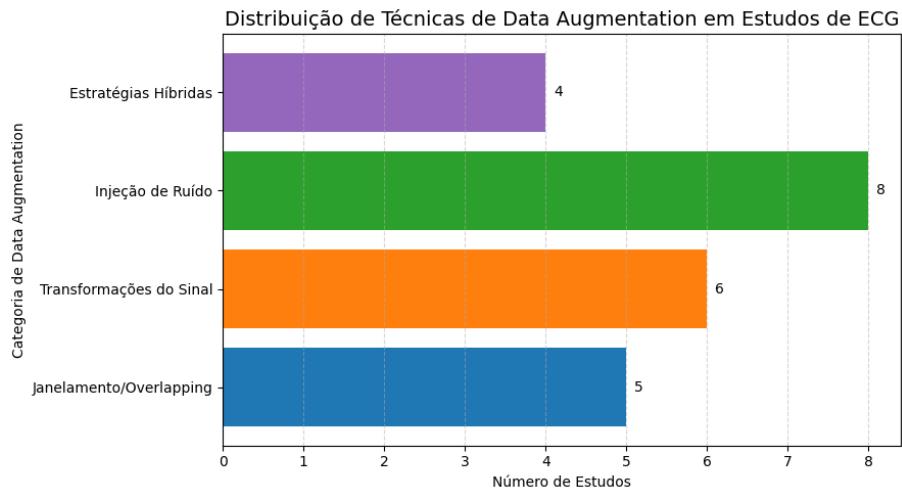


Figura 4.4: Distribuição das categorias de *data augmentation* utilizadas nos estudos de ECG analisados.

Fonte: Elaborada pelo autor(a).

Tabela 4.1: Resumo das estratégias de *data augmentation* empregadas nos estudos selecionados.

Estudo	Base de Dados	Estratégias de Data Augmentation
Camps et al. (2018)	QTDB	Janelamento com 99% de sobreposição; adição de constantes.
Sodmann et al. (2018)	QTDB + NSTDB	Ruído gaussiano, <i>baseline wander</i> , deslocamento.
Zahid et al. (2021)	MIT-BIH + NSTDB	Injeção de <i>baseline wander</i> e ruído sintético.
Habib et al. (2022)	MIT-BIH	Replicação de batimentos minoritários, deslocamentos.
Li et al. (2023)	LUDB + CPSC 2020	Translação temporal, escalonamento, inversão.
Wang et al. (2023)	LUDB	Blindagem de derivações (<i>random shielding</i>), <i>cut-out</i> .
Bioengineering (2023)	LUDB	Janelas de 4 s com sobreposição de 75%.

NSTDB: Noise Stress Test Database; **CPSC:** China Physiological Signal Challenge.

a classificação correta da morfologia quanto a precisão temporal.

Identificou-se uma predominância absoluta de métricas que ponderam o equilíbrio entre a detecção de eventos e a confiabilidade da predição. Conforme detalhado na Tabela 4.2, as métricas mais recorrentes foram:

- **F1-Score (ou Coeficiente de Dice):** Identificada como a métrica padrão-ouro para comparação de desempenho (> 90% dos estudos).
- **Sensibilidade e Precisão:** A maioria dos autores opta por reportar esse par de métricas em conjunto, visando fornecer um diagnóstico clínico de segurança e eficiência.
- **IoU (Jaccard):** Sua presença foi notada especialmente em trabalhos que utilizam arquiteturas de visão computacional (U-Nets) para validar a qualidade geométrica.

- **Erro Temporal (MAD):** Estudos focados na aplicabilidade médica tendem a incluir métricas de erro temporal para validar a precisão dos limites (*onsets/offsets*).

Tabela 4.2: Principais métricas de avaliação identificadas na revisão sistemática.

Métrica	Aplicação Principal Identificada	Frequência Estimada
F1-Score / DSC	Avaliação global de desempenho em dados desbalanceados.	Alta (> 90%)
Sensibilidade (Se)	Avaliação de segurança clínica (não perder eventos).	Alta (> 90%)
Precisão (PPV)	Avaliação de robustez a ruído (não gerar falsos alarmes).	Alta (> 85%)
Acurácia (Acc)	Visão geral, mas frequentemente citada com ressalvas sobre viés.	Média (~ 60%)
IoU (Jaccard)	Avaliação geométrica da sobreposição de segmentos.	Média (~ 50%)
MAD (ms)	Precisão temporal fina para delimitação de <i>onsets/offsets</i> .	Média-Baixa (~ 40%)

4.4 Síntese da Revisão

A revisão sistemática permitiu mapear o estado da arte e identificar que a área de segmentação de ECG caminha para a consolidação de arquiteturas híbridas (como U-Nets associadas a Transformers), o uso mandatório de múltiplas bases de dados para garantir generalização e a aplicação de técnicas robustas de *data augmentation*. Além disso, o levantamento das métricas confirmou a necessidade de uma avaliação multidimensional (classificação + precisão temporal).

Para consolidar os achados teóricos deste capítulo, a Tabela 4.3 sintetiza as respostas para as questões de pesquisa fundamentadas na literatura (P1, P2 e P3). Esses achados embasam diretamente as escolhas metodológicas da etapa experimental (Capítulo 5).

Tabela 4.3: Síntese das Questões de Pesquisa e Respostas Obtidas

P1	Quais arquiteturas de <i>deep learning</i> têm sido aplicadas ao delineamento de ondas P, QRS e T?
	A literatura evidencia uma evolução tecnológica clara: partindo de CNNs 1D simples para modelos U-Net 1D, incorporando posteriormente conexões residuais e blocos SE para maior profundidade. O estado da arte atual explora arquiteturas híbridas que fundem a capacidade local das CNNs com o contexto global dos <i>Transformers</i> (mecanismos de atenção).
P2	Quais bases de dados e estratégias de <i>data augmentation</i> são mais utilizadas?
	QTDB, LUDB e MIT-BIH permanecem sendo as bases mais correntes, embora haja crescimento no uso da PTB-XL. O <i>data augmentation</i> tornou-se mandatório para lidar com o desbalanceamento de classes e variabilidade morfológica, utilizando estratégias como janelamento com sobreposição, transformações do sinal e injeção de ruído realista.
P3	Quais métricas são mais empregadas para avaliação de segmentação de ECG?
	Predominam F1-score/DSC para avaliação global, complementadas pelo par Sensibilidade/Precisão para diagnóstico clínico do modelo. O erro temporal (MAD em ms) é essencial para validar a precisão dos limites de onda (onsets/offsets).

5 Delineamento Experimental e Arquiteturas

Após o mapeamento do estado da arte realizado no capítulo 4, esta etapa apresenta uma análise experimental aplicada. O objetivo é observar, em um cenário prático e controlado, o comportamento das arquiteturas convolucionais e híbridas quando submetidas a limitações de dados, verificando aspectos de reproduzibilidade e consistência dos resultados.

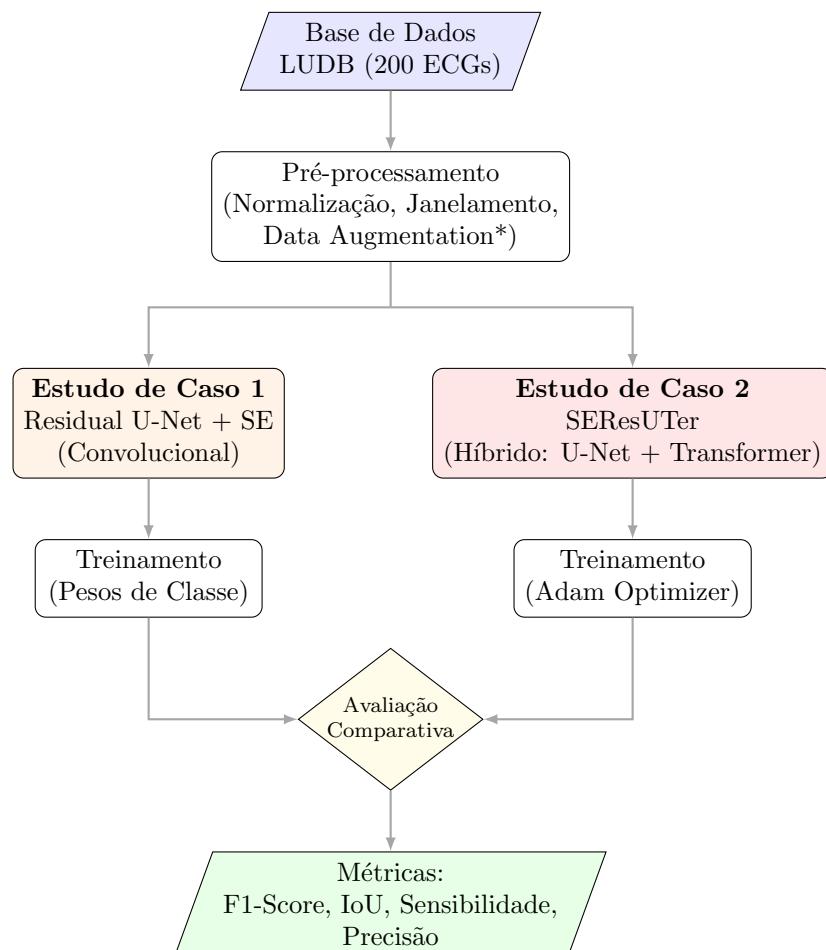
5.1 Fluxo do Experimento

Para garantir a comparabilidade entre os dois paradigmas, foi desenhado um fluxo de trabalho unificado, onde ambos os modelos são submetidos aos mesmos processos de ingestão de dados e critérios de avaliação, conforme ilustrado na Figura 5.1.

5.2 Definição dos Estudos de Caso

Para testar a viabilidade de replicação, foram selecionadas duas arquiteturas que representam paradigmas opostos identificados no estado da arte:

- **Caso 1 (Convolucional):** Baseado em Duraj et al. (2022), utilizando uma Residual U-Net. Representa a abordagem clássica, eficiente e focada em características locais.
- **Caso 2 (Híbrido):** Baseado no modelo SEResUTer de Li et al. (2023), que incorpora mecanismos de atenção (*Transformers*). Representa a tendência moderna de captura de contexto global.



*Data Augmentation aplicado especificamente no Estudo de Caso 2.

Figura 5.1: Fluxo metodológico da etapa de replicação experimental. Ambos os modelos foram submetidos a pipelines similares de pré-processamento e avaliação para garantir a comparabilidade dos resultados.

Fonte: Elaborada pelo autor (2025).

5.3 Base de Dados Utilizada (LUDB)

O estudo utilizou integralmente a LUDB. A escolha desta base justifica-se pela sua alta qualidade de anotação e pela diversidade de morfologias, essenciais para testar a generalização dos modelos. A Figura 5.2 ilustra a complexidade do sinal e o padrão de segmentação (*ground truth*) disponível na base.

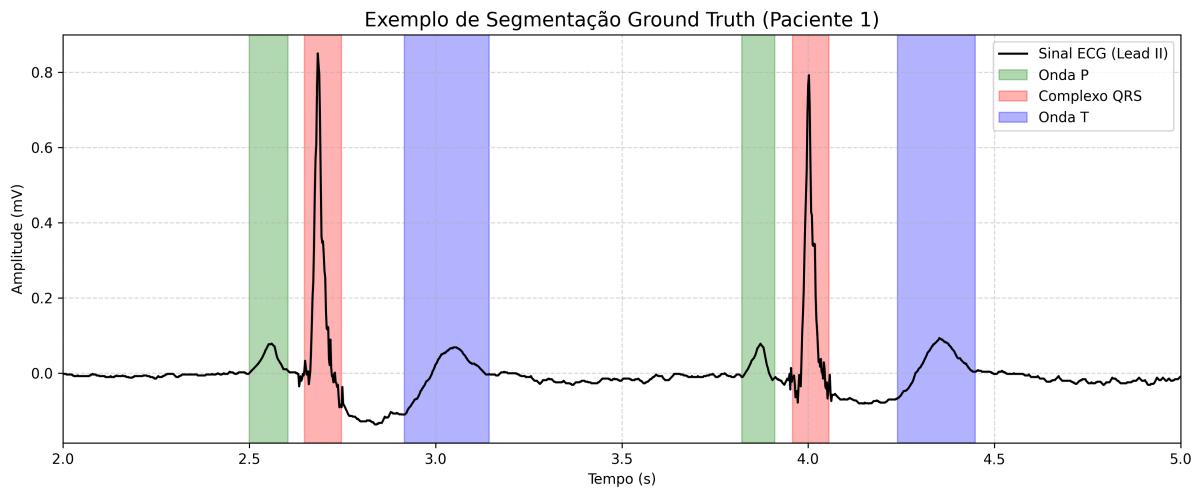


Figura 5.2: Exemplo de um ciclo cardíaco segmentado na base LUDB. As áreas coloridas representam as anotações de especialistas para Onda P, Complexo QRS e Onda T.

Fonte: Elaborada pelo autor (2025).

5.3.1 Caracterização dos Dados

A base é composta por registros de 200 pacientes, totalizando 2400 sinais individuais (considerando as 12 derivações padrão por paciente: I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5 e V6).

- **Perfil Demográfico:** A base de dados apresenta distribuição heterogênea em relação ao sexo e à idade. Observa-se predominância de indivíduos do sexo masculino (57,5%) em relação ao feminino (42,5%). A faixa etária é ampla, variando aproximadamente dos 10 aos 90 anos, com maior concentração de pacientes entre 50 e 70 anos. Essa variabilidade demográfica contribui para que o modelo seja treinado em sinais provenientes de diferentes perfis biológicos, reduzindo o risco de viés associado a sexo ou idade específicos.

- **Diversidade Clínica e Patológica:** Um diferencial crítico da base LUDB é que ela não se restringe a indivíduos saudáveis, incluindo uma variedade representativa de patologias. A Tabela 5.1 detalha a distribuição dos ritmos cardíacos presentes na base.

Embora haja predominância do Ritmo Sinusal (padrão normal), destaca-se a presença de 15 casos de Fibrilação Atrial (7,5% da base). A inclusão desta patologia é fundamental para o teste de robustez, pois a Fibrilação Atrial frequentemente suprime a Onda P e introduz irregularidades no intervalo R-R, desafiando a capacidade de generalização das redes neurais em cenários clínicos reais.

- **Desbalanceamento de Classes:** A análise exploratória confirmou um grande desbalanceamento. Como evidenciado na Figura 5.3, a classe “Fundo” (linha isoelétrica) representa a vasta maioria dos pontos, enquanto ondas críticas como a Onda P representam uma fração minoritária. A visualização desta disparidade fundamentou a adoção da estratégia de *Class Weights* (detalhada na Seção 5.5.2).

Tabela 5.1: Distribuição dos diagnósticos de ritmo cardíaco presentes nos registros da base LUDB.

Diagnóstico / Ritmo	Qtd.	Pacientes	Percentual
Ritmo Sinusal (Normal)	145		72,5%
Bradicardia Sinusal	25		12,5%
Fibrilação Atrial	15		7,5%
Outros (Arritmias Diversas)	11		5,5%
Taquicardia Sinusal	4		2,0%
Total	200		100,0%

Fonte: Elaborada pelo autor (2025).

5.4 Protocolo de Avaliação e Ferramentas

Para garantir a reprodutibilidade e a comparabilidade entre os estudos de caso, estabeleceu-se o seguinte protocolo experimental:

- **Entrada Unificada:** Diferentemente dos estudos originais que utilizavam janelas temporais distintas, neste trabalho ambas as arquiteturas (Convolucional e Híbrida)

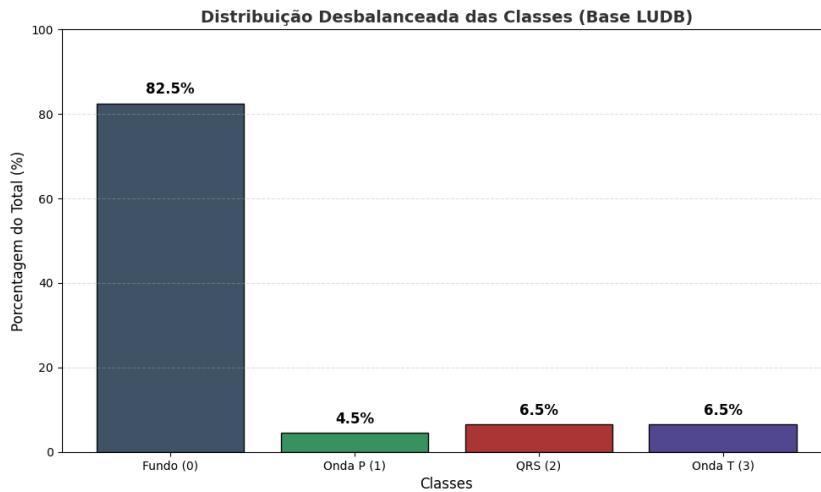


Figura 5.3: Distribuição percentual das classes na base de treinamento, evidenciando o predomínio da classe Fundo.

Fonte: Elaborada pelo autor (2025).

foram adaptadas para receber o sinal completo de 10 segundos (5000 pontos) a 500 Hz. Isso elimina o viés de pré-processamento, garantindo que qualquer diferença de desempenho seja atribuída exclusivamente à capacidade da arquitetura.

- **Ambiente de Execução:** Os experimentos foram conduzidos em ambiente controlado utilizando GPU NVIDIA RTX 3060, com implementações em *Python* (TensorFlow/Keras).
- **Divisão Estratificada:** O conjunto de dados (2400 sinais) foi dividido seguindo a proporção 80/20:
 - **Treino:** 1920 sinais (utilizados para ajuste de pesos).
 - **Teste:** 480 sinais (reservados estritamente para a avaliação final das métricas).
- **Métricas Estatísticas:** Os resultados finais são reportados como Média \pm Desvio Padrão ($\mu \pm \sigma$) de 5 execuções independentes, permitindo avaliar a estabilidade de convergência.
- **Prevenção de Sobreajuste (*Early Stopping*):** Para evitar a memorização dos dados (*overfitting*), monitorou-se a função de perda no conjunto de validação. O treinamento foi configurado para ser interrompido caso não houvesse melhoria (decaimento da Loss) após um número pré-definido de épocas (paciência), restaurando-

se automaticamente os pesos da melhor época observada. Os valores de paciência específicos para cada modelo constam na Tabela 5.2.

- **Critério de Viabilidade:** A replicação foi considerada bem-sucedida se o modelo atingisse convergência estável (baixo desvio padrão entre as rodadas) e demonstrasse capacidade de generalização na base de teste, validando o comportamento teórico esperado para cada arquitetura (viés indutivo).

Tabela 5.2: Hiperparâmetros e especificidades das arquiteturas comparadas.

Parâmetro	Configuração
<i>Configurações de Treinamento (Comuns)</i>	
Otimizador	Adam
Taxa de Aprendizado	5×10^{-4}
Tamanho do Lote (Batch Size)	32
Critério de Parada (Paciência)	12 épocas (Duraj) / 6 épocas (SEResUTer)
Divisão Treino / Validação	80% / 20%
<i>Diferenças nas Arquiteturas</i>	
Caso 1: Residual U-Net (Duraj)	Kernel = 9 — Sem Atenção
Caso 2: SEResUTer (Híbrido)	Kernel = 5 — Com Transformer
<i>Detalhes da Arquitetura</i>	
Filtros Iniciais	32
Taxa de Dropout	0,2 (Convoluçãoes) / 0,1 (Transformer)
Função de Ativação	ReLU (Ocultas) / Softmax (Saída)
Profundidade (Encoder)	4 Níveis (Duraj) / 3 Níveis (SEResUTer)

5.5 Fundamentação Matemática da Implementação

Antes de apresentar as especificidades de cada arquitetura, é fundamental detalhar as estratégias matemáticas adotadas para garantir o aprendizado dos modelos, especificamente a escolha da função de perda e o tratamento do desbalanceamento de classes.

5.5.1 Função de Perda: Entropia Cruzada Categórica

Para o problema de segmentação semântica multiclasse, onde cada amostra temporal do ECG deve ser classificada como uma de quatro classes exclusivas (Fundo, Onda P, Complexo QRS, Onda T), utilizou-se a *Categorical Cross-Entropy* (CCE). Matematicamente,

essa função mede a divergência entre a distribuição real y (vetor *one-hot*) e a distribuição predita \hat{y} :

$$L_{CCE} = - \sum_{c=1}^C y_{i,c} \cdot \log(\hat{y}_{i,c}) \quad (5.1)$$

Essa formulação penaliza severamente o modelo quando ele atribui uma probabilidade baixa à classe verdadeira, forçando o ajuste dos pesos para aumentar a confiança na segmentação correta.

5.5.2 Estratégia de Ponderação de Classes (Class Weights)

Devido ao severo desbalanceamento da base LUDB, onde a linha isoelétrica (fundo) predomina, a rede tende a convergir para um mínimo local trivial (classificar tudo como fundo). Para corrigir isso, aplicou-se a técnica de Ponderação de Classes Suavizada.

Diferentemente da ponderação inversa padrão, optou-se por aplicar uma suavização pela raiz quadrada nos pesos calculados. Essa estratégia evita que as classes minoritárias recebam pesos excessivamente altos (o que poderia desestabilizar os gradientes), mantendo, contudo, a penalidade para erros nas ondas de interesse:

$$w_c = \sqrt{\frac{N_{total}}{N_{classes} \cdot N_c}} \quad (5.2)$$

A aplicação desta fórmula sobre o conjunto de treinamento resultou nos seguintes pesos, incorporados à função de perda: a classe majoritária ‘Fundo’ recebeu peso de $\approx 0,60$, enquanto as classes clínicas foram priorizadas: Onda P ($\approx 1,67$), Complexo QRS ($\approx 1,30$) e Onda T ($\approx 1,91$). Esses valores garantem que o modelo priorize a detecção das morfologias cardíacas sem ignorar completamente a estabilidade da linha de base.

5.6 Arquiteturas Implementadas

5.6.1 Caso 1: Residual U-Net (Abordagem Convolucional)

A primeira arquitetura replicada segue a proposta de (DURAJ et al., 2022). Ela caracteriza-se pelo uso de blocos residuais e mecanismos de *Squeeze-and-Excitation* (SE) para recalibrar

brar os canais de características. A estrutura, apresentada na Figura 5.4, utiliza apenas operações de convolução, o que a torna computacionalmente eficiente e estável durante o treinamento.

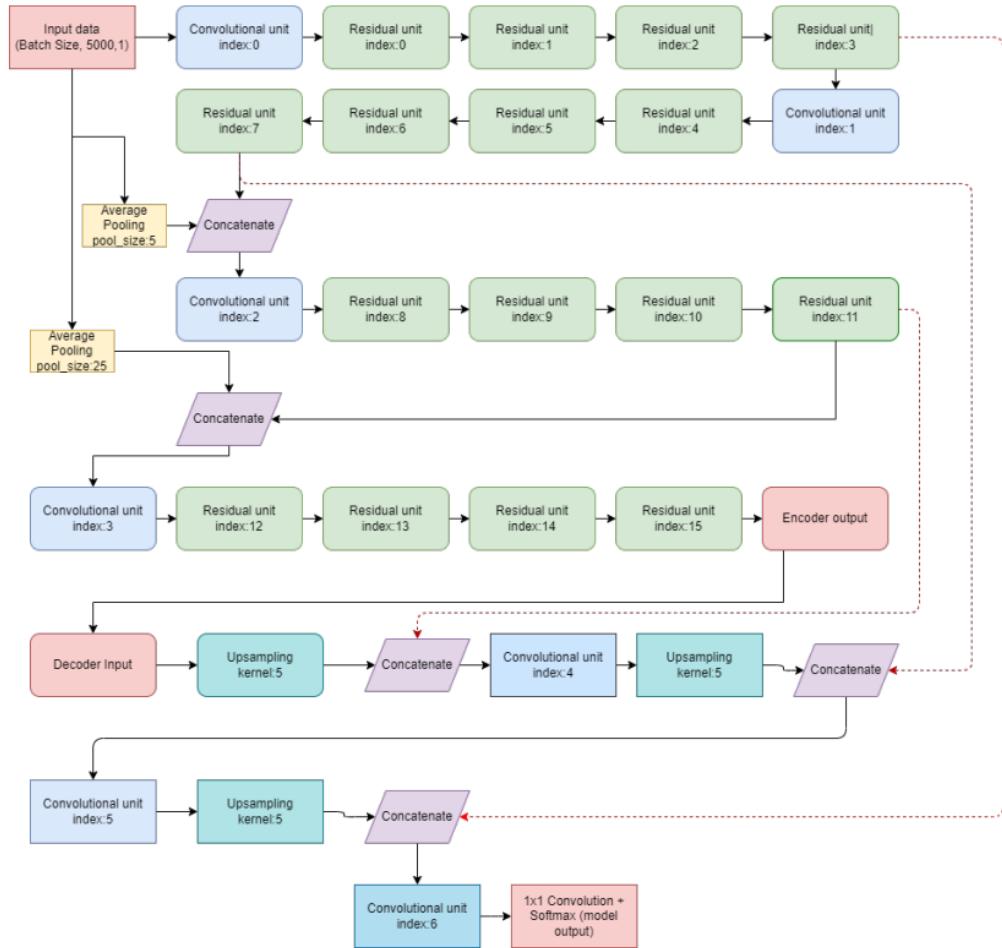


Figura 5.4: Diagrama da arquitetura Residual U-Net com blocos SE utilizada no Estudo de Caso 1.

Fonte: Duraj et al. (2022).

5.6.2 Caso 2: SEResUTer (Abordagem Híbrida)

O segundo modelo, proposto por (LI et al., 2023), introduz um módulo *Transformer* no gargalo da rede (Figura 5.5). O objetivo é capturar o contexto global da série temporal através de mecanismos de autoatenção, superando a limitação do campo receptivo local das convoluções.

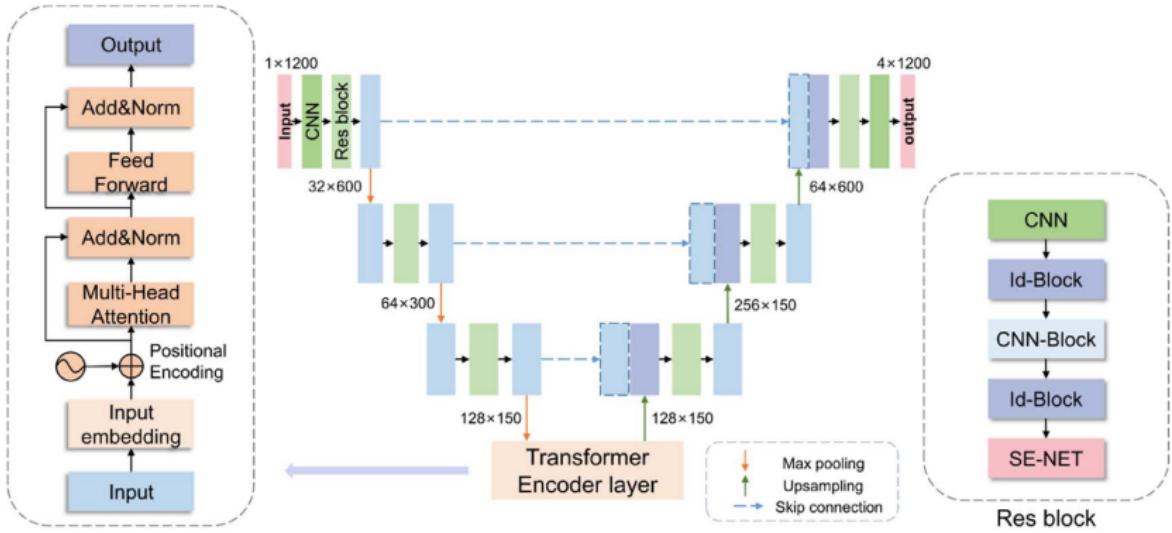


Figura 5.5: Diagrama da arquitetura híbrida SEResUTer, integrando U-Net com Transformer Encoder no gargalo.

Fonte: Li et al. (2023).

5.6.3 Estratégias de Estabilização para o Modelo Híbrido

Durante a implementação do modelo híbrido (Caso 2), foram necessárias adaptações metodológicas para viabilizar o treinamento, não exigidas no modelo convolucional puro.

A primeira medida foi o Ajuste Fino da Taxa de Aprendizado. A arquitetura híbrida revelou instabilidade inicial com as taxas padrão (10^{-3}). Para solucionar isso, reduziu-se a taxa de aprendizado para 5×10^{-4} no otimizador Adam, permitindo uma descida de gradiente mais suave e evitando divergências nas primeiras épocas.

Além da estabilização numérica, dada a alta complexidade paramétrica do Transformer e o tamanho limitado da base LUDB, implementou-se um pipeline de *Data Augmentation* para atuar como regularizador. As transformações aplicadas dinamicamente durante o treino foram:

- **Adição de Ruído Gaussiano:** Simulação de interferências elétricas e artefatos de aquisição de alta frequência;
- **Escalonamento de Amplitude:** Multiplicação do sinal por fatores aleatórios (entre 0.9 e 1.1) para simular variações de ganho do amplificador;
- **Desvio de Linha de Base (Baseline Wander):** Adição de ondas senoidais de baixa frequência para simular a oscilação causada pela respiração do paciente,

forçando a rede a ser robusta a flutuações não-cardíacas.

Os resultados obtidos a partir destas implementações são detalhados no capítulo 6.

6 Resultados Experimentais e Discussão

Este capítulo detalha os resultados quantitativos obtidos na execução dos estudos de caso. Ressalta-se que, para garantir a robustez estatística, todos os valores apresentados a seguir representam a média e o desvio padrão consolidados de 5 rodadas independentes de treinamento e avaliação, conferindo maior confiabilidade à análise de reproduzibilidade.

6.1 Resultados do Caso 1: Residual U-Net

O treinamento da arquitetura convolucional (Caso 1) demonstrou alta estabilidade. A aplicação dos pesos de classe foi decisiva para a detecção correta das ondas.

Uma observação metodológica crucial refere-se à composição da base de dados. O estudo original de (DURAJ et al., 2022) relata a remoção manual de 23 registros ruidosos. Nesta replicação, optou-se por manter a base LUDB íntegra (2400 sinais) para evitar viés de seleção. Mesmo neste cenário mais desafiador, o modelo obteve desempenho robusto e baixa variância entre as execuções, conforme detalhado na Tabela 6.1.

Tabela 6.1: Desempenho Estatístico do Modelo Convolucional (Média de 5 Rodadas).

Métrica	Replicação ($\mu \pm \sigma$)	Original (DURAJ et al., 2022)
Acurácia	$95,44\% \pm 0,27\%$	95,00%
Precisão	$95,45\% \pm 0,26\%$	95,00%
Recall (Sensibilidade)	$95,42\% \pm 0,27\%$	99,00%
F1-Score	$0,9544 \pm 0,0027$	0,96

Fonte: Elaborada pelo autor com dados de Duraj et al. (2022).

Em relação à diferença observada no *Recall* (95,42% na replicação contra 99,00% no estudo original), esta é consequência direta da decisão metodológica de manter os sinais ruidosos na base de teste. Diferentemente da abordagem original, que removeu registros de baixa qualidade, esta replicação optou por preservar a integridade do *dataset* para evitar viés de seleção. Embora essa escolha impacte a métrica absoluta, ao obrigar a rede a lidar com artefatos de aquisição, ela reflete um cenário mais realista de aplicação clínica, onde a pré-seleção manual de sinais ideais nem sempre é viável.

6.2 Resultados do Caso 2: SEResUTer

A replicação do modelo híbrido confirmou os desafios de generalização para arquiteturas baseadas em *Transformers* em bases de médio porte. Para isolar o impacto das estratégias de regularização, o modelo foi avaliado em dois cenários: com e sem *Data Augmentation* (DA).

Tabela 6.2: Impacto do Data Augmentation no Modelo Híbrido (Média de 5 Rodadas).

Métrica	Com Data Augmentation	Sem Data Augmentation
Acurácia Global	83,77% \pm 0,75%	82,84% \pm 0,71%
F1-Score	0,8373	0,8278
Recall (Sensibilidade)	83,23%	82,17%
Precisão	84,22%	83,40%

Observa-se que, embora a aplicação de *Data Augmentation* tenha resultado em ganho de desempenho, este foi marginal (inferior a 1 ponto percentual). Atribui-se esse fenômeno à natureza multicanal da base LUDB. Como o treinamento processou as 12 derivações simultaneamente, o modelo foi exposto a 12 perspectivas espaciais distintas do mesmo evento cardíaco. Essa variabilidade intrínseca atua como uma forma de '*Data Augmentation* natural', mitigando a necessidade de dados sintéticos e tornando a injeção artificial de ruído menos impactante do que seria em bases de derivação única.

Contudo, nota-se que o desempenho global do modelo híbrido (83,77%) permaneceu significativamente inferior ao da abordagem convolucional (95,44%), confirmando a dificuldade dos *Transformers* em extraír padrões eficientes sem volumes massivos de dados.

6.3 Análise Comparativa e Conclusão dos Experimentos

A comparação direta das abordagens (Tabela 6.3) sintetiza os achados desta pesquisa, destacando o *trade-off* entre complexidade e eficiência.

Essa discrepância evidencia dois compromissos arquiteturais importantes:

1. **Eficiência de Dados (Viés Indutivo):** As CNNs beneficiam-se de propriedades

Tabela 6.3: Comparativo final: Residual U-Net vs. SEResUTer (Melhores Configurações).

Dimensão	Residual U-Net (Conv.)	SEResUTer (Híbrida com DA)
Acurácia Média	Alto (95,44%)	Médio (83,77%)
Estabilidade (σ)	Alta ($\pm 0,27\%$)	Média ($\pm 0,75\%$)
Generalização (LUDB)	Excelente	Limitada (Data Hunger)

Fonte: Elaborada pelo autor (2025).

arquiteturais intrínsecas, especificamente a conectividade local e a invariância à translação temporal. Diferentemente dos modelos baseados em Transformers, que precisam inferir todas as relações espaciais a partir do zero (o que exige volumes massivos de dados), as CNNs já possuem essas regras de “como olhar a morfologia do sinal” embutidas em sua estrutura. Esse viés arquitetural funcionou como um catalisador no aprendizado, permitindo que o modelo de Duraj (Caso 1) capturasse padrões robustos de forma eficiente, mesmo diante da escassez de amostras da base LUDB.

2. **Estabilidade de Convergência:** O desvio padrão três vezes maior no modelo híbrido ($\pm 0,75\%$ vs $\pm 0,27\%$) indica que Transformers são mais sensíveis à inicialização dos pesos, tornando o treinamento mais imprevisível em cenários de bases menores.

A replicação demonstrou que, para o cenário específico de segmentação de ECG com a base LUDB, a arquitetura convolucional clássica (Residual U-Net) é a escolha superior em termos de precisão, estabilidade e simplicidade.

7 Conclusão e Trabalhos Futuros

Este trabalho apresentou uma abordagem integrada, combinando uma revisão sistemática da literatura com um estudo experimental de replicação, para investigar o estado da arte e a viabilidade prática do uso de técnicas de *deep learning* na identificação automática das ondas do ECG. A estratégia adotada permitiu mapear tendências na pesquisa recente e avaliar, em condições reais, os desafios de reproduzibilidade de arquiteturas avançadas.

7.1 Síntese dos Resultados e Objetivos

O objetivo principal de investigar o estado da arte e avaliar a reproduzibilidade de técnicas de *deep learning* foi atingido através da execução dos eixos teóricos e práticos da pesquisa.

No que tange aos objetivos específicos de mapeamento teórico (arquiteturas, bases e métricas), a Revisão Sistemática realizada no Capítulo 4 permitiu identificar uma clara evolução tecnológica na última década. Observou-se que a literatura partiu de CNNs simples para modelos profundos (U-Net, ResNets) e, mais recentemente, convergiu para arquiteturas híbridas com mecanismos de Atenção. O estudo também confirmou a predominância de bases públicas padronizadas (QTDB, LUDB) e evidenciou que o *Data Augmentation* consolidou-se como um componente crítico — e não apenas acessório — para mitigar o desbalanceamento de classes em cenários de base de dados menores.

Quanto ao objetivo específico experimental, a análise comparativa entre os modelos Residual U-Net e SEResUTer forneceu evidências práticas sobre a viabilidade dessas técnicas. A replicação da arquitetura convolucional comprovou-se altamente viável, atingindo acurácia superior a 95% com alta estabilidade, confirmando que o viés indutivo das CNNs (conectividade local) é extremamente vantajoso para bases de tamanho moderado. Por outro lado, a replicação do modelo híbrido evidenciou os desafios práticos de arquiteturas baseadas em *Transformers*, que apresentaram maior instabilidade e desempenho inferior (83,77%) mesmo após otimização, corroborando a hipótese de que mecanismos de atenção exigem volumes massivos de dados para justificar sua complexidade.

Dessa forma, conclui-se que a pesquisa cumpriu seu propósito ao demonstrar que, embora arquiteturas complexas sejam a tendência teórica, modelos convolucionais robustos permanecem como a escolha mais pragmática, estável e eficiente para aplicações clínicas com escassez de dados rotulados.

7.2 Limitações do Estudo

Apesar dos avanços alcançados, algumas limitações foram identificadas tanto na literatura quanto na execução experimental:

- **Heterogeneidade dos dados:** Bases públicas apresentam diferentes protocolos de anotação e características populacionais, dificultando a criação de modelos generalizáveis.
- **Reprodutibilidade de arquiteturas complexas:** A replicação de modelos baseados em *Transformers* revelou uma dependência crítica de detalhes de implementação (como inicialização de pesos e estratégias de otimização) que raramente são documentados com precisão nos artigos originais.
- **Desafios com dados parcialmente anotados:** Durante os experimentos preliminares com bases de anotações incompletas, observou-se a falha das funções de perda tradicionais. Como esses mecanismos dependem matematicamente da comparação entre a predição e um rótulo real para calcular a penalidade (erro), a ausência de anotações em trechos do sinal inviabiliza o cálculo correto do gradiente, levando-o a zero ou a valores instáveis e impedindo a convergência da rede.
- **Interpretabilidade restrita:** A natureza “caixa-preta” dos modelos de *Deep Learning* permanece uma barreira para a adoção clínica, onde a justificativa morfológica para uma segmentação é tão importante quanto o resultado final.

7.3 Trabalhos Futuros

Com base nas lacunas identificadas e nas dificuldades observadas durante a replicação, sugerem-se as seguintes direções para pesquisas futuras:

- **Funções de Perda Robustas a Rótulos Parciais:** Desenvolver ou adaptar funções de custo que consigam ignorar (mascarar) regiões não anotadas do sinal durante o cálculo do erro, permitindo o aproveitamento de bases de dados massivas que possuem apenas anotações parciais.
- **Pré-treinamento Auto-supervisionado:** Investigar o uso de grandes volumes de ECGs sem anotação para pré-treinar os pesos das redes (aprendendo a estrutura do sinal), realizando o ajuste fino (*fine-tuning*) apenas com os dados rotulados disponíveis, técnica que tem mostrado sucesso em modelos de linguagem.
- **Benchmarks de Robustez a Ruído:** Estabelecer protocolos padronizados de teste que avaliem não apenas a acurácia em sinais limpos, mas a degradação do desempenho sob diferentes níveis de ruído sintético e real, simulando condições de dispositivos vestíveis (*wearables*).
- **Explainable AI (XAI) para ECG:** Implementar e avaliar visualmente mapas de atenção e relevância que indiquem quais partes do complexo P-QRS-T foram determinantes para a segmentação, aumentando a confiança do especialista médico na ferramenta.

De modo geral, os resultados empíricos sugerem que, embora a tecnologia de segmentação esteja madura, a fronteira da pesquisa deve se deslocar da simples criação de novas arquiteturas para a melhoria da qualidade dos dados, robustez a ruídos e interpretabilidade dos modelos existentes.

Bibliografia

- ACHARYA, U. R. et al. A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, v. 89, p. 389–396, 2017.
- ALAM, R.; AGUIRRE, A.; STULTZ, C. M. Detecting QT prolongation from a single-lead ECG with deep learning. *PLOS Digital Health*, Public Library of Science San Francisco, CA USA, v. 3, n. 6, p. e0000539, 2024.
- BAHDANAU, D.; CHO, K.; BENGIO, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- Brasil. Ministério da Saúde. *Cerca de 400 mil pessoas morreram em 2022 no Brasil por problemas cardiovasculares*. 2022. Biblioteca Virtual em Saúde. Acesso em: 13 dez. 2025. Disponível em: <<https://bvsms.saude.gov.br/cerca-de-400-mil-pessoas-morreram-em-2022-no-brasil-por-problemas-cardiovasculares/>>.
- CHOU, C.-C. et al. Comparing artificial intelligence-enabled electrocardiogram models in identifying left atrium enlargement and long-term cardiovascular risk. *Canadian Journal of Cardiology*, Elsevier, v. 40, n. 4, p. 585–594, 2024.
- CLARIVATE. *Web of Science*. 2025. Acessado em: 10 jun. 2025. Disponível em: <<https://www.webofscience.com>>.
- DUBIN, D. *Interpretação Rápida do ECG*. 6. ed. São Paulo: Editora Publicações Médicas, 2000.
- DURAJ, K. et al. Semantic segmentation of 12-lead ECG using 1d residual U-Net with squeeze-excitation blocks. *Applied sciences*, MDPI, v. 12, n. 7, p. 3332, 2022.
- FAUST, O. et al. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 161, p. 1–13, 2018.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.
- GOOVAERTS, G. et al. A machine-learning approach for detection and quantification of QRS fragmentation. *IEEE journal of biomedical and health informatics*, IEEE, v. 23, n. 5, p. 1980–1989, 2018.
- HALL, J. E.; GUYTON, A. C. *Tratado de Fisiologia Médica*. 13. ed. Rio de Janeiro: Elsevier, 2017.
- HANNUN, A. Y. et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, Nature Publishing Group, v. 25, n. 1, p. 65–69, 2019.
- HAYKIN, S. *Neural networks and learning machines*. 3. ed. Upper Saddle River, NJ: Pearson Education, 2009.

- HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.]: IEEE, 2016. p. 770–778.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997.
- HONG, S. et al. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine*, Elsevier, v. 122, p. 103801, 2020.
- HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.]: IEEE, 2018. p. 7132–7141.
- IEEE. *IEEE Xplore Digital Library*. 2025. Acessado em: 10 jun. 2025. Disponível em: <https://ieeexplore.ieee.org>.
- Instituto de Métrica e Avaliação em Saúde (IHME). *Carga Global de Doenças 2023: Achados do Estudo GBD de 2023*. Seattle, WA, 2025. Acesso em: 13 dez. 2025. Disponível em: <http://www.healthdata.org>.
- KALYAKULINA, A. I. et al. LUDB: A new open-access validation tool for electrocardiogram delineation algorithms. *IEEE Access*, IEEE, v. 8, p. 186181–186190, 2020.
- KIRANYAZ, S.; INCE, T.; GABBOUJ, M. Real-time patient-specific ECG classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, v. 63, n. 3, p. 664–675, 2016.
- LAGUNA, P. et al. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In: IEEE. *Computers in cardiology 1997*. [S.l.], 1997. p. 673–676.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LI, X. et al. SEResUTer: a deep learning approach for accurate ECG signal delineation and atrial fibrillation detection. *Physiological Measurement*, IOP Publishing, v. 44, n. 12, p. 125005, 2023.
- Manuais MSD. *ECG: Como interpretar as ondas*. 2024. Acesso em: 24 jan. 2026. Disponível em: <https://www.msmanuals.com/pt/casa/multimedia/image/ecg-como-interpretar-as-ondas>.
- MARTÍNEZ, J. P. et al. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on Biomedical Engineering*, IEEE, v. 51, n. 4, p. 570–581, 2004.
- MEHRI, M. et al. A deep learning architecture using 3d vectorcardiogram to detect r-peaks in ecg with enhanced precision. *Sensors*, MDPI, v. 23, n. 4, p. 2288, 2023.
- MOODY, G. B.; MARK, R. G. The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, IEEE, v. 20, n. 3, p. 45–50, 2001.
- National Center for Biotechnology Information. *PubMed*. 2025. Acessado em: 10 jun. 2025. Disponível em: <https://pubmed.ncbi.nlm.nih.gov>.

- NURMAINI, S. et al. Beat-to-beat electrocardiogram waveform classification based on a stacked convolutional and bidirectional long short-term memory. *IEEE Access*, IEEE, v. 9, p. 92600–92613, 2021.
- OUZZANI, M. et al. Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, BioMed Central, v. 5, n. 1, p. 210, 2016.
- PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, British Medical Journal Publishing Group, v. 372, p. n71, 2021.
- PAN, J.; TOMPKINS, W. J. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, IEEE, n. 3, p. 230–236, 1985.
- PEIMANKAR, A.; PUTHUSERYPADY, S. DENS-ECG: A deep learning approach for ECG signal delineation. *Expert systems with applications*, Elsevier, v. 165, p. 113911, 2021.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. [S.l.: s.n.], 2015. p. 234–241.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, American Psychological Association, v. 65, n. 6, p. 386–408, 1958.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986.
- SHARMA, N.; JOSHI, D. Ecg_deepnet: A novel approach for delineation and classification of electrocardiogram signal based on ensemble deep-learning. *IEEE Transactions on Artificial Intelligence*, IEEE, 2025.
- SILVA, M. d. P. D. et al. Semantic segmentation of QRS complex in 12-lead ECG signals. In: *IEEE. 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.], 2024. p. 1–5.
- SODMANN, P. et al. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms. *Physiological measurement*, IOP Publishing, v. 39, n. 10, p. 104005, 2018.
- TIHAK, A.; KONJICIJA, S.; BOSKOVIC, D. Deep learning models for atrial fibrillation detection: A review. In: *IEEE. 2022 30th Telecommunications Forum (TELFOR)*. [S.l.], 2022. p. 1–4.
- VASWANI, A. et al. Attention is all you need. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2017. v. 30.
- VIJAYA, G.; KUMAR, V.; VERMA, H. K. ANN-based QRS-complex analysis of ECG. *Journal of Medical Engineering & Technology*, Taylor & Francis, v. 22, n. 4, p. 160–167, 1998.
- WASIMUDDIN, M. et al. Stages-based ECG signal analysis from traditional signal processing to machine learning approaches: A survey. *IEEE Access*, IEEE, v. 8, p. 177782–177803, 2020.

- XU, X. et al. A 12-lead ECG delineation algorithm based on a quantized CNN-BiLSTM auto-encoder with 1-12 mapping. In: IEEE. *2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. [S.l.], 2023. p. 1–5.
- YILDIRIM, O. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology and Medicine*, v. 96, p. 189–202, 2018.
- ZHANG, Z.; LIU, Q.; WANG, Y. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, IEEE, v. 15, n. 5, p. 749–753, 2018.