Classificação de Eventos Sonoros Baseada em Redes Neurais para Monitoramento de Idosos

Victor Gomes Dias, Danilo Brito Teixeira de Almeida, Fabrício Braga Soares de Carvalho e Waslon Terllizzie Araújo Lopes

***Resumo*— Este artigo trata do desenvolvimento de um sistema inteligente capaz de identificar situações críticas envolvendo idosos no ambiente doméstico. Utilizando Redes Neurais Convolucionais treinadas com um banco de dados subdividido em classes de eventos sonoros relevantes, o sistema classifica os áudios captados do ambiente e emite alertas, usando o protocolo MQTT (*Message Queuing Telemetry Transport*), caso uma situação crítica seja identificada.**

***Palavras-Chave*— Reconhecimento de Voz, Redes Neurais, Detecção de Eventos Sonoros, Idosos, MQTT, Sistema de Alerta**

***Abstract*— This article addresses the development of an intelligent system capable of identifying critical situations involving elderly individuals in a home environment. Using Convolutional Neural Networks trained with a database subdivided into relevant sound event classes, the system classifies audio captured from the environment and issues alerts via the MQTT (Message Queuing Telemetry Transport) protocol if a critical situation is detected.**

***Keywords*— Voice Recognition, Neural Networks, Sound Event Detection, Elderly, MQTT, Alert System**

I. Introdução

O crescimento da população idosa tem motivado a criação de tecnologias assistivas voltadas à melhoria da qualidade de vida e segurança, principalmente em ambientes domésticos. Situações como quedas, gemidos de dor ou pedidos de socorro representam riscos sérios, especialmente para idosos que vivem sozinhos. Nesse contexto, sistemas automáticos de reconhecimento de voz e sons ambientais tornam-se ferramentas promissoras para detectar eventos críticos de forma precoce.

Diversos estudos recentes têm explorado a aplicação de redes neurais na classificação de eventos sonoros. Em [7], os autores propõem o uso de redes neurais convolucionais (RNCs) combinadas com diferentes tipos de espectrogramas Mel como entrada. De modo similar, em [6] também é utilizada uma arquitetura RNC para classificação de sons curtos em múltiplos conjuntos públicos (ESC-10, ESC-50 e UrbanSound8K). Por sua vez, em [5], os autores realizam uma análise comparativa entre diferentes arquiteturas de redes neurais, incluindo RNCs, redes neurais recursivas (RNRs) e redes de memórias de longo e purto prazo, aplicadas a múltiplos conjuntos de dados (Audioset, FSD50K, UrbanSound8K e ESC-50). Essas abordagens mostram a eficácia de arquiteturas baseadas em RNCs para reconhecer padrões acústicos em cenários complexos e ruidosos, como os encontrados em ambientes residenciais.

Este artigo descreve o desenvolvimento de um sistema baseado em redes neurais convolucionais, treinadas para reconhecer sons indicativos de risco. Ao detectar um som crítico, o sistema envia um alerta utilizando o protocolo de comunicação MQTT [9] [10]. O modelo do sistema é ilustrado na Figura 1.

As etapas de pré-processamento e classificação de áudio adotadas neste trabalho seguem um fluxo metodológico similar ao descrito em [4], com ênfase específica na identificação de classes sonoras típicas de ambientes domésticos, uma vez que estas são as categorias relevantes para a aplicação proposta. Além disso, a escolha do protocolo de comunicação MQTT foi fundamentada na discussão apresentada em [9], destacando sua leveza, confiabilidade e adequação para sistemas embarcados e aplicações de monitoramento contínuo.

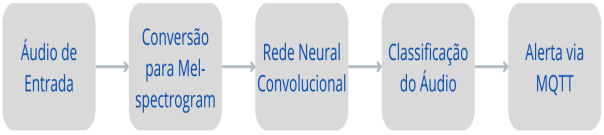

Fig. 1. Modelo do sistema proposto

II. Fundamentação Teórica

As ferramentas necessárias para a implementação do sistema proposto serão descritas a seguir.

A. Espectrograma Mel

Antes de serem entregues à entrada da rede neural, os áudios precisam ser convertidos em *Mel-spectrograms*, uma representação do sinal de áudio no domínio do tempo-frequência, baseada na escala Mel, que se aproxima da percepção auditiva humana. Ele é obtido aplicando a Transformada de Fourier em janelas do sinal de áudio e, em seguida, convertendo as frequências para a escala Mel [2] [3]. O resultado é uma imagem bidimensional na qual o eixo horizontal representa o tempo, o eixo vertical representa as bandas de frequência, e a intensidade das cores indica a energia (ou potência) em cada faixa.

Victor Gomes Dias, Centro de Energias Alternativas e Renováveis (CEAR), UFPB, João Pessoa-PB, e-mail: victorg.dias@estudante.cear.ufpb.br; Danilo Brito Teixeira de Almeida, CEAR, UFPB, João Pessoa-PB, e-mail: danilo.almeida@ee.ufcg.edu.br; Fabrício Braga Soares de Carvalho, CEAR, UFPB, João Pessoa-PB, e-mail: fabricio@cear.ufpb.br; Waslon Terllizzie Araújo Lopes, CEAR, UFPB, João Pessoa-PB, e-mail: waslon@cear.ufpb.br.
Este trabalho foi parcialmente financiado pelo CNPq e CAPES. Os autores agradecem aos entes financiadores, à Universidade Federal da Paraíba e à todos aqueles que contribuíram para o desenvolvimento deste trabalho.

B. Redes Neurais Convolucionais

As Redes Neurais Convolucionais (RNCs) são modelos amplamente utilizados para reconhecimento de padrões em dados com estrutura espacial, como imagens e espectrogramas de áudio. Sua arquitetura é composta por camadas convolucionais, que extraem automaticamente características locais relevantes, seguidas por camadas de *pooling*, que reduzem a dimensionalidade, e camadas totalmente conectadas, responsáveis pela classificação [1]. Durante o treinamento, a rede ajusta seus pesos por meio de retropropagação e otimização, associando padrões extraídos aos rótulos corretos. Após esse processo, a RNC é capaz de realizar a classificação de novos dados, sendo uma abordagem eficiente e precisa em tarefas como classificação de imagens e sons.

C. Comunicação via MQTT

O MQTT é um protocolo de comunicação leve, baseado no modelo publicador/assinante, ideal para aplicações com recursos limitados, como sistemas embarcados e IoTs (*Internet of Things*) [10]. Ele utiliza um *broker* para intermediar a troca de mensagens entre os dispositivos, que publicam e assinam mensagens em tópicos específicos [9].

III. Metodologia

A seguir será abordada a metodologia empregada neste trabalho.

A. Preparação do Conjunto de Dados

Foi utilizado um conjunto de dados composto por áudios organizados nas diferentes classes de interesse, tais como "queda", "gemido", "fala" e "passos". Esses áudios foram obtidos a partir de repositórios públicos disponíveis na internet, como ESC-50 e *Common Voice*, assim como extraídos de vídeos do *YouTube*.

As classes adotadas, e seus respectivos números de amostras, estão dispostos na Tabela I.

TABELA I

Classes e número de amostras adotados.

Classe	Número de Amostras
aspirados	40
bater_porta	40
cao	40
chuva	40
descarga_sanitario	40
escovar_dentes	40
espirro	40
fala	70
gato	40
gemido	40
passos	40
queda	40
tosse	40

Após a separação das amostras nas classes de interesse, os áudios foram previamente processados para extração de características relevantes, sendo convertidos em *Mel-spectrograms*, que capturam a energia do sinal ao longo de frequências na escala Mel. Com a finalidade de limtar o vetor da dados de entrada na rede neural, cada espectrograma gerado foi redimensionado para uma forma fixa com 128 faixas Mel e 256 quadros de tempo, determinando a dimensão do vetor de recursos na RNC [4].

B. Treinamento da Rede Neural

Um modelo de rede neural convolucional foi implementado no ambiente *Python* utilizando a biblioteca *PyTorch*. A arquitetura da rede inclui duas camadas convolucionais *Conv2d*, com 32 e 64 filtros respectivamente, seguidas de operações de *pooling* 2×2 e funções de ativação ReLu. Posteriormente, são aplicadas duas camadas totalmente conectadas *Linear* para a etapa final de classificação [1]. O fluxograma da rede neural com suas operações pode ser visto na Figura 2.

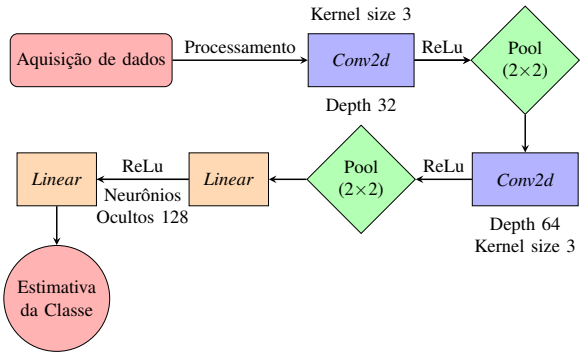

Fig. 2. Fluxograma da rede neural utilizada.

O treinamento foi conduzido com o otimizador *Adam* e a função de perda *CrossEntropyLoss*. Foram adotados 90% dos dados disponíveis para o treinamento, e reservados 10% dos dados para testes. Os dados treinados foram então agrupados em lotes de 32 amostras e apresentados à rede a cada época.

Para a avaliação da estabilidade da rede neural, foram realizados cinco experimentos independentes. Em cada experimento, os dados atribuídos aos grupos de teste e treinamento foram alterados aleatoriamente. As métricas utilizadas para análise dos resultados foram a perda (*loss*) e a acurácia (*accuracy*), permitindo observar a capacidade de generalização do modelo em diferentes execuções.

C. Transmissão de Alertas com MQTT

Para simular a comunicação do sistema com uma central de atendimento, foi utilizado o protocolo MQTT. Um *script* emissor foi implementado para publicar uma mensagem de alerta em um tópico específico (SOS/alerta) sempre que uma situação crítica fosse detectada pela rede neural. Do outro lado, um *script* receptor foi responsável por manter uma conexão aberta com o *broker* MQTT público e aguardar mensagens. Ao receber uma nova publicação no tópico de alerta, o receptor imprime no terminal uma notificação informando a natureza do problema.

IV. Resultados e Discussão

Os resultados dos cinco experimentos realizados são exibidos na Figura 3. Observa-se que, em todos os casos, houve

uma redução significativa da função de perda (*Loss*) ao longo das épocas, indicando uma convergência satisfatória dos pesos sinápticos da rede neural.

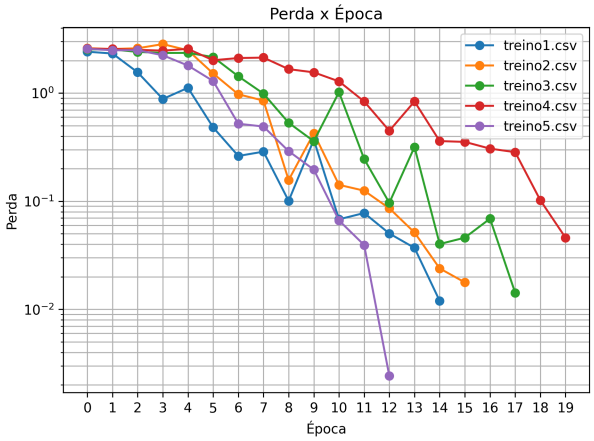

Fig. 3. Curvas de perda em função das épocas.

Ao final dos testes, o segundo experimento (Treino 2) apresentou os melhores resultados. Nesse caso, o modelo obteve uma acurácia de 72,73% e uma perda final de 1,2076 sobre o conjunto de teste, indicando um desempenho satisfatório considerando a simplicidade da arquitetura e o tamanho do conjunto de dados. Esses resultados indicam que o modelo é capaz de distinguir, com boa confiabilidade, eventos sonoros críticos (como quedas e gemidos) e sons cotidianos (como fala ou passos).

A integração entre o sistema de classificação e a comunicação via protocolo MQTT, implementada utilizando a biblioteca *paho* do *Python*, se mostrou eficaz como uma solução viável de alerta remoto. A cada detecção de um evento crítico, o emissor publica uma mensagem de alerta no tópico configurado. O receptor, conectado ao mesmo *broker*, imprime em tempo real a mensagem recebida, simulando o recebimento de uma notificação por uma central de atendimento.

A emissão e recepção de mensagens foram realizadas com sucesso simulando de forma funcional a atuação de um sistema de monitoramento capaz de detectar situações críticas e comunicar uma central de atendimento.

V. Conclusão

Este artigo apresentou o desenvolvimento de um sistema baseado em redes neurais convolucionais para reconhecer sons críticos em ambientes com idosos, cujo objetivo é detectar situações de risco, como quedas e pedidos de socorro. O modelo, treinado com áudios organizados em classes distintas, atingiu uma acurácia de 72,73% nos testes. Além disso, foi implementada a integração com o protocolo MQTT para envio automático de alertas. Os resultados demonstram a viabilidade de uma solução acessível, embora melhorias futuras incluam a ampliação do conjunto de dados, o aperfeiçoamento do modelo e sua aplicação prática em dispositivos embarcados.

Como trabalhos futuros, a perspectiva é melhorar o conjunto de dados, aumentando o número de amostras e incluindo um maior número de classes. Refinar a arquitetura da rede neural para tratamento de dados *online* e incluir uma etapa de aquisição de dados em tempo real na camada de aplicação.

Referências

[1] M. Krichen, “Convolutional neural networks: a survey,” *Computers*, Basel, v. 12, n. 8, art. 151, 2023.

[2] K. Zaman, M. Sah, C. Direkoglu and M. Unoki, “A survey of audio classification using deep learning,” *IEEE Access*, v. 11, p. 106620–106649, 2023.

[3] Q. Zhou, J. Shan, W Ding, C. Wang, S. Yuan, F. Sun, H. Li and F. Fang, “Cough recognition based on mel-spectrogram and convolutional neural network,” *Frontiers in Robotics and AI*, v. 8, 2021.

[4] B. Zhang, J. Leitner and S. Thorton, “Audio recognition using mel spectrograms and convolution neural networks,” *Noiselab University of California*, San Diego, CA, USA, 2019.

[5] G. Mkrtchian and Y. Furletov, “Classification of Environmental Sounds Using Neural Networks, 2022 Systems of Signal Synchronization,” Generating and Processing in Telecommunications (SYNCHROINFO), Arkhangelsk, Russian Federation, 2022, pp. 1-4.

[6] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, USA, 2015, pp. 1-6.

[7] A. Chaturvedi, S. A. Yadav, H. M. Salman, H. R. Goyal, H. Gebregziabher and A. K. Rao, “Classification of Sound using Convolutional Neural Networks,” 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1015-1019.

[8] M. Massoudi, S. Verma and R. Jain,“Urban Sound Classification using CNN,” 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 583-589.

[9] Naik, N, “ Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP,” In: *IEEE International Systems Engineering Symposium (ISSE)*, 2017, Vienna, Austria. IEEE, 2017. p. 1–7.

[10] E. Al-Masri, K. R. Kalyanam, J. Batts, J. Kim, S. Singh, T. Vo, and C. Yan, “Investigating Messaging Protocols for the Internet of Things (IoT),” in IEEE Access, vol. 8, pp. 94880-94911, 2020.