Uma Metodologia Baseada em Modelos Transformer para Reconhecimento de Atividades Humanas Independente de Pessoa Usando Dados Wi-Fi CSI

Allan Costa Nascimento dos Santos [1,2], Iandra Galdino[1], Julio C. H. Soto[1], Taiane C. Ramos[1], Celio V. N. de Albuquerque[1], Raphael Guerra[1], Cledson de Sousa[1], Natalia C. Fernandes[1], Débora Muchaluat-Saade[1], Gheorghita Ghinea[2]

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

[2]Department of Computer Science, Brunel University London – UK

{allans, igar, jsoto, taiane, celio, cledson, debora}@midiacom.uff.br

rguerra@ic.uff.br, nataliacf@id.uff.br, george.ghinea@brunel.ac.uk

Abstract. *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. Therefore, this paper proposes a methodology called* `MPA-CSI` *to identify the activity of a person in a room through the analysis of CSI data and a dataset used for its evaluation.* `MPA-CSI` *uses Transformer models developed to process time series data featuring a structure that allows capturing temporal dependencies.* `MPA-CSI` *is capable of identifying activities of people who did not participate in the training phase. The movement identification accuracy is 96.67% using a dataset with CSI data from 59 volunteers.*

Resumo. *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. Portanto, este artigo propõe uma metodologia chamada* `MPA-CSI` *para identificar a atividade de uma pessoa em uma sala por meio da análise de dados CSI e um conjunto de dados usado para sua avaliação. O* `MPA-CSI` *usa modelos Transformer desenvolvidos para processar dados de séries temporais. O* `MPA-CSI` *é capaz de identificar a atividade de pessoas que não participaram da fase de treinamento do modelo. A acurácia da identificação de movimento é de 96,67% usando um conjunto de dados CSI de 59 voluntários.*

1. Introdução

A análise dos padrões de atividade física em idosos, pacientes, pessoas com deficiência, doenças crônicas ou em tratamento assistido é cada vez mais importante devido ao seu impacto direto na área de assistência médica e na qualidade de vida e bem-estar dessas populações [Caballero et al. 2023, Chen et al. 2019]. O monitoramento da atividade física ajuda a identificar comportamentos sedentários e promover a adoção de hábitos

mais ativos [Soto et al. 2022a, Caballero et al. 2023]. Essa abordagem contribui para a prevenção de doenças crônicas como diabetes tipo 2, doenças cardíacas, obesidade e osteoporose, além de melhorar a saúde mental e as funções físicas em geral [Soto et al. 2022b, Gouveia et al. 2024]. Para pessoas com deficiência ou idosos em recuperação de lesões ou cirurgias, o monitoramento contínuo da atividade física pode fornecer informações valiosas sobre o progresso da reabilitação. Isso permite que os profissionais de saúde adaptem os planos de tratamento e intervenções com base em dados precisos sobre a mobilidade individual [Santos et al. 2020, dos Santos et al. 2024, dos Santos et al. 2022].

O custo de dispositivos de alta qualidade e infraestrutura de monitoramento remoto geralmente constitui um obstáculo à adoção generalizada, especialmente entre populações de baixa renda. Além disso, a coleta de dados sobre atividade física em ambientes não controlados pode resultar em dados ruidosos ou incompletos, dificultando a interpretação precisa dos padrões [Soto et al. 2022a, Caballero et al. 2023]. A análise precisa requer técnicas robustas de filtragem de dados e um profundo entendimento dos contextos de movimento.

A tecnologia CSI (Channel State Information) possibilita o monitoramento da atividade física através da análise das informações do estado do canal das ondas eletromagnéticas de uma rede Wi-Fi. Essa abordagem inovadora permite a reutilização da infraestrutura existente, reduzindo custos em comparação com o uso de dispositivos dedicados. Adicionalmente, o CSI oferece um monitoramento não invasivo e confortável, dispensando a necessidade de contato físico ou o uso de dispositivos acoplados ao corpo humano. Este artigo propõe uma metodologia baseada em modelos Transformer para reconhecimento de atividades humanas utilizando dados de informação do estado do canal (CSI - *Channel State Information*) em redes Wi-Fi, chamada `MPA-CSI`- *Monitoring Physical Activity using Channel State Information*. O `MPA-CSI` objetiva superar o desafio da adesão à tecnologia por parte da população idosa, uma vez que se trata de um modelo que não demanda o uso de dispositivos vestíveis ou qualquer tipo de ação por parte do indivíduo. O modelo analisa o sinal Wi-Fi transmitido por pontos de acesso sem fio de prateleira, frequentemente presente em ambientes indoor. As atividades diárias do indivíduo e seus movimentos resultam em alterações no sinal coletado. Por meio da captura e interpretação de dados CSI, `MPA-CSI` é capaz de detectar a atividade física ou movimentos diários do paciente, possibilitando que cuidadores e profissionais de saúde monitorem sua condição sem a necessidade de sensores vestíveis ou câmeras invasivas [Gouveia et al. 2024].

O `MPA-CSI` utiliza modelos Transformer [Rothman 2021] desenvolvidos 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. Os modelos de `MPA-CSI` foram treinados, validados e testados utilizando dados CSI de 59 voluntários de diferentes gêneros, idades e características físicas coletados em um ambiente controlado [Galdino et al. 2023], produzindo resultados muito promissores. Diferentemente de outros modelos baseados em aprendizado de máquina na literatura, o `MPA-CSI` possui modelos capazes de identificar a atividade de pessoas que não participaram da fase de treinamento. O `MPA-CSI` identifica a presença de uma pessoa em um ambiente e indica se a pessoa está se movendo. Caso a pessoa esteja se movendo,

o `MPA-CSI` identifica se a pessoa está andando ou correndo. Caso contrário, classifica a posição como deitada ou sentada.

O restante do artigo está organizado da seguinte maneira. A Seção 2 apresenta conceitos básicos sobre Wi-Fi CSI. A Seção 3 descreve trabalhos relacionados, enquanto a Seção 4 apresenta a metodologia proposta. A descrição dos experimentos, resultados e discussões é fornecida na Seção 5. Finalmente, as conclusões e sugestões para trabalhos futuros são apresentadas na Seção 6.

2. Dados CSI

Os dados CSI trazem algumas das propriedades fundamentais dos canais de comunicação. Com isso é possível descrever como o sinal é alterado à medida que se propaga do transmissor até o receptor.

Na especificação IEEE 802.11ax [Soto et al. 2022a, Caballero et al. 2023, Santos et al. 2020], a camada física das redes Wi-Fi utiliza a técnica de multiplexação por divisão ortogonal de frequência (OFDM). OFDM é uma técnica de multiplexação que divide a largura de banda disponível em vários subcanais ortogonais [Soto et al. 2022b, Santos et al. 2020]. Dessa forma, a informação pode ser transmitida de forma independente em diferentes subportadoras [Lee et al. 2018]. Além disso, as subportadoras são canais ortogonais, cada um dos quais pode fornecer dados únicos de CSI; portanto, cada subportadora pode ser tratada como um sensor independente capaz de coletar dados de CSI.

Em um sistema Wi-Fi MIMO sob a especificação IEEE 802.11n, com P antenas transmissoras e Q antenas receptoras, o sinal contendo os dados estimados de CSI para cada fluxo de dados pode ser expresso como:

$$h_{p,q} = |h|e^{j\theta}, \tag{1}$$

onde $h_{p,q}$ representa o CSI entre a p-ésima antena transmissora e a q-ésima antena receptora, $|h|$ é a magnitude do sinal de CSI, relacionada à atenuação do sinal na propagação, e $e^{j\theta}$ representa a fase do CSI, relacionada às mudanças de fase do sinal na propagação.

Como o canal é dividido em várias subportadoras no OFDM, a representação do sinal recebido será um vetor. Supondo que c seja o número de subportadoras, o CSI entre um par de antenas (p, q) pode ser representado como um vetor com c elementos:

$$\boldsymbol{h}_{p,q} = [h_1, h_2, \ldots, h_c]^{\mathrm{T}}. \tag{2}$$

Quando uma pessoa se encontra entre os dispositivos transmissor e receptor, ela atua como um obstáculo, afetando a propagação do sinal eletromagnético. As alterações causadas pela presença da pessoa são observadas nos dados CSI. Essas variações, analisadas ao longo do tempo, podem ser usadas para detectar a presença humana e os movimentos corporais.

3. Trabalhos Relacionados

Dispositivos Wi-Fi estão atualmente difundidos em praticamente todos os ambientes, com suas características de sinal influenciadas por diversos fatores ambientais,

incluindo a presença e o movimento humano [Wang et al. 2017, Galdino et al. 2023, dos Santos et al. 2024]. Essas variações podem ser detectadas em dados de Informação do Estado do Canal (CSI), que fornecem detalhes da camada física (PHY), como amplitude e fase [Soto et al. 2022a]. O uso dessas informações extraídas de dados CSI para reconhecimento de atividades humanas apresenta grande potencial na área da saúde, especialmente no monitoramento remoto de pacientes [Santos et al. 2020, dos Santos et al. 2024].

Xiao et al. [Xiao et al. 2019] propuseram uma rede generativa adversarial semi-supervisionada (GAN) para o reconhecimento de atividades baseado em dados CSI. O método proposto emprega um gerador complementar, capaz de utilizar dados não rotulados limitados para gerar diversas amostras sintéticas para o treinamento de um discriminador robusto. Para o discriminador, eles modificam o número de probabilidades de saídas, o que pode auxiliar na obtenção do limite de decisão correto. Eles propuseram uma regularização de variedade, que pode estabilizar o processo de aprendizado. A ideia principal da regularização de variedade é que o subconjunto relevante de dados, que vem de pontos próximos, deve receber rótulos semelhantes. De acordo com essa ideia, eles projetaram a regularização específica do manifold que é mais adequada para a situação com amostras limitadas não rotuladas. Esse termo pode melhorar ainda mais a estabilidade do treinamento, bem como o desempenho preditivo final. 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, nome da proposta segundo os autores, 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. 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.

Wang et al. [Wang et al. 2021] propuseram um sistema de reconhecimento de atividades baseado em estado do canal multimodal (MCBAR) que aproveita as infraestruturas Wi-Fi existentes e monitora atividades humanas a partir de medições de dados CSI. O MCBAR aplica um gerador multimodal para aproximar a distribuição de dados CSI em diferentes configurações ambientais com dados CSI medidos limitados. Os dados CSI gerados usando o gerador multimodal podem fornecer diversidade para a transferência de conhecimento. O gerador de tradução no MCBAR é de uma estrutura multimodal, que melhora técnicas de transferência de estilo. Nessa proposta, uma matriz de interferência gerada aleatoriamente é usada para simular a dinâmica ambiental que pode afetar os dados CSI. O gerador de tradução pode traduzir dados CSI de uma atividade do domínio de origem para o domínio de destino com diferentes matrizes de interferência. Como resultado, os dados CSI dessa atividade interferida por diferentes dinâmicas ambientais no domínio de destino podem ser simulados. Com a aleatoriedade trazida por essa matriz de interferência, os dados CSI podem ser traduzidos em diferentes domínios de diversas maneiras. Uma GPU 2080ti foi usada em seus experimentos. No entanto, os dados usados para treinar o modelo se limitaram a apenas alguns candidatos (10), com algumas atividades realizadas, faltando um modelo generalizado independente da pessoa monitorada.

Li et al. [Li et al. 2021] propuseram um esquema de reconhecimento de atividades humanas baseado na colaboração entre visão e Wi-Fi. Eles coletaram dados CSI e pontos

do esqueleto humano do vídeo. Desenvolveram uma rede Transformer de longo e curto prazo para estabelecer a colaboração entre os dados CSI e os pontos do esqueleto. O método proposto utiliza a rede neural Transformer para compor a rede Transformer de longo e curto prazo (LSTT) de modo que o método possa obter os pontos do esqueleto humano a partir dos dados CSI. Seu método atinge uma precisão de 96%, no entanto, foi testado em dados CSI de apenas uma pessoa, limitando o desempenho de generalização do método para mais pessoas. Seu método também requer dados de vídeo para funcionar, causando sérias limitações, como custo, segurança dos dados gravados e privacidade.

Caballero et al. [Caballero et al. 2023] propuseram uma abordagem para obter reconhecimento de atividades humanas (HAR) por meio do uso de dispositivos Wi-Fi comerciais. Utilizando sua proposta, é possível inferir a posição de uma pessoa monitorada em um ambiente interno. Para alcançar isso, eles limparam e processaram a amplitude dos dados CSI coletados, selecionaram e avaliaram cinco diferentes algoritmos de classificação para inferir a posição dos indivíduos e compararam seu desempenho. Os dados foram coletados enquanto uma pessoa realizava uma variedade de atividades em uma sala. Para o cenário e conjunto de dados considerados neste estudo, os resultados mostraram que o algoritmo Random Forest (RF) apresentou o melhor desempenho em todos os testes, atingindo uma precisão média de $93,03\%$. No entanto, um modelo RF deve ser treinado individualmente para cada pessoa para detectar sua posição. Portanto, o modelo proposto é específico de cada pessoa, o que difere da solução proposta deste trabalho.

A Tabela 1 apresenta uma comparação do modelo proposto com outros estudos que empregam técnicas de aprendizado de máquina para monitorar atividades humanas por meio de dados CSI. O `MPA-CSI` se destaca por ter envolvido um número significativamente maior de participantes (59), 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 `MPA-CSI` cobre 17 posturas distintas, além de considerar uma condição de ambiente vazio. Essa diversidade permite ao modelo discriminar uma ampla gama de comportamentos e situações. Os estudos relacionados empregam uma variedade de métodos de aprendizado de máquina. A proposta deste artigo utiliza modelos Transformer, reconhecidos por sua eficácia no aprendizado de padrões temporais complexos [Li et al. 2021, Rothman 2021]. Essa característica pode ser particularmente vantajosa para capturar as sutilezas das atividades posturais, em comparação com métodos tradicionais como GAN e Random Forest, utilizados em trabalhos anteriores. O `MPA-CSI` proposto é único por utilizar exclusivamente dados CSI de coletas reais, ou seja, não são dados gerados por IA, independentemente das características individuais dos participantes. 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 `MPA-CSI` alcança uma acurácia de $96,67\%$, 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.

Tabela 1. Tabela comparativa de trabalhos relacionados.

Ref.	Partic.	Atividades	Método ML	Utiliza apenas dados reais coletados pelo CSI em tempo real independente da pessoa	Acurácia
[Xiao et al. 2019]	3	Cair, andar, pular, pegar, abrir e fechar portas, levantar as mãos, sentar e levantar	GAN	Não	86,27%
[Wang et al. 2021]	10	Correr, caminhar, cair, boxear, girar os braços e limpar o chão	GAN	Não	92,90%
[Li et al. 2021]	3	Caminhar, acenar com as mãos, pegar, pular, levantar as mãos e agachar	Long-short-term Transformer	Não	96%
[Caballero et al. 2023]	125	Sentado, em pé, deitado, caminhando, correndo e varrendo	Random Forest	Não	93,03%
MPA-CSI	**59**	**17 posturas distintas mais dados CSI da sala vazia**, como sentado, em pé, deitado, andando, varrendo e correndo realizadas também alternadamente, com a pessoa de frente ou de costas para o aparelho e pausando a respiração	**Transformer**	**Sim**	**96, 67%**

O CsiGAN apresentado em [Xiao et al. 2019] utilizou um conjunto de dados (FallDeFi Data) com a participação de apenas 3 voluntários para a coleta de dados, enquanto o MCBAR [Wang et al. 2021] contou com 10 voluntários. O conjunto de dados empregado pelo `MPA-CSI` abrange dados CSI de 59 voluntários, assegurando maior variabilidade. O `MPA-CSI` demonstra maior acurácia na identificação de movimentos ($96,67\%$) em comparação às propostas apresentadas em [Xiao et al. 2019] ($86,27\%$), em [Wang et al. 2021] ($92,90\%$) e em [Caballero et al. 2023] ($93,03\%$). Embora o desempenho seja similar ao alcançado em [Caballero et al. 2023], a proposta deste trabalho apresenta a vantagem de ser suficientemente geral para reconhecer atividades de indivíduos não incluídos no conjunto de dados de treinamento. Ademais, diferentemente de outros trabalhos correlatos, o `MPA-CSI` é capaz de detectar a atividade da pessoa em tempo real, demandando aproximadamente $0,56$ s de processamento.

4. Metodologia

Nesta seção é apresentada a metodologia proposta. O `MPA-CSI` [1] implementa um fluxo de treinamento e avaliação de interações para otimizar a precisão, ajustando hiperparâmetros com base em perdas e registrando métricas. Após isso, o `MPA-CSI` pode identificar a atividade humana em ambientes internos, analisando dados CSI mesmo de uma pessoa nunca vista pelo modelo durante o treino. `MPA-CSI` envolve o desenvolvimento de quatro modelos de classificação binária, cada um abordando um aspecto específico da atividade humana: presença, movimento, postura (deitado ou sentado) e locomoção (correndo ou andando). O fluxo geral de processamento é ilustrado na Figura 1 de maneira simplificada. 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 avaliação' 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.

4.1. Base de Dados

Os dados CSI[2] usados nesta pesquisa são parte do dataset eHealth CSI [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

[1]Os detalhes de implementação, os códigos e arquivos estão disponíveis no GitHub: https://github.com/mestrelan/MPA-CSI

[2]Este projeto foi aprovado pelo Comitê de Ética da UFF sob número CAAE 54359221.4.0000.5243.

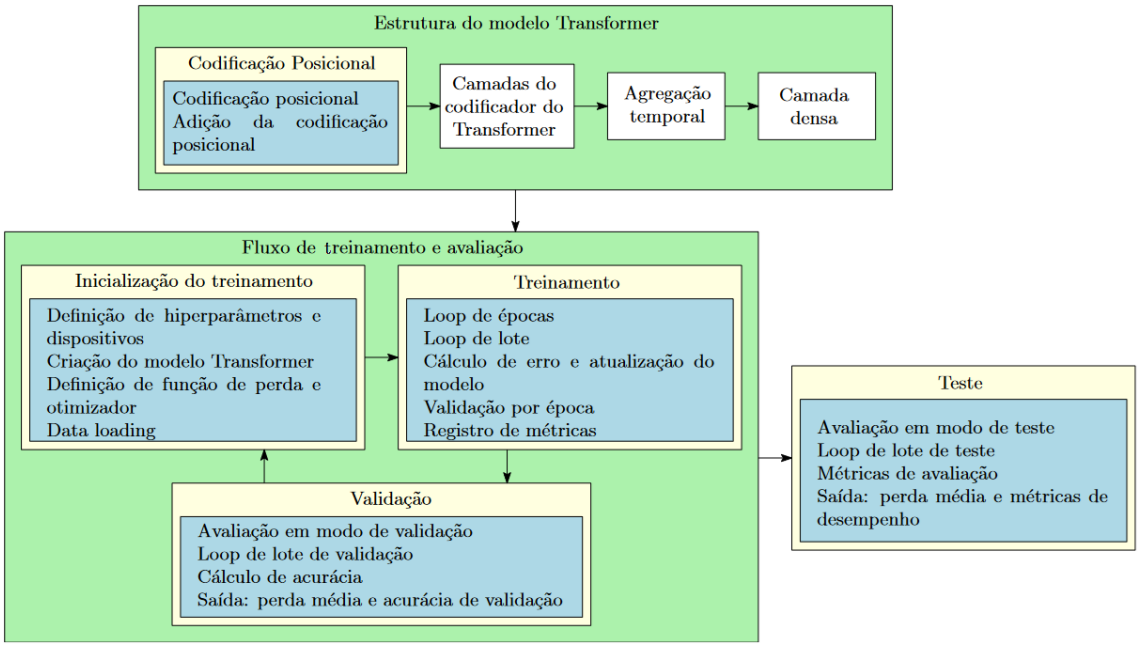


Figura 1. Diagrama de blocos do `MPA-CSI`.

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. Entre as diferentes posições como sentado, em pé, deitado, andando, varrendo e correndo. As coletas foram realizadas com a pessoa realizando essas atividades alternadamente e com a pessoa de frente ou de costas para o aparelho.

4.2. Processamento de dados CSI

A análise apresentada em [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, 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.

4.3. Geração e Treinamento de Modelos

Foram desenvolvidos quatro modelos independentes que recebem como entrada os dados CSI coletados e processados anteriormente. Os modelos foram desenvolvidos para detectar e classificar uma variedade de atividades humanas em um ambiente indoor, incluindo detecção de presença, detecção de movimento, classificação da marcha (caminhar versus correr) e classificação de postura (sentado versus deitado). Cada um dos modelos gerados utiliza posições específicas do conjunto de dados. Esses modelos realizam classificação binária, produzindo apenas dois resultados possíveis. Foi empregado o algoritmo Transformer com várias configurações de hiperparâmetros para os modelos propostos. A especificação de hiperparâmetros e dispositivos é essencial para a proposta dos modelos. Hiperparâmetros, incluindo o número de cabeças, camadas e taxa de aprendizado, juntamente com a escolha de CPU ou GPU, definem a arquitetura do modelo. Os modelos Transformer são instanciados e colocados no dispositivo designado. Foi utilizado *BCEWithLogitsLoss* para classificação binária, *Adam* como nosso otimizador e *ReduceLROnPlateau* para o escalonamento da taxa de aprendizado. Os dados são rotulados, particionados e alimentados aos modelos, conforme detalhado a seguir.

Para criar e treinar o Modelo 1 (presença), foi preciso dividir os dados CSI em segmentos. Foram utilizados dados de salas vazias e de salas com pessoas. Foram coletados dados extras de salas vazias e foram usados os dados de todas as 17 posições registradas para salas com pessoas. Os dados foram marcados como **sem presença** para salas vazias e **presença** para salas com pessoas. Para o Modelo 2 (movimento), a modelagem de dados foi realizada segmentando os dados CSI das 17 posições. Especificamente, foram consideradas posições com e sem movimento. Posições estáticas, como sentado, em pé e deitado, foram classificadas como **sem movimento**, enquanto posições envolvendo caminhar, correr e varrer foram categorizadas como **movimento**. Esses rótulos foram então codificados como 0 (sem movimento) e 1 (movimento) para o algoritmo Transformer. Para o Modelo 3 (Caminhar/Correr), foram utilizados dados das posições de caminhar e correr dentro das 17 posições. A posição de caminhar foi rotulada como 0 e a posição de correr foi rotulada como 1. Para o Modelo 4 (Sentado/Deitado), foram utilizados dados das posições sentado e deitado. Vale ressaltar que o conjunto de dados contém múltiplas coletas para as posições sentado e deitado para considerar diferentes orientações (por exemplo, de frente ou de costas, deitado de barriga para baixo ou para cima). Dados das posições sentadas (independentemente da orientação) foram rotulados como 0 e os das posições deitadas (independentemente da orientação) foram rotulados como 1 para o algoritmo Transformer. Os modelos propostos usam loops para treinar, repetindo o processo por um número determinado de épocas. Os dados são fornecidos aos modelos em grupos (batches). Após cada grupo, é calculado o erro e atualizado o modelo usando um critério e um otimizador específicos. Por fim, são armazenadas a perda e a precisão do treinamento.

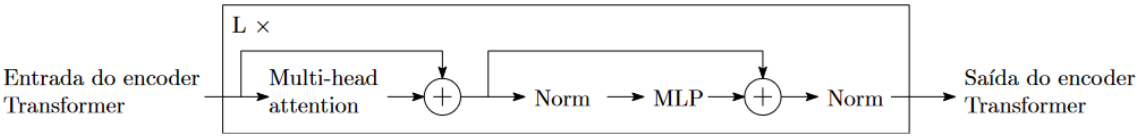


Figura 2. Camadas do enconder transformer.

A Figura 2 apresenta as camadas Encoder Transformer. Os dados entram no encoder no formato [batch_size, seq_len, num_channels]. O parâmetro batch_size representa o número de amostras processadas simultaneamente. O seq_len indica a quantidade de instantes no tempo, enquanto num_channels corresponde ao número de variáveis medidas em cada instante. No encoder, os dados passam por várias camadas idênticas empilhadas, conforme o valor de num_layers. Cada uma dessas camadas executa um conjunto de operações essenciais para o aprendizado da representação da sequência temporal. O primeiro passo é a aplicação do mecanismo de autoatenção multi-cabeças (Multi-Head Self Attention) [Rothman 2021], que permite ao modelo capturar relações entre diferentes instantes da sequência. Cada cabeça de atenção, definida pelo parâmetro num_heads, processa informações de maneira independente.

Para calcular a autoatenção, são gerados três tensores: Q, K e V. O tensor Q, chamado de Query, representa os elementos que consultam outros instantes da sequência. O tensor K, conhecido como Key, contém os elementos que podem fornecer informações relevantes. Já o tensor V, denominado Value, armazena os valores reais das características a serem processadas. A similaridade entre Q e K é então calculada para identificar quais instantes no tempo são mais relevantes para cada outro instante. A saída dessa operação é uma combinação ponderada dos valores em V, onde instantes mais importantes recebem pesos maiores. Após a autoatenção, a saída passa por uma camada de normalização, conhecida como LayerNorm, que estabiliza o treinamento. Além disso, uma conexão residual adiciona a entrada original à saída da autoatenção, garantindo gradientes mais estáveis. Em seguida, a saída é processada por uma rede feed-forward, que consiste em um perceptron multicamadas. Essa rede inclui uma camada densa, seguida por uma função de ativação ReLU, e outra camada densa para ajustar a dimensão da saída. Depois da rede feed-forward, uma nova camada de normalização estabiliza a saída, enquanto outra conexão residual adiciona a entrada do bloco à sua saída final. O processo se repete ao longo de num_layers blocos, refinando progressivamente a representação dos instantes da sequência. Ao final do encoder, cada instante da série temporal contém informações globais sobre toda a sequência. A saída mantém o formato [batch_size, seq_len, num_channels], mas agora os valores representam representações mais ricas e informativas da sequência temporal.

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].

4.4. Validação e teste dos modelos

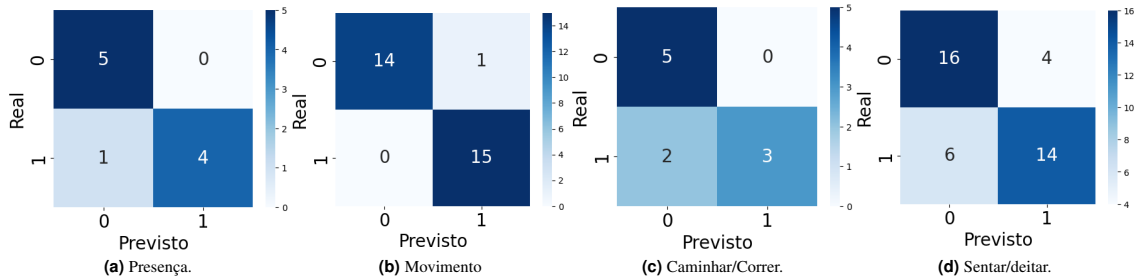
O conjunto de dados foi dividido em conjuntos de treinamento 75%, utilizado na Seção 4.3, validação, 15%, e teste, 10%. O conjunto de validação é usado para avaliar o desempenho dos modelos treinados, calculando métricas de perda e acurácia sem alterar os parâmetros do modelo. Para garantir uma avaliação determinística, comportamentos específicos de treinamento, como *Dropout* e normalização por *Batch*, são desativados, e os cálculos de gradiente são interrompidos. O modelo itera sobre lotes de validação, proces-

sando dados de entrada e rótulos de verdade fundamental, e acumula métricas de perda e acurácia. Foi definida a configuração dos hiperparâmetros do modelo que alcançou a maior acurácia e a menor perda. O teste avalia o desempenho do modelo treinado em dados nunca vistos. Aqui, foram calculadas várias métricas, incluindo perda, acurácia, *recall*, precisão, pontuação F1 e uma matriz de confusão.

5. Resultados e Discussão

Dados de 59 participantes, incluindo homens e mulheres, foram utilizados nos experimentos realizados. Desses, 44 participantes (75%) foram utilizados para treinamento dos modelos, 10 participantes (15%) para validação e 5 participantes (10%) para teste. Conforme detalhado na Seção 4.1, configurações específicas foram empregadas para a coleta de dados, que ocorreu em uma sala dedicada no Instituto de Computação da Universidade Federal Fluminense. Os dados foram coletados usando um Raspberry Pi B4 equipado com um chipset 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.

A Figura 3 apresenta as matrizes de confusão dos resultados de teste de cada um dos 4 modelos descritos anteriormente, ou seja, (a) detecção de presença, (b) detecção de movimento, (c) reconhecimento da atividade de caminhar ou correr e (d) reconhecimento da posição sentada ou deitada. A Figura 3 reflete um conjunto de teste balanceado, composto por 5 voluntários. Devido ao equilíbrio dos dados, as tarefas de detecção de presença e caminhada/corrida apresentam matrizes de confusão com 5 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. Para as tarefas de detecção de movimento e postura (como sentar/deitar), o conjunto de teste inclui mais amostras, permitindo que o modelo execute vários testes por voluntário. Isso aumenta as capacidades de validação e análise de desempenho para esses cenários, uma vez que o modelo pode capturar uma variedade mais ampla de instâncias para cada voluntário, fornecendo uma avaliação mais abrangente do desempenho em atividades complexas.



(a) Presença. **(b)** Movimento **(c)** Caminhar/Correr. **(d)** Sentar/deitar.

Figura 3. Matriz de confusão de cada um dos 4 modelos de reconhecimento de atividades propostos.

A Tabela 2 apresenta os resultados gerais obtidos na identificação de diferentes atividades dos voluntários de teste. Além disso, são apresentadas as configurações utilizadas em cada um dos 4 modelos Transformer. A análise proposta começa com a detecção de

Tabela 2. Resultados gerais dos testes.

Modelo	Heads	Layers	L. Rate	Epochs	Batch Size	Accurácia	Precisão	Recall	F1 Score	Test Loss
Detecção de Presença	3	6	0,0001	60	24	90%	100%	80%	0,89	0,5667
Detecção de Movimento	3	2	0,0001	100	4	96,67%	93,75%	100%	0,97	0,1542
Reconhecer Caminhada/corrida	3	11	0.00001	100	24	80%	100%	60%	0.75	0,6923
Reconhecer Sentado/deitado	8	4	0,0001	20	4	75%	77,78%	70%	0,74	0,9584

presença humana em uma sala. O modelo alcançou uma acurácia de 90% e uma precisão perfeita de 100%, indicando que todas as instâncias de presença humana foram corretamente identificadas. O F1-Score de 0.89 corrobora ainda mais a robustez dos resultados. O modelo de detecção de movimento humano alcançou uma acurácia ligeiramente superior a 96% com uma precisão de $93,75\%$. A sensibilidade, ou *recall*, que mede a taxa de verdadeiros positivos, foi de 100%, com um F1-Score correspondente de $0,97$. Esses resultados indicam que o modelo é capaz de detectar com precisão o movimento humano em um ambiente fechado. Quando associada à identificação de presença, essa tecnologia tem o potencial de ser aplicada a uma variedade de aplicações cotidianas não críticas. Na detecção de atividades específicas, como caminhar/correr, e as posições sentada/deitada, os modelos obtiveram acurácias de 80% e 75%, respectivamente. Para detecção de atividade de Caminhada/Corrida, a pontuação F1 de $0,75$ apresentou um desempenho mais modesto. O *recall* de 60% sugere que ele tem dificuldade em capturar todos os casos, abrindo espaço para melhorias futuras. Na detecção da posição Sentado/Deitado, alcançou-se uma precisão de $77,78\%$ e um *recall* de 70%. O F1-Score de $0,74$ mostra um desempenho razoável, embora também haja espaço para melhoria, especialmente no *recall*. É importante destacar que a detecção de posições com altos níveis de movimento, como correr e caminhar, apresenta um desafio considerável. A natureza intrínseca desses movimentos introduz ruído nos dados CSI, dificultando a detecção precisa.

As colunas **Heads** e **Layers** na Tabela 2 especificam o número de cabeças de atenção e camadas do modelo Transformer para cada tarefa. Em geral, tarefas mais complexas, como a detecção de sentar/deitar, utilizam mais cabeças (8) e camadas intermediárias (4), indicando a necessidade de maior capacidade de atenção para capturar nuances nas posturas. Por outro lado, tarefas mais simples, como a detecção de presença e movimento, utilizam uma configuração mais leve, com 3 cabeças e entre 2 e 6 camadas. A taxa de aprendizado é ajustada para cada tarefa. Tarefas mais difíceis, como a classificação de caminhar/correr, possuem uma taxa de aprendizado mais baixa $(0,00001)$, o que permite que o modelo se ajuste gradualmente e evite o overfitting. Para tarefas mais simples, como a detecção de presença e movimento, uma taxa mais alta $(0,0001)$ é suficiente. O número de épocas varia entre 20 e 100, dependendo da tarefa. 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 para cada tarefa, variando de 4 a 24. Atividades mais simples, como presença e movimento, utilizam um lote maior, enquanto atividades mais complexas exigem lotes menores para otimizar a capacidade de aprendizado e o uso da memória.

A Figura 4 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.

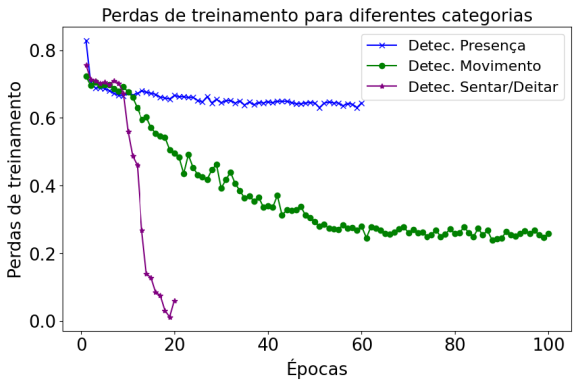


Figura 4. Perdas de treinamento para diferentes categorias do `MPA-CSI`.

A Figura 5 apresenta as perdas e as acurácias da validação ao longo das épocas. As acurácias no conjunto de validação representam a proporção de previsões corretas feitas pelo modelo. Taxas mais altas indicam um modelo mais eficaz. Durante a validação do modelo para detecção de presença, a Figura 5 mostra que a perda inicial apresentou flutuações, estabilizando e atingindo valores menores a partir da época 40. Esse padrão se repetiu na acurácia, que também se estabilizou após a época 40. Embora o modelo tenha demonstrado um aprendizado mais lento, ele finalmente convergiu para resultados satisfatórios. A Figura 5 mostra que essa dificuldade inicial pode estar relacionada à similaridade entre os dados CSI de salas vazias. Em contraste com a detecção de presença, a detecção de movimento apresentou um aprendizado mais rápido e estável, atingindo valores desejáveis de perda e acurácia antes da época 20. Essa performance superior pode ser atribuída à maior distinção entre os dados CSI dos participantes presentes na sala.

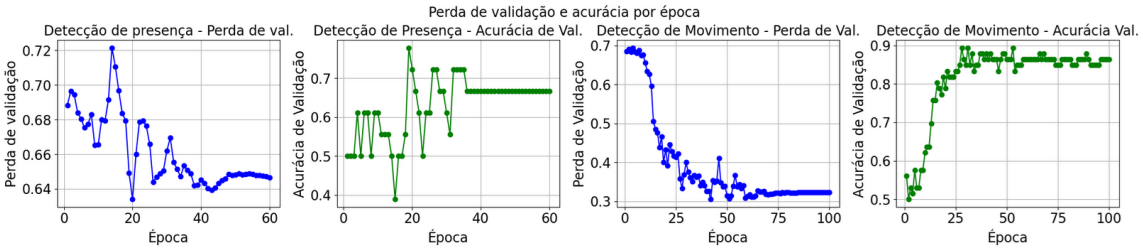


Figura 5. Perda de validação e acurácia por época do `MPA-CSI`.

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.

6. 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 MPA-CSI para identificar a atividade de uma pessoa em um cômodo através da análise de dados CSI. Considerando um conjunto de dados de 59 voluntários utilizados para sua avaliação, foi atingida uma acurácia de $96,67\%$ para identificar movimento 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 MPA-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 MPA-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 o reaproveitamento da infraestrutura já existente, o que gera uma redução de custos em relação ao emprego de dispositivos específicos. Além disso, o MPA-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 detectar sinais fisiológicos, como a frequência respiratória, para detectar apneia usando dados CSI quando um indivíduo estiver dormindo e aprimorar o desempenho do MPA-CSI em tarefas desafiadoras (caminhar/correr e sentar/mentir) refinando o pré-processamento de dados e explorando modelos híbridos que combinam Transformers com outras abordagens de aprendizado. Portanto, pretende-se analisar o desempenho do modelo em ambientes com níveis variados de interferência (por exemplo, vários dispositivos Wi-Fi ou paredes), o que pode afetar sua confiabilidade em situações cotidianas. Sendo assim, avaliar possíveis estratégias de mitigação da interferência.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, CAPES Print, CNPq, FAPERJ, FINEP, INCT-ICONIoT e INCT-MACC.

Referências

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.

Chen, Z., Zhang, L., Jiang, C., Cao, Z., and Cui, W. (2019). Wifi csi based passive human activity recognition using attention based blstm. *IEEE Transactions on Mobile Computing*, 18(11):2714–2724.

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 *Proceedings of the IEEE Virtual Conference on Communications*. Accepted for publication.

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*, 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.

Gouveia, B. G., Galdino, I., Caballero, E., Soto, J. C. H., Ramos, T. C., Guerra, R., Muchaluat-Saade, D., and Albuquerque, C. V. N. (2024). Parameter tuning for accurate heart rate measurement using wi-fi signals. In *2024 International Conference on Computing, Networking and Communications (ICNC)*, pages 407–411.

Lee, S., Park, Y. D., Suh, Y. J., and Jeon, S. (2018). Design and implementation of monitoring system for breathing and heart rate pattern using WiFi signals. *IEEE Annual Consumer Communications and Networking Conference*, pages 1–7.

Li, S., Ge, Y., Shentu, M., Zhu, S., Imran, M., Abbasi, Q., and Cooper, J. (2021). Human activity recognition based on collaboration of vision and wifi signals. In *2021 International Conference on UK-China Emerging Technologies (UCET)*, pages 204–208.

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. *Sociedade Brasileira de Computação*.

Soto, J. C., Galdino, I., Caballero, E., Ferreira, V., Muchaluat-Saade, D., and Albuquerque, C. (2022a). A survey on vital signs monitoring based on wi-fi csi data. *Computer Communications*, 195:99–110.

Soto, J. C. H., Galdino, I., Gouveia, B. G., Caballero, E., Ferreira, V., Muchaluat-Saade, D., and Albuquerque, C. (2022b). Wi-fi csi-based human presence detection using dtw features and machine learning. In *2022 IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6.

Wang, D., Yang, J., Cui, W., Xie, L., and Sun, S. (2021). Multimodal csi-based human activity recognition using gans. *IEEE Internet of Things Journal*, 8(24):17345–17355.

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.