



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E
COMPUTAÇÃO

ISAÍAS RAMOS MOREIRA

**Filtragem de Ruído em Sinais de
Eletrocardiograma e Áudio utilizando
Autoencoders e Mecanismos de
Monitoramento**

Goiânia
2025



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC no 1240/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei no 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG)

Nome(s) completo(s) do(a)(s) autor(a)(es)(as): **Isaías Ramos Moreira**

Título do trabalho: **Filtragem de Ruído em Sinais de Eletrocardiograma e Áudio utilizando Autoencoders e Mecanismos de Monitoramento**

2. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador) Concorda com a liberação total do documento [x] SIM [] NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: a) consulta ao(à)(s) autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro.

Obs.: Este termo deve ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Alisson Assis Cardoso, Professor do Magistério Superior**, em 30/06/2025, às 08:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Isaías Ramos Moreira, Discente**, em 30/06/2025, às 10:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5468397** e o código CRC **45098E12**.

Referência: Processo nº 23070.013367/2025-21

SEI nº 5468397

ISAÍAS RAMOS MOREIRA

Filtragem de Ruído em Sinais de Eletrocardiograma e Áudio utilizando Autoencoders e Mecanismos de Monitoramento

Trabalho de Conclusão apresentado à Coordenação do Curso de Engenharia Elétrica do Escola de Engenharia Elétrica, Mecânica e Computação da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Bacharel em Engenharia Elétrica.

Orientador: Prof. Dr. Álisson Assis Cardoso

Goiânia
2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Moreira, Isaías Ramos

Filtragem de Ruído em Sinais de Eletrocardiograma e Áudio utilizando Autoencoders e Mecanismos de Monitoramento [manuscrito] / Isaías Ramos Moreira. - 2025.

xvi, 16 f.

Orientador: Prof. Dr. Álisson Assis Cardoso.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de Computação (EMC), Engenharia Elétrica, Goiânia, 2025.

1. Remoção de Ruído em ECG. 2. Reconstrução de Sinais. 3. Reconstrução de Espectrogramas. 4. Redes Neurais Profundas. 5. Autoencoders. I. Cardoso, Álisson Assis, orient. II. Título.

CDU 621.3



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Ao(s) **25** dia(s) do mês de **junho** do ano de **2025** iniciou-se a sessão pública de defesa do Trabalho de Conclusão de Curso (TCC) intitulado “**Filtragem de Ruído em Sinais de Eletrocardiograma e Áudio utilizando Autoencoders e Mecanismos de Monitoramento**”, de autoria de **Isaías Ramos Moreira**, do curso de Engenharia Elétrica, do(a) Escola de Engenharia Elétrica, Mecânica e Computação (EMC) da UFG. Os trabalhos foram instalados pelo(a) Prof. Dr. Alisson Assis Cardoso com a participação dos demais membros da Banca Examinadora: Prof. Dr. Rodrigo Pinto Lemos (EMC/UFG) e Dr. Henrique Pires Corrêa (EMC-UFG). Após a apresentação, a banca examinadora realizou a arguição do(a) estudante. Posteriormente, de forma reservada, a Banca Examinadora atribuiu a nota final de **10,0**, tendo sido o TCC considerado **aprovado**.

Proclamados os resultados, os trabalhos foram encerrados e, para constar, lavrou-se a presente ata que segue assinada pelos Membros da Banca Examinadora.



Documento assinado eletronicamente por **Alisson Assis Cardoso, Professor do Magistério Superior**, em 25/06/2025, às 21:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodrigo Pinto Lemos, Professor do Magistério Superior**, em 25/06/2025, às 21:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Henrique Pires Correa, Técnico**, em 25/06/2025, às 21:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5462579** e o código CRC **5FAEF736**.

Filtragem de Ruído em Sinais de Eletrocardiograma e Áudio utilizando Autoencoders e Mecanismos de Monitoramento

Isaías Ramos Moreira¹, Álisson Assis Cardoso²

Escola de Engenharia Elétrica, Mecânica e de Computação, Universidade Federal de Goiás
Goiânia, Goiás, Brasil.

E-mails: ¹moreira.isaias@discente.ufg.br, ²alsnac@ufg.br

Resumo— Este estudo propõe uma abordagem baseada em *autoencoders* com blocos residuais e mecanismos de monitoramento para a filtragem de sinais de eletrocardiograma contaminados por ruídos tanto de forma isolada quanto combinada, bem como de espectrogramas contaminados com ruídos comuns do ambiente urbano, criados a partir de arquivos de áudio. Foram avaliadas duas arquiteturas distintas: uma utilizando um bloco *Squeeze-and-Excitation* e outra incorporando um mecanismo de monitoramento denominado *Attention-Block*.

Os modelos foram submetidos a testes em dados de sinais de eletrocardiograma e espectrogramas contaminados por ruídos a fim de avaliar seu desempenho na tarefa de reconstrução dos sinais e remoção dos ruídos. O *Attention-Block* demonstrou desempenho superior nas duas tarefas. Concluiu-se que essa vantagem se deve à sua capacidade de integrar mecanismos de atenção por canal e detecção local de ruído, o que se mostrou pertinente quando existem múltiplos tipos de interferência com características particulares atuando simultaneamente.

Palavras-Chave— Remoção de Ruído em ECG, Reconstrução de Sinais, Reconstrução de Espectrogramas, Redes Neurais Profundas, Autoencoders, Mecanismos de Monitoramento.

Abstract— This study proposes an approach based on *autoencoders* with residual blocks and monitoring mechanisms for filtering electrocardiogram signals contaminated by noise, both in isolation and in combination, as well as spectrograms corrupted by common urban environmental noise generated from audio files. Two distinct architectures were evaluated: one using a *Squeeze-and-Excitation* block and another incorporating a monitoring mechanism called the *Attention Block*.

The models were tested on electrocardiogram signals and spectrograms contaminated with noise to assess their performance in signal reconstruction and noise removal tasks. The *Attention Block* demonstrated superior performance in both tasks. This advantage is attributed to its ability to integrate channel-wise attention mechanisms and local noise detection, which proved relevant when multiple types of interference with distinct characteristics occur simultaneously.

Keywords— Noise Removal in ECG, Signal Reconstruction, Spectrogram Reconstruction, Deep Neural Networks, Autoencoders, Monitoring Mechanisms.

I. INTRODUÇÃO

Conforme o avanço das técnicas de inteligência computacional, a modernização dos sistemas de processamento de sinais, especialmente em aplicações que demandam alta precisão, como nas áreas biomédica e de telecomunicações, se impulsionou significativamente [1], [2]. Nesse contexto, a

natureza não estacionária e ampla variabilidade temporal e espectral de certos ruídos representa um desafio substancial para técnicas tradicionais de filtragem que operam com base em pressupostos fixos de frequência e linearidade, como o filtro passa-altas de Butterworth, o filtro notch e o filtro de média móvel.

Dessa maneira, modelos baseados em Redes Neurais Profundas (*Deep Neural Networks* – DNNs) aliados a técnicas já consagradas na remoção de ruídos têm sido propostos. Em [3] apresenta-se uma metodologia a partir de Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs) integradas à Transformada Wavelet Discreta (*Discrete Wavelet Transform* – DWT) em que os autores utilizam camadas convolucionais para maior extração de características locais do sinal de eletrocardiograma (ECG) e a DWT é empregada para a redução da dimensionalidade, separando o sinal em componentes de diferentes bandas de frequência, permitindo a remoção do ruído e conservação do sinal útil mediante uma reconstrução a partir dos coeficientes de *wavelet* de cada camada. Essa abordagem garante que propriedades morfológicas do sinal sejam preservadas devido à sua capacidade de extração de características discriminantes e à sua flexibilidade para modelar relações não lineares.

A. Sinais de Eletrocardiograma

A Organização Mundial da Saúde (OMS) concluiu que as doenças cardiovasculares representam a maior causa de óbitos pelo mundo, ao passo que a razão mais comum de infartos fulminantes está relacionada a arritmias não diagnosticadas [4]. Nesse contexto, o ECG fornece informações essenciais com relação à atividade elétrica do coração, permitindo a identificação de distúrbios no ritmo cardíaco, bloqueios de condução, isquemias e outras anormalidades funcionais e estruturais, possibilitando inferir o estado de saúde do paciente com alto grau de sensibilidade [5]. Esse exame da atividade cardíaca configura-se como uma análise de baixo custo operacional, não invasiva e de rápida aplicação [6]. Apesar de sua execução frequentemente realizada em ambientes clínicos controlados, os sinais de ECG ainda estão sujeitos a diversas formas de interferência que podem comprometer sua qualidade e interpretação.

Nesse sentido, é possível classificar os principais tipos de interferência em três categorias: *baseline wander* (BW), *muscle artifact* (MA) e *electrode motion* (EM). O BW é caracterizado por componentes de baixa frequência, geralmente associados a movimentos respiratórios do paciente [7], que produzem variações lentas na linha de base do ECG e dificultam a correta identificação das ondas cardíacas por conta de flutuações e mitigação de amplitudes relevantes. Já o MA decorre da atividade elétrica dos músculos esqueléticos, apresentando conteúdo espectral concentrado em faixas de alta frequência [8]. O ruído do tipo EM é causado por deslocamentos dos eletrodos sobre a pele, o que altera a impedância na interface eletrodo-pele, resultando em distorções transitórias e abruptas no sinal [9].

B. Sinais de Áudio

Sinais de áudio desempenham um papel central em diversas aplicações modernas, abrangendo desde sistemas de comunicação e entretenimento até tecnologias de acessibilidade e automação, como reconhecimento automático de fala, identificação de eventos acústicos, compressão de dados e monitoramento de ambientes [10]. A integridade desses sinais pode ser comprometida por ruídos introduzidos durante a aquisição, transmissão ou armazenamento. Causas comuns de interferência incluem sobreposição de múltiplas fontes sonoras, reverberações, distorções de microfones e perdas de dados em processos de compressão. Além disso, sons provenientes de ambientes cotidianos, como conversas em multidões, tráfego de veículos, ruídos de ventilação, constituem obstáculos recorrentes, especialmente em aplicações que dependem de sinais de áudio capturados em espaços urbanos não controlados [11].

Dada a natureza não estacionária dos sinais de áudio, uma modalidade prudente para sua análise envolve a conversão para o domínio tempo-frequência, o que permite a identificação de características espectrais que variam ao longo do tempo. Entre as técnicas mais utilizadas para essa finalidade destaca-se a Transformada de Fourier de Tempo Curto (*Short-Time Fourier Transform* - STFT) [12]–[14], que possibilita a geração de espectrogramas — representações bidimensionais nas quais o eixo horizontal corresponde ao tempo, o eixo vertical às faixas de frequência, e a intensidade de cor representa a magnitude da energia do sinal em cada ponto [15].

C. Trabalhos Relacionados

Vários estudos propõem novas arquiteturas de autoencoders para a tarefa de remoção de ruído. Em [16], os autores implementaram um autoencoder removedor de ruídos composto totalmente por CNNs, visando a filtragem de sinais de ECG contaminados simultaneamente por três ruídos reais: *baseline wander*, *muscle artifact* e *electrode motion*. Essa arquitetura foi comparada a outras, incluindo uma DNN *fully-connected* e um *Denoising Autoencoder* (DAE) baseado em redes neurais convolucionais. Concluiu-se que, em comparação aos demais, o modelo totalmente composto por CNNs apresentou maior fidelidade e melhores métricas em comparação aos demais.

No campo de *Speech Enhancement*, os autores de [17] contornaram problemas como: demasiada adaptação do modelo

para um certo tipo de ruído e desempenho discrepante quando submetido a ruídos não vistos durante o treinamento, mediante a aplicação de um modelo de autoencoder com um codificador baseado em uma árvore de decisão que formula um mapeamento especializado de entrada ruidosa para fala limpa, e seus resultados são posteriormente integrados no decodificador. O processo de inferência envolve a passagem do sinal ruidoso por todos os ramos do codificador, permitindo que a reconstrução final utilize diferentes representações paralelas.

Ao invés de fazer com que o modelo aprendesse características do sinal limpo e filtrasse-o, os autores em [18] sugeriram um procedimento em que a arquitetura do autoencoder foi construída com a intenção de que o modelo observasse o ruído e seus atributos de modo que seus parâmetros fossem ajustados pela minimização da função de perda que comparava a saída do autoencoder com o ruído original. Dessa maneira, após a reconstrução do ruído pelo modelo, ocorria a subtração entre o sinal ruidoso e a saída do autoencoder. Os resultados demonstraram uma boa performance e levantam a possibilidade de que, dadas as características estocásticas do ruído e do sinal limpo, treinar o modelo para aprender o ruído pode ser uma melhor escolha do que o procedimento mais usual de reconstrução do sinal original.

Em todos os trabalhos relatados, demonstra-se que os autoencoders possuem desempenho consistente em tarefas de remoção de ruído. Dada sua flexibilidade estrutural, esses modelos ainda oferecem um vasto espaço a ser explorado em soluções mais especializadas, adaptativas e orientadas a diferentes domínios do processamento de sinais.

D. Objetivo Geral do Estudo

O presente trabalho integra a arquitetura de autoencoders com blocos residuais, bloco *Squeeze-and-Excitation* (*SE-Block*) e mecanismos de monitoramento. São separados dois modelos de autoencoder com blocos residuais: um contendo o mecanismo de monitoramento *Attention-Block* proposto e outro utilizando o *SE-Block*. A proposta busca explorar a capacidade desses mecanismos em realçar características relevantes dos sinais e suprimir componentes ruidosos em dois domínios de aplicação: sinais biomédicos de eletrocardiograma e sinais acústicos representados por espectrogramas. Os conjuntos de dados são contaminados com ruídos sintéticos, no caso dos sinais de ECG, e ruídos reais para os espectrogramas, representando os artefatos mais comuns para cada situação.

O objetivo central deste trabalho é expor e avaliar o desempenho dos dois modelos, medindo sua capacidade de recuperação do sinal original frente a diferentes tipos de degradação. Espera-se, com isso, contribuir para o desenvolvimento de soluções em pré-processamento de sinais, com potencial aplicação em dispositivos embarcados de monitoramento contínuo, sistemas de assistência médica e tecnologias inteligentes de reconhecimento de áudio.

II. FUNDAMENTAÇÃO TEÓRICA

A. Características do Sinal de ECG

O sinal de ECG da Figura 1 é a representação temporal da atividade elétrica do coração, registrada por meio de eletrodos

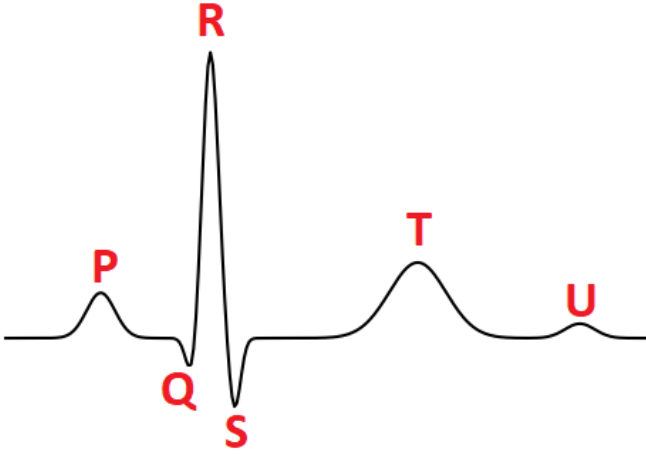


Fig. 1: Representação e identificação dos principais componentes do sinal de ECG.

dispostos na superfície do corpo, compilando a sequência de despolarização e repolarização do músculo cardíaco [5]. O traçado do ECG é composto por uma série de componentes morfológicos de interesse, entre os quais se destacam a onda P, o complexo QRS, as ondas T e U. Cada um desses elementos reflete etapas distintas do ciclo cardíaco e, portanto, carrega informações essenciais para o diagnóstico de condições fisiopatológicas.

Inicialmente, a onda P é a primeira deflexão do traçado, representando a despolarização dos átrios, ou seja, o processo de ativação elétrica que precede a contração atrial [19]. Por essa razão, habitualmente possui baixa amplitude e curta duração, sendo que sua morfologia pode fornecer indícios sobre sobrecargas atriais e distúrbios de condução.

Logo em seguida, ocorre o complexo QRS [20], que representa a despolarização ventricular. Esse complexo é caracterizado por três deflexões distintas: a onda Q, negativa inicial que representa a ativação do septo interventricular da esquerda para a direita; a onda ou pico R, uma deflexão positiva pronunciada que reflete a propagação do impulso elétrico através das paredes ventriculares principais; e a onda S [6], uma deflexão negativa consecutiva, correspondente à ativação das regiões basais dos ventrículos.

Após o complexo QRS, a onda T marca a fase de repolarização ventricular, processo pelo qual as células miocárdicas retornam ao seu potencial de repouso. A análise de distúrbios eletrolíticos e isquêmicos passa pela sua avaliação [19]. Em algumas circunstâncias, é possível observar uma pequena deflexão adicional após a onda T, conhecida como onda U. Apesar de seu mecanismo ainda não ser totalmente compreendido, as hipóteses mais aceitas denotam que a onda U está associada à repolarização das regiões mais tardias da parede ventricular [21]. Sua presença ou ausência, bem como alterações em sua morfologia, podem indicar anormalidades como hipocalemia ou efeitos colaterais de fármacos.

Nestes sinais, as arritmias cardíacas [22] se manifestam causando irregularidades como ausência da onda P, complexos QRS alargados, ondas T invertidas ou presença de ondas

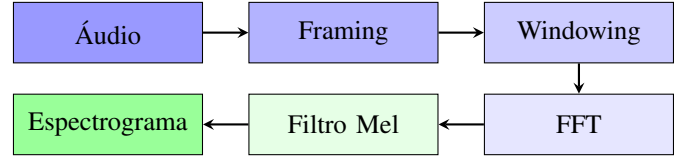


Fig. 2: Esquema em blocos da transformação de um sinal de áudio qualquer em um espectrograma Mel.

fibrilatórias. Por exemplo, na fibrilação atrial [23], observa-se uma perda da regularidade e da morfologia típica da onda P, enquanto na taquicardia ventricular, o complexo QRS tende a apresentar-se largo e com morfologia aberrante, refletindo a condução anormal nos ventrículos [24].

B. Representação de Sinais de Áudio em Espectrogramas Mel

Sinais de áudio possuem características não estacionárias e uma ampla faixa de frequências, comumente entre 20 Hz e 20 kHz [25]. Sabendo-se disso, os Espectrogramas Mel são utilizados para representar a energia espectral dos sinais de áudio em uma escala de frequência não linear inspirada na percepção auditiva humana [26]. Dessa maneira, sua representação facilita a extração de padrões e o processo de reconstrução mediante CNNs que operam sobre essas representações espectrais para realizar tarefas como classificação, detecção de eventos sonoros e *denoising*.

Nessa orientação, a transformação esquematizada na Figura 2 inicia-se com as amostras $x(n)$ do sinal de áudio contínuo $x(t)$, que são divididas em pequenos quadros (*frames*) de curta duração, geralmente de 20 a 40 ms, assumindo-se, por aproximação, que o sinal é estacionário nesse intervalo. Esses quadros são então multiplicados por uma função de janela, usualmente do tipo *Hamming*, a fim de reduzir os efeitos de descontinuidade nas bordas. O sinal segmentado é suavizado e, então, transformado para o domínio da frequência por meio da Transformada Rápida de Fourier (*Fast Fourier Transform* - FFT), produzindo um espectro $X[k]$:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot w[n] \cdot e^{-j \frac{2\pi k n}{N}}, \quad (1)$$

onde $w[n]$ representa os pontos discretos da função de janelamento, n o número de pontos da FFT e k representa o índice da frequência discreta no domínio da FFT, com $k \in \{0, 1, \dots, N-1\}$.

O espectro obtido é então filtrado utilizando um banco de filtros triangulares distribuídos na escala Mel. A conversão da frequência em Hertz para a escala Mel é dada por:

$$m = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right), \quad (2)$$

onde f é a frequência em Hertz associada às posições k do espectro e m é sua correspondente na escala Mel. Essa etapa prioriza as faixas de frequência mais relevantes para a audição humana, enquanto comprime as frequências mais altas e expande as mais baixas. Após a aplicação dos filtros Mel, obtém-se um espectrograma de potência com resolução perceptualmente adaptada. Em seguida, é comum aplicar-se uma

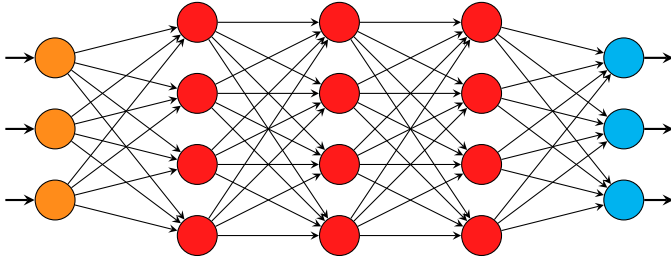


Fig. 3: Representação de uma rede neural profunda totalmente conectada com 3 neurônios de entrada, 3 camadas ocultas com 4 neurônios cada e 3 neurônios de saída.

escala logarítmica aos coeficientes resultantes, expressando-os em decibéis (dB). O resultado final é uma matriz bidimensional que facilita o treinamento de modelos de CNNs.

C. Redes Neurais Profundas

A ideia das redes neurais artificiais surgiu como uma tentativa de replicar, de forma simplificada, o comportamento dos neurônios biológicos. Um neurônio no cérebro humano é capaz de receber sinais de múltiplas fontes, processar esses sinais e, eventualmente, emitir uma resposta se certos critérios forem satisfeitos. Essa lógica inspirou o modelo de um neurônio artificial, originalmente proposto por McCulloch e Pitts na década de 1940, e refinado posteriormente por Rosenblatt com o modelo de *Perceptron* [27], [28].

Na forma mais simples, um neurônio artificial realiza uma combinação linear dos seus sinais de entrada e aplica uma função de ativação ao resultado. Isso pode ser representado pela equação:

$$y = \phi \left(\sum_{i=1}^n w_i x_i + b \right) \quad (3)$$

onde:

- x_i são os sinais de entrada,
- w_i são os pesos associados a cada entrada,
- b é o viés,
- $\phi(\cdot)$ é a função de ativação, responsável por introduzir não linearidades no sistema,
- y é a saída do neurônio.

Em uma rede neural artificial, as chamadas camadas ocultas são aquelas situadas entre a camada de entrada e a camada de saída, sendo compostas por um conjunto de neurônios que não interagem diretamente com os dados de entrada nem com as saídas finais da rede [29]. Por forma que, ao se empilhar múltiplas camadas ocultas, conforme estruturado na Figura 3, a rede se torna capaz de decompor problemas complexos em uma sequência de transformações particulares, cada uma realizada por uma camada distinta. Essa capacidade de extração progressiva de características é o que diferencia as redes profundas das redes rasas.

Do ponto de vista matemático, uma DNN com L camadas ocultas, processando uma entrada vetorial $\mathbf{X} \in \mathbf{R}^n$, realiza uma sequência de transformações hierárquicas, conforme des-

crita a seguir:

$$\begin{aligned} \mathbf{h}^{(0)} &= \mathbf{X} \\ \mathbf{h}^{(1)} &= \sigma \left(\mathbf{W}^{(1)} \mathbf{h}^{(0)} + \mathbf{b}^{(1)} \right) \\ \mathbf{h}^{(2)} &= \sigma \left(\mathbf{W}^{(2)} \mathbf{h}^{(1)} + \mathbf{b}^{(2)} \right) \\ &\vdots \\ \hat{\mathbf{X}} &= \sigma \left(\mathbf{W}^{(L)} \mathbf{h}^{(L-1)} + \mathbf{b}^{(L)} \right) \end{aligned}$$

onde:

- $\mathbf{h}^{(i)}$ representa a ativação (ou *embedding*) da i -ésima camada da rede, para $i = 1, 2, \dots, L$,
- $\mathbf{W}^{(i)}$ é a matriz de pesos da camada i ,
- $\mathbf{b}^{(i)}$ é o vetor de vieses da camada i ,
- $\sigma(\cdot)$ é a função de ativação não linear, aplicada elemento a elemento após a combinação linear,
- $\hat{\mathbf{X}}$ é a saída final da rede, resultante das transformações sucessivas aplicadas à entrada original.

Nota-se que, com o aumento da profundidade da rede neural, torna-se possível reutilizar e combinar representações extraídas por camadas anteriores. Dessa forma, redes mais profundas tendem a apresentar maior capacidade de generalização. No entanto, para seu treinamento, uma DNN requer mais dados, maior capacidade computacional e técnicas de regularização já que são mais suscetíveis a problemas como o *overfitting* [30], situação em que o modelo passa a memorizar os dados de treinamento, perdendo sua capacidade de generalizar para dados não vistos.

O treinamento de uma DNN consiste no ajuste iterativo de seus parâmetros internos, representados pelas matrizes de pesos e vetores de vieses de cada camada, com o objetivo de minimizar uma função de perda. Essa função mede o erro entre as previsões da rede e os valores esperados, orientando o processo de aprendizagem por meio de algoritmos de otimização, como o gradiente descendente estocástico (*Stochastic Gradient Descent - SGD*) [31] e o *Adaptive Moment Estimation* (Adam) [32]. Formalmente, o objetivo do treinamento pode ser descrito pela minimização da função de perda \mathcal{L} , definida sobre o conjunto de parâmetros $\theta = \{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}\}_{i=1}^L$, conforme a seguinte equação:

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\hat{\mathbf{X}}, \mathbf{X}) \quad (4)$$

onde:

- \mathbf{X} representa os valores verdadeiros (alvos) do conjunto de dados,
- $\hat{\mathbf{X}}$ são as saídas previstas pela rede após o processamento da entrada \mathbf{X} ,
- θ^* representa o conjunto de parâmetros otimizados após o processo de treinamento.

Nesse processo, o algoritmo de otimização, como por exemplo o SGD, depende diretamente da derivada da função de ativação para calcular o gradiente da função de perda em relação aos pesos e vieses de cada camada, ou seja, tal função tem de ser diferenciável para que o erro possa ser propagado de volta pelas camadas da rede.

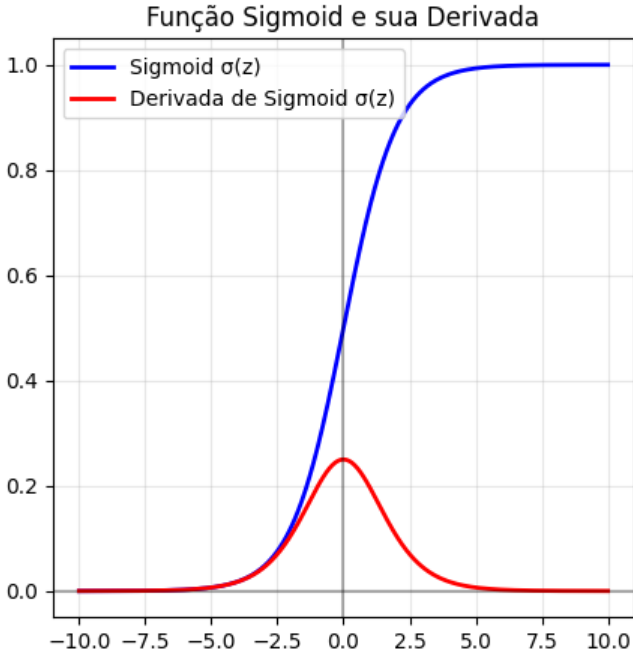


Fig. 4: Representação gráfica da função *sigmoid* e sua derivada.

Considerando uma camada i , a atualização da matriz de pesos $\mathbf{W}^{(i)}$ depende da aplicação da regra da cadeia, sendo o gradiente da função de perda \mathcal{L} em relação a esse peso dado por:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(i)}} = \frac{\partial \mathcal{L}}{\partial \mathbf{h}^{(i)}} \cdot \frac{\partial \mathbf{h}^{(i)}}{\partial z^{(i)}} \cdot \frac{\partial z^{(i)}}{\partial \mathbf{W}^{(i)}} \quad (5)$$

onde:

- $z^{(i)} = \mathbf{W}^{(i)}\mathbf{h}^{(i-1)} + \mathbf{b}^{(i)}$ é a saída linear da camada antes da ativação,
- $\mathbf{h}^{(i)} = \sigma(z^{(i)})$ é a saída não linear após a aplicação da função de ativação $\sigma(\cdot)$,
- $\frac{\partial \mathbf{h}^{(i)}}{\partial z^{(i)}}$ é a derivada da função de ativação.

É evidente que, caso essa derivada se aproxime de zero, o treinamento será comprometido. Esse fenômeno é denominado *vanishing gradient* [33], desaparecimento de gradiente, e ocorre com funções que saturam, a exemplo da *sigmoid* que apresenta regiões onde suas derivadas se tornam muito pequenas, conforme representado na Figura 4.

Com a intenção de mitigar essa problemática, a função de ativação ReLU (*Rectified Linear Unit*) [29] introduz não linearidade no modelo ao aplicar uma função que zera todos os valores negativos e mantém os positivos inalterados, permitindo que os neurônios se tornem seletivamente ativos. Sua derivada constante e igual a 1, denotada na Figura 5, favorece a propagação dos gradientes. Sua forma matemática é definida por:

$$\delta(z) = \text{ReLU}(z) = \max(0, z) \quad (6)$$

Após o cálculo dos gradientes, os pesos da camada i são atualizados de acordo com a seguinte equação:

$$\mathbf{W}^{(i)} \leftarrow \mathbf{W}^{(i)} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(i)}} \quad (7)$$

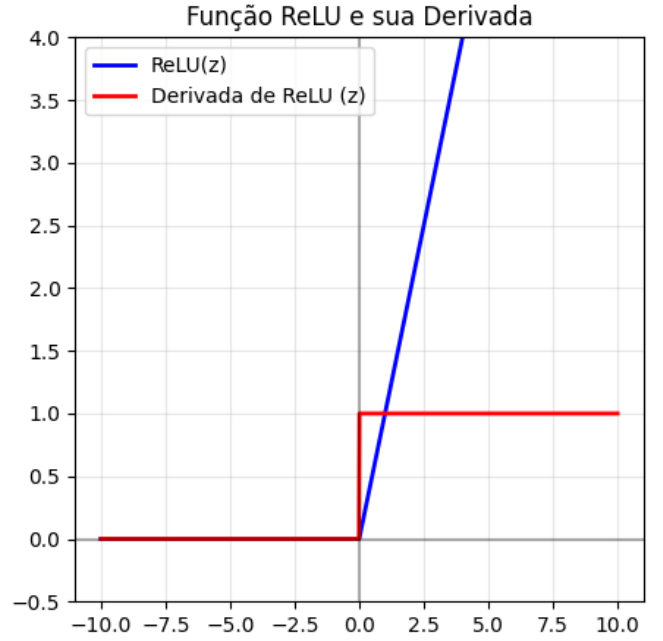


Fig. 5: Representação gráfica da função ReLU e sua derivada.

onde η é a taxa de aprendizado (*learning rate*).

Essa atualização ocorre iterativamente ao longo das épocas de treinamento, ajustando os pesos na direção que minimiza o erro de predição.

D. Autoencoders, Blocos Residuais e Mecanismos de Monitoramento

Comumente, um autoencoder qualquer, conforme o representado na Figura 6, é dividido em três partes: codificador (*encoder*), gargalo (*bottleneck*) e decodificador (*decoder*). Seja $\tilde{\mathbf{x}} \in \mathbf{R}^n$ um vetor de entrada ruidoso. O *encoder* é responsável por transformar $\tilde{\mathbf{x}}$ em uma representação de menor dimensão, $\mathbf{z} \in \mathbf{R}^m$ com $m < n$, conforme:

$$\mathbf{z} = f_{\theta_e}(\tilde{\mathbf{x}}) = \sigma(\mathbf{W}_e \tilde{\mathbf{x}} + \mathbf{b}_e), \quad (8)$$

onde \mathbf{W}_e e \mathbf{b}_e são os pesos e vieses do codificador.

A camada do *bottleneck* trata-se da camada com menor dimensionalidade da rede, onde se espera que a rede retenha as informações mais relevantes da entrada. O *decoder*, por sua vez, tenta reconstruir a entrada original a partir da codificação \mathbf{z} , gerando uma saída $\hat{\mathbf{x}} \in \mathbf{R}^n$:

$$\hat{\mathbf{x}} = g_{\theta_d}(\mathbf{z}) = \sigma(\mathbf{W}_d \mathbf{z} + \mathbf{b}_d), \quad (9)$$

onde \mathbf{W}_d e \mathbf{b}_d são os parâmetros do decodificador.

O treinamento do autoencoder é feito de forma que a saída $\hat{\mathbf{x}}$ se assemelhe o máximo possível ao valor verdadeiro original \mathbf{x} . Sendo assim, busca-se o ajuste dos parâmetros $\theta = \{\mathbf{W}_e, \mathbf{b}_e, \mathbf{W}_d, \mathbf{b}_d\}$ da rede que mais se aproxima desse resultado.

Justamente por essa característica de compreensão dos dados de entrada, autoencoders costumam ser usados em tarefas de remoção de ruído, já que essa arquitetura aprende uma representação simplificada dos dados e busca reconstruí-los

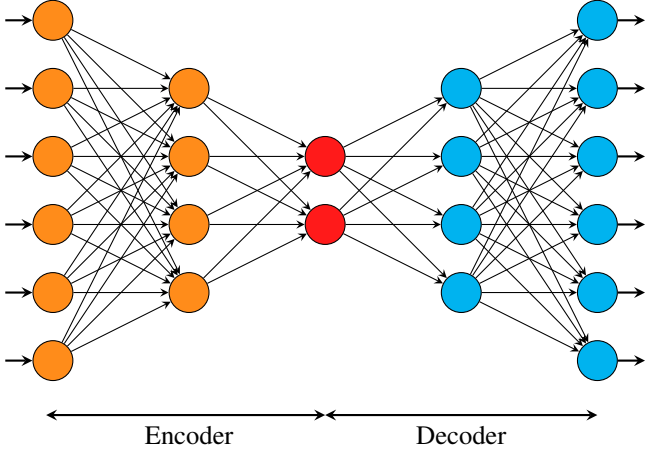


Fig. 6: Representação esquemática de um modelo de auto-encoder.

de modo a descartar o ruído presente. Essa representação intermediária atua como um filtro, ajudando a separar o conteúdo útil das interferências.

Em sentido similar, blocos residuais estimulam que o modelo aprenda apenas as diferenças, ou resíduos, entre a entrada e a saída desejada. No contexto de tarefas de *denoising*, isso significa que a rede pode direcionar seu foco em aprender as pequenas correções necessárias para eliminar interferências, sem modificar aquilo que já está correto na entrada. Isso aumenta a probabilidade de uma convergência mais rápida e em um desempenho melhor em tarefas como classificação, segmentação e reconstrução de sinais.

Um bloco residual é construído de forma a permitir que o sinal de entrada seja propagado diretamente para a saída, através de uma conexão de atalho (*skip connection*) [34]. Formalmente, o bloco residual da Figura 7 pode ser representado por:

$$\hat{\mathbf{x}} = \mathcal{F}(\mathbf{x}, \theta) + \mathbf{x} \quad (10)$$

onde $\mathbf{x} \in \mathbf{R}^{C \times H \times W}$ é o tensor de entrada, $\mathcal{F}(\cdot)$ representa a transformação não-linear, que é frequentemente composta por duas camadas convolucionais seguidas de funções de ativação e normalização, e $\hat{\mathbf{x}}$ é a saída do bloco. A operação $\mathcal{F}(\mathbf{x}, \theta)$ representa o mapeamento residual que a rede precisa aprender.

Ainda nesse contexto, é possível aliar os blocos supracitados à implementação de mecanismos de monitoramento [35] baseados em DNNs que ajustam dinamicamente o comportamento da rede mediante multiplicadores escalares aprendidos, *gates* condicionais ou redes auxiliares que inferem características relevantes com base em informações externas (como contexto, rótulos auxiliares ou indicadores) ou internas (como estatísticas intermediárias do próprio sinal). Tais blocos podem atuar como dispositivos de modulação, adaptando a resposta da rede conforme a natureza ou intensidade da perturbação presente na entrada. O princípio por trás dessa abordagem está relacionado à ideia de que o sistema ajusta seus parâmetros operacionais com base em uma estimativa corrente do ambiente ou do próprio estado.

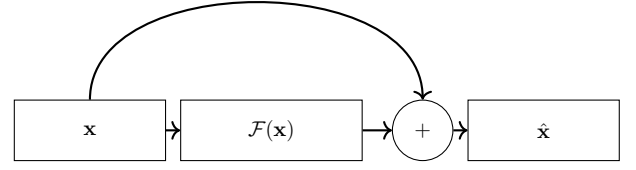


Fig. 7: Esquema de um bloco residual com conexão de atalho (*skip connection*).

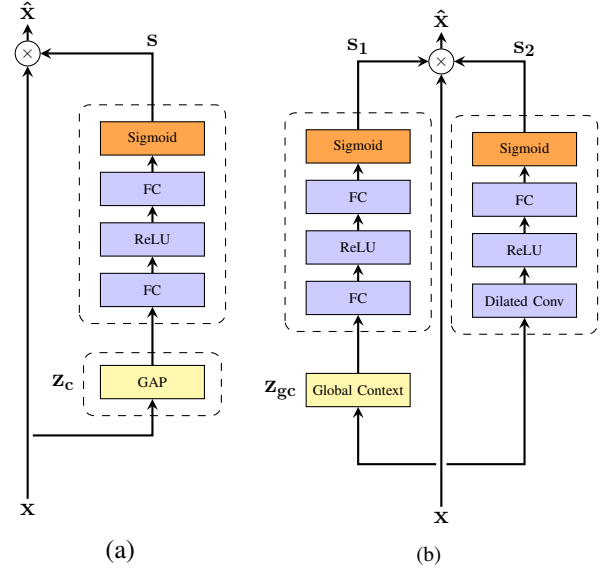


Fig. 8: Representação esquemática dos mecanismos de monitoramento utilizados. (a) Squeeze-and-Excitation. (b) Attention-Block.

III. METODOLOGIA

A. Mecanismos de Monitoramento

O bloco *Squeeze-and-Excitation* (SE) [36] constituído na Figura 8a tem como objetivo ajustar dinamicamente a importância de cada canal de ativação em DNNs. Esses canais correspondem aos diferentes mapas de características produzidos por uma camada convolucional, representando distintas informações extraídas do dado de entrada. Cada canal pode destacar padrões, como bordas, texturas ou outras características relevantes. O bloco SE atua, permitindo que a rede realce as informações mais úteis e atenuie as menos relevantes para a tarefa em questão. Sua estrutura é dividida em duas etapas principais: *squeeze* e *excitation*. Na fase de *squeeze*, busca-se condensar a informação espacial de cada canal em um único valor escalar que representa uma média global do canal, por meio de uma operação de *Global Average Pooling* (GAP), dada por:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j), \quad (11)$$

onde $x_c(i, j)$ representa o valor da ativação no canal c na posição espacial (i, j) , e z_c é o valor agregado correspondente ao canal c . O vetor resultante $\mathbf{z} \in \mathbf{R}^C$ representa uma descrição compacta dos C canais da entrada.

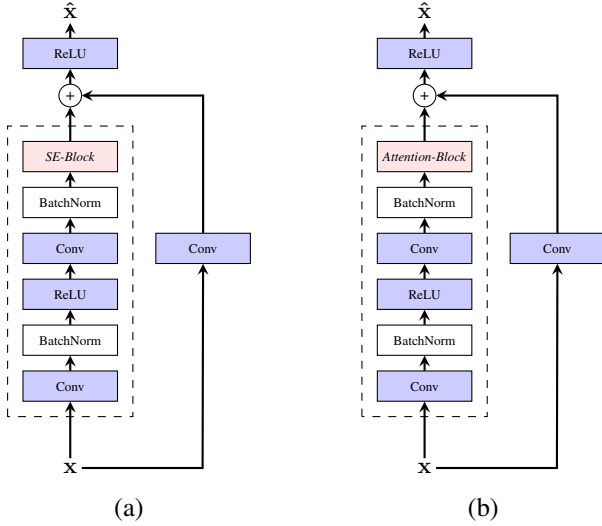


Fig. 9: Esquematização dos blocos residuais utilizados para comparação dos modelos. (a) Primeira variante. (b) Segunda variante.

A etapa seguinte, denominada *excitation*, gera um conjunto de pesos a serem aplicados às ativações originais. Essa etapa é implementada por meio de um pequeno bloco de rede neural composto por duas camadas totalmente conectadas (*Fully-Connected* - FC). Sendo a operação descrita como:

$$\mathbf{s} = \sigma(\mathbf{W}_2 \cdot \delta(\mathbf{W}_1 \cdot \mathbf{z})), \quad (12)$$

onde $\delta(\cdot)$ representa a função ReLU, $\sigma(\cdot)$ é a função *sigmoid*, $\mathbf{W}_1 \in \mathbf{R}^{\frac{C}{r} \times C}$ e $\mathbf{W}_2 \in \mathbf{R}^{C \times \frac{C}{r}}$ são os pesos das camadas FC, e r é um fator de redução que controla a dimensionalidade intermediária e o custo computacional do bloco.

Por fim, os coeficientes $\mathbf{s} = [s_1, s_2, \dots, s_C]$ obtidos são aplicados de forma canal-a-canal às ativações originais da entrada, realizando uma reponderação:

$$\hat{\mathbf{x}}_c(i, j) = s_c \cdot \mathbf{x}_c(i, j), \quad (13)$$

resultando nas ativações recalibradas $\hat{\mathbf{x}}$, que seguem para as próximas camadas da rede.

De forma análoga, o *Attention-Block* esquematizado na Figura 8b foi projetado com o objetivo de aplicar uma modulação adaptativa sobre os canais da entrada. A estrutura combina informação contextual global e mecanismos de detecção local de ruído; o bloco utiliza a soma de duas operações de agregação: média e máximo adaptativos.

Dado o tensor de entrada \mathbf{x} , o *Global Context* \mathbf{z}_{gc} é definido como:

$$\mathbf{z}_{\text{gc}} = \text{AvgPool}(\mathbf{x}) + \text{MaxPool}(\mathbf{x}), \quad (14)$$

sendo que este é aplicado a uma sequência idêntica à etapa de *excitation* do bloco SE.

Paralelamente, é realizada uma análise localizada do sinal por meio de convoluções dilatadas, que levam à identificação de padrões em diferentes escalas temporais, desenvolvendo uma máscara de ruído, constatando o que é sinal útil e o que deve ser descartado.

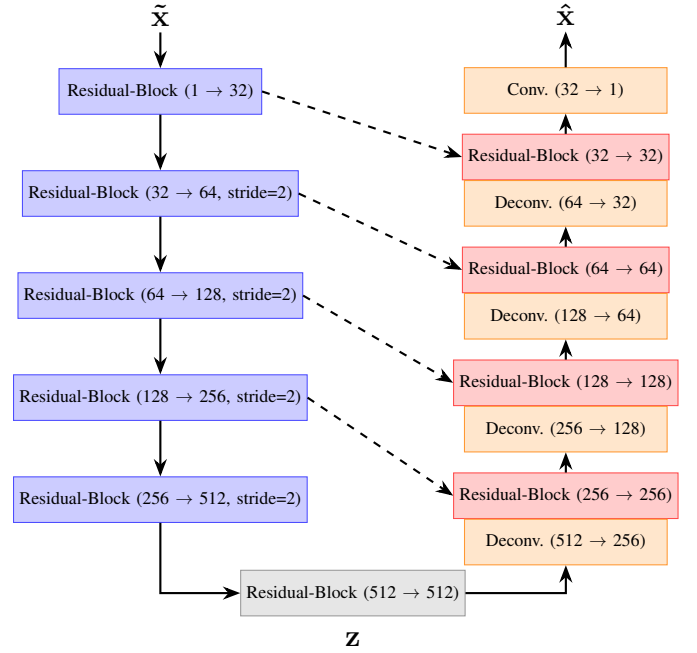


Fig. 10: Arquitetura do Autoencoder com blocos residuais e skip-connections. As seções do encoder, bottleneck e decoder são mostradas da esquerda para a direita.

A saída do *Attention-Block* é gerada com a multiplicação ponto a ponto entre a entrada original, a máscara estimada de ruído e os coeficientes de atenção por canal. Esse processo aplica filtros temporais e espaciais sobre o sinal, atenuando regiões ruidosas e reduzindo a influência de canais com baixa relevância para a reconstrução.

É possível observar na Figura 8 a semelhança estrutural entre os dois blocos, ambos incorporando mecanismos de monitoramento. Porém, o *SE-Block* realiza modulação exclusivamente no domínio dos canais, com base em estatísticas globais. Já o *Attention-Block* proposto expande esse tratamento ao combinar o contexto geral com características salientes, extraídas por meio de operações de *Max Pooling* e convoluções dilatadas dinâmicas.

B. Arquitetura do Autoencoder

Os blocos residuais da Figura 9 empregados são inspirados na arquitetura *ResNet* [34], sendo compostos por duas convoluções seguidas de uma conexão de atalho (*skip connection*). Os mecanismos de monitoramento supracitados foram inseridos como etapas desse bloco residual, posicionando-se logo antes da operação de somatório com a entrada original. Essa decisão foi motivada pela preservação da identidade dos dados enquanto ocorre a incorporação de um refinamento dinâmico orientado ao conteúdo do sinal.

O autoencoder da Figura 10 implementado possui um *encoder* composto por cinco blocos residuais consecutivos, que progressivamente aumentam a profundidade do tensor de entrada enquanto reduzem sua dimensionalidade temporal e um *decoder*, que reconstrói a entrada original a partir dessa representação comprimida, construído com camadas de *transposed convolution* (deconvoluções), seguidas por blocos

residuais que refinam a saída por meio da concatenação com os mapas de ativação correspondentes do *encoder*, em uma estrutura típica de *skip-connections*. No centro da arquitetura, o *bottleneck*, representado por um bloco residual com profundidade de 512 canais, captura as representações latentes de maior abstração do sinal e atua como um ponto de compressão informacional entre o *encoder* e o *decoder*.

C. Dados Experimentais

1) *Preparação dos Dados de Sinal de ECG*: A base de dados utilizada para essa primeira parte foi construída a partir do *MIT-BIH Arrhythmia Database* [37]. De cada um dos 48 registros de pacientes disponíveis no *dataset*, com uma taxa de amostragem de 360 Hz, foram extraídos 600 segmentos com pontos de início selecionados aleatoriamente e comprimento fixo de 1024 amostras, normalizados no intervalo $[0, 1]$ ou $[-1, 1]$ para o treinamento focado no ruído MA. A base final foi dividida em 80% para o treinamento do modelo, 10% para validação e 10% para o teste final, totalizando 23040 sinais para treinamento, 2880 para validação e 2880 para teste.

Foram gerados três tipos de ruídos sintéticos, dois deles modelados como sinais senoidais com frequência variável no intervalo de 20 a 80 Hz no caso do ruído MA e entre 0,05 e 0,7 Hz para o ruído BW, refletindo movimentos lentos e variações respiratórias que afetam o sinal de ECG. Para esses dois artefatos, as amplitudes dos ruídos foram amostradas aleatoriamente entre 0,05 e 0,1. O modelo matemático seguido para a geração dos ruídos sintéticos foi:

$$n_s[k] = A[k] \cdot \sin(2\pi f t_k + \phi) + \eta[k], \quad (15)$$

onde t_k representa os instantes de amostragem uniformemente espaçados, f é a frequência escolhida aleatoriamente para cada segmento, ϕ é uma fase aleatória uniforme em $[0, 2\pi]$, $A[k]$ é a amplitude variável ao longo do segmento e $\eta[k] \sim \mathcal{N}(0; 0,05^2)$ é o ruído gaussiano branco.

Pontua-se que a modelagem descrita na equação (15) permite a geração de um ruído BW muito próximo à realidade [7], no entanto, para o ruído do tipo MA, a modelagem adotada apresenta boa fidelidade em relação às características espectrais observadas na literatura, mas adota maior simplicidade de implementação. A introdução de uma amplitude ponto a ponto variável em $A[k]$, somada à presença de ruído branco aditivo $\eta[k]$, confere simulação de variações fisiológicas e mecânicas da atividade muscular. Porém, a representação senoidal contínua com frequência e fase fixas ao longo de cada segmento não consegue capturar adequadamente a natureza abrupta e não estacionária dos artefatos musculares [9]. Ainda assim, essa simplificação foi adotada devido à sua facilidade computacional e à possibilidade de controlar os parâmetros do ruído e avaliar seus efeitos de forma sistemática durante os experimentos.

Optou-se por não incluir uma simulação do ruído EM devido à sua complexidade e variabilidade intrínseca, que dependem fortemente de fatores externos como o tipo de eletrodo [9].

2) *Preparação dos Dados de Áudio*: Nesta etapa, foi realizada a preparação dos dados a partir dos arquivos *.wav* do conjunto *VoiceBank + DEMAND* [38]. Esse dataset combina as falas limpas do *CSTR VCTK Corpus* [39] com os ruídos de multidões, ventilação, tráfego de carros, músicas de fundo, dentre outros comuns em áreas urbanas que fazem parte do pacote *DEMAND* [40]. Assim sendo, os sinais de áudio, tanto limpos quanto contaminados por ruído, foram convertidos em espectrogramas Mel, normalizados e com dimensões padronizadas para entrada em CNNs.

Cabe destacar que os áudios do *dataset* representam 56 falantes — 28 masculinos e 28 femininos — com amostragem de 48 kHz e duração variável. Para padronização, os áudios com menos de 3 segundos foram preenchidos com zeros, enquanto os mais longos foram truncados para exatamente 3 segundos. Ao todo, 23075 áudios foram utilizados para treinamento e 824 para teste.

Para representar os sinais de áudio no domínio tempo-frequência, foi utilizada a transformação em *Mel Spectrogram*. Dessa forma, foi definido um número-alvo de 272 *frames* temporais, comprimento da janela de análise igual a 4096 amostras, sendo este o mesmo número de pontos para realização da FFT com tamanho de 4096 amostras, resultando em espectros de magnitude que são então filtrados pelas 512 bandas Mel escolhidas para a análise. Por fim, o espectrograma final de cada sinal de áudio tem dimensões de 512×272 .

Esses parâmetros foram ajustados de modo a garantir uma boa resolução espectral e uma representação temporal suficientemente detalhada, o que possibilita a posterior reconstrução do áudio em formato *.wav* com poucas perdas perceptíveis que comprometam sua compreensão. Destaca-se que os valores obtidos nos espectrogramas foram normalizados individualmente para o intervalo $[0, 1]$ por meio da técnica de normalização *min-max*.

D. Treinamento dos Modelos de Autoencoder

Durante o processo de treinamento, os modelos recebem como entrada tensores $\tilde{\mathbf{X}}$ representando espectrogramas Mel ou sinais de ECG ruidosos. O objetivo é ajustar os parâmetros da rede de forma a minimizar a função de perda do Erro Quadrático Médio (MSE — *Mean Squared Error*), que mede a discrepância entre o sinal ou espectrograma original \mathbf{X} e o *output* $\hat{\mathbf{X}}$ do modelo.

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2, \quad (16)$$

onde: X_i é o valor do i -ésimo elemento do tensor de referência que representa o dado alvo original anterior à adição do ruído; \hat{X}_i é o valor correspondente previsto pela rede neural; N representa o número total de elementos nos tensores comparados.

É notável que erros maiores possuem impacto proporcionalmente maior na função de perda que penaliza desvios ponto a ponto entre o tensor previsto e o tensor original.

Para a otimização dos parâmetros dos modelos durante o treinamento, foi utilizado o algoritmo AdamW (*Adam with Weight Decay*) [41], uma variação do otimizador Adam que

inclui um termo de regularização baseado em decaimento de peso.

O AdamW atualiza os parâmetros θ da rede com base nas estimativas de primeira e segunda ordem dos momentos dos gradientes, retirando a regularização do cálculo do gradiente, promovendo estabilidade e também maior velocidade na convergência dos modelos. A expressão de atualização dos parâmetros é descrita abaixo:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} + \lambda \theta_t \right) \quad (17)$$

onde: θ_t representa os parâmetros da rede no instante t ; \hat{m}_t e \hat{v}_t são as médias móveis dos gradientes e de seus quadrados, respectivamente, ambas com correção de viés; η é a taxa de aprendizado; ϵ é um pequeno valor para evitar divisão por zero; λ representa o fator de decaimento de peso, utilizado para melhorar a capacidade de generalização do modelo.

IV. RESULTADOS E DISCUSSÕES

A. Critérios de Avaliação para Sinais ECG Reconstruídos

Para avaliar o desempenho do modelo na reconstrução de sinais ECG contaminados por ruídos, foram utilizadas três métricas principais: a Raiz do Erro Quadrático Médio (RMSE – Root Mean Squared Error), a Diferença Percentual da Raiz do Erro Quadrático Médio (PRD – Percentage Root-mean-square Difference) e a Relação Sinal-Ruído (SNR – Signal-to-Noise Ratio), utilizada tanto para o sinal contaminado (SNR_{in}) quanto para o sinal reconstruído (SNR_{out}).

Nesse contexto de avaliação, a RMSE oferece uma métrica eficaz para quantificar o erro médio entre os sinais reconstruído e original, expressando-o na mesma unidade do sinal, o que facilita sua interpretação prática. Embora também penalize discrepâncias maiores devido ao uso do erro quadrático, a aplicação da raiz quadrada suaviza a penalização em relação à MSE, tornando a RMSE uma medida mais intuitiva do desvio médio. Seu cálculo é realizado por meio da raiz quadrada da média dos quadrados das diferenças entre os valores originais e os reconstruídos, conforme expressa a equação:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

Dessa forma, em aplicações que envolvem a reconstrução de componentes, a RMSE destaca-se como uma métrica fundamental para a detecção de distorções que possam comprometer a integridade e fidelidade dos sinais reconstruídos em relação aos originais.

Outrossim, a PRD fornece uma medida relativa da preservação do sinal original após o processo de reconstrução pelo modelo. Essa métrica é calculada com base na razão entre o erro quadrático entre os sinais original e reconstruído e a energia do próprio sinal original, consoante ao descrito na equação:

$$\text{PRD} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2}} \times 100 \quad (19)$$

Tanto para a PRD quanto para a RMSE, quanto menores forem os seus valores, melhor é o desempenho do modelo na reconstrução dos sinais, indicando maior proximidade em relação ao sinal original.

Além disso, os valores da SNR entre o sinal ruidoso, reconstruído e limpo foram comparados, permitindo quantificar o ganho de qualidade promovido pelo modelo na reconstrução dos sinais. A SNR de um sinal é calculada com base na razão entre a potência do sinal limpo e a potência do ruído, expressa em decibéis (dB) por meio de:

$$\text{SNR} = 10 \times \log_{10} \left(\frac{\sum_{i=1}^n y_i^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \right) \quad (20)$$

Neste caso, quanto maior for o valor da SNR_{out} em relação a SNR_{in} , melhor foi o desempenho do modelo na remoção dos ruídos presentes.

Por fim, com o objetivo de obter uma validação estatística dos sinais reconstruídos pelo autoencoder, os sinais limpos e os reconstruídos foram comparados por meio do teste de Kolmogorov-Smirnov (KS) para duas amostras. Esse teste avalia se dois conjuntos de dados provêm da mesma distribuição, comparando suas Funções de Distribuição Acumulada (FDA) e calculando a maior diferença absoluta entre elas:

$$D_{n,m} = \sup_x |F_n(x) - G_m(x)| \quad (21)$$

Consequentemente, um valor baixo da distância $D_{n,m}$ é desejável, pois representa maior similaridade entre as distribuições dos valores do sinal reconstruído e do sinal original.

B. Resultados por Tipo de Ruído nos Sinais de ECG

Os experimentos que se seguem foram feitos no modelo com o *SE-Block*, comparando seu desempenho em diferentes cenários de ruídos.

1) *Sinal Contaminado por Ruído Muscle Artifact*: De primeiro momento, destaca-se que o sinal com ruído representado na Figura 11 possui uma $\text{SNR}_{\text{in}} = 1,81$ dB, ou seja, quase 40% do sinal é composto por ruído de alta frequência representando movimentações musculares do paciente durante o exame médico.

Nesse sentido, o ruído sintético MA causa deformação principalmente nas ondas P do sinal ECG, tornando-as praticamente irreconhecíveis. Ademais, o efeito do ruído nos complexos QRS se manifesta pela adição de componentes inexistentes ou mesmo mitigação dos originais, como visualizado no segundo pico R do sinal ruidoso da Figura 11. Em um cenário real, a distorção dessas ondas prejudica a análise do ritmo sinusal e compromete algoritmos de detecção automática, que podem interpretar oscilações artificiais como eventos cardíacos reais.

No que concerne o desempenho do modelo, o sinal foi reconstruído com uma $\text{SNR}_{\text{out}} = 22,42$ dB um crescimento de 20,61 dB em relação ao sinal de entrada, ao passo que somente 0,57% do sinal de saída é formado por ruído.

Além disso, o *output* apresentou uma $\text{RMSE} = 0,026$ em relação ao sinal original. Levando em conta que os sinais ECG para o caso do ruído MA foram previamente

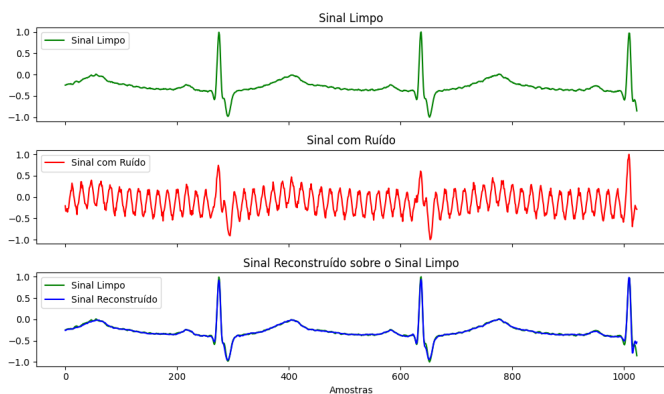


Fig. 11: Sinal limpo, ruidoso e reconstruído pelo modelo com *SE-Block* para o ruído MA.

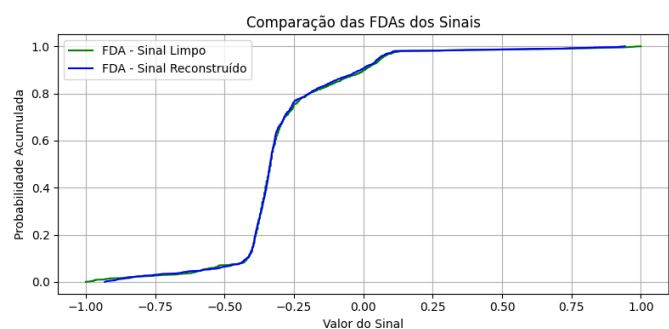


Fig. 12: Funções de Distribuição Acumulada dos sinais limpo e reconstruído pelo modelo com *SE-Block* a partir do sinal contaminado com o ruído MA.

normalizados no intervalo $[-1, 1]$, esse valor indica um erro médio correspondente a aproximadamente 1,3% da amplitude total do sinal. Conjuntamente, a $PRD = 7,57\%$ corrobora o bom desempenho do modelo com baixo nível de discrepância relativa, mantendo as principais características morfológicas.

Por último, submetendo-se os sinais reconstruído e original ao teste KS a distância absoluta entre as suas FDAs foi $D = 0,033$ de modo que, no pior caso, a FDA do sinal reconstruído difere apenas 3,3% da FDA do sinal original. Tal métrica fica ainda mais evidente ao se analisar os gráficos das respectivas FDAs representadas na Figura 12. A sobreposição entre as curvas de FDA dos sinais limpo e reconstruído, corroborada por uma distância KS reduzida, demonstra que o modelo efetuou uma reconstrução com manutenção das características globais da distribuição do sinal.

Durante o processo de treinamento do modelo, notou-se que a normalização dos dados para o intervalo $[-1, 1]$ acarretou em uma rápida convergência, bem como em um melhor desempenho de reconstrução. Nesse sentido, considerando que a função de geração do ruído MA é caracterizada por componentes de alta frequência aliados a variações abruptas na amplitude, a manifestação no sinal ECG ocorre por meio de picos intermitentes que, possivelmente, em um intervalo $[0, 1]$, são comprimidos de modo a acentuar a dificuldade do modelo em identificar o que seria ruído e o que seria sinal útil.

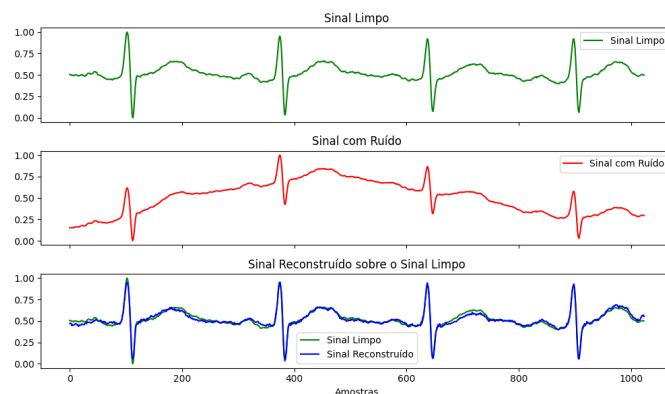


Fig. 13: Sinal limpo, ruidoso e reconstruído pelo modelo com *SE-Block* para o ruído BW.

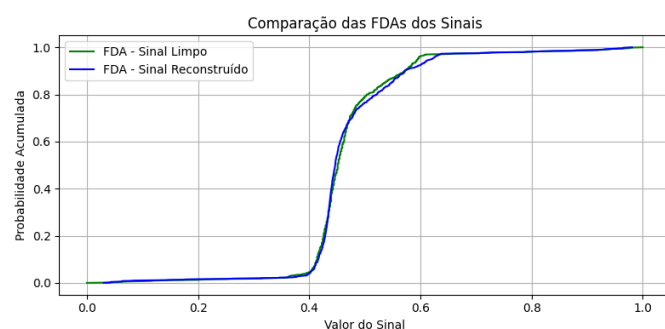


Fig. 14: Funções de Distribuição Acumulada dos sinais limpo e reconstruído a partir do sinal contaminado com o ruído BW.

Sob outra perspectiva, observou-se que, em alguns momentos, o modelo tende a interpretar erroneamente segmentos relacionados à deflexão S do complexo QRS. Essa conjuntura ocorre no sinal reconstruído da Figura 11. Após o valor de 1000 amostras, logo em seguida do último pico R, o começo da deflexão original foi diretamente perturbado pelo ruído e, por sua vez, o modelo acabou por seguir a propensão existente no sinal ruidoso. Esse comportamento pode ser atribuído à arquitetura do autoencoder com os blocos residuais que, embora promova a preservação da morfologia do sinal, também pode induzir a rede a manter padrões ruidosos se estes forem recorrentes ou similares às componentes do sinal original.

2) *Sinal Contaminado por Ruído Baseline Wander*: Na Figura 13, o sinal contaminado com ruído do tipo *Baseline Wander* apresenta uma razão sinal-ruído de entrada $SNR_{in} = 9,29$ dB, o que indica que aproximadamente 10,5% do conteúdo do sinal corresponde ao ruído.

Nesse cenário, as ondas P e T, que naturalmente possuem menor amplitude, quase que completamente desaparecem e acabam se incorporando com a flutuação ruidosa, o que promove a impossibilidade de, por exemplo, identificar arritmias, ao comprometer a análise do ritmo e da repolarização atrial. Além da elevação, pode ocorrer também o declive do segmento ST que acontece logo em seguida a um pico R, representando o momento em que os ventrículos estão totalmente despolarizados, ou seja, não há movimento elétrico. Sendo assim, a

alteração deste segmento pode simular isquemia ou infarto.

Depois de passar pelo modelo, o sinal de saída obteve uma $SNR_{out} = 27,44$ dB, onde somente 0,179% do sinal corresponde a ruído, uma melhora de 18,15 dB.

A comparação do sinal de saída com o original resultou em uma $RMSE = 0,022$, o que, para sinais normalizados no intervalo $[0, 1]$, corresponde a um erro quadrático médio de aproximadamente 2,2% da amplitude total do sinal. Adicionalmente, a PRD obtida foi de 4,25%. Esse valor está em conformidade com o comportamento visual observado: o modelo conseguiu reduzir o desvio de linha de base e preservar com aceitável fidelidade a morfologia dos complexos QRS e das ondas P e T, porém com erros para mais ou para menos nos valores das amplitudes destes.

O teste KS de duas amostras entre o sinal de entrada e o de saída retornou um valor $D = 0,055$, indicando que a diferença máxima entre as funções de distribuição acumulada dos dois sinais é de 5,5%. Tal valor aponta uma alta similaridade estatística entre os sinais. Essa interpretação é perceptível no gráfico da Figura 14, onde observa-se que as curvas das FDAs estão quase totalmente sobrepostas ao longo de todo o domínio. As maiores discrepâncias entre as curvas ocorrem na região de transição da distribuição, entre os valores normalizados de aproximadamente 0,4 e 0,6, justamente onde o sinal apresenta maior variação.

Ao contrário do que foi observado no treinamento para remoção do ruído MA, a normalização no intervalo $[-1, 1]$ comprometeu o desempenho do modelo. Esse efeito pode ser atribuído ao fato de que, ao distribuir os valores do sinal de forma simétrica em torno do zero, a normalização reduziu a resolução relativa das pequenas flutuações de baixa frequência características do ruído BW, comprimindo suas representações.

No sinal reconstruído, observa-se, de forma geral, um desempenho satisfatório, com a preservação do padrão característico do sinal de ECG original. Contudo, a atenuação de amplitude provocada pela flutuação associada ao ruído BW, especialmente evidente na região do primeiro pico R, contribuiu para que o modelo aprendesse parcialmente a tendência do ruído, em detrimento da morfologia real do sinal. O padrão consistente observado no ruído, por sua baixa frequência, amplitude relativamente estável e persistência ao longo do tempo, pode ter levado o modelo a interpretar tais variações como parte legítima da estrutura do sinal.

3) *Sinal Contaminado por Ruído Gaussiano*: A Figura 15 apresenta uma abordagem distinta em relação aos dois primeiros experimentos com o modelo. O ruído gaussiano adicionado não segue os mesmos padrões de regularidade dos ruídos anteriores, não possuindo intervalos comuns de frequência ou amplitude. Nos experimentos anteriores, os ruídos simulavam condições fisiológicas reais observadas em sinais biológicos, com características espectrais bem definidas, o que permitiu ao modelo adaptar-se parcialmente ao ruído e manter um nível de performance aceitável. Em contraste, o ruído gaussiano introduzido neste experimento é não correlacionado e possui distribuição de energia uniforme em todo o espectro de frequências.

O sinal ruidoso destacado possui uma $SNR_{in} = 2,89$ dB, de modo que aproximadamente 34% de seu conteúdo é ruído.

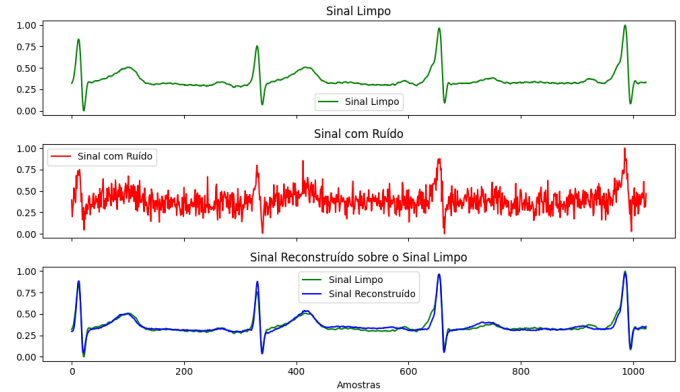


Fig. 15: Sinal limpo, ruidoso e reconstruído pelo modelo com *SE-Block* para o ruído gaussiano.

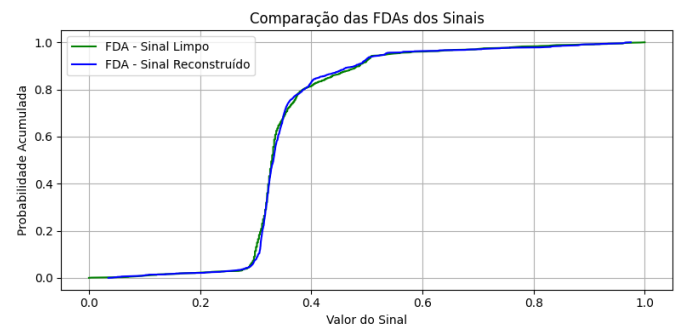


Fig. 16: Funções de Distribuição Acumulada dos sinais limpo e reconstruído pelo modelo com *SE-Block* a partir do sinal contaminado com o ruído gaussiano.

Visualmente, é possível identificar que a atuação do ruído, embora presente em praticamente todo o intervalo do sinal, não é capaz de inteiramente danificar a estrutura do mesmo; por exemplo, a primeira variação de amplitude após o pico R, representada pela onda T, ocorrida por volta das primeiras 100 amostras do ECG, se mantém ainda com a adição do ruído, o que também acontece na onda T subsequente. Isto posto, pode-se dizer que o ruído em questão age como uma máscara que obscurece detalhes mais sutis do sinal, como possíveis ondas P ou U, mas preserva as características principais como os complexos QRS e as ondas T.

Nesse cenário, constatou-se uma $RMSE = 0,029$ juntamente com uma $PRD = 7,25\%$, evidenciando-se que, apesar das dificuldades em acompanhar as variações do sinal original, o modelo, em geral, desempenha de forma moderadamente fidedigna na reconstrução, sendo que somente uma pequena porcentagem da energia do ECG original é perdida ou distorcida, enquanto os aspectos característicos das ondas e complexos possuem discrepâncias que não são exageradas. Ainda nessa perspectiva, a $SNR_{out} = 22,8$ dB destaca, novamente, a mitigação da potência de ruído presente no sinal de entrada.

A análise estatística, por intermédio do teste KS, resultou no valor $D = 0,081$ ratificando que mesmo com a presença do ruído gaussiano e sua natureza não correlacionada, o modelo conseguiu preservar a distribuição estatística, no entanto, com mais dificuldades. A análise da Figura 16 apoia a métrica

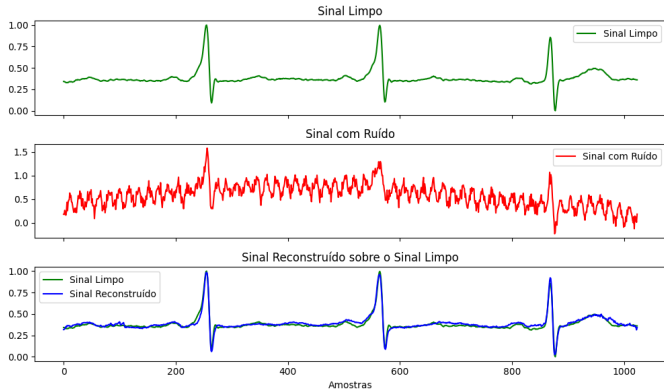


Fig. 17: Sinal limpo, ruidoso e reconstruído pelo modelo com *SE-Block* para os três ruídos conjuntamente.

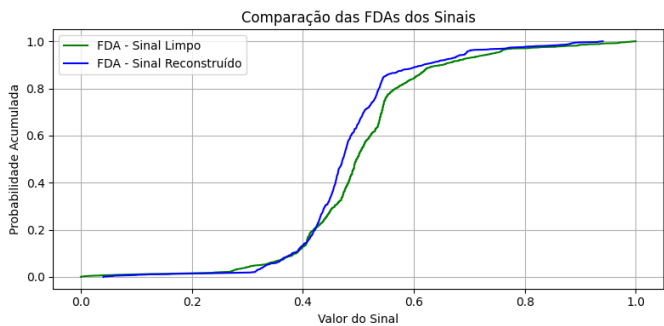


Fig. 18: Funções de Distribuição Acumulada dos sinais limpo e reconstruído pelo modelo com *SE-Block* a partir do sinal contaminado com todos os ruídos simultaneamente.

calculada, mas também revela maiores ocorrências de discordâncias em trechos de baixa energia do sinal, onde transições locais sutis acabam por não serem identificadas corretamente.

4) *Sinal Contaminado pelos Três Ruídos Simultaneamente*: Na Figura 17, é viável identificar a ação dos três ruídos sobre o sinal de ECG. Preliminarmente, nota-se a flutuação do sinal provocada pelo ruído BW, uma alta concentração de alternâncias de frequência moderada advindas do MA e, por fim, o ruído gaussiano que age de forma uniforme durante toda a amostragem. Evidentemente, este é o cenário de maior adversidade para o modelo, onde ruídos de diferentes particularidades – baixa frequência, alta frequência – são combinados.

Dessarte, a $SNR_{in} = -0,19$ dB indica que cerca de 55,4% do sinal é ruído. As ondas relacionadas ao segmento ST, bem como o segundo pico R, são deformadas tanto pela atenuação de suas amplitudes, quanto pela adição de resíduos e deslocamento temporal. Com isso em mente, o processo de reconstrução resultou em um $SNR_{out} = 19,53$ dB, reduzindo o ruído para cerca de 1,1% do sinal de ECG, provocando o ajuste necessário nas regiões com maior densidade de distorção, e restabelecendo os principais componentes do traçado.

Em conjunto a isso, o valor calculado da $RMSE = 0,051$ reflete uma reconstrução por boa parte coerente, porém com erros notáveis. No sinal reconstruído da Figura 17, conforme em alguns momentos dos casos anteriores, as partes do sinal

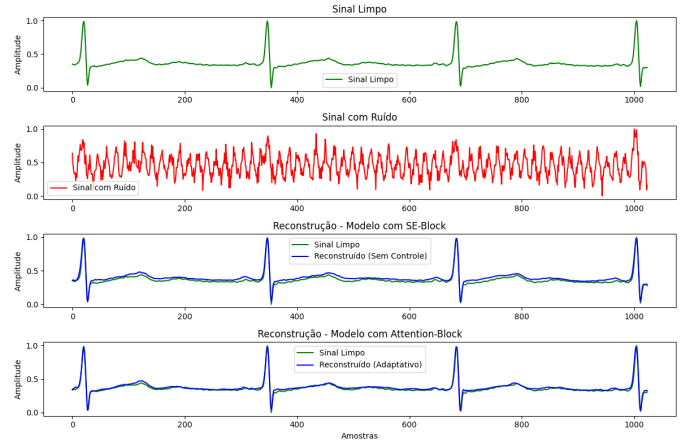


Fig. 19: Comparação entre sinal limpo, ruidoso e reconstruído para os dois modelos propostos.

TABELA I: Comparação de desempenho entre o *SE-Block* e *Attention-Block* na filtragem do sinal ruidoso da Figura 19.

Método	SNR_{out} (dB)	RMSE	PRD (%)
<i>SE-Block</i>	21,07	0,034	8,84
<i>Attention-Block</i>	25,62	0,019	5,23

que possuem rápidas variações sutis, denotam-se como as mais difíceis para o modelo, onde existem pequenas discrepâncias localizadas que apontam para erros residuais. Além disso, nas regiões de transição entre segmentos isovolumétricos e ondas, como no final do complexo QRS e início da onda T, logo após o ponto de 800 amostras, observam-se diferenças no perfil de curvatura, um indício de que o modelo pode acabar suavizando em excesso tais transições em prol da atenuação de ruído.

A métrica $PRD = 9,66\%$ confere justamente tal observação, já que os maiores desvios ocorrem em regiões de alta variabilidade temporal; tais discrepâncias contribuem diretamente para o valor da PRD. Ainda que, do ponto de vista funcional, não comprometam a identificação clínica dos eventos cardíacos. Em contraste, a preservação dos segmentos expressa o equilíbrio por parte do modelo em propor a supressão de ruídos, garantindo a integridade morfológica.

Todavia, apesar da competente correspondência morfológica observada na reconstrução ponto a ponto, a análise estatística com o auxílio da comparação entre as FDAs revela desigualdades notáveis entre as distribuições dos valores do sinal limpo e reconstruído. Como se observa na Figura 18, há uma divergência visível entre as curvas na região entre 0,4 e 0,65, indicando que a reconstrução não preserva integralmente a distribuição de probabilidade dos valores do sinal original. O valor decorrente do teste KS $D = 0,155$ corrobora essas afirmações e expressa uma discrepância estatisticamente relevante entre as distribuições acumuladas do sinal limpo e do sinal reconstruído.

C. Comparação dos Modelos

Os resultados da Tabela I demonstram a superioridade do *Attention-Block* em relação ao *SE-Block*, sendo que o primeiro

proporciona melhor qualidade na filtragem dos sinais de ECG, evidenciada por uma maior estabilidade, com menos oscilações e reduzidos artefatos residuais, conforme representado. Contudo, na reconstrução do modelo destacado na Figura 19 com *SE-Block*, são perceptíveis flutuações na linha de base e uma leve distorção nos complexos de baixa energia. Uma hipótese para esse comportamento reside no fato de que o *SE-Block* utiliza um mecanismo simples de excitação por canal baseado em uma redução dimensional fixa; em contraste, o *Attention-Block* adota um módulo de ajuste que integra mecanismos globais e locais de forma complementar.

Ainda nesse sentido, o *Attention-Block* permite ao modelo ajustar dinamicamente sua resposta conforme a distribuição do ruído ao longo do sinal por meio da máscara de ruído que permite uma resposta mais seletiva. Desta maneira, a convolução dilatada usada para detectar o ruído local amplia o campo receptivo sem perda de resolução, o que é particularmente útil neste caso onde existem três ruídos com diferentes escalas e padrões temporais. Ocorre, então, a geração de um mecanismo de supressão, que consegue mitigar múltiplos ruídos simultâneos sem comprometer significativamente a integridade morfológica do sinal original.

A análise estatística do teste KS revelou um valor $D_1 = 0,4648$ e $D_2 = 0,2412$ para o modelo com *SE-Block* e com *Attention-Block* respectivamente.

D. Limitações dos Modelos

Observa-se na Figura 20 que o sinal de ECG em questão possui particularidades que diferem dos demais sinais limpos analisados. Nesse ponto de vista, distingue-se uma menor amplitude nos complexos QRS e maior presença de variações nas ondas P e T, identificando níveis de arritmia, além de que certas seções do sinal limpo em questão apresentam semelhança com os ruídos analisados aqui. Nessa perspectiva, pode-se inferir que o modelo acaba por identificar partes morfológicas do sinal como ruidosas em vez de sinal útil, e, por isso, descarta-as.

É notável, também, que suas características morfológicas como os complexos QRS e as ondas P estão deterioradas antes mesmo da aplicação do ruído. Portanto, já que se trata de um evento de arritmia raro, em comparação com os outros do *dataset*, a reconstrução com maior fidelidade exigiria um treinamento específico, com enfoque nessas peculiaridades.

E. Métricas de Avaliação para os Espectrogramas

Com o intuito de realizar uma avaliação objetiva da qualidade de reconstrução dos espectrogramas, além da utilização da RMSE e PRD, foram incorporadas a Relação Sinal-Ruído de Pico (*Peak Signal-to-Noise Ratio* - PSNR) e o Índice de Medida da Similaridade Estrutural (*Structural Similarity Index Measure* - SSIM).

O PSNR é uma métrica baseada no Erro Médio Quadrático (MSE) entre os espectrogramas comparados e fornece uma noção da diferença absoluta média entre os valores dos pontos ou intensidades do espectrograma. A fórmula do MSE para

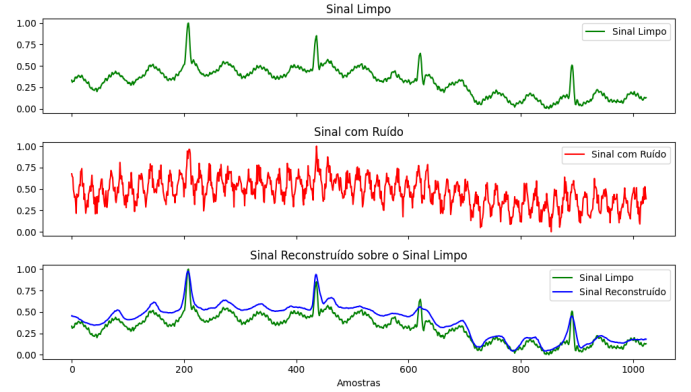


Fig. 20: Exemplo de sinal onde os modelos não conseguem demonstrar uma reconstrução satisfatória.

este caso é dada por:

$$\text{MSE} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - \hat{I}(i, j)]^2 \quad (22)$$

onde m e n são as dimensões do espectrograma. A partir do MSE, o PSNR é calculado como:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{L^2}{\text{MSE}} \right) \quad (23)$$

em que L representa o valor máximo possível da intensidade do espectrograma. Neste experimento, esse valor é igual 0 dB, uma vez que os espectrogramas estão compreendidos na faixa de -80 dB até 0 dB. Valores mais elevados de PSNR indicam menor erro de reconstrução em termos absolutos.

Por outro lado, o SSIM avalia a qualidade da reconstrução levando em conta não apenas diferenças pontuais, mas também aspectos perceptuais como luminância, contraste e, principalmente, a estrutura. A equação do SSIM entre duas janelas x e y extraídas dos espectrogramas I e \hat{I} é expressa por:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (24)$$

onde μ_x e μ_y são as médias, σ_x^2 e σ_y^2 as variâncias, e σ_{xy} a covariância das janelas x e y . As constantes C_1 e C_2 são parâmetros de estabilização para evitar divisões por valores muito pequenos. O SSIM varia de 0 a 1, sendo 1 a similaridade estrutural perfeita.

F. Resultados do Modelo em Espectrogramas

Em um primeiro instante, na Figura 21a é possível notar uma concentração clara de energia nas regiões de frequência mais baixas. Por outro lado, no espectrograma ruidoso, essas estruturas tornam-se visivelmente menos nítidas. Há uma elevação generalizada da densidade espectral de energia em praticamente toda a faixa de frequências, o que indica a presença de componentes ruidosos espalhados. As regiões entre os harmônicos, antes mais escuras e silenciosas, agora apresentam intensidades elevadas, ou seja, ocorreu um preenchimento artificial causado pelo ruído. Sendo assim, a ação do ruído sobre o espectrograma se manifesta afetando tanto

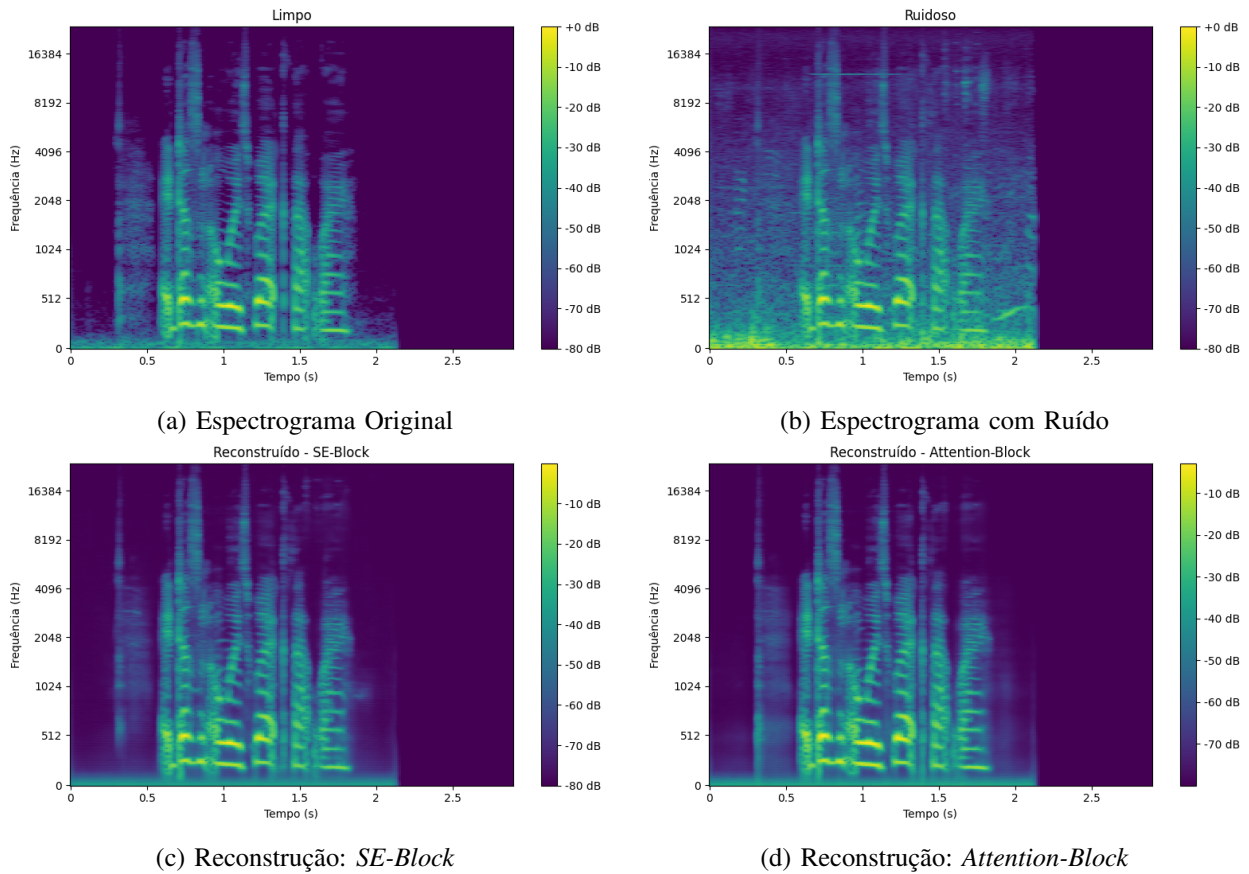


Fig. 21: Espectrogramas: (a) limpo, (b) ruidoso, (c) reconstruído pelo modelo com *SE-Block* e (d) reconstruído pelo modelo com *Attention-Block*.

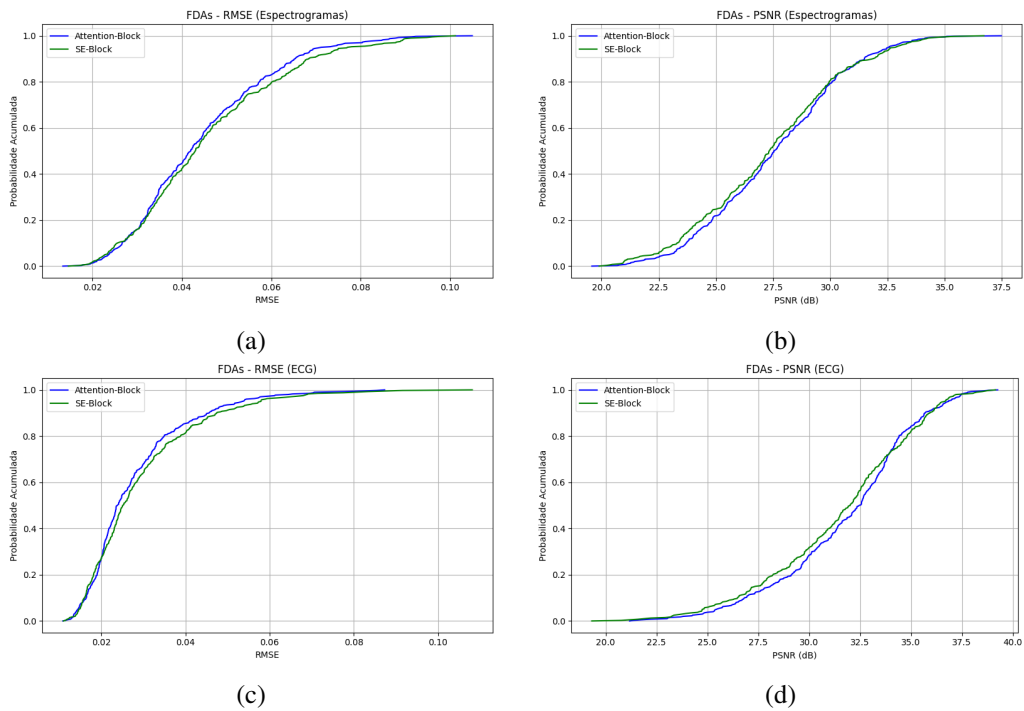


Fig. 22: Comparação das Funções de Distribuição Acumulada das métricas RMSE e PSNR para os dois modelos após 400 testes: (a)-(b) espectrogramas; (c)-(d) sinais de ECG.

a distribuição de energia no tempo quanto a clareza das informações espectrais.

Nas reconstruções, algumas diferenças sutis podem ser observadas entre os dois modelos. O modelo com *SE-Block* apresenta uma reconstrução um pouco mais suave nas regiões de baixa frequência, o que pode indicar um leve alisamento das informações temporais. Ainda assim, mantém a maior parte das estruturas do sinal original.

Já o modelo com *Attention-Block*, por outro lado, preserva, com um pouco mais de fidelidade, os detalhes em torno de 1 s e 1,5 s, onde há maior concentração de energia no sinal limpo. Da mesma maneira, observa-se que a energia nas regiões de ruído residual é ligeiramente menor, especialmente nas faixas inferiores da escala de frequência, o que pode indicar que o mecanismo de atenção concentrou-se nas partes mais relevantes do espectro.

A RMSE do modelo com *Attention-Block* apresentou o menor valor de 3,03 dB, enquanto o valor do modelo *SE-Block* foi de 3,34 dB. No caso do SSIM, os modelos obtiveram resultados aproximadamente idênticos. O modelo *Attention-Block* apresentou um valor igual 0,8786 contra 0,8735, indicando uma preservação marginalmente melhor das estruturas espectrais.

A métrica PSNR favorece o modelo com *Attention-Block*, que obteve 28,42 dB valor superior ao do *SE-Block* 27,62 dB. Um PSNR mais alto sugere que o sinal reconstruído se aproxima mais do sinal original em termos de intensidade e qualidade geral. Por fim, a PRD mostra valores baixos para os dois modelos: 4,65% para *SE-Block* e 4,24% para *Attention-Block*.

De fato, tanto a análise visual como as métricas demonstraram que os modelos praticamente atuaram de forma igual, com uma sutil vantagem de desempenho por meio do *Attention-Block* que perdura em todas as métricas avaliadas.

G. Comparação Geral Final

Na etapa final do estudo, os modelos foram submetidos a 400 testes tanto para o *dataset* de testes dos espectrogramas quanto para o dos sinais de ECG; no caso deste último, os três ruídos já comentados atuaram simultaneamente nesta etapa. Para cada teste realizado, calcularam-se a RMSE e a PSNR. Garantiu-se que os dois modelos recebessem exatamente os mesmos dados. Desse modo, os gráficos das respectivas FDAs da Figura 22 foram plotados.

Uma análise preliminar, referente ao desempenho do modelo nos espectrogramas, permite identificar rapidamente o que foi considerado na seção anterior. Os valores da RMSE e da PSNR, em boa parte do gráfico, se mantêm próximos, com diferenças sutis, e em um momento o *SE-Block* consegue um desempenho levemente superior. No gráfico da Figura 22a, o modelo com *Attention-Block* apresenta valores de RMSE ligeiramente menores em boa parte da distribuição, especialmente na região intermediária da curva. Isso indica que o *Attention-Block* mantém uma reconstrução mais consistente com erros absolutos mais baixos, em média.

Quanto à representação gráfica da FDA referente ao PSNR, o *Attention-Block* apresenta valores superiores de forma con-

sistente, porém com uma margem pequena. Aponta-se, no entanto, que os valores de PSNR apresentam um comportamento de crescimento com poucas variações bruscas. O *SE-Block* possui um crescimento mais lento durante o teste ao mesmo tempo em que oscila bastante, mostrando uma adaptabilidade limitada quando a complexidade dos ruídos aumenta.

Na segunda parte, referente aos sinais de ECG, nota-se um comportamento gráfico similar tanto para a RMSE quanto para o PSNR. No entanto, é importante destacar que a curva do modelo com *Attention-Block* se posiciona consistentemente à esquerda da curva do bloco SE. Essa resiliência do *Attention-Block* indica que o mecanismo de atenção, ao focar seletivamente nas regiões mais relevantes do sinal e suprimir regiões ruidosas, é melhor em cenários complexos. Já o bloco SE, por depender de um mecanismo mais global, apresenta maior vulnerabilidade frente à interferência cruzada entre ruídos de diferentes características.

A Figura 22d apresenta que ambos os modelos possuem distribuições semelhantes, com valores de PSNR variando entre 25 dB e 38 dB. No entanto, o *Attention-Block* apresenta desempenho ligeiramente superior nas regiões de PSNR mais elevadas. Isso é evidenciado pelo fato de que, para uma mesma probabilidade acumulada acima de 0,6, o *Attention-Block* tende a alcançar valores de PSNR maiores que o bloco SE. Sendo assim, novamente, tal comportamento sugere que o *Attention-Block* preserva a qualidade do sinal reconstruído em casos de ruído mais severo.

Além disso, observa-se que a curva do *Attention-Block* cresce de forma mais gradual, indicando uma distribuição mais homogênea dos resultados ao longo dos diferentes níveis de ruído. Em contraste, a curva do *SE-Block* é um pouco mais inclinada no intervalo central, o que sugere uma maior concentração de amostras em uma faixa específica de PSNR, com menos variabilidade.

V. CONCLUSÕES

Este trabalho identificou que, em ambos os modelos, com *SE-Block* e *Attention-Block*, os autoencoders apresentam um desempenho satisfatório na tarefa de reconstrução e remoção de ruído. Ao longo deste trabalho, foi possível explorar o uso dessa arquitetura para esta tarefa de remoção de ruído em sinais biomédicos e de áudio. A proposta central foi desenvolver um modelo eficiente, capaz de lidar com diferentes tipos de interferências simultaneamente.

Os experimentos realizados mostraram que aliar a arquitetura com blocos residuais e mecanismos de monitoramento garantiu adaptabilidade, levando o modelo a aprender, de forma dinâmica, a realçar as regiões mais confiáveis do sinal útil e suprimir as mais degradadas. A estrutura do autoencoder se mostrou pertinente para atenuar ruídos em representações temporais-frequenciais e evidencia sua possível aplicabilidade em sistemas de comunicação e processamento de áudio.

A tarefa de *denoising*, por sua natureza, exige que o modelo preserve informações relevantes do sinal original enquanto elimina o ruído. Os autoencoders se destacam nesse cenário justamente por sua capacidade de mapear o sinal corrompido. Sendo assim, trabalhos futuros podem explorar estratégias de

atenção temporal e espacial mais refinadas, talvez baseadas em *Transformers* [42], para aprimorar ainda mais a capacidade de adaptação do modelo a diferentes tipos e intensidades de ruído. Também seria interessante estudar a aplicação dessas abordagens em dispositivos embarcados, visando soluções em tempo real para monitoramento contínuo de sinais fisiológicos.

REFERÊNCIAS

- [1] V. A. Ardeti, V. R. Kolluru, G. T. Varghese, and R. K. Patjoshi, "An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to ai-based approaches," *Expert Systems with Applications*, vol. 217, p. 119561, 2023.
- [2] K. Kumar, R. Pandey, S. S. Bhattacharjee, and N. V. George, "Exponential hyperbolic cosine robust adaptive filters for audio signal processing," *IEEE Signal Processing Letters*, vol. 28, pp. 1410–1414, 2021.
- [3] Y. Jin, C. Qin, J. Liu, Y. Liu, Z. Li, and C. Liu, "A novel deep wavelet convolutional neural network for actual ecg signal denoising," *Biomedical Signal Processing and Control*, vol. 87, p. 105480, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809423009138>
- [4] W. H. Organization *et al.*, "Global status report on noncommunicable diseases 2010," 2012.
- [5] M. S. Thaler, *ECG essencial: eletrocardiograma na prática diária*. Artmed Editora, 2023.
- [6] D. Geselowitz, "On the theory of the electrocardiogram," *Proceedings of the IEEE*, vol. 77, no. 6, pp. 857–876, 1989.
- [7] J. van Alsté, W. van Eck, and O. Herrmann, "Ecg baseline wander reduction using linear phase filters," *Computers and Biomedical Research*, vol. 19, no. 5, pp. 417–427, 1986. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0010480986900376>
- [8] V. de Pinto, "Filters for the reduction of baseline wander and muscle artifact in the ECG," *Journal of Electrocardiology*, vol. 25, pp. 40–48, 1992.
- [9] Z. Wang, C. M. Wong, J. N. da Cruz, F. Wan, P.-I. Mak, P. U. Mak, and M. I. Vai, "Muscle and electrode motion artifacts reduction in ecg using adaptive fourier decomposition," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2014, pp. 1456–1461.
- [10] R. Prabhavalkar, T. Hori, T. N. Sainath, R. Schlüter, and S. Watanabe, "End-to-end speech recognition: A survey," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 325–351, 2024.
- [11] S. A. Shahriyar, M. A. H. Akhand, N. Siddique, and T. Shimamura, "Speech enhancement using convolutional denoising autoencoder," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1–5.
- [12] F. Owens and M. Murphy, "A short-time fourier transform," *Signal Processing*, vol. 14, no. 1, pp. 3–10, 1988. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0165168488900400>
- [13] R. Tao, Y.-L. Li, and Y. Wang, "Short-time fractional fourier transform and its applications," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2568–2580, 2010.
- [14] M. Parchami, W.-P. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time fourier transform domain," *IEEE Circuits and Systems Magazine*, vol. 16, no. 3, pp. 45–77, 2016.
- [15] E. Cordero, G. Giacchi, and L. Rodino, "A unified approach to time-frequency representations and generalized spectrograms," *Journal of Fourier Analysis and Applications*, vol. 31, no. 1, p. 9, Jan 2025. [Online]. Available: <https://doi.org/10.1007/s00041-024-10142-7>
- [16] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, and S.-Y. Chien, "Noise reduction in ecg signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60 806–60 813, 2019.
- [17] C. Yu, R. E. Zezario, S.-S. Wang, J. Sherman, Y.-Y. Hsieh, X. Lu, H.-M. Wang, and Y. Tsao, "Speech enhancement based on denoising autoencoder with multi-branched encoders," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2756–2769, 2020.
- [18] W.-H. Lee, M. Ozger, U. Challita, and K. W. Sung, "Noise learning-based denoising autoencoder," *IEEE Communications Letters*, vol. 25, no. 9, pp. 2983–2987, 2021.
- [19] J. W. Hurst, "Naming of the waves in the ecg, with a brief account of their genesis," *Circulation*, vol. 98, no. 18, pp. 1937–1942, 1998. [Online]. Available: <https://www.ahajournals.org/doi/abs/10.1161/01.CIR.98.18.1937>
- [20] L. S. Rnmo and P. Laguna, "Electrocardiogram (ecg) signal processing," in *Wiley Encyclopedia Biomed. Eng.*, 2006, pp. 1–16.
- [21] T. Hlaing, T. DiMino, P. R. Kowey, and G.-X. Yan, "Ecg repolarization waves: Their genesis and clinical implications," *Annals of Noninvasive Electrocardiology*, vol. 10, no. 2, pp. 211–223, 2005. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1542-474X.2005.05588.x>
- [22] D.-g. Fu, "Cardiac arrhythmias: Diagnosis, symptoms, and treatments," *Cell Biochemistry and Biophysics*, vol. 73, no. 2, pp. 291–296, Nov 2015. [Online]. Available: <https://doi.org/10.1007/s12013-015-0626-4>
- [23] C. R. Wyndham, "Atrial fibrillation: the most common arrhythmia," *Texas Heart Institute Journal*, vol. 27, no. 3, p. 257, 2000.
- [24] H. J. Wellens, "Ventricular tachycardia: diagnosis of broad qrs complex tachycardia," *Heart*, vol. 86, no. 5, pp. 579–585, 2001.
- [25] K. Umopathy, B. Ghoraani, and S. Krishnan, "Audio signal processing using time-frequency approaches: Coding, classification, fingerprinting, and watermarking," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 451695, Jun 2010. [Online]. Available: <https://doi.org/10.1155/2010/451695>
- [26] S. S. Stevens and J. Volkman, "The relation of pitch to frequency: A revised scale," *The American Journal of Psychology*, vol. 53, no. 3, pp. 329–353, 1940.
- [27] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.
- [28] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [29] S. J. Prince, *Understanding Deep Learning*. The MIT Press, 2023. [Online]. Available: <http://udlbook.com>
- [30] X. Ying, "An overview of overfitting and its solutions," in *Journal of physics: Conference series*, vol. 1168. IOP Publishing, 2019, p. 022022.
- [31] N. Ketkar, "Stochastic gradient descent," in *Deep learning with Python: A hands-on introduction*. Springer, 2017, pp. 113–132.
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [33] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [35] G. Brauwuers and F. Frasinca, "A general survey on attention mechanisms in deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3279–3298, 2023.
- [36] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [37] G. Moody and R. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.
- [38] C. Valentini-Botinhao, "Noisy speech database for training speech enhancement algorithms and its models, 2016 [sound]," 2017, dataset. [Online]. Available: <https://doi.org/10.7488/ds/2117>
- [39] J. Yamagishi, C. Veaux, and K. MacDonald, "Cstr vctk corpus: English multi-speaker corpus for cstr voice cloning toolkit (version 0.92)," Sound, 2019, <https://datashare.ed.ac.uk/handle/10283/3443>.
- [40] J. Thiemann, N. Ito, and E. Vincent, "The diverse environments multi-channel acoustic noise database (demand): A database of multichannel environmental noise recordings," in *Proceedings of Meetings on Acoustics*, vol. 19, no. 1. AIP Publishing, 2013.
- [41] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [42] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.