

Detecção de Quedas e Crises Epilépticas com Dispositivo Vestível utilizando TinyML e ESP32

Francisco Thiago Barbosa Quinto¹, Thiago de Sena Lima¹, Fábio Timbó Brito¹, José Daniel de Alencar Santos¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Campus Maracanaú
Departamento da Indústria – Maracanaú – CE – Brasil

francisco.thiago.barbosa07@aluno.ifce.edu.br,
thiago.sena.lima07@aluno.ifce.edu.br, fabio@ifce.edu.br,
jdaniel@ifce.edu.br

Abstract. *Epilepsy is a chronic neurological condition affecting approximately 50 million people worldwide, requiring solutions that ensure safety and rapid emergency response. This study presents the development of an ESP32-based wearable system for the continuous monitoring of elderly individuals and patients with epilepsy. The primary goal is the autonomous detection of falls and seizures using inertial sensors. The methodology consisted of generating a dataset based on acceleration profiles from specialized literature, followed by training an artificial intelligence classifier via the Edge Impulse platform. By employing a TinyML approach, the model was deployed directly onto the hardware, enabling real-time inference with low latency and eliminating the need to transmit user data to external servers, thus enhancing privacy. Results indicate that this edge computing approach is highly effective, allowing the device to operate without an internet connection. In conclusion, the proposed solution provides an additional layer of safety, mitigating risks of delayed assistance and promoting greater patient autonomy.*

Resumo. *A epilepsia é uma condição neurológica crônica que afeta cerca de 50 milhões de pessoas globalmente, demandando soluções que garantam a segurança e a agilidade no atendimento emergencial. Este trabalho apresenta o desenvolvimento de um sistema vestível baseado no microcontrolador ESP32 para o monitoramento contínuo de idosos e pacientes epiléticos. O objetivo central é a detecção autônoma de quedas e crises convulsivas por meio de sensores inerciais. A metodologia envolveu a criação de um conjunto de dados baseado em perfis de aceleração da literatura especializada, seguido pelo treinamento de um classificador de inteligência artificial na plataforma Edge Impulse. Utilizando a abordagem de TinyML, o modelo foi embarcado diretamente no hardware, permitindo inferências em tempo real com baixa latência e sem a necessidade de envio de dados a servidores, o que preserva a privacidade do usuário. Os resultados indicam que a computação de borda é eficaz para esta aplicação, permitindo o funcionamento do dispositivo de forma offline. Conclui-se que a solução oferece uma camada adicional de segurança, mitigando riscos de atraso no socorro e promovendo maior autonomia aos usuários.*

1. Introdução

A epilepsia e as quedas em idosos representam desafios globais de saúde pública que comprometem severamente a segurança e a autonomia dos indivíduos (WORLD HEALTH ORGANIZATION - WHO, 2024; WHO, 2021). Pacientes epiléticos sofrem frequentemente com lesões físicas decorrentes de crises tônico-clônicas generalizadas (GTCS), caracterizadas pela perda súbita de consciência e sequências de contrações musculares severas (ASADI-POOYA et al., 2012; BENICZKY et al., 2022; FISHER et al., 2017). Paralelamente, o declínio do equilíbrio e da cognição associado ao envelhecimento torna os idosos maiores de 60 anos altamente vulneráveis a quedas graves, resultando em fraturas, lesões cerebrais e um forte impacto psicológico que limita sua independência (WORLD HEALTH ORGANIZATION - WHO, 2021).

Em ambos os cenários, a incapacidade do indivíduo de solicitar socorro de forma consciente agrava os desfechos clínicos. Idosos frequentemente não conseguem se reerguer ou ativar sistemas tradicionais de assistência voluntária após um acidente, permanecendo longos períodos no chão sem ajuda (FLEMING et al., 2008). No contexto da epilepsia crônica, a morte súbita inesperada (SUDEP) é responsável por cerca de 17% dos óbitos (FÉRRER; BARROS, 2021) e ocorre majoritariamente durante o repouso noturno, evidenciando a falta de vigilância como um fator de risco letal (JONES; THOMAS, 2017). Dessa forma, a literatura indica que depender de uma ação voluntária para pedir ajuda não é confiável em sistemas de assistência e monitoramento.

Para garantir um atendimento rápido e autônomo, este projeto propõe um sistema vestível de monitoramento baseado no ESP32/M5StickC que integra Internet das Coisas (IoT - *Internet of Things*) e computação de borda (*edge computing*). Por meio de um modelo de Inteligência Artificial otimizado para microcontroladores (TinyML, do inglês *Tiny Machine Learning*) embarcado no dispositivo, é possível analisar dados inerciais em tempo real para detectar crises epiléticas e quedas, diminuindo a latência e a dependência de servidores externos. Complementarmente, a arquitetura estabelece uma comunicação resiliente com uma estação base, operando primariamente via protocolos locais de baixa latência com chaveamento automático para MQTT (*Message Queuing Telemetry Transport*) via internet em caso de falhas ou distanciamento, garantindo redundância e o disparo de alertas.

2. Metodologia e Implementação

Nas últimas décadas, o desenvolvimento de tecnologias para o monitoramento da epilepsia evoluiu significativamente, abrangendo métodos como eletroencefalograma, eletrocardiograma e acelerometria (RAMGOPAL et al., 2014). Embora os sistemas baseados em acelerometria sejam eficazes na detecção de crises motoras tônico-clônicas, eles enfrentam o desafio de distinguir esses eventos das Atividades de Vida Diária (ADLs - *Activities of Daily Living*) (ZIA et al., 2021). Algoritmos tradicionais baseados puramente em limiares (*thresholds*) são altamente suscetíveis a interpretar movimentos rotineiros de forma errônea, gerando falsos positivos que causam estresse aos cuidadores (NIJSEN, 2008; RAMGOPAL et al., 2014). Devido à variação individual na execução dessas tarefas rotineiras, torna-se imperativa a criação de sistemas discretos e confiáveis,

capazes de aprender de forma autônoma os padrões específicos que diferenciam uma crise ou queda de uma atividade normal (RAMGOPAL et al., 2014; ZIA et al., 2021).

Para superar essas limitações, o sistema proposto adota o microcontrolador ESP32 (integrado à plataforma vestível M5StickC), equipado com uma unidade de medição inercial (IMU) de seis eixos. A solução emprega uma rede neural convolucional unidimensional (1D CNN), estruturada na plataforma Edge Impulse para processar os dados brutos e extrair automaticamente os padrões temporais complexos dos movimentos. O treinamento do modelo utilizou um conjunto de dados próprio, que se baseou nos perfis de aceleração mostrados por Arends (2018) para simular os movimentos de convulsão. Por fim, a arquitetura foi otimizada e embarcada por meio uma biblioteca C++, permitindo a inferência local e em tempo real diretamente na borda, eliminando a dependência de processamento em nuvem.

2.1 Arquitetura Geral e Hardware

A arquitetura baseia-se em IoT e computação de borda, e é composta por um dispositivo vestível e uma estação base. Conforme a Figura 1, a lógica prioriza detecção local e comunicação redundante via protocolos ESP-NOW e MQTT. O núcleo do projeto é o M5StickC (baseado no ESP32-PICO), escolhido por seu formato reduzido e ergonomia para aplicações vestíveis, que são fatores determinantes para a adesão dos usuários (M5STACK, 2020; MOZAFFARI et al., 2020). A Figura 1 apresenta o fluxograma de funcionamento do sistema.

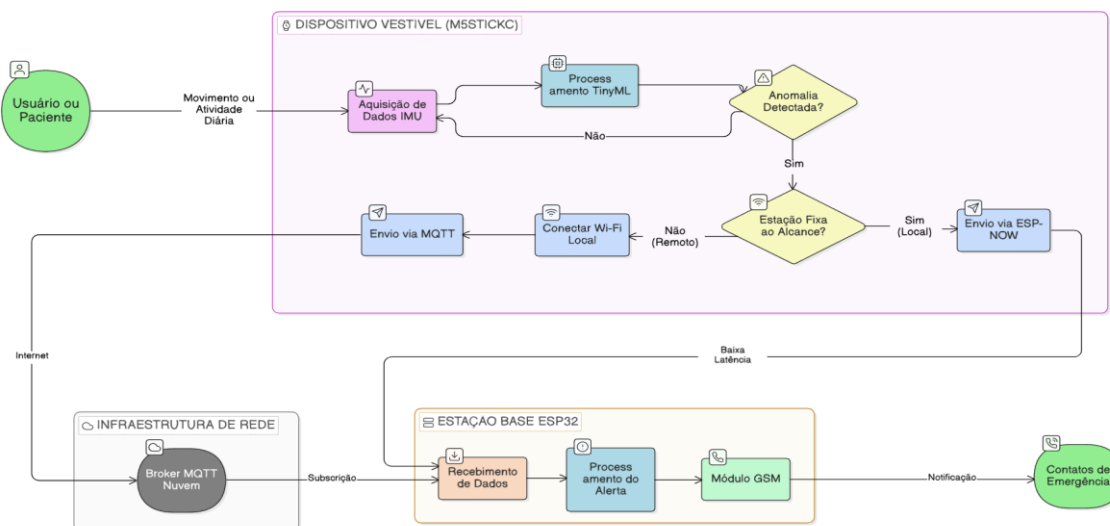


Figura 1. Fluxograma de funcionamento do sistema.

O sensor IMU MPU-6886 combina giroscópio e acelerômetro, ambos de 3 eixos, para capturar impactos de quedas e vibrações rítmicas de crises (M5STACK, 2020). Para a coleta de dados sem fios, utilizou-se um módulo ESP-01 como um *gateway* de aquisição, transmitindo dados via *socket* (UDP/TCP) para o computador. A estação base utiliza um ESP32 conectado a um modem GSM SIM800L, garantindo o envio de SMS e chamadas de emergência caso o Wi-Fi local falhe. O sistema emprega o protocolo ESP-NOW para baixa latência doméstica (ESPRESSIF SYSTEMS, 2026) e MQTT para redundância via internet.

2.2 Desenvolvimento do Classificador TinyML

O desenvolvimento do classificador seguiu as etapas de aquisição de dados, treinamento da rede neural e otimização para o sistema embarcado. Para diminuir ruídos mecânicos e garantir a liberdade de movimento do usuário na etapa de aquisição, a coleta de dados foi realizada sem fio, transmitindo as informações da pulseira para o computador via um ESP-01 funcionando como *gateway*, conforme mostrado na Figura 2.



Figura 2. Comunicação da pulseira com o ESP-01 na etapa de coleta de dados.

No total, foram coletados 29 minutos e 27 segundos de dados brutos a uma frequência de 47 Hz, categorizados em “queda”, ADLs (“parado”, “andando” e “escadas”) e “convulsão”, com as simulações desta última baseando-se nos perfis cinemáticos de Arends (2018). Na plataforma Edge Impulse, os sinais brutos foram segmentados em janelas de 3 segundos com avanço (*stride*) de aproximadamente 446 milissegundos, gerando as instâncias de entrada. Posteriormente, os dados foram normalizados e particionados em 80% para treinamento (com uma subdivisão interna de 20% para validação) e 20% para testes. A distribuição dos dados brutos e a quantidade de janelas geradas para cada classe é mostrada na Tabela 1.

Tabela 1. Distribuição quantitativa e temporal dos dados brutos e janelas geradas.

Classe	Nº de Arquivos Brutos	Duração Total (mm:ss)	Total de Janelas Geradas	Janelas de Treino	Janelas de Teste	Proporção da Classe
Convulsão	69	05:34	323	263	60	19.34%
Queda	125	06:15	125	100	25	7.49%
Andando	68	05:31	322	260	62	19.29%
Parado	81	06:34	383	307	76	22.93%
Escadas	37	5:33	517	414	103	30.96%
Total	380	29:27	1670	1344	326	100%

Embora a plataforma Edge Impulse, que é uma solução robusta de MLOps (*Machine Learning Operations*) voltada para o ecossistema TinyML (HYMEL et al., 2022), ofereça a criação de modelos via blocos visuais, optou-se pela programação direta do treinamento utilizando Python (TensorFlow/Keras). Essa abordagem proporcionou

maior controle sobre a arquitetura e os hiperparâmetros, permitindo a implementação de estratégias customizadas não disponíveis no fluxo visual padrão.

O modelo concebido baseia-se em uma Rede Neural Convolutacional Unidimensional (1D CNN), arquitetura eficaz na extração de características espaciais e temporais de sinais inerciais complexos (REUSCH et al., 2022). A rede consiste em três blocos convolucionais (Conv1D) com 8, 16 e 32 filtros, respectivamente, utilizando kernels de tamanhos 5 e 3. Cada bloco é seguido por camadas de *Batch Normalization*, *MaxPooling1D* e *Dropout* para mitigar os riscos de *overfitting* (GOODFELLOW et al., 2016; IOFFE; SZEGEDY, 2015). Após as convoluções, utilizou-se uma camada de *GlobalAveragePooling1D* para a redução de parâmetros seguida por uma camada densa de 32 neurônios com regularização L2. As configurações da arquitetura e os hiperparâmetros de treinamento são apresentados na Tabela 2.

Tabela 2. Especificações da arquitetura e parâmetros de configuração do modelo.

Descrição	Valor / Configuração
Total de épocas	300 (máximo)
Tamanho do Lote	32 instâncias
Critério de parada	<i>Early Stopping</i> (paciência de 50 épocas)
Taxa de Aprendizado	0.001 (com decaimento via <i>ReduceLROnPlateau</i>)
Pesos de classe	Intervalo de 1 a 30 (ajustados conforme a severidade)
Arquitetura Base	Rede Neural Convolutacional 1D (1D CNN)
Camadas Convolucionais	3x <i>Conv1D</i> (16, 32, 64 filtros)
Regularização	<i>Dropout</i> progressivo (0.1, 0.2, 0.3)
Camada de Redução	<i>GlobalAveragePooling1D</i>
Camada Densa	32 neurônios com regularização L2
Funções de Ativação	ReLU e Softmax (saída)
Otimizador	Adam ($\beta_1=0.9$, $\beta_2=0.999$, $\epsilon=10^{-7}$)

A convergência do modelo foi otimizada por meio da técnica de parada antecipada (*Early Stopping*) e do *callback ReduceLROnPlateau*, que ajusta a taxa de aprendizado dinamicamente, conforme o desempenho da validação. O treinamento utilizou o otimizador Adam e a aplicação de pesos de classe diferenciados, variando em uma escala de 1 a 30, para compensar o desequilíbrio do conjunto de dados e priorizar a detecção de eventos críticos (GOODFELLOW et al., 2016).

Concluída a validação, o classificador foi submetido a um processo de quantização e otimização para viabilizar sua execução restrita na arquitetura do ESP32 (HYMEL et al., 2022; RAY, 2022). O modelo final foi exportado como uma biblioteca C++ e integrado nativamente ao firmware da pulseira.

2.3. Integração IoT e Comunicação

Como a integração do modelo TinyML exigiu um *firmware* capaz de gerenciar periféricos e múltiplos canais de comunicação em tempo real, utilizou-se o FreeRTOS para garantir a constância da taxa de amostragem, necessário para a validade da inferência. O processamento foi distribuído nos núcleos do ESP32: o Núcleo 0 prioriza a leitura do

sensor inercial, enquanto o Núcleo 1 realiza a inferência local, descartando dados de atividades rotineiras. Ao detectar anomalias, o dispositivo utiliza comunicação redundante, transmitindo prioritariamente via ESP-NOW, protocolo ponto a ponto de baixa latência ideal para ambientes domiciliares (ESPRESSIF SYSTEMS, 2026). Em caso de falha ou distanciamento, o sistema alterna automaticamente para publicação MQTT via Wi-Fi.

A estação base, formada por um segundo ESP32 e um módulo GSM SIM800L, atua como central de despacho. O *hardware* foi encapsulado em uma carcaça impressa em 3D, adaptado de um projeto de código aberto da plataforma Thingiverse (SEVAGJB, 2024). Ao receber o alerta, a estação utiliza comandos AT seriais para enviar SMS e ligar para os contatos de emergência (Figura 3). Embora o SIM800L possua limitações de hardware, como operação exclusiva em 2G e incapacidade de transmitir áudio nas ligações, a lógica do projeto contorna isso de forma eficaz, com o toque persistente da chamada e o SMS atuando como o sinalizador principal para concluir a cadeia de alerta.

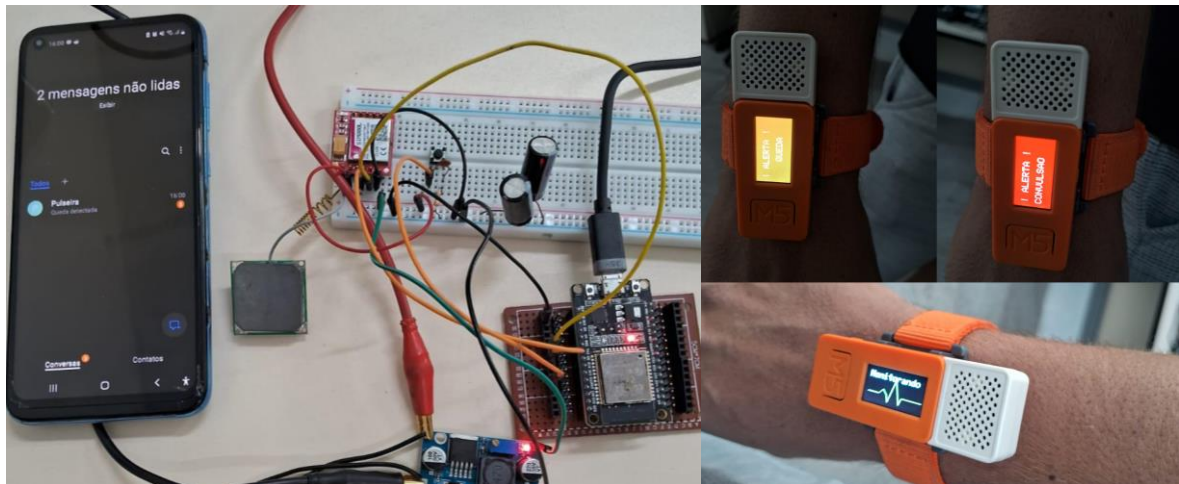


Figura 3. Protótipo da estação base montada na protoboard para testes de envio de mensagens e ligações telefônicas para cuidadores/familiars.

3. Resultados e Discussões

A avaliação do sistema confirmou que o processamento na borda garantiu baixa latência de inferência e preservação da privacidade, processando e descartando dados brutos localmente. A demonstração dos resultados iniciais obtidos na detecção de crises epilépticas e quedas pode ser vista no vídeo https://youtu.be/Ogyl_dJJPR8.

3.1. Resultados do Modelo

A avaliação da arquitetura de rede neural convolucional demonstrou eficácia na classificação dos sinais inerciais. Com base nos dados do conjunto de teste, o modelo alcançou uma sensibilidade de 100% tanto para a classe “queda” quanto para a classe “convulsão”. Este resultado é importante para a aplicação médica, pois garante a

identificação de dos eventos críticos testados, eliminando a ocorrência de falsos negativos.

Abaixo, a Figura 4 apresenta o desempenho por meio da matriz de confusão, que correlaciona as previsões do algoritmo com as classes reais.

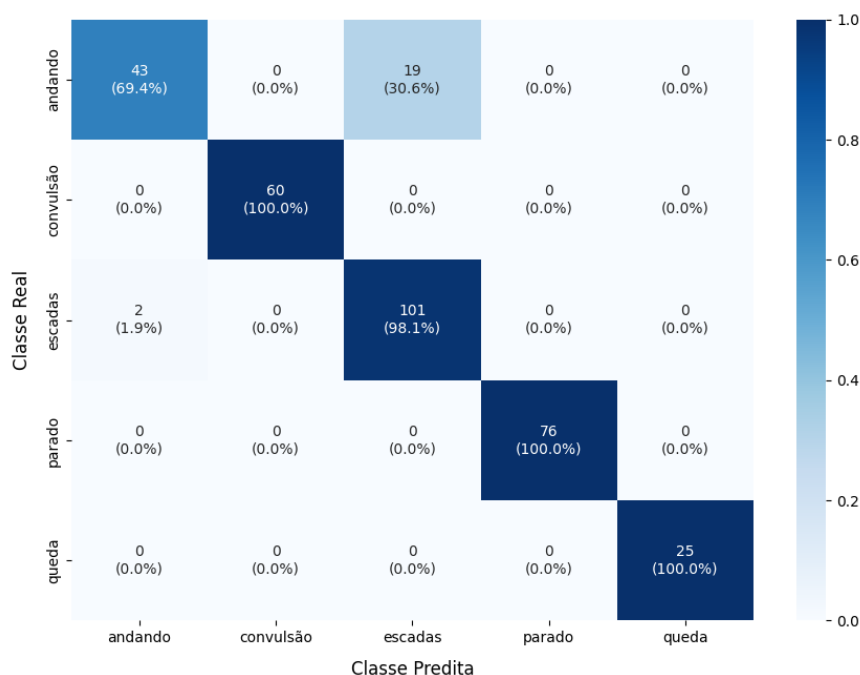


Figura 4. Matriz de confusão do conjunto de teste.

A análise das métricas revela bom desempenho, a acurácia, que indica os acertos globais entre as previsões e os eventos reais, manteve-se elevada, com 93%. A sensibilidade (*recall*) confirmou a capacidade do modelo em identificar corretamente as classes de interesse (“queda” e “convulsão”). Observou-se, contudo, que a classe “andando” obteve uma sensibilidade de aproximadamente 69,4%, confundindo essas instâncias com a classe “escadas”. Essa divergência é esperada, uma vez que ambos os movimentos compartilham padrões semelhantes. No contexto deste projeto, essa confusão entre ADLs não compromete a segurança do usuário, pois o foco principal é a distinção clara entre movimentos rotineiros e eventos de crise.

Destaca-se que, embora os resultados sejam promissores e comprovem o funcionamento da solução, o sistema demanda uma contínua expansão do conjunto de dados de ADLs para refinar a distinção entre movimentos similares. Além disso, o refinamento definitivo exigirá a substituição das simulações de convulsão por dados clínicos reais de pacientes epiléticos. Todavia, os testes realizados validam a arquitetura proposta como uma solução de monitoramento robusta e com implementação funcional.

O processamento na borda provou-se fluido e livre de gargalos. Com taxa de amostragem de 47 Hz, o ESP32 executou a rede neural em apenas 30 ms (segundo a plataforma Edge Impulse), tempo inferior ao de acúmulo da janela deslizante. Por fim, o modelo quantizado confirmou a viabilidade da abordagem TinyML, consumindo apenas

59 KB de memória Flash e 9,9 KB de RAM, operando com ampla margem de segurança no hardware.

3.2 Avaliação do Sistema

A avaliação do protótipo validou tanto a eficiência da arquitetura de comunicação quanto o processamento em tempo real do dispositivo. Para aumentar a eficiência energética do vestível, o rádio transmissor operou apenas quando necessário, permanecendo em estado de hibernação ou escuta passiva durante as atividades normais e ativando-se exclusivamente mediante a detecção de uma anomalia. O mecanismo de redundância de rede se mostrou funcional: a comunicação primária via ESP-NOW demonstrou estabilidade na entrega de pacotes em ambientes internos com obstáculos, enquanto o chaveamento automático para o protocolo Wi-Fi/MQTT assegurou a transmissão do alerta à estação base sempre que o distanciamento físico exigiu, garantindo a notificação contínua aos cuidadores. Simultaneamente, o processamento na borda provou-se possível.

4. Conclusões e Recomendações

Este trabalho comprovou a viabilidade técnica de um sistema de monitoramento de saúde de baixo custo e não invasivo, baseado em inteligência artificial embarcada no microcontrolador ESP32. A descentralização do processamento para a borda resultou em uma arquitetura resiliente a falhas de conexão e capaz de fornecer respostas em tempo real. Os testes realizados validaram a capacidade do dispositivo em executar redes neurais convolucionais de forma eficiente e autônoma, garantindo a privacidade dos dados e baixa latência na detecção de eventos críticos.

Embora o módulo GSM SIM800L tenha se mostrado funcional para a prova de conceito e demonstração do protótipo, sua utilização impõe desafios estratégicos para o futuro. A dependência de redes 2G, que enfrentam um processo de desativação global, compromete a longevidade do sistema. Além disso, os elevados picos de corrente durante a transmissão exigem fontes de alimentação externas, embora este fator não tenha sido um limitador para a estação base deste projeto.

Para a evolução do sistema, recomenda-se o refinamento do modelo por meio da incorporação de dados clínicos reais de pacientes epiléticos e da expansão contínua do conjunto de ADLs, visando superar as limitações das simulações motoras e aumentar a especificidade do diagnóstico. Paralelamente, propõe-se a implementação de técnicas de fusão de sensores com o uso de fotopletismografia (PPG) para correlacionar padrões inerciais com a taquicardia ictal (GLASSTETTER et al., 2021), além da substituição do módulo 2G por tecnologias celulares de maior longevidade, como módulos 4G LTE Cat-1 ou NB-IoT, ou ainda pela eliminação de hardware adicional através do uso do Wi-Fi integrado a APIs de mensagens para o envio de alertas. Por fim, a realização de um estudo piloto com usuários reais torna-se essencial para validar o conforto do dispositivo vestível, a aceitabilidade da solução por parte dos cuidadores e a robustez dos mecanismos de alerta em condições de uso prolongado no cotidiano.

Agradecimentos

Os autores deste trabalho agradecem o apoio desta pesquisa ao laboratório Espaço Maker do IFCE campus Maracanaú e à Pró-reitora de Pesquisa e Inovação do IFCE.

Referências

ARENDS, Johan BAM. Movement-based seizure detection. *Epilepsia*, v. 59, p. 30-35, 2018.

ASADI-POOYA, Ali A. et al. Physical injuries in patients with epilepsy and their associated risk factors. *Seizure*, v. 21, n. 3, p. 165-168, 2012.

BENICZKY, Sándor et al. Seizure semiology: ILAE glossary of terms and their significance. *Epileptic Disorders*, v. 24, n. 3, p. 447-495, 2022.

ESPRESSIF SYSTEMS. ESP-NOW Wireless Communication Protocol, [S.d.]. Disponível em: <https://www.espressif.com/en/solutions/low-power-solutions/esp-now>. Acesso em: 5 fev. 2026.

FÉRRER, H.; BARROS, H. Desenvolvimento de pulseira inteligente para detecção e prevenção de ataques epiléticos . *Revista do Fórum Gerencial*, v. 1, n. 2, p. 11, 20 out. 2021.

FISHER, Robert S. et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia*, v. 58, n. 4, p. 522-530, 2017.

FLEMING, Jane; BRAYNE, Carol. Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. *Bmj*, v. 337, 2008.

GLASSTETTER, Martin et al. Identification of ictal tachycardia in focal motor-and non-motor seizures by means of a wearable PPG sensor. *Sensors*, v. 21, n. 18, p. 6017, 2021.

GOODFELLOW, Ian et al. Deep learning. Cambridge: MIT press, 2016.

HYMEL, Shawn et al. Edge impulse: An ml ops platform for tiny machine learning. arXiv preprint arXiv:2212.03332, 2022.

IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pmlr, 2015. p. 448-456.

JONES, Lliwen A.; THOMAS, Rhys H. Sudden death in epilepsy: Insights from the last 25 years. *Seizure*, v. 44, p. 232-236, 2017.

M5STACK. M5StickC: Documentation and Specifications. 2020. Disponível em: <https://docs.m5stack.com/en/core/m5stickc>. Acesso em: 3 fev. 2026.

MOZAFFARI, N. et al. IoT-based Activity Recognition with Machine Learning from Smartwatch. *International Journal of Wireless & Mobile Networks*, v. 12, n. 1, p. 29–38, 29 fev. 2020.

NIJSEN, Tamara Mathea Elisabeth. Accelerometry based detection of epileptic seizures. 2008.

RAMGOPAL, Sriram et al. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & behavior*, v. 37, p. 291-307, 2014.

RAY, Partha Pratim. A review on TinyML: State-of-the-art and prospects. *Journal of King Saud University-Computer and Information Sciences*, v. 34, n. 4, p. 1595-1623, 2022.

REUSCH, Rafael Schild; JURACY, Leonardo Rezende; MORAES, Fernando Gehm. Assessment and optimization of 1D CNN model for human activity recognition. In: 2022 XII Brazilian symposium on computing systems engineering (SBESC). IEEE, 2022. p. 1-7.

SEVAGJB. WT32-ETH01 + PEO Enclosure. Thingiverse, 4 nov. 2024. Disponível em: <https://www.thingiverse.com/thing:6820274/files>. Acesso em: 23 fev. 2026.

WORLD HEALTH ORGANIZATION. Epilepsy. Genebra: WHO, 7 fev. 2024. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/epilepsy>. Acesso em: 10 dez. 2025.

WORLD HEALTH ORGANIZATION. Step Safely: strategies for preventing and managing falls across the life-course. Geneva: World Health Organization, 2021. Disponível em: <https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/step-safely>. Acesso em: 15 jan. 2025.

ZIA, S. et al. Detection of Generalized Tonic Clonic Seizures and Falls in Unconstraint Environment Using Smartphone Accelerometer. *IEEE Access*, v. 9, p. 39432–39443, 2021.