

Uma Metodologia Baseada em Modelo Transformer para Reconhecimento da Atividade de Varrer Independente de Pessoa Usando Dados Wi-Fi CSI

Allan Costa Nascimento dos Santos^{1,2}, Iandra Galdino¹, Julio C. H. Soto¹,
Taiane C. Ramos¹, Celio V. N. de Albuquerque¹, Raphael Guerra¹, Cledson de
Sousa¹, Natalia C. Fernandes¹, Débora Muchaluat-Saade¹, Gheorghita Ghinea²

¹Laboratório MídiaCom, Instituto de Computação
Universidade Federal Fluminense (UFF) – Niterói – RJ – Brasil

²Department of Computer Science, Brunel University London – UK

{allans, igar, jsoto, taiane, celio, cledson, debora}@midiacon.uff.br
rguerra@ic.uff.br, nataliacf@id.uff.br, george.ghinea@brunel.ac.uk

Abstract. *The use of channel state information (CSI) for human activity recognition holds promise in healthcare, especially for remote patient monitoring. By capturing and interpreting Wi-Fi signals in indoor environments, CSI can be used to detect physical activity, falls, or daily movements of a patient, allowing caregivers and healthcare professionals to monitor patients without the need for wearable sensors or invasive cameras. CSI also has great potential in elderly care. Therefore, this paper proposes a methodology called DVC-CSI to identify the floor sweeping activity of a person in a room through the analysis of CSI data and a dataset used for its evaluation. DVC-CSI uses Transformer models developed to process time series data featuring a structure that allows capturing temporal dependencies. DVC-CSI is capable of identifying activities of people who did not participate in the training phase. The accuracy of floor sweeping activity identification is 88.89% using a CSI dataset of 86 volunteers (60 participants for training, 17 participants for validation, and 9 participants for testing).*

Resumo. *O uso de informações de estado do canal para reconhecimento de atividades humanas é promissor na área da saúde, especialmente para monitoramento remoto de pacientes. Ao capturar e interpretar sinais Wi-Fi em ambientes internos, o CSI pode ser usado para detectar atividade física, quedas ou movimentos diários de um paciente, permitindo que cuidadores e profissionais de saúde monitorem pacientes sem a necessidade de sensores vestíveis ou câmeras invasivas. O CSI tem grande potencial no atendimento a idosos. Portanto, este artigo propõe uma metodologia chamada DVC-CSI para identificar a atividade de varrer de uma pessoa em uma sala por meio da análise de dados CSI e um conjunto de dados usado para sua avaliação. O DVC-CSI usa modelos Transformer desenvolvidos para processar dados de séries temporais com uma estrutura que permite capturar dependências temporais. O DVC-CSI é capaz de identificar a atividade de pessoas que não participaram da fase de treinamento do modelo. A acurácia da identificação da atividade de varrer dentro de um*

conjunto de dados de 17 atividades é de 88,89% usando um conjunto de dados CSI de 86 voluntários (60 participantes para treinamento, 17 participantes para validação e 9 participantes para teste).

1. Introdução

No contexto de aplicações em saúde, o CSI (*Channel State Information*) é um conjunto de dados que descreve como os sinais Wi-Fi se propagam dentro de um ambiente [Soto et al. 2022, Caballero et al. 2023]. O CSI captura informações detalhadas sobre a amplitude e a fase dos sinais Wi-Fi em várias subportadoras de frequência [Caballero et al. 2023]. Quando uma pessoa se move dentro de um espaço, ela causa mudanças sutis nos sinais Wi-Fi devido ao efeito Doppler e à atenuação do sinal. Analisando os padrões no CSI, é possível detectar essas variações, permitindo o monitoramento de movimentos corporais e atividades físicas sem a necessidade de dispositivos vestíveis ou câmeras.

Sendo assim, é possível monitorar a atividade física para avaliar o nível de independência de um idoso ou paciente, identificando atividades como caminhar, correr, sentar, deitar ou até mesmo tarefas domésticas, como varrer o chão. Esse método não é intrusivo e preserva a privacidade, pois usa apenas mudanças nos sinais Wi-Fi, sem recorrer à gravação de vídeo ou áudio. Em sistemas de monitoramento de idosos, detectar atividades como varrer o chão pode ajudar a garantir que a pessoa está ativa e realizando tarefas domésticas com segurança, além de avaliar o seu grau de independência, visto que essa é uma atividade que demanda tempo e esforço. Em monitoramento remoto de pacientes, essa detecção pode ajudar a acompanhar o nível de autonomia de um idoso, verificando se ele consegue realizar atividades diárias como varrer o chão. Em casas inteligentes, saber que alguém está varrendo o chão pode ativar funcionalidades automáticas, como pausar robôs aspiradores ou ajustar a iluminação. Em ambientes públicos ou comerciais, isso pode ajudar a analisar padrões de comportamento para fins de segurança ou otimização de espaço (por exemplo, verificando se funcionários estão realizando tarefas de limpeza em horários programados). Assistentes pessoais e dispositivos IoT podem adaptar suas interações com base nas atividades identificadas, como ajustar o volume de notificações quando a pessoa está ocupada varrendo [Santos et al. 2020, de Almeida et al. 2023, dos Santos et al. 2022].

Detectar a atividade de varrer o chão usando dados CSI e aprendizado de máquina apresenta vários desafios. Varrer envolve movimentos repetitivos, geralmente sutis, de braços e tronco que podem ser difíceis de distinguir de outras atividades [Caballero et al. 2023], tornando a distinção de características únicas dos dados CSI um desafio. Em conjuntos de dados de reconhecimento de atividade, a baixa frequência dessa atividade em comparação com atividades como andar ou sentar pode levar ao desequilíbrio de classes, o que pode enviesar o modelo para atividades mais comuns, reduzindo a precisão da detecção para varredura [Santos et al. 2020]. Além disso, as pessoas têm diferentes estilos de varredura, velocidades e mãos dominantes. Um modelo treinado em um grupo de usuários pode não generalizar bem para outros, sem diversidade suficiente nos dados de treinamento.

Este artigo propõe uma metodologia baseada em Transformer [Rothman 2021] para reconhecimento da atividade humana de varrer o chão utilizando dados de

informação do estado do canal (CSI - *Channel State Information*) em redes Wi-Fi, chamada DVC-CSI- Detecção de Varrer o Chão usando CSI. No cenário experimental, foram realizadas coletas de dados CSI de 86 voluntários (60 participantes para treinamento, 17 participantes para validação e 9 participantes para teste) realizando diversas atividades em uma sala, incluindo a atividade de varrer e os dados CSI da sala vazia. O DVC-CSI foi treinado, validado e testado utilizando dados CSI de voluntários varrendo de diferentes formas, de diferentes gêneros, idades e características físicas, coletados em um ambiente controlado [Galdino et al. 2023], produzindo resultados muito promissores em um conjunto de dados com ampla variedade. A acurácia da identificação da atividade de varrer dentro de um conjunto de dados de 17 atividades é de 88,89%. O conjunto de dados coletado inclui 17 posturas e atividades diárias diferentes. O modelo Transformer binário desenvolvido é capaz de identificar se a pessoa está varrendo ou não, dentre todas as outras atividades como correr, sentar, deitar, etc. O modelo Transformer desenvolvido é robusto o suficiente para extrair características únicas e reconhecer quando a pessoa está varrendo o chão. O modelo analisa o sinal Wi-Fi transmitido por pontos de acesso sem fio de prateleira, frequentemente presentes em ambientes indoor, onde as atividades diárias do indivíduo e seus movimentos resultam em alterações no sinal coletado. Por meio da captura e interpretação dessas alterações nos dados CSI, o DVC-CSI é capaz de detectar a atividade física.

O DVC-CSI utiliza um modelo Transformer [Rothman 2021] desenvolvido para processar séries temporais, com codificação posicional, atenção multi-cabeça e uma camada totalmente conectada, proporcionando aprendizado sofisticado de relações temporais e gerando previsões refinadas com base nos padrões encontrados na sequência de entrada. Diferentemente de outros modelos baseados em aprendizado de máquina na literatura, o DVC-CSI possui um modelo capaz de identificar a atividade de pessoas que não participaram da fase de treinamento.

O restante do artigo está organizado da seguinte maneira. A Seção 2 descreve trabalhos relacionados, enquanto a Seção 3 apresenta a metodologia proposta. A descrição dos experimentos, resultados e discussões é fornecida na Seção 4. Finalmente, as conclusões e sugestões para trabalhos futuros são apresentadas na Seção 5.

2. Trabalhos Relacionados

O monitoramento de pessoas portadoras de deficiências durante atividades domésticas básicas, como dormir e sentar, ganhou interesse no desenvolvimento de tecnologias de inteligência ambiental [Wang et al. 2017]. O reconhecimento de atividade humana pode ser realizado usando sistemas baseados em visão computacional, sensores vestíveis e/ou sinais de radiofrequência [Santos et al. 2020, de Almeida et al. 2023]. No entanto, os sistemas baseados em visão computacional exigem uma linha de visão direta e levantam preocupações de privacidade, enquanto os sensores vestíveis podem ser desconfortáveis para idosos [Galdino et al. 2023, dos Santos et al. 2024].

Caballero *et al.* (2023) propuseram um sistema baseado em CSI para identificar a atividade de varrer o chão. O estudo analisa um grande volume de dados e compara cinco algoritmos de aprendizado de máquina para avaliar o desempenho no reconhecimento de atividade humana. Uma classificação multiclasse foi realizada para identificar seis diferentes atividades humanas: ficar em pé, sentar, deitar, andar, correr e varrer.

Seus resultados mostram que o algoritmo *Random Forest* obteve o maior desempenho de classificação, atingindo uma acurácia geral de 86,65% e uma precisão variando de 85,68% a 88,35% em todas as classes. Dentre estas, as classes correr e varrer obtiveram a melhor precisão, com 87,45% e 88,35%, respectivamente. No entanto, um modelo RF deve ser treinado individualmente para cada pessoa para detectar sua posição. Portanto, o modelo proposto é específico de cada indivíduo, diferindo da solução proposta deste trabalho.

Galdino *et al.* (2023) descrevem a estrutura geral da detecção de presença humana baseada em Wi-Fi e apresentam as ferramentas de processamento de sinal usadas para avaliar o desempenho do sistema. O desempenho do conjunto de dados apresentado na detecção de presença humana para exemplificar sua aplicabilidade. O trabalho fornece um conjunto de dados específico para o desenvolvimento de novas aplicações de CSI, além de uma descrição detalhada do experimento, equipamentos e protocolos. O conjunto de dados CSI inclui coletas da atividade de varrer o chão, mas não foi utilizado no modelo proposto no artigo para detecção de presença humana.



Figura 1. Gesto de varrer conforme Zheng *et al.* [Zheng *et al.* 2019].

Os trabalhos de Isack Bulugu (2023) e Zheng *et al.* (2019) propõem um método de reconhecimento de gestos. Os métodos propostos são capazes de analisar os dados CSI para identificar o movimento da mão e do braço conforme a Figura 1. Esse movimento foi caracterizado como o movimento de varrer. Embora os trabalhos de Isack Bulugu (2023) e Zheng *et al.* (2019) possam identificar esse gesto de varrer, nem todas as pessoas varrem o chão dessa maneira. Assim, os métodos não foram avaliados em situações reais de uma pessoa varrendo o chão de diferentes formas e se movimentando pelo ambiente. Além disso, os trabalhos não levam em consideração a movimentação do corpo, mas apenas das mãos, limitando o reconhecimento da atividade de varrer em situações reais. O trabalho de Isack Bulugu (2023) propõe um método de reconhecimento de gestos entre domínios usando redes neurais convolucionais 3D (3DCNN). O reconhecimento de gestos é amplamente usado na interação humano-computador, incluindo casas inteligentes e realidade virtual, mas métodos tradicionais que dependem de câmeras, sensores vestíveis ou equipamentos especializados enfrentam limitações, como preocupações com privacidade e alcance de detecção restrito. Os resultados do artigo mostram que no reconhecimento de cenas conhecidas, a taxa de reconhecimento de varrer é 85% e que no reconhecimento de cenas desconhecidas, a taxa de reconhecimento de varrer é 86%.

Zheng *et al.* (2019) propuseram um sistema de reconhecimento de gestos baseado em Wi-Fi projetado para atingir alta acurácia, que introduz um novo recurso independente de domínio chamado perfil de velocidade de coordenada corporal (BVP) para capturar padrões de movimento humano em presença de um nível de sinal baixo. A principal contribuição do Wistar3.0 é estimar perfis de velocidade de gestos no nível de sinal mais baixo, que representam características cinéticas únicas de gestos e são independentes de domínios. Contudo, o modelo funciona apenas em linha de visada direta. O sistema depende da existência de caminhos de propagação indo diretamente do transmissor Wi-Fi

para o humano e do humano para os receptores Wi-Fi, ou seja, o LOS (Line of Sight) entre dispositivos Wi-Fi e humanos. Isso ocorre porque o sistema requer que os caminhos LOS estabeleçam geometricamente a relação entre a propagação do sinal e a velocidade de movimento dos refletores. Embora a condição LOS seja comum ao realizar a interação com dispositivos inteligentes em casas inteligentes, o sistema não funcionará quando o LOS não for garantido em alguns outros cenários. Para obter um Espectro Doppler distinto para estimativa de BVP, o sistema requer que os usuários fiquem dentro de uma zona predefinida com alguma flexibilidade em locais e orientações. Embora o sistema tenha potencial para aplicações mais específicas, ainda existe uma lacuna entre o cenário avaliação proposto e a aplicabilidade no mundo real.

Xiao *et al.* (2019) propuseram uma rede generativa adversarial semi-supervisionada (GAN) para o reconhecimento de atividades baseado em dados CSI. Eles desenvolveram um modelo para abordar a degradação do desempenho da validação *leave-one-subject-out* para o reconhecimento de atividades baseado em CSI. Os autores avaliaram eficácia do CsiGAN em dois conjuntos de dados em cenários semi-supervisionados e supervisionados. Três componentes são propostos e integrados ao CsiGAN para lidar com a escassez de dados não rotulados e aprimorar o desempenho de reconhecimento de atividades humanas. Para dados SignFi, a acurácia da CNN é de 77%, o que é quase o mesmo que o resultado da validação *leave-one-subject-out* usando CNN, enquanto a acurácia do CsiGAN atinge mais de 84%, o que é cerca de 9% maior do que a da CNN. Da mesma forma, para dados FallDeFi, a precisão do CsiGAN, 86%, também é significativamente maior do que a do SVM, 80%. Entretanto, dados de apenas três voluntários foram utilizados para treinar o modelo e, ainda, foi utilizado um conjunto limitado de atividades, resultando em baixa precisão.

A Tabela 1 apresenta uma comparação do DVC-CSI com outros estudos que empregam técnicas de aprendizado de máquina para monitorar atividades humanas por meio de dados CSI. O DVC-CSI se destaca por ter envolvido um número significativamente maior de participantes (86), em contraste com a maioria dos estudos comparados. Esse fator é crucial, pois uma maior diversidade na amostragem de dados contribui para a robustez e generalização do modelo. A proposta se diferencia pela abrangência e variedade das atividades monitoradas. Enquanto os estudos comparados geralmente incluem entre 5 e 6 atividades, o DVC-CSI cobre 17 atividades distintas, além de considerar uma condição de ambiente vazio, sendo capaz de distinguir a atividade de varrer o chão dentre todas as outras atividades. Essa diversidade permite ao modelo discriminar dentro de uma ampla gama de comportamentos e situações. Os estudos relacionados empregam uma variedade de métodos de aprendizado de máquina. A proposta deste artigo utiliza modelo Transformer, reconhecido por sua eficácia no aprendizado de padrões temporais complexos. Essa característica pode ser particularmente vantajosa para capturar as sutilezas das atividades humanas, em comparação com métodos tradicionais como GAN e Random Forest, utilizados em trabalhos anteriores. O DVC-CSI é único por utilizar exclusivamente dados CSI de coletas reais, ou seja, não são dados gerados por IA, independentemente das características individuais dos participantes varrendo o chão. Essa abordagem apresenta vantagens significativas para aplicações práticas, eliminando a necessidade de calibração do modelo para cada usuário e, consequentemente, ampliando sua aplicabilidade em cenários reais. O DVC-CSI alcança uma acurácia de 88,89%, demonstrando competitividade em relação a outros métodos. É importante destacar que essa elevada acurácia foi obtida sem

a utilização de dados personalizados para cada indivíduo. Essa característica, combinada com a alta precisão, evidencia a eficiência e adaptabilidade do modelo proposto. Ademais, diferentemente de outros trabalhos correlatos, o DVC-CSI é capaz de detectar a atividade da pessoa em tempo real, demandando aproximadamente 0.56s de processamento.

Tabela 1. Tabela comparativa de trabalhos relacionados.

Ref.	Partic.	Atividades	Método ML	Dados reais coletados em tempo real com detecção independente da pessoa enquanto varre e se move	Resultado
[Xiao et al. 2019]	3	Cair, andar, pular, abrir portas, levantar mãos, sentar	GAN	Não	86.27% acurácia (não inclui varrer)
[Bulugu 2023]	16	Empurrar, puxar, varrer, bater palmas, deslizar, desenhar círculos	3DCNN	Não	86% reconhecimento de varrer
[Zheng et al. 2019]	16	Empurrar, puxar, varrer, bater palmas, deslizar, desenhar círculos	BVP	Não	80% acurácia de varrer
[Caballero et al. 2023]	125	Sentado, em pé, deitado, caminhando, correndo, varrendo	<i>Random Forest</i>	Não	88,35% precisão para varrer
DVC-CSI	86	17 posturas distintas e sala vazia	Transformer	Sim	88, 89% precisão para varrer

3. Metodologia

O DVC-CSI¹ implementa uma sequência de testes com base em diferentes hiperparâmetros. Cada combinação é avaliada no conjunto de validação de acordo com as métricas de precisão, *recall*, acurácia e F1-score e o modelo final é avaliado no conjunto de dados de teste. DVC-CSI explora a otimização de um modelo *Transformer* para classificação binária de séries temporais de dados CSI, avaliando diferentes combinações de hiperparâmetros. O objetivo é identificar as configurações que maximizam o desempenho em termos de precisão, *recall* e F1. O modelo permite explorar como o número variável de cabeças, número de camadas, taxa de aprendizado, número de épocas e tamanho do lote afetam as métricas de desempenho. Após isso, o DVC-CSI pode identificar a atividade humana de varrer em ambientes internos, analisando dados CSI mesmo de uma pessoa nunca vista pelo modelo durante o treino. As seguintes combinações de hiperparâmetros foram testadas: número de cabeças: [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16], número de camadas: [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12], taxa de aprendizagem (lr): [0,00001, 0,0001, 0,001, 0,01], número de épocas: [20, 40, 60, 100], tamanho do lote: [4, 8, 12, 16, 20, 24]. A Tabela 2 mostra as combinações de hiperparâmetros do modelo final.

Tabela 2. As combinações de parâmetros do modelo final.

Modelo	Heads	Camadas	L. Rate	Épocas	Batch Size
Reconhecer Varrendo	6	4	0.0001	40	20

O DVC-CSI envolve o desenvolvimento de um modelo de classificação binário, cujo fluxo geral de processamento é ilustrado na Figura 2. O bloco ‘Coleta de dados CSI’

¹Os detalhes de implementação, os códigos e arquivos estão disponíveis no GitHub: <https://github.com/mestrelan/gloss-blue>.

ilustra o processo de coleta e processamento dos dados CSI. O bloco ‘Estrutura do modelo Transformer’ ilustra o modelo com sua codificação posicional, camadas do *Encoder*, agregação temporal e a camada totalmente conectada. O bloco ‘Fluxo de treinamento e validação’ ilustra a inicialização do treinamento, o treinamento e a validação do modelo. O bloco ‘Teste’ ilustra a avaliação final do modelo no conjunto de dados de teste.

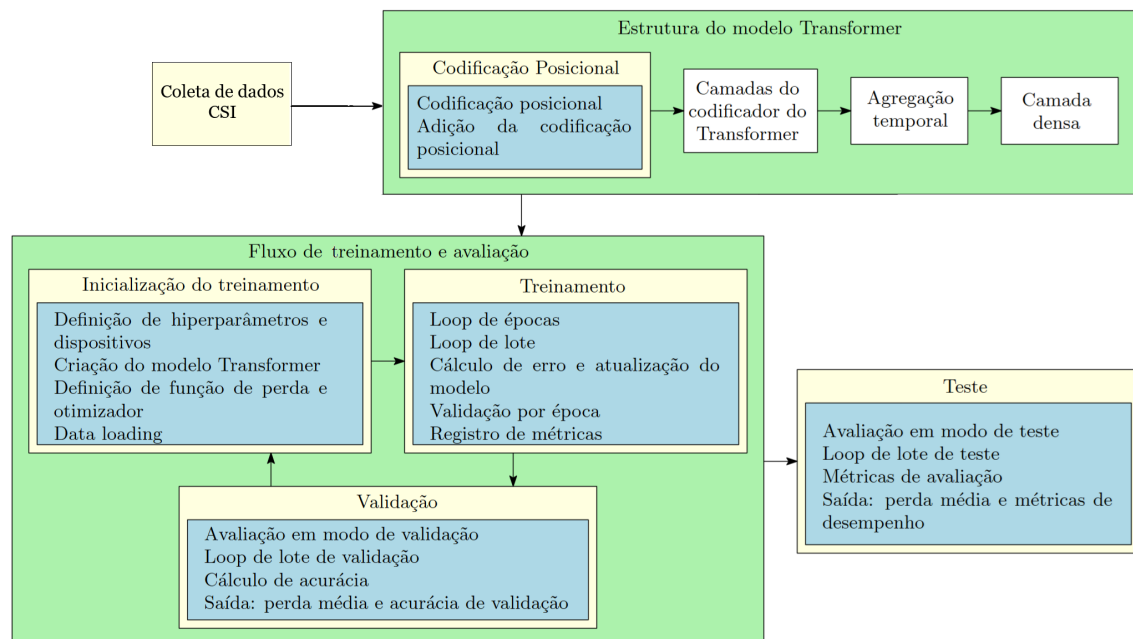


Figura 2. Diagrama de blocos do DVC-CSI.

Dados de 86 participantes, incluindo homens e mulheres, foram utilizados nos experimentos realizados. Desses, 60 participantes (75%) foram utilizados para treinamento do modelo, 17 participantes (15%) para validação e 9 participantes (10%) para teste. Os dados foram coletados usando um Raspberry Pi B4 equipado com um *chip-set* bcm43455c0. Os dispositivos utilizados nos experimentos foram posicionados a aproximadamente um metro dos participantes. Além disso, os participantes não tiveram restrições quanto ao uso de roupas ou dispositivos eletrônicos, e o ambiente experimental foi projetado para simular um ambiente doméstico.

3.1. Base de Dados

Os dados CSI² usados nesta pesquisa foram cedidos pelos autores do trabalho [Galdino et al. 2023]. Foram coletados de uma rede Wi-Fi operando a 5 GHz com uma largura de banda de 80 MHz. Para garantir a qualidade ideal dos dados, foi realizada uma configuração meticulosa da montagem experimental, incluindo análise espectral para identificar um canal desocupado dentro da banda ISM de 5 GHz. Os dados foram coletados usando um Raspberry Pi equipado com NEXMON [Galdino et al. 2023] capturando 33–34 amostras por segundo durante 60 segundos, resultando em 2000 amostras em cada uma das 256 subportadoras. O conjunto de dados abrange 17 posturas distintas mais uma coleção de sala vazia, abrangendo uma gama mais ampla de comportamentos humanos em comparação com estudos anteriores.

²Projeto aprovado pelo Comitê de Ética sob número CAAE 54359221.4.0000.5243.

3.2. Processamento de dados CSI

A análise apresentada por Sousa *et al.* [de Sousa et al. 2024] mostrou que as subportadoras abaixo do índice 60 exibiram amplitudes significativamente maiores do que aquelas acima. Portanto, foram utilizadas as primeiras 60 subportadoras, resultando em uma matriz complexa de dimensões 2000×60 para cada uma das coletas das 17 posições de cada participante. Em seguida, foram removidos os dados de 12 subportadoras nulas e piloto dentre as 60, que não carregam dados significativos, resultando em 48 subportadoras. Dessa forma, a matriz complexa de dados final utilizada foi de dimensões 2000×48 . Posteriormente, foram calculadas as amplitudes dos sinais coletados em cada uma das subportadoras. Esse cálculo foi realizado considerando os componentes dos números complexos obtidos a partir dos sinais coletados. Como um número complexo é definido por suas partes real e imaginária (amplitude e fase), foi determinada a amplitude com base no módulo do número complexo. Assim, foram obtidas amplitudes para todas as 2000 amostras ao longo das 48 subportadoras resultantes para cada coleta de dados. Assim, os dados CSI são processados em uma série temporal de um minuto de duração (tempo de execução de uma atividade do conjunto de dados), que será usada como entrada para o modelo *Transformer*.

3.3. Geração & Treinamento do Modelo

Foi desenvolvido um modelo independente que recebe como entrada os dados CSI coletados e processados anteriormente para detectar e classificar a atividade humana de varrer em um ambiente *indoor*. O modelo gerado utiliza todas as atividades do conjunto de dados e compara com a atividade específica de varrer para concluir se a pessoa está varrendo ou não. Esse modelo realiza uma classificação binária, produzindo apenas dois resultados possíveis. Foi empregado o algoritmo *Transformer* com várias configurações de hiperparâmetros para o modelo proposto. A especificação de hiperparâmetros e dispositivos é essencial para a proposta do modelo. Como mencionado anteriormente, tais hiperparâmetros incluem o número de cabeças, camadas e taxa de aprendizado, juntamente com a escolha de CPU ou GPU, conforme apresentado na Tabela 2. O modelo *Transformer* é instanciado e colocado no dispositivo designado.

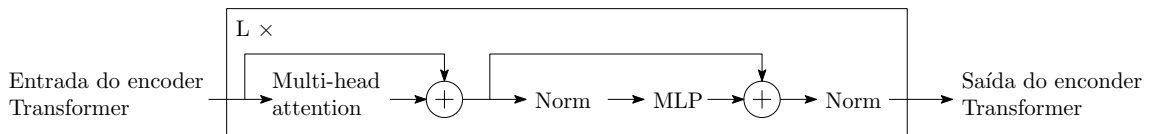


Figura 3. Camadas do encoder transformer.

A Figura 3 apresenta as camadas do Encoder: Os dados entram com o formato $[\text{batch_size}, \text{seq_len}, \text{num_channels}]$, onde: batch_size = Número de amostras processadas simultaneamente (20). seq_len = Número de amostras no tempo (2000 amostras). num_channels = Número de variáveis medidas em cada instante (48 subportadoras).

Passagem pelos num_layers Blocos do Encoder: O encoder consiste em várias camadas idênticas empilhadas (conforme num_layers), onde cada camada aplica os seguintes passos: Mecanismo de Autoatenção Multi-cabeças (Multi-Head Self Attention) [Rothman 2021]: Permite que o modelo aprenda relações entre diferentes instantes da sequência. Cada cabeça de atenção (definido por num_heads) processa uma parte diferente da informação.

Ainda na Figura 3, o bloco ‘Multi-head attention’ realiza o cálculo dos Tensores Q, K e V: Q (Query) que representa os elementos consultando outros instantes da sequência. K (Key): representa os elementos que podem fornecer informações relevantes. V (Value): Contém os valores reais das características a serem processadas.

Cálculo da Atenção, conforme a Equação 1: A similaridade entre Q e K é calculada para identificar quais instantes no tempo são mais relevantes para cada outro instante. A saída é uma combinação ponderada de V, onde instantes mais importantes recebem pesos maiores. $\sqrt{d_k}$ é a dimensão das chaves. Cada cabeça de atenção (head) é calculada usando a fórmula de Atenção Escalonada [Rothman 2021] e depois concatenada e transformada por um peso (W^O).

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (1)$$

Resumindo as etapas do processo, a entrada dos dados com codificação posicional, seguida por Autoatenção multi-cabeças para encontrar dependências temporais, então, Normalização e conexões residuais para estabilização, para então a Rede Feed-Forward para capturar padrões mais complexos, mais normalização e conexão residual e repetição do processo num layers vezes. Essa estrutura permite que o Transformer aprenda padrões globais e locais em séries temporais, sem a limitação da recorrência (como em RNNs) com a saída refinada com relações temporais destacadas [Rothman 2021].

3.4. Validação & teste dos modelos

O conjunto de dados foi dividido em conjuntos de treinamento 75%, utilizado na Seção 3.3, validação, 15%, e teste, 10%. O modelo itera sobre lotes de validação, processando dados de entrada e rótulos de verdade fundamental, e acumula métricas de perda e acurácia. O teste avalia o desempenho do modelo treinado em dados nunca vistos com a configuração dos hiperparâmetros do modelo que alcançou a maior acurácia e a menor perda na etapa de validação. Aqui, calculam-se várias métricas, incluindo perda, acurácia, *recall*, precisão, F1-score e uma matriz de confusão.

4. Resultados e Discussão

A Figura 4 reflete um conjunto de teste balanceado, composto por 9 voluntários considerando atividades de varrer e outras atividades. Portanto a base de testes tem 18 instâncias. Devido ao equilíbrio dos dados, a tarefa de detecção de varrer apresenta matriz de confusão com 9 casos de teste, representando um teste por voluntário. Isso ocorre porque cada voluntário contribui com uma amostra específica para essas atividades, refletindo um equilíbrio controlado entre as classes.

A Tabela 3 apresenta os resultados gerais obtidos na identificação da atividade de varrer dos voluntários da base de teste.

Tabela 3. Resultados gerais dos testes.

Modelo	Voluntários	Acurácia	Precisão	Recall	F1-Score	Test Loss
Reconhecer Varrendo	86	88.89%	88.89%	88.89%	88.89	0.38589

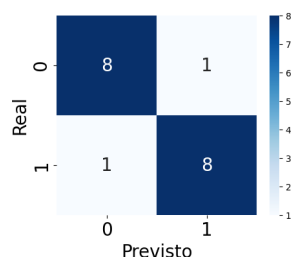


Figura 4. Matriz de confusão.

A sensibilidade, ou *recall*, que mede a taxa de verdadeiros positivos, foi de 88.89%, com um F1-Score correspondente de 88.89. Esses resultados indicam que o modelo é capaz de detectar com precisão a atividade de varrer em um ambiente fechado. Essa tecnologia tem o potencial de ser aplicada a uma variedade de aplicações cotidianas não críticas.

As colunas **Heads** e **Layers** na Tabela 3 especificam o número de cabeças de atenção e camadas do modelo Transformer para cada tarefa. Em geral, tarefas mais complexas utilizam mais cabeças (6) e camadas intermediárias (4), indicando a necessidade de maior capacidade de atenção para capturar nuances nas posturas. A taxa de aprendizado é ajustada para cada tarefa. Tarefas mais difíceis possuem uma taxa de aprendizado mais baixa, o que permite que o modelo se ajuste gradualmente e evite o overfitting. Para tarefas mais simples uma taxa mais alta é suficiente. Tarefas mais complexas exigem mais épocas para alcançar bons resultados, permitindo que o modelo aprenda padrões sutis. Os tamanhos dos lotes (*batch size*) são ajustados. Atividades mais simples, utilizam um lote maior, enquanto atividades mais complexas exigem lotes menores para otimizar a capacidade de aprendizado e o uso da memória.

A Figura 5 apresenta as perdas de treinamento ao longo das épocas. A perda é uma métrica utilizada para avaliar o desempenho do modelo durante o treinamento, sendo valores menores indicativos de melhor desempenho. O valor inicial da perda de treinamento, aproximadamente 0,8288, diminui para aproximadamente 0,6442. Isso é um bom sinal, pois indica que o modelo está aprendendo e melhorando ao longo do tempo.

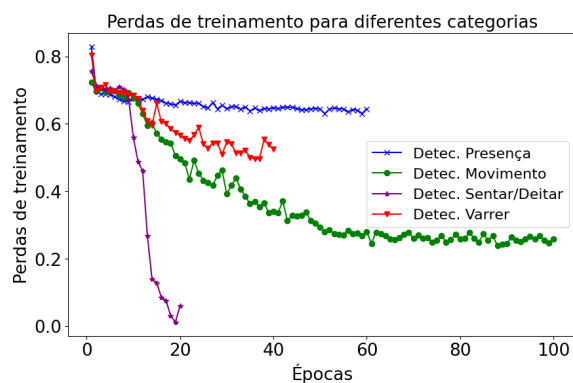


Figura 5. Perdas de treinamento para diferentes categorias do DVC–CSI.

Os valores de Perda de Teste variam significativamente entre as atividades, com

a detecção de movimento apresentando a menor perda e a detecção de sentado/deitado a maior. Isso sugere que o modelo é mais eficiente em atividades mais simples ou com padrões claros, enquanto tarefas com maior variabilidade ou maior complexidade postural, como sentado/deitado e caminhar/correr, exigem ajustes para reduzir a Perda de Teste e aumentar a generalização.

5. Conclusões

As características do sinal Wi-Fi podem ser afetadas pelo ambiente, com suas características de sinal influenciadas por vários fatores ambientais, incluindo presença e movimento humanos. Essas variações podem ser detectadas em dados de CSI, que fornecem detalhes da camada física (PHY), como amplitude e fase. Este artigo propôs o DVC-CSI para identificar a atividade de uma pessoa de varrer em um cômodo através da análise de dados CSI. Considerando um conjunto de dados de 86 voluntários utilizados para sua avaliação, foi atingida uma acurácia de 88.89% para identificar a atividade em um conjunto de dados treinado com diferentes pessoas. A detecção de movimento apresentou um aprendizado rápido e estável, atingindo valores desejáveis de perda e acurácia mesmo sem a necessidade de muitas épocas de treinamento. A proposta do DVC-CSI se destaca pelo número de participantes envolvidos nos experimentos, pela variedade de atividades monitoradas, pelo uso de modelos Transformer e pela sua independência do usuário. Essas qualidades tornam o DVC-CSI mais robusto e adequado para aplicações em cenários reais de monitoramento de atividade humana. Esses resultados são promissores e espera-se que sejam usados para monitorar idosos em suas atividades diárias. Essa abordagem inovadora possibilita a reaproveitamento da infraestrutura já existente, o que gera uma redução de custos em relação ao emprego de dispositivos específicos. Além disso, o DVC-CSI proporciona um monitoramento de dados CSI não invasivo e confortável, eliminando a necessidade de contato físico ou do uso de dispositivos conectados ao corpo.

Como trabalhos futuros, pretende-se aprimorar o desempenho do DVC-CSI em tarefas desafiadoras (caminhar/correr e sentar/deitar) refinando o pré-processamento de dados e explorando modelos híbridos que combinam Transformers com outras abordagens de aprendizado. Análise do desempenho do modelo em ambientes com níveis variados de interferência (por exemplo, vários dispositivos Wi-Fi ou paredes), o que pode afetar a confiabilidade do modelo em situações cotidianas. Sendo assim, avaliar possíveis estratégias de mitigação da interferência.

Agradecimentos

Este trabalho foi parcialmente apoiado pelo CNPq, CAPES, FAPERJ, INCT-MACC, INCT-ICONIoT e FINEP.

Referências

- Bulugu, I. (2023). Gesture recognition system based on cross-domain csi extracted from wi-fi devices combined with the 3d cnn. *Signal, Image and Video Processing*, 17(6):3201–3209.
- Caballero, E., Galdino, I., Soto, J. C., Ramos, T. C., Guerra, R., Muchaluat-Saade, D., and Albuquerque, C. (2023). Human Activity Recognition Using Wi-Fi CSI. In *International Conference on Pervasive Computing Technologies for Healthcare*, pages 309–321. Springer.

- de Almeida, G. C., dos Santos, A. C., Soares, C. L. d. A., Pinto, P. C. A., Dal Bello, F. d. S., Boechat, Y. E. M., Seixas, F. L., dos Santos, A. A. S., Mesquita, C. T., Mesquita, E. T., et al. (2023). Nova geração da telessaúde: Oportunidades, tendências e desafios. In *Minicursos do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*.
- de Sousa, C., Fernandes, V., Coimbra, E. A., and Huguenin, L. (2024). Subcarrier Selection for HAR Using CSI and CNN: Reducing Complexity and Enhancing Accuracy. In *2024 IEEE Virtual Conference on Communications (VCC)*, pages 1–7.
- dos Santos, A. C. N., de Paula, K., Vidal, M. T. L., da Silva, J. M. M., de Sousa, C., Fernandes, L. A. F., de Castro, T. B., Bedo, M., Kohwalter, T. C., Bastos, C. A. M., Seixas, F. L., Fernandes, N. C., Muchaluat-Saade, D. C., and Ghinea, G. (2024). A computer vision model to support individuals with disabilities within university campuses. In *2024 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, pages 1–7.
- dos Santos, A. C. N., Seixas, F. L., and Fernandes, N. C. (2022). Provendo um modelo automático de detecção de quedas baseado em rede adversária generativa para assistência de idosos. In *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 120–131. SBC.
- Galdino, I., Soto, J. C. H., Caballero, E., Ferreira, V., Ramos, T. C., Albuquerque, C., and Muchaluat-Saade, D. C. (2023). eHealth CSI: A Wi-Fi CSI Dataset of Human Activities. *IEEE Access*, 11:71003–71012.
- Rothman, D. (2021). *Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more*. Packt Publishing Ltd.
- Santos, A. C., Firmino, R. M., Soto, J. C., Medeiros, D. S., Mattos, D. M., Albuquerque, C. V., Seixas, F., Muchaluat-Saade, D. C., and Fernandes, N. C. (2020). Aplicações em redes de sensores na área da saúde e gerenciamento de dados médicos: tecnologias em ascensão. In *Minicursos do XX Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 137–186.
- Soto, J. C., Galdino, I., Caballero, E., Ferreira, V., Muchaluat-Saade, D., and Albuquerque, C. (2022). A survey on vital signs monitoring based on wi-fi csi data. *Computer Communications*, 195:99–110.
- Wang, Y., Wu, K., and Ni, L. M. (2017). Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594.
- Xiao, C., Han, D., Ma, Y., and Qin, Z. (2019). Csigan: Robust channel state information-based activity recognition with gans. *IEEE Internet of Things Journal*, 6(6):10191–10204.
- Zheng, Y., Zhang, Y., Qian, K., Zhang, G., Liu, Y., Wu, C., and Yang, Z. (2019). Zero-effort cross-domain gesture recognition with wi-fi. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services*, pages 313–325.