Classificação de Cenários Acústicos em Dispositivos Auditivos: Uma Revisão Sistemática da Literatura sobre Desafios, Avanços e Perspectivas

Thiago Vinícius Macegoza
thiago.2000.macegoza@gmail.com
Universidade Tecnológica Federal do Paraná (UTFPR)
Departamento Acadêmico de Eletrotécnica (DAELT)
Curitiba, Paraná, BRA

José Ricci Neto
jneto.2020@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná (UTFPR)
Departamento Acadêmico de Eletrotécnica (DAELT)
Curitiba, Paraná, BRA

Lívia M. A. Companhoni
Livia.companhoni@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná (UTFPR)
Departamento Acadêmico de Eletrotécnica (DAELT)
Curitiba, Paraná, BRA

Wemerson Delcio Parreira
wemerson.delcio@puc-campinas.edu.br
Pontifícia Universidade Católica de Campinas (PUC-Campinas)
Escola Politécnica
Campinas, São Paulo, BRA

Fábio Pires Itturriet
fabioitturriet@utfpr.edu.br
Universidade Tecnológica Federal do Paraná (UTFPR)
Programa de Pós-Graduação em Engenharia Biomédica (PPGEB), Curitiba, Paraná, BRA

Renata Coelho Borges
renatacoelho@utfpr.edu.br
Universidade Tecnológica Federal do Paraná (UTFPR)
Programa de Pós-Graduação em Engenharia Biomédica (PPGEB), Curitiba, Paraná, BRA

ABSTRACT

Acoustic Scene Classification (ASC) is an area of growing relevance, with applications ranging from assistive devices, such as hearing aids, to advanced wearable technologies (hearables). This paper presents a Systematic Literature Review (SLR) that analyzes the main adaptive and machine learning-based methods used in ASC, with a focus on hearing devices. The challenges related to computational resource limitations, energy consumption and real-time operation, especially in dynamic environments, are discussed. The review highlights recent advances, such as the use of generative probabilistic models and convolutional neural networks, as well as hybrid approaches that combine cloud computing and edge computing for greater efficiency. The results show that, despite significant progress, there are still important technical barriers, such as the need for more efficient, customizable and robust algorithms to operate in real conditions. This study contributes by identifying gaps in the literature and suggesting future directions to improve the integration of ASC in hearing devices.

KEYWORDS

Acoustic Scene Classification, Hearing Aids, Neural Networks

1 INTRODUÇÃO

A Classificação de Cenários Acústicos (*Acoustic Scene Classification* – ASC) tem se destacado com uma área crucial na pesquisa em processamento de áudio, impulsionada pelos avanços recentes em inteligência artificial [1]. As técnicas que a compõem envolvem a identificação e categorização de diferentes ambientes sonoros a partir de gravações de áudio. Em termos práticos, sistemas de ASC são capazes de distinguir ambientes como ruas movimentadas, ambientes internos tranquilos como escritórios ou bibliotecas, ou até locais religiosos, como igrejas, baseando-se nos padrões acústicos únicos de cada contexto. O processo de classificação inicia com a captura do áudio e avança para um processamento de sinal robusto capaz de analisar características sonoras, como frequências, reverberações e a presença de sons específicos que caracterizam o ambiente em questão [2].

Os avanços nas técnicas de ASC se dão principalmente devido ao uso cada vez maior de inteligência artificial, em particular com o uso de redes neurais profundas (*deep learning*) e outras técnicas de aprendizado supervisionado, que permitem um alcance maior em eficácia e precisão na classificação de variados ambientes [3–5], identificando padrões complexos e sutis que seriam difíceis de serem capturados por métodos tradicionais. O resultado disso é uma melhora significativa na eficácia da classificação, que tem se mostrado cada vez maior com o uso de algoritmos mais robustos [6–8], mesmo em cenários dinâmicos e complexos.

Neste contexto, a utilização de estratégias de processamento como *cloud computing* e *edge computing* podem ser promissoras para superar as limitações computacionais dos dispositivos auditivos [9]. O *cloud computing* permite que o processamento intensivo seja transferido para servidores remotos, reduzindo a carga nos dispositivos locais e possibilitando o uso de modelos mais complexos e precisos. Já o *edge computing*, por realizar o processamento mais próximo do dispositivo, oferece vantagens como baixa latência, respostas em tempo real, operação eficiente mesmo em ambientes com conectividade limitada e maior flexibilidade e escalabilidade. Embora implementações baseadas em *edge computing* apresentem acurácia e precisão comparáveis a aplicações locais, sua capacidade de operar em tempo real, mesmo em contextos dinâmicos e desafiadores, faz com que essa abordagem seja ideal para sistemas de Classificação de Cenários Acústicos [10].

O DCASE (*Challenge on Detection and Classification of Acoustic Scenes and Events*) tem sido um marco importante no desenvolvimento da ASC, promovendo avanços consideráveis ao estimular a inovação em técnicas de aprendizado de máquina aplicadas à detecção e classificação de cenários e eventos acústicos. A competição anual tem permitido uma melhoria constante nos modelos de classificação e na eficácia dos sistemas, destacando a importância crescente dessa área de pesquisa [11].

Desse modo, este trabalho busca analisar o estado da arte atual da ASC, pontuar os principais desafios encontrados nesta área e detectar soluções inovadoras que permitem a implementação dessa tecnologia em dispositivos auditivos e *hearables*, de forma a tornar a ASC uma funcionalidade viável e eficiente. Este levantamento foi pautado em uma Revisão Sistemática da Literatura (RSL), que tem como objetivo identificar, avaliar e interpretar a pesquisa relevante e recente relacionada ao tema de estudo.

2 CLASSIFICAÇÃO DE CENÁRIOS ACÚSTICOS

A capacidade de reconhecer e classificar ambientes com base nos padrões sonoros presentes tem aplicações que vão desde a melhoria de sistemas de assistência auditiva até a criação de dispositivos inteligentes, como *hearables*, que são capazes de adaptar sua operação a diferentes contextos acústicos. Com o avanço das tecnologias, os métodos de ASC têm alcançado índices de precisão superiores a 90% em ambientes controlados, utilizando abordagens como aprendizado supervisionado e modelos de *deep learning*, o que demonstra a eficácia da técnica em identificar e categorizar cenários acústicos de maneira bastante precisa [1].

O mercado de *wearables* e dispositivos auditivos tem se expandido significativamente, com o setor de *wearables* projetado para atingir em US$ 186,14 bilhões em 2030, crescendo a uma taxa composta anual de 14,6% entre 2023 e 2030 [12]. O mercado de aparelhos auditivos, por sua vez, deve alcançar US$ 12,57 bilhões até 2030, com uma taxa de crescimento de 15% ao ano, a partir de 2024 [13]. Esse crescimento reflete uma necessidade na demanda por tecnologias que possam ser integradas a esses dispositivos, incluindo soluções de ASC, que oferecem o potencial de melhorar a experiência auditiva, oferecendo uma percepção mais rica e adaptativa do ambiente sonoro ao redor do usuário.

No entanto, a introdução de ASC em dispositivos com recursos limitados, como aparelhos auditivos e *hearables*, exige uma adaptação dos algoritmos para que possam operar de forma eficiente dentro das restrições de bateria e potência de processamento. Em dispositivos auditivos, por exemplo, o consumo de energia deve ser mantido em níveis baixos para garantir uma longa duração de uso, enquanto a capacidade de processar informações em tempo real é essencial para que a ASC funcione de maneira eficaz. A maioria dos métodos de ASC baseados em *deep learning* exige uma quantidade significativa de processamento computacional, o que representa um desafio técnico importante. Dados recentes apontam que a eficiência energética é uma das principais barreiras para a adoção de algoritmos de inteligência artificial em dispositivos vestíveis e auditivos, e ainda há uma grande lacuna em otimizar essas tecnologias para que possam funcionar sem comprometer a duração da bateria e o desempenho, mesmo que implementadas utilizando técnicas de *cloud/edge computing* [14–16].

Apesar desses desafios técnicos, a ASC oferece uma série de benefícios potenciais, especialmente no contexto de saúde auditiva. Hoje, mais de 1,5 bilhão de pessoas no mundo sofrem de alguma forma de deficiência auditiva, e 466 milhões de pessoas têm perda auditiva incapacitante, segundo dados da Organização Mundial da Saúde [17]. Para essas pessoas, a ASC poderia melhorar significativamente a qualidade de vida, permitindo que aparelhos auditivos ou dispositivos vestíveis possam detectar e ajustar automaticamente o som ambiente, dando prioridade a sons de interesse, como conversas humanas, e atenuando o ruído de fundo. Em um cenário urbano, por exemplo, a ASC poderia identificar a aproximação de um veículo e alertar o usuário com uma vibração ou um aumento de volume no fone de ouvido, proporcionando maior segurança e conforto. Já em ambientes como escritórios ou salas de aula, a tecnologia poderia otimizar o ambiente sonoro, ajustando automaticamente o volume do dispositivo de acordo com o nível de ruído ambiente, melhorando a concentração do usuário.

3 REVISÃO SISTEMÁTICA DA LITERATURA

O objetivo fundamental desta revisão sistemática da literature (RSL), é identificar, avaliar e interpretar as pesquisas disponíveis relevantes para uma questão de pesquisa específica, área temática ou fenômeno de interesse [18]. A RSL, com base nos preceitos de Kitchenham, adotada neste trabalho, compreende três fases distintas, sendo elas: planejamento, condução e relato da revisão.

A metodologia utilizada para organizar esta RSL foi pautada na plataforma Parsifal[1] (*Perform Systematic Literature Reviews*), uma aplicação *on-line* que permite estruturar e classificar os critérios e resultados de pesquisa, oferecendo assim recursos que permitem ao pesquisador acelerar as etapas iniciais de estudo. Nesta plataforma, as etapas desta RSL serão abordadas individualmente.

3.1 Planejamento

Na fase de planejamento, foi adotado um protocolo detalhado, que incluiu a definição clara dos objetivos da pesquisa, a construção da estratégia PICOC, a formulação das questões de pesquisa e a seleção de palavras-chave e sinônimos. Além disso, a sequência de busca foi delineada, juntamente com as bases de dados a serem consultadas, e os critérios de inclusão e exclusão dos estudos foram estabelecidos.

Assim, foi estabelecida uma descrição objetiva e sucinta para a RSL: "identificar, analisar e comparar os métodos adaptativos e baseados em *machine learning* utilizados na classificação de cenas acústicas em aparelhos auditivos, a fim de entender a relação entre os métodos de classificação e a eicácia nos resultados de classificação". Após a definição do objetivo da pesquisa, foi elaborado o PICOC —- estratégia utilizada para construir perguntas de pesquisa em diversas área para ajudar a se recordar o que esta deve solucionar e especificar e encontrar informações relevantes para a pergunta alvo —- conforme a estrutura apresentada na Tabela 1.

Uma vez que o PICOC foi finalizado, foram definidas as questões pertinentes ao tema de pesquisa, com o propósito de delimitar a área de busca. Essas questões estão apresentadas na Tabela 2. Na etapa seguinte, foram elaboradas as palavras-chave, sinônimos e

[1] https://parsif.al/

Tabela 1: Descrição da estratégia PICOC.

Ac.	Definição	Descrição
P	*Population*	Aparelhos auditivos, *hearables*
I	*Intervention*	Abordagens e métodos de classificação de cenários acústicos
C	*Comparison*	Abordagens e algoritmos de classificação de cenários acústicos já utilizados ou propostos para dispositivos auditivos
O	*Outcomes*	Precisão na classificação dos cenários acústicos, aplicabilidade prática dos métodos, impacto na experiência auditiva dos usuários, e desempenho dos modelos em condições reais
C	*Context*	Cenários acústicos variados (ex.: ambientes com fala, ruído ambiente, música), com foco em ambientes reais e condições práticas para dispositivos auditivos

definidos os critérios de relacionamento delas na *string* de pesquisa a ser utilizada nas bases de dados. As palavras-chave, juntamente com seus sinônimos, foram criteriosamente selecionadas para abranger toda a área de interesse, sendo descritas na Tabela 3.

Tabela 2: Perguntas de pesquisa.

Ordem	Pergunta
1	Como integrar a classificação de cenários acústicos com aparelhos auditivos ou *hearables*?
2	Como os métodos adaptativos de classificação de cenários acústicos contribuem para melhorar a experiência auditiva, especialmente em ambientes dinâmicos?
3	Quais os melhores métodos para programar uma inteligência artificial neste cenário?
4	O autor explica a metodologia utilizada no estudo?
5	O autor compara técnicas do estudo aplicadas ao contexto dos aparelhos auditivos?

Tabela 3: Palavras-chave e sinônimos.

Palavra-chave	Sinônimo
Hearing aids	Earphones, Hearables
Acoustic Scene	Acoustic scenario, Audio scene, Auditory scene, Scene-based sound
Classification	Analysis, Detection

Essas palavras-chave, juntamente com seus respectivos sinônimos, foram utilizadas para compor a frase de busca. Esse processo exigiu cuidado, pois a frase foi utilizada para direcionar as bases de dados ao foco da pesquisa. Por isso, ela foi formulada de maneira precisa, com o uso dos conectores 'OR' e 'AND', resultando no seguinte formato final: ("*Acoustic Scene*" OR "*Acoustic scenario*" OR "*Audio scene*" OR "*Auditory scene*" OR "*Scene-based sound*") AND ("*Classification*" OR "*Analysis*" OR "*Detection*") AND ("*Hearing Aids*" OR "*Earphones*" OR "*Hearable*"). Essa frase de busca foi então aplicada nas bases de dados descritas na Tabela 4, com exceção da ScienceDirect. Esta base, especificamente, possui uma particularidade na construção da frase de busca, pois permite no máximo 8 conectores booleanos. Desse modo, foi necessário reformular a *string* para que se adequasse a este padrão, ocasionando no seguinte formato: ("*Acoustic Scene*" OR "*Acoustic scenario*" OR "*Scene-based sound*") AND ("*Classification*" OR "*Analysis*" OR "*Recognition*") AND ("*Hearing Aids*").

Por fim, foram definidos os critérios de inclusão e exclusão, conforme apresentado na Tabela 5. Esses itens foram utilizados de modo a servir como base para a seleção dos artigos obtidos das bases de dados.

Tabela 4: Bases de dados.

Bases	Link
ACM Digital	http://dl.acm.org/
Engineering Village	https://engineeringvillage.com/
IEEE Xplore	http://ieeexplore.ieee.org/
ScienceDirect	http://sciencedirect.com/
Scopus	http://scopus.com
Springer	http://springer.com/
Web of Science	https://webofknowledge.com/

Tabela 5: Critérios de inclusão e exclusão.

Inclusão	Exclusão
Estudos relacionados a classificação de ambientes acústicos	Estudos duplicados
Estudos completos	Literatura cinzenta
Estudos primários	Publicação anterior a 2019
	Fora do escopo de estudo

3.2 Condução

Com a conclusão da fase de planejamento, foram feitas as buscas nas bases apresentadas na Tabela 4. Essa etapa resultou em 139 artigos, distribuídos percentualmente como apresentado na Figura 1. Observa-se que a base ELSEVIER contribuiu com a maior parte dos artigos coletados, representando 38,1% do total, seguida pela Scopus (23%), Engineering Village (14,4%), IEEE Digital Library (13,7%) e ISI Web of Science (10,8%). As bases ACM digital e Springer não retornaram artigos com nenhuma das *strings* de busca geradas. Com os artigos selecionados, iniciou-se um protocolo de triagem. Essa etapa tem como objetivo refinar o processo inicial de busca de artigos, de tal forma que se eliminasse os estudos fora de contexto. Este processo esta definido na Figura 2.

O primeiro passo dessa triagem foi estabelecer uma base de dados interna. Essa base foi constituída com os artigos obtidos na fase de

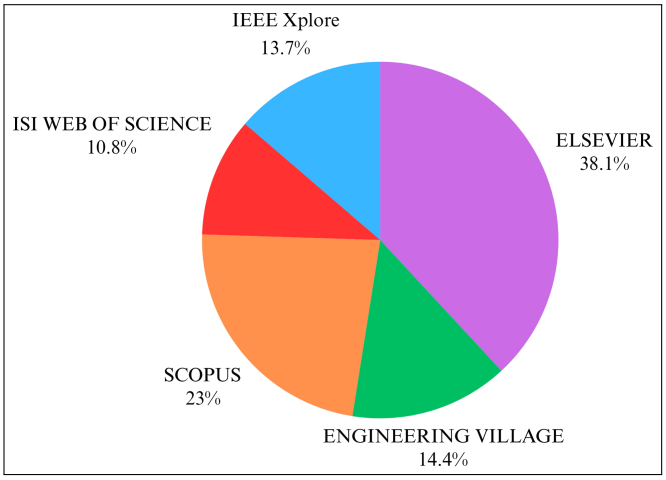

Figura 1: Porcentagem de artigos por base

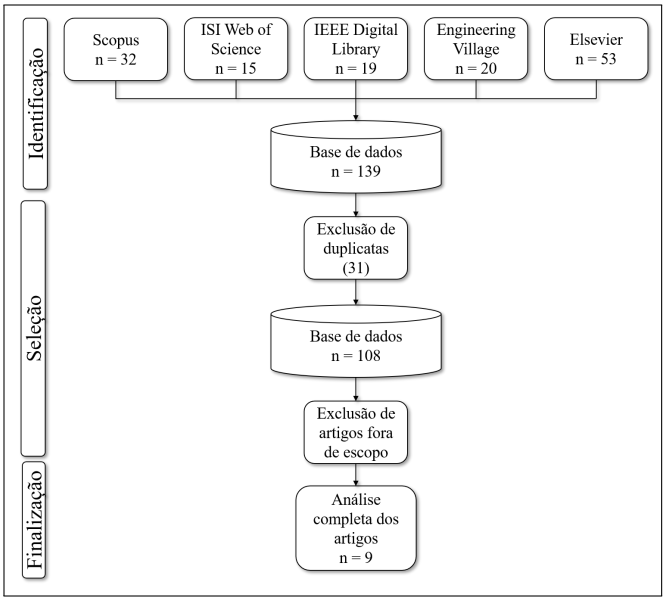

Figura 2: Fluxograma das decisões tomadas

busca. Uma vez que essa base estava estruturada, foi realizado um processo de análise de duplicatas, que contabilizou 31 estudos. Com isso, a base interna passou a conter 108 artigos restantes.

Na etapa seguinte, foi iniciado o processo de leitura cuidadosa dos títulos, *abstracts* e resultados desses artigos, de forma a determinar a relevância de cada um deles em relação ao tema central. Esse processo eliminou todos os artigos que estavam fora do escopo e que não possuíam contribuições relevantes para essa revisão, resultando em 9 artigos. A Figura 3 mostra a relação entre os artigos inicias e os que foram aceitos após a triagem, separados por base de dados.

Com a conclusão dessa etapa, o próximo, e último passo, foi pautado em uma busca por informações cruciais para o tema contidas em cada artigo, a partir de uma leitura completa. Para isso, foram

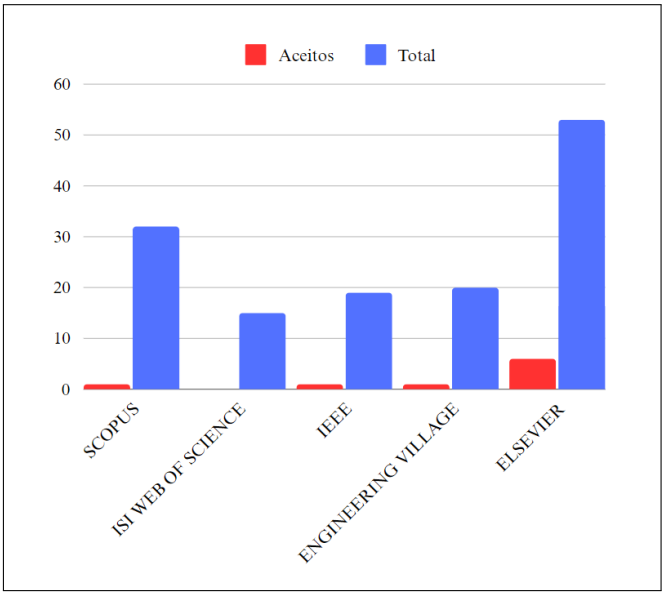

Figura 3: Relação entre total e aceitos por base.

definidas as perguntas de pesquisa, descritas na Tabela 6. Nesta etapa não houve exclusão de artigos.

Tabela 6: Perguntas de busca nos artigos.

Ordem	Pergunta
1	Qual o objetivo principal do estudo?
2	Qual o método de classificação de cenários acústicos abordado neste estudo?
3	Qual foi o algoritmo ou técnica utilizada?
4	Quais foram os tipos de cenários acústicos utilizados para classificação?
5	Quais foram as métricas reportadas para avaliar o desempenho do classificador utilizado?
6	Os autores identificam limitações do estudo?

3.3 Resultados

Nesta seção, são apresentados os resumos dos 9 artigos que avançaram para a etapa de leitura e análise completa, contemplando os principais objetivos, metodologias adotadas e limitações de cada um deles.

A - *A new pyramidal concatenated CNN approach for environmental sound classification* (2020) [6]:

O artigo explora o reconhecimento de som ambiental por meio da comparação e detecção da melhor abordagem de CNN profunda: primeiro, com a remoção de ruído dos sinais de entrada, que foram convertidos em imagens sonoras pelo método *Short-Time Fourier Transform* (STFT); em seguida, um modelo CNN pré-treinado foi utilizado para extração de características das imagens de forma piramidal. Por último, um classificador baseado em aprendizado de máquina foi usado para ESC. Foram utilizados três bancos de dados

diferentes: ESC-10, ESC-50 e UrbanSound8K. Também foram empregadas três CNNs pré-treinadas distintas: VGGNet16, VGGNet19 e DenseNet201, com diferentes divisões de classificação: 10 classes, 50 classes e 10 classes, respectivamente. Além disso, os sinais foram divididos em diferentes classificações para a obtenção dos resultados de precisão dos três modelos de CNN. Dessa forma, diferentes classificadores foram executados, e a precisão média variou entre 40,4% e 72,5%. Os resultados mostraram que os modelos de fusão tardia em ASC podem produzir modelos mais precisos em comparação com CNNs comuns. No entanto, trabalhar com dados reais será um desafio futuro para o método.

B - *Late fusion for acoustic scene classification using swarm intelligence* (2022) [7]:

Introduz um método de fusão tardia que visa melhorar o desempenho da ASC e extrair as maiores vantagens de cada modelo, utilizando as previsões existentes. Conforme demonstrado na Tabela 5, o estudo difere dos demais por aplicar o Algoritmo de Embaralhamento de Salto Aleatório de Sapos (*Shuffled Frog Leaping Algorithm* - SFLA) como o principal nesta aplicação. Os autores também apontam que, embora as áreas de processamento de áudio como reconhecimento de fala e classificação musical estejam em patamares mais avançados de estudos, o desempenho da ASC ainda é muito inferior comparada a elas, o que inevitavelmente ocasiona certas limitações. Além disso, por se tratar de um método de fusão tardia, existe ainda a possibilidade de uma degradação do desempenho devido a disputa nas previsões sobre os mesmos dados entre os modelos. Apesar das limitações apontadas pelos autores, a metodologia proposta neste estudo obteve uma acurácia de 87,64% com a base de dados do DCASE (*Challenge on Detection and Classification of Acoustic Scenes and Events*) 2019, o que representa uma melhoria de 2,37% em relação ao melhor modelo individual e de 25,63% em comparação com o sistema de referência do DCASE 2019.

C - *A novel acoustic scene classification model using the late fusion of convolutional neural networks and different ensemble classifiers* (2020) [8]:

Apresenta um modelo de ASC que utiliza a fusão tardia entre Redes Neurais Convolucionais (*Convolution Neural Networks* - CNNs) e diferentes classificadores em conjunto (*ensemble classifiers*), visando uma melhor classificação em comparação com o uso de apenas um modelo. Além disso, é possível observar que o conjunto de dados DCASE 2017, da base de dados *TUT Acoustic Scenes* 2017, utilizado no estudo, contém um número relativamente pequeno de gravações acústicas e amostras de áudio com duração muito longa em relação aos outros estudos, como demostrado na 10. Neste estudo, a acurácia final média das CNNs foi de 72,9% com um desvio padrão de ±20%, demostrando o menor percentual entre os artigos estudados, conforme mostrado na Tabela 7. Além disso, a acurácia média dos modelos de classificadores em conjunto variou entre 40,4% e 76,5%. O classificador de subespaço aleatório obteve a maior acurácia média (76,5% com desvio padrão de ±18%) e, quando combinado com o modelo CNN, alcançou uma acurácia média de 80%. Este modelo de fusão apresentou um aumento médio na precisão de 10% em relação à acurácia média do modelo CNN individual. Além disso, os autores mencionam que quando comparado a estudos anteriores, que utilizaram modelos de fusão precoce de CNN, o modelo de fusão tardia mostrou uma melhoria de pelo menos 7% na acurácia média.

D - *Deep learning based source identification of environmental audio signals using optimized convolutional neural networks* (2023) [9]:

Esse artigo apresenta como proposta o aumento da precisão e da eficiência na classificação de sons ambientais utilizando técnicas de aprendizado profundo. O modelo explorado é baseado em CNN e são introduzidas novas técnicas para extração de características e otimização dos sinais, como a Transformada Discreta de Fourier (I-DFT). O banco de dados usado é o *Urban Sound8K* (US8K) e para este metodo foi aplicado em *Deep Learning* com suporte de técnicas de *Machine Learning* e com 10 categorizações de sons ambientais. No estudo observa-se uma alta complexidade computacional e é mencionado a dificuldade de se adequar a cenários reais. No entanto, os resultados de acurácia apresentados das simulações foram de 94,15%, demonstrando a eficácia do método.

E - *Acoustic scene classification from few examples* (2019) [19]:

Desenvolve um modelo Probabilístico Generativo (PG) baseado em um Modelo Semi-Markov Oculto (*Hidden Semi-Markov Model* - HSMM) para a classificação de cenas acústicas personalizáveis, adequado para ser executado em condições *in-situ* em dispositivos de baixo consumo de energia, como aparelhos auditivos. O modelo é projetado para funcionar com um número muito reduzido de exemplos de treinamento, permitindo a classificação eficaz de ambientes acústicos a partir de uma única observação de 10 segundos por ambiente, no entanto, foram relatadas dificuldades em distinguir entre categorias acústicas muito semelhantes, como, por exemplo, entre uma biblioteca e um escritório. Além disso, a aprendizagem limitada do classificador acústico a partir de uma única ou poucas gravações de ondas acústicas é um fator limitante apontado por este estudo. Ainda assim, a acurácia obtida pelo método de classificação proposto foi de 51% em um modo de aprendizado de um único exemplo (*one-shot learning*) e a acurácia do classificador HSMM aumentou gradualmente com o número de exemplos de treinamento, chegando a 71% quando foram apresentados 5 exemplos rotulados para cada classe.

F - *Hierarchical classification for acoustic scenes using deep learning* (2023) [20]:

O artigo divide a ASC em tarefas de 3 classes(interna, externa e transporte), que são chamadas de alto nível, e em 10 classes, que são chamadas de nível baixo e são as subclasses dentro das categorias de alto nível. E utiliza métodos de classificação hierárquica em ASC para otimizar seu desempenho em níveis baixos. As *frameworks* da ASC são descritas de acordo com o uso da hierarquia, e são investigados 14 métodos de aumento de dados e suas combinações. O estudo explora ainda a fusão híbrida para avaliar classificações incorretas e apresenta o método ASC hierárquico baseado em fusão tardia otimizada. Os pesquisadores aplicaram os metodos em dois bancos de dados diferentes para testar a eficacia TAU Urban Acoustic Scenes 2019 e 2020. No entando houve uma parcela de dados (cerca de 10%) que não foram usados no estudo, pois eram incompatíveis entre os dois bancos de dados que reduziu o desempenho do metodo aplicado em ASC.

G - *Two-level fusion-based acoustic scene classification* (2020) [21]:

Tabela 7: Métricas reportadas.

Artigos	*Acurácia*	*Matriz de Confusão*	*Outras Métricas*	*Resultado*
[6]	✓	✓	ROC, AUC	40,4% a 72,5 %
[7]	✓	✓		87,64%
[8]	✓	✓	Cross-validation (10 dobras)	72,9%
[9]	✓		Precisão, Sensibilidade	94,15%
[19]	✓			71%
[20]	✓	✓	Log loss	
[21]	✓	✓		81,0% / 70%
[22]	✓		Probabilidade de classificação correta	
[23]	✓			93,84%

Desenvolve um sistema hierárquico de cenas acústicas, partindo de uma visão ampla com ambientes conhecidos para uma visão específica. Ou seja, utiliza uma abordagem híbrida, como demonstrado na Tabela 9, usando métodos baseados em características acústicas e *machine learning*. É observado que a propagação de erros e imprecisões são fatores que dão ao estudo uma margem de resultados inconsistentes em relação a outras bases de dados e a utilização em cenários reais, diminuindo a eficiência da pesquisa. Apesar disso, a acurácia obtida pelo método de classificação proposto neste estudo foi de 81,0% no conjunto de avaliação TUTAS16 e 70,0% no conjunto de avaliação TUTAS17, conforme mostrado na Tabela 7.

H - *Environmental Classification in Hearing Aids* (2021) [22]:

O estudo utiliza-se de métodos como o modelo oculto de Markov (*Hidden Markov Model* – HMM), análise de *cluster* e Bayesianos e examina a precisão e o desempenho dos classificadores automáticos em aparelhos auditivos. Utiliza a base de dados simulados e reais de aparelhos auditivos, como demonstrado na Tabela 8, para comparar a aplicação dos métodos a um referencial humano e mapeia como esse humano reage a estímulos ambientais reais para alcançar um resultado mais preciso. Ainda assim, há dificuldades relacionadas à variabilidade entre diferentes aparelhos auditivos, além de suscetibilidade a classificações incorretas em cenários complexos.

I - *Acoustic Scene Classification in Hearing aid using Deep Learning* (2020) [23]:

Desenvolve um sistema de classificação de cenas acústicas para aparelhos auditivos utilizando *deep learning* e ajustes automáticos para diferentes ambientes. Como o estudo utiliza uma divisão em apenas 5 classes distintas (música, ruído, fala, fala com ruído e silêncio) existe a redução de precisão em classes com características similares, mas ainda assim o modelo conseguiu atingir uma acurácia de 93,84%. Ainda que o modelo tenha obtido uma acurácia elevada, os autores chamam a atenção quanto as restrições de processamento pela aplicação do modelo em *hardware* de aparelhos auditivos.

3.4 Discussões

Ao analisar os artigos selecionados, é possível agrupá-los em quatro categorias principais: Algoritmo Utilizado, Método de Classificação, Métricas Reportadas e Dados de Entrada (Banco de Dados, Frequência e Duração por Amostra), como demonstrado nas Tabelas 7, 8, 9, e 10. Essa estrutura permite uma comparação mais organizada dos diferentes enfoques adotados pelos estudos, facilitando a identificação de tendências e pontos de melhoria em futuras pesquisas.

Tabela 8: Algoritmos utilizados.

Artigos	*CNN*	*SVM*	*HMM*	*SFLA*
[6]	✓	✓		
[7]				✓
[8]	✓			
[9]	✓			
[19]			✓	
[20]	✓			
[21]		✓		
[22]			✓	
[23]	✓			

Tabela 9: Método de Classificação.

Artigos	*ML*	*DL*	*Híbrido*	*PG*
[6]		✓		
[7]			✓	
[8]		✓	✓	
[9]		✓		
[19]				✓
[20]		✓	✓	
[21]			✓	
[22]	✓			
[23]		✓		

A utilização das CNNs destaca-se como a abordagem predominante na maioria dos estudos sobre ASC. Os artigos [6, 8, 9, 20, 23] aplicam essa técnica, sendo frequentemente combinada com outros modelos para melhorar a classificação dos sinais acústicos. O estudo [6], por exemplo, combina CNN com SVM, alcançando resultados superiores a 80% de acurácia nos testes realizados. A combinação dessas abordagens demonstra a eficácia da fusão de modelos, mas também evidencia que, em alguns casos, a CNN sozinha pode não ser suficiente para alcançar um desempenho ideal, sugerindo

[1] Probabilístico/Generativo

Tabela 10: Dados de Entrada: Bancos de Dados, Frequência e Duração por Amostra.

Artigos	*Bancos de Dados Utilizado*	*Frequência (kHz)*	*Duração por Amostra*
[6]	ESC-10, ESC-50, UrbanSound8K	44.1	< 4s
[7]	TAU Urban Acoustic Scenes 2019	44.1	10s
[8]	TUT Acoustic Scenes 2017	44.1	3-5 minutos
[9]	UrbanSound8K	22.05	4s
[19]	TUT Acoustic Scenes 2017	44.1	10s
[20]	TAU Urban Acoustic Scenes 2019 & 2020 Mobile	44.1	10s
[21]	DCASE 2016/2017(TUTAS16/TUTAS17)	44.1	30s/10s
[22]	Dados simulados e reais de aparelhos auditivos	22.05	Não especificado
[23]	Freesound, Million Song Database, LibriSpeech ASR Corpus	Não especificado	Não especificado

a necessidade de explorar outras alternativas ou otimizações. A comparação entre os resultados dos diferentes artigos revela que, embora as CNNs sejam poderosas, o desempenho pode ser sensível a fatores como a escolha dos dados e a configuração dos modelos.

Outro ponto interessante é o uso de IA em conjunto com HMM, como é o caso do artigo [22]. Embora o estudo apresente uma proposta inovadora, a ausência de dados completos sobre a acurácia final dificulta a avaliação quantitativa da eficiência do método. Isso aponta para uma lacuna importante nos estudos, onde a transparência dos resultados precisa ser melhorada para garantir comparações justas e a reprodutibilidade dos experimentos. A falta de dados claros sobre a acurácia compromete a robustez da análise e a confiança nas conclusões do estudo.

Quanto aos métodos de classificação, observa-se uma prevalência das abordagens de *Deep Learning* (DL) e híbridas, conforme mostrado nas Tabelas 9 e 7. Em [8, 20], por exemplo, ambos os métodos foram aplicados simultaneamente, buscando aprimorar a precisão dos modelos. No entanto, os resultados obtidos por [8] ficaram aquém das expectativas, o que pode estar relacionado a diversos fatores, como a escolha das amostras de áudio, já que o estudo utilizou amostras mais longas. Isso sugere que a combinação de métodos de *deep learning* e híbridos pode ser mais eficaz quando aplicada a conjuntos de dados de maior duração, mas também revela que, em certos contextos, pode ser necessário ajustar as abordagens para alcançar melhores resultados, especialmente quando lidamos com amostras menores ou de menor qualidade.

Os algoritmos híbridos se destacam por serem uma abordagem em inteligência artificial que combina diferentes técnicas e algoritmos para resolver problemas complexos, fazendo uso de métodos baseados em regras ou aprendizado de máquina. Essa solução visa alcançar resultados mais precisos e eficientes, chegando a lugares que antes não seriam possíveis com o uso das técnicas isoladamente, pois a fusão entre diferentes algoritmos extrai as maiores vantagens de cada um [7, 8].

As métricas utilizadas para avaliação, particularmente a acurácia, são um ponto de destaque, como apresentado na Tabela 7. Embora todos os estudos analisem a acurácia, nem todos fornecem dados completos sobre os resultados, o que prejudica a interpretação comparativa. Os artigos analisados mostram uma faixa de desempenho que vai de 40,4% a 93,84% de acurácia, o que sugere que, embora os resultados sejam promissores, ainda há uma margem significativa de variação dependendo do modelo e dos dados utilizados. A análise das métricas poderia ser complementada com outras abordagens, como a análise de precisão, revocação ou F1-score, para uma avaliação mais robusta dos métodos.

Para finalizar, observou-se que a utilização e escolha dos bancos de dados utilizados têm grande impacto nos resultados, especialmente em termos de duração das amostras e frequência. Como observado nos artigos [8, 19], que utilizaram os mesmos bancos de dados, mas com durações diferentes de amostras, a comparação entre os modelos fica mais precisa ao observar o impacto da duração nas métricas. Embora a frequência utilizada nos estudos seja predominantemente uniforme, com exceção dos artigos [9, 22, 23], que reportaram ou utilizaram frequências significativamente mais baixas, a uniformidade na escolha da frequência também facilita a comparação entre os resultados. Contudo, a falta de informações detalhadas sobre as frequências em alguns artigos sugere uma necessidade de maior padronização na coleta de dados para garantir a reprodutibilidade dos experimentos.

4 CONCLUSÕES

A revisão sistemática da literatura realizada neste estudo teve como objetivo organizar e apresentar os estudos mais recentes sobre a classificação de cenários acústicos aplicados a aparelhos auditivos, com o intuito de fornecer uma base sólida para o desenvolvimento de melhores análises de sinais acústicos. O foco desse desenvolvimento é melhorar a qualidade de vida da comunidade surda por meio de algoritmos mais elaborados, permitindo uma percepção mais precisa e satisfatória dos ambientes sonoros ao seu redor. Para alcançar esse objetivo, a metodologia de revisão adotada seguiu etapas rigorosamente definidas e apontadas na Seção 3.1, o que proporcionou uma análise consistente e relevante dos artigos selecionados.

Por meio da metodologia de revisão, observou-se que a integração da ASC com aparelhos auditivos têm grande potencial para melhorar a experiência do usuário, pois são capazes de ajustar os classificadores conforme o contexto acústico, otimizando o desempenho do aparelho auditivo em diferentes cenários, conforme apontado por [19]. Este é um ponto fundamental para a personalização das experiências auditivas, essencial para proporcionar a melhor qualidade de vida aos usuários de aparelhos auditivos.

Os resultados da revisão também indicam que os métodos baseados em Redes Neurais Convolucionais são os mais prevalentes entre os estudos revisados, como destacado na Tabela 9. Isso sugere que

as CNNs são uma abordagem promissora para classificação acústica em dispositivos auditivos e devem ser consideradas como um caminho central para futuras pesquisas. No entanto, a análise também revelou desafios como a complexidade computacional, dificuldade de aplicar os métodos em ambientes reais e limitação dos bancos de dados, que muitas vezes apresentam uma quantidade reduzida de amostras dificultando a diferenciação de cenários acústicos muito semelhantes.

Apesar desses desafios, os estudos analisados apontam que as soluções atuais têm um grande potencial para ser integradas em diferentes estágios do processo de classificação, criando sistemas mais robustos. A combinação de modelos e algoritmos pode, assim, proporcionar uma experiência mais realista e dinâmica para os usuários. Isso permitiria, por exemplo, a participação ativa em shows, eventos religiosos, festivais e outras experiências sociais, que são muitas vezes inacessíveis para a comunidade surda. Portanto, as futuras pesquisas devem focar em superar as limitações atuais, como a complexidade computacional, a personalização dos classificadores e a utilização de bases de dados mais representativas e variadas. Além disso, um direcionamento importante para a pesquisa futura é a exploração de métodos híbridos e modelos probabilísticos generativos que consigam equilibrar o desempenho dos classificadores com as restrições de hardware, especialmente em dispositivos como aparelhos auditivos.

REFERÊNCIAS

[1] Biyun Ding, Tao Zhang, Chao Wang, Ganjun Liu, Jinhua Liang, Ruimin Hu, Yulin Wu, and Difei Guo. Acoustic scene classification: a comprehensive survey. *Expert Systems with Applications*, page 121902, 2023.

[2] Yizhou Tan, Haojun Ai, Shengchen Li, and Mark D Plumbley. Acoustic scene classification across cities and devices via feature disentanglement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[3] Nisan Aryal and Sang-Woong Lee. Frequency-based cnn and attention module for acoustic scene classification. *Applied Acoustics*, 210:109411, 2023.

[4] Yun-Fei Shao, Xin-Xin Ma, Yong Ma, and Wei-Qiang Zhang. Deep semantic learning for acoustic scene classification. *EURASIP Journal on Audio, Speech, and Music Processing*, 2024(1):1, 2024.

[5] Bandhav Veluri, Malek Itani, Justin Chan, Takuya Yoshioka, and Shyamnath Gollakota. Semantic hearing: Programming acoustic scenes with binaural hearables. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–15, 2023.

[6] Fatih Demir, Muammer Turkoglu, Muzaffer Aslan, and Abdulkadir Sengur. A new pyramidal concatenated cnn approach for environmental sound classification. *Applied Acoustics*, 170:107520, 2020.

[7] Biyun Ding, Tao Zhang, Ganjun Liu, Lingguo Kong, and Yanzhang Geng. Late fusion for acoustic scene classification using swarm intelligence. *Applied Acoustics*, 192:108698, 2022.

[8] Mahmoud A Alamir. A novel acoustic scene classification model using the late fusion of convolutional neural networks and different ensemble classifiers. *Applied Acoustics*, 175:107829, 2021.

[9] Krishna Presannakumar and Anuj Mohamed. Deep learning based source identification of environmental audio signals using optimized convolutional neural networks. *Applied Soft Computing*, 143:110423, 2023.

[10] Chiun-Li Chin, Chia-Chun Lin, Jing-Wen Wang, Wei-Cheng Chin, Yu-Hsiang Chen, Sheng-Wen Chang, Pei-Chen Huang, Xin Zhu, Yu-Lun Hsu, and Shing-Hong Liu. A wearable assistant device for the hearing impaired to recognize emergency vehicle sirens with edge computing. *Sensors*, 23(17):7454, 2023.

[11] DCASE. Challenge on detection and classification of acoustic scenes and events (DCASE) 2024, 2024. URL https://dcase.community/challenge2024/. Accessed: 2024-12-12.

[12] Grand View Research. Wearable technology market size, share & trends analysis report by product (head & eyewear, wristwear), by application (consumer electronics, healthcare), by region (asia pacific, europe), and segment forecasts, 2023 - 2030. Report 978-1-68038-165-8, Grand View Research, April 2023. URL https://www.grandviewresearch.com/industry-analysis/wearable-technology-market. Horizon Databook.

[13] Grand View Research. Hearing aids market size, share & trends analysis report by product type (bte, canal hearing aids), by technology (digital, analog), by sales channel, by region, and segment forecasts, 2024 - 2030. Report 978-1-68038-166-5, Grand View Research, April 2023. URL https://www.grandviewresearch.com/industry-analysis/hearing-aids-market. Horizon Databook.

[14] Waleed Bin Qaim, Aleksandr Ometov, Antonella Molinaro, Ilaria Lener, Claudia Campolo, Elena Simona Lohan, and Jari Nurmi. Towards energy efficiency in the internet of wearable things: A systematic review. *IEEE Access*, 8:175412–175435, 2020.

[15] Kalin Penev, Alexander Gegov, Olufemi Isiaq, and Raheleh Jafari. Energy efficiency evaluation of artificial intelligence algorithms. *Electronics*, 13(19):3836, 2024.

[16] Sahalu Balarabe Junaid, Abdullahi Abubakar Imam, Abdullateef Oluwagbemiga Balogun, Liyanage Chandratilak De Silva, Yusuf Alhaji Surakat, Ganesh Kumar, Muhammad Abdulkarim, Aliyu Nuhu Shuaibu, Aliyu Garba, Yusra Sahalu, et al. Recent advancements in emerging technologies for healthcare management systems: a survey. In *Healthcare*, volume 10, page 1940. MDPI, 2022.

[17] Shelly Chadha, Kaloyan Kamenov, and Alarcos Cieza. The world report on hearing, 2021. *Bulletin of the World Health Organization*, 99(4):242, 2021.

[18] Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[19] Ivan Bocharov, Tjalling Tjalkens, and Bert De Vries. Acoustic scene classification from few examples. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 862–866. IEEE, 2018.

[20] Biyun Ding, Tao Zhang, Ganjun Liu, and Chao Wang. Hierarchical classification for acoustic scenes using deep learning. *Applied Acoustics*, 212:109594, 2023.

[21] Shefali Waldekar and Goutam Saha. Two-level fusion-based acoustic scene classification. *Applied Acoustics*, 170:107502, 2020.

[22] Donald Hayes. Environmental classification in hearing aids. In *Seminars in Hearing*, volume 42, pages 186–205. Thieme Medical Publishers, Inc., 2021.

[23] VS Vivek, S Vidhya, and P Madhanmohan. Acoustic scene classification in hearing aid using deep learning. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, pages 0695–0699. IEEE, 2020.