



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
CURSO DE ENGENHARIA MECÂNICA**

ALEXANDRE SAMPAIO DE FIGUEIREDO

**ANÁLISE DE SOM PARA DETECTAR DESBALANCEAMENTO EM
UM SISTEMA DE HÉLICE USANDO REDES NEURAIAS
CONVOLUCIONAIS**

**JOÃO PESSOA/PB
2025**

ALEXANDRE SAMPAIO DE FIGUEIREDO

**ANÁLISE DE SOM PARA DETECTAR DESBALANCEAMENTO EM UM SISTEMA
DE HÉLICE USANDO REDES NEURAS CONVOLUCIONAIS**

Monografia apresentada ao Curso de Bacharelado em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal da Paraíba, como requisito obrigatório para a obtenção do Título de Bacharel em Engenharia Mecânica.

Orientador(a): Prof. Dr. Marcelo Cavalcanti Rodrigues.

**JOÃO PESSOA/PB
2025**

Catálogo na publicação
Seção de Catalogação e Classificação

F475a Figueiredo, Alexandre Sampaio de.

Análise de som para detectar desbalanceamento em um sistema de hélice usando redes neurais convolucionais / Alexandre Sampaio de Figueiredo. - João Pessoa, 2025.
53 f.

Orientação: Marcelo Cavalcanti Rodrigues.
TCC (Graduação) - UFPB/Tecnologia.

1. Redes neurais convolucionais. 2. Áudio. 3. Falhas mecânicas. 4. Manutenção preditiva. I. Rodrigues, Marcelo Cavalcanti. II. Título.

UFPB/BSCT

CDU 621(043.2)

ALEXANDRE SAMPAIO DE FIGUEIREDO

ANÁLISE DE SOM PARA DETECTAR DESBALANCEAMENTO EM UM SISTEMA DE HÉLICE USANDO REDES NEURAIS CONVOLUCIONAIS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal da Paraíba, como requisito obrigatório para a obtenção do título de Bacharel em Engenharia Mecânica, defendido e aprovado pela banca examinadora constituída pelos docentes:

Documento assinado digitalmente
 **MARCELO CAVALCANTI RODRIGUES**
Data: 24/09/2025 21:09:42-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Marcelo Cavalcanti Rodrigues - UFPB
Orientador

Documento assinado digitalmente
 **JOSE CARLOS DE LIMA JUNIOR**
Data: 25/09/2025 16:51:37-0300
Verifique em <https://validar.iti.gov.br>

José Carlos de Lima Junior - UFPB
Membro da Banca Examinadora

Documento assinado digitalmente
 **KOJE DANIEL VASCONCELOS MISHINA**
Data: 25/09/2025 09:10:57-0300
Verifique em <https://validar.iti.gov.br>

Koje Daniel Vasconcelos Mishina - UFPB
Membro da Banca Examinadora

JOÃO PESSOA/PB
2025

DECLARAÇÃO DE AUTORIA PRÓPRIA

Eu, Alexandre Sampaio de Figueiredo, matrícula n.º 20180053797, autor(a) do Trabalho de Conclusão de Curso intitulado ANÁLISE DE SOM PARA DETECTAR DESBALANCEAMENTO EM UM SISTEMA DE HÉLICE USANDO REDES NEURAS CONVOLUCIONAIS, orientado(a) pelo(a) professor(a) Marcelo Cavalcanti Rodrigues, como parte das avaliações do Curso de Engenharia Mecânica no período letivo 2025.1 e requisito obrigatório à obtenção do grau de Bacharelado, declaro que o trabalho em referência é de minha total autoria, não tendo sido copiado ou extraído, seja parcial ou integralmente, de forma ilícita de nenhuma fonte, além daquelas públicas consultadas e corretamente referenciadas ao longo do trabalho, obedecendo aos padrões nacionais para referências diretas e indiretas, ou daquelas cujos dados resultaram de investigações empíricas por mim realizadas para fins de produção deste trabalho. Afirmando que em hipótese alguma representa plágio de material disponível em qualquer meio, e declaro, estar ciente das penalidades previstas nos artigos 184 e 298 do Decreto-Lei n.º 2.848/1940 – Código Penal Brasileiro, como também declaro não infringir nenhum dispositivo da Lei n.º 9.610/98 – Lei dos Direitos Autorais.

Assim, se houver qualquer trecho do texto em questão que configure o crime de plágio ou violação aos direitos autorais, assumo total responsabilidade, ficando a Instituição, o orientador e os demais membros da banca examinadora isentos de qualquer ação negligente da minha parte, ou pela veracidade e originalidade desta obra, cabendo ao corpo docente responsável pela sua avaliação não aceitá-lo como Trabalho de Conclusão de Curso da Universidade Federal da Paraíba - UFPB, no Curso de Engenharia Mecânica, e, por conseguinte, considerar-me reprovado no Trabalho de Conclusão de Curso.

Por ser verdade, firmo a presente.

João Pessoa, 26 de setembro de 2025.

Documento assinado digitalmente
 ALEXANDRE SAMPAIO DE FIGUEIREDO
Data: 26/09/2025 16:14:36-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do(a) discente

AGRADECIMENTOS

Aos meus pais e irmã, Maria Gilneide, José Alves e Maria Cecília que sempre foram meu alicerce, exemplos de dedicação e resiliência, pelo incentivo constante e pelo apoio em cada etapa desta jornada acadêmica.

Aos meus familiares e amigos, pela ajuda, pelas palavras de encorajamento e pela companhia em todos os momentos.

Aos professores que, ao longo da minha formação, contribuíram com seus conhecimentos, paciência e dedicação. Em especial, expresso minha gratidão ao meu orientador, Prof. Dr. Marcelo Cavalcanti Rodrigues, pela orientação, pelas valiosas contribuições acadêmicas e pela confiança depositada em meu trabalho. E ao Grupo de Pesquisa de Integridade e Inspeção (GPII).

A todos que, de forma direta ou indireta, colaboraram para a concretização deste trabalho, deixo registrado meu mais sincero agradecimento.

Somos um caminho para o cosmos
conhecer a si mesmo.

Carl Sagan 1980

RESUMO

O diagnóstico precoce de falhas é tradicionalmente realizado por técnicas de análise de vibração, acústica e inspeção visual; entretanto, tais métodos apresentam limitações quando aplicados em larga escala ou em ambientes com alto nível de ruído. Nesse cenário, o avanço da inteligência artificial, em particular das Redes Neurais Convolucionais (CNNs) tem possibilitado novas abordagens para a detecção automática de anomalias em sinais de áudios e imagens. Este trabalho apresenta a aplicação das CNNs para o tratamento de áudios com o objetivo de identificar falhas em máquinas rotativas. O experimento consistiu em coletar dados em duas condições: em estado normal e com desbalanceamento em uma das pás da hélice. Os sinais de áudio foram tratados e transformados em representações espectrográficas, permitindo a utilização nas CNNs para classificação automática. A metodologia aplicada visa contribuir com o avanço da manutenção preditiva em máquinas, oferecendo uma abordagem baseada em inteligência artificial para diagnóstico de falhas mecânicas. Resultados apontam para alta acurácia do modelo (acurácia 100%, f1-score = 100% precisão 100%, recall = 100%) demonstrando a viabilidade da proposta.

Palavras-chave: Redes neurais convolucionais. Áudio. Falhas mecânicas. Manutenção preditiva.

ABSTRACT

Early fault diagnosis is traditionally performed using vibration analysis, acoustics, and visual inspection techniques; however, these methods have limitations when applied on a large scale or in environments with high noise levels. In this scenario, advances in artificial intelligence, particularly Convolutional Neural Networks CNNs, have enabled new approaches for the automatic detection of anomalies in signals of audio and images. This work presents the application of CNNs to the processing of audio to identify faults in rotating machinery. The experiment consisted of collecting data under two conditions: normal state and with imbalance in one of the propeller blades. The audio signals were processed and transformed into spectrographic representations, allowing them to be used in CNNs for automatic classification. The applied methodology aims to contribute to the advancement of predictive maintenance in machines by offering an artificial intelligence-based approach for diagnosing mechanical faults. Results indicate high model accuracy (100% accuracy, 100% f1-score, 100% precision, 100% recall), demonstrating the viability of the proposal.

Keywords: Convolutional neural networks. Audio. Mechanical failures. Predictive maintenance.

LISTA DE FIGURAS

FIGURA 1 - COMPONENTES DO AEROGERADOR.....	17
FIGURA 2 - REPRESENTAÇÃO ESQUEMÁTICA DA TRANSFORMADA DE FOURIER (TEMPO → FREQÜÊNCIA)	20
FIGURA 3 - ESPECTROGRAMA STFT, ÁUDIO NA CONDIÇÃO NORMAL DE FUNCIONAMENTO .	22
FIGURA 4 - ESPECTROGRAMA TME, ÁUDIO NA CONDIÇÃO NORMAL DE FUNCIONAMENTO...	22
FIGURA 5 - ESPECTROGRAMA STFT, ÁUDIO NA CONDIÇÃO DE DEFEITO DE FUNCIONAMENTO	23
FIGURA 6 - ESPECTROGRAMA TME, ÁUDIO NA CONDIÇÃO DE DEFEITO DE FUNCIONAMENTO	23
FIGURA 7 - CAMADAS EM UMA RNA.....	26
FIGURA 8 - PRINCÍPIO COMPUTACIONAL DE UM NEURÔNIO ARTIFICIAL	27
FIGURA 9 - REPRESENTAÇÃO DA ARQUITETURA DE UMA CNN CLÁSSICA (LENET-5).....	31
FIGURA 10 - MATRIZ DE CONFUSÃO, PRIMEIRA ANÁLISE.....	38
FIGURA 11 - ACURÁCIA DE TREINO E DE VALIDAÇÃO NAS ÉPOCAS, PRIMEIRA ANÁLISE	39
FIGURA 12 - LOSS DO TREINO E DA VALIDAÇÃO NAS ÉPOCAS, PRIMEIRA ANÁLISE	39
FIGURA 13 - MATRIZ DE CONFUSÃO, SEGUNDA ANÁLISE	40
FIGURA 14 - ACURÁCIA DE TREINO E DE VALIDAÇÃO NAS ÉPOCAS, SEGUNDA ANÁLISE	41
FIGURA 15 - LOSS DO TREINO E DA VALIDAÇÃO NAS ÉPOCAS, SEGUNDA ANÁLISE	41
FIGURA 16 - MATRIZ DE CONFUSÃO, TERCEIRA ANÁLISE.....	43
FIGURA 17 - ACURÁCIA DE TREINO E DE VALIDAÇÃO NAS ÉPOCAS, TERCEIRA ANÁLISE.....	43
FIGURA 18 - LOSS DO TREINO E DA VALIDAÇÃO NAS ÉPOCAS, TERCEIRA ANÁLISE.....	44
FIGURA 19 - MATRIZ DE CONFUSÃO, QUARTA ANÁLISE	44
FIGURA 20 - ACURÁCIA DE TREINO E DE VALIDAÇÃO NAS ÉPOCAS, QUARTA ANÁLISE	45
FIGURA 21 - LOSS DO TREINO E DA VALIDAÇÃO NAS ÉPOCAS, QUARTA ANÁLISE	45
FIGURA 22 - MATRIZ DE CONFUSÃO, QUINTA ANÁLISE.....	47
FIGURA 23 - ACURÁCIA DE TREINO E DE VALIDAÇÃO NAS ÉPOCAS, QUINTA ANÁLISE	47
FIGURA 24 - LOSS DO TREINO E DA VALIDAÇÃO NAS ÉPOCAS, QUINTA ANÁLISE	48

LISTA DE TABELAS

TABELA 1 - ANÁLISE DE SITUAÇÕES	27
TABELA 2 - PARÂMETRO E RESULTADOS, PRIMEIRA ANÁLISE	38
TABELA 3 - PARÂMETRO E RESULTADOS, SEGUNDA ANÁLISE	40
TABELA 4 - PARÂMETRO E RESULTADOS, TERCEIRA ANÁLISE	42
TABELA 5 - PARÂMETRO E RESULTADOS, QUARTA ANÁLISE	44
TABELA 6 - PARÂMETRO E RESULTADOS, QUINTA ANÁLISE	47

LISTA DE ABREVIATURAS E SIGLAS

CNN - Redes Neurais Convolucionais

DP - Deep Learning

EM - Escala Mel

MLP - Multilayer Perceptron

ReLU - Rectified Linear Unit

RNA - Redes Neurais Artificiais

RNN - Redes Recorrentes

STFT - Transformada de Fourier de Curto Prazo

Tanh - Tangente Hiperbólica

TME - Transformada Mel-Espectral

SUMÁRIO

1	INTRODUÇÃO	14
1.1	OBJETIVOS	15
1.1.1	Objetivo geral	15
1.1.2	Objetivos específicos.....	15
1.2	JUSTIFICATIVA.....	16
1.3	REVISÃO LITERÁRIA	17
1.3.1	Manutenção preditiva e falhas mecânicas.....	17
1.3.2	Processamento de sinais e imagens	19
1.3.3	Redes neurais	24
2	PROCEDIMENTO METODOLÓGICO.....	33
2.1	COLETA DOS DADOS	33
2.2	ARQUITETURA DA REDE NEURAL CONVOLUCIONAL	33
2.3	MÉTRICAS DE ANÁLISES	35
3	RESULTADOS E DISCUSSÃO	38
3.1	INTERPRETAÇÃO DOS RESULTADOS.....	49
4	CONCLUSÃO.....	50
	REFERÊNCIAS.....	51

1 INTRODUÇÃO

A identificação precoce de falhas em máquinas rotativas é um tema de grande relevância na Engenharia Mecânica, sobretudo no contexto da manutenção preditiva. Entre essas máquinas, destacam-se os aerogeradores, que desempenham papel estratégico na matriz energética mundial. Por serem responsáveis pela conversão da energia cinética do vento em energia elétrica, sua confiabilidade impacta diretamente a eficiência da geração, a redução de custos operacionais e a expansão das fontes renováveis. Um aerogerador parado devido a falhas mecânicas pode representar não apenas prejuízos financeiros significativos, mas também perda de competitividade no setor energético (LEE, Jay, 2014).

Além dos aerogeradores, outros sistemas de refrigeração, climatização transporte, mineração, metalurgia também são exemplos que contém máquinas rotativas, amplamente utilizados em ambientes industriais. As máquinas rotativas de ventilação, por exemplo, é um equipamento cuja confiabilidade está associada ao conforto térmico, à eficiência energética e à segurança operacional. Tanto em aplicações domésticas quanto industriais, a manutenção preventiva e preditiva desses equipamentos é essencial para garantir sua disponibilidade e prolongar sua vida útil.

Historicamente, técnicas convencionais como a análise de vibrações, medições acústicas, Termografia Infravermelha, inspeções visuais, entre outras têm sido utilizadas para monitorar o estado de operação de máquinas rotativas (RANDALL, 2011). No entanto, essas técnicas muitas vezes exigem equipamentos especializados e análise manual, o que limita a escalabilidade em aplicações industriais de grande porte, como parques eólicos para os aerogeradores.

Com o avanço das técnicas de inteligência artificial e aprendizado profundo, novas abordagens vêm sendo aplicadas com sucesso no diagnóstico automático de falhas. Nesse cenário, destacam-se as Redes Neurais Convolucionais CNNs, cuja arquitetura é particularmente eficiente na análise de padrões em imagens e sinais espectrais. Aplicações recentes demonstram o potencial das CNNs na classificação de sons ambientais (NORDBY, 2019), na inspeção visual de equipamentos industriais (SANVEZZO, 2022), e até na avaliação de processos agrícolas (MUNDIM FILHO, 2021).

Neste trabalho, propõe-se a utilização de CNNs para o tratamento de áudios

coletados de uma máquina rotativa de ventilação em funcionamento. Inicialmente, registraram-se os sinais em condição normal e, posteriormente, introduziu-se um desbalanceamento em uma das pás da hélice para simular o cenário de falha. Embora aplicado a uma máquina rotativa de pequeno porte, o estudo busca demonstrar a viabilidade do uso de técnicas de aprendizado profundo na detecção automática de anomalias em sistemas mecânicos, com potencial aplicação em máquinas de maior complexidade, como aerogeradores.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Desenvolver e avaliar um modelo baseado em Redes Neurais Convolucionais para a identificação de falhas em motores rotativos, a partir da análise de áudios obtidos em condições normais e desbalanceadas.

1.1.2 Objetivos específicos

- a) Coletar dados experimentais de áudio de uma máquina rotativa em condições de funcionamento normais e com desbalanceamento.
- b) Pré-processar os sinais acústicos, gerando espectrogramas representativos para análise por CNN.
- c) Implementar e treinar a CNN para classificação das condições de operação.
- d) Avaliar o desempenho do modelo utilizando métricas como acurácia, *loss* (perda), precisão, *recall*, *f1-score*, *support*.
- e) Discutir a aplicabilidade da abordagem para sistemas de manutenção preditiva em engenharia mecânica.

1.2 JUSTIFICATIVA

A crescente demanda por sistemas de manutenção preditiva reflete o papel estratégico da confiabilidade dos equipamentos na redução de custos e na garantia de segurança operacional. Na indústria, falhas não detectadas em máquinas rotativas podem gerar paradas não programadas, desperdício de energia, aumento de custos de reparo e até acidentes (Randall, 2011).

Nesse contexto, o emprego de técnicas baseadas em inteligência artificial torna-se um diferencial competitivo, uma vez que possibilita diagnósticos mais precisos e automáticos. Trabalhos como o de Sanvezzo (2022), mostraram o uso de CNNs para inspeção de equipamentos industriais, indicando maior confiabilidade em relação a métodos tradicionais de monitoramento. De forma semelhante, Nordby (2019) demonstrou a viabilidade da classificação de sons ambientais por CNNs em hardwares de baixo custo, destacando o potencial dessas técnicas em aplicações reais, mesmo em ambientes com restrição de recursos computacionais.

Na área agrícola, Mundim Filho (2021), aplicou redes convolucionais na classificação de imagens de frutos de café em diferentes estágios de maturação, evidenciando que modelos de aprendizado profundo são capazes de extrair padrões visuais complexos, mesmo em cenários de variação natural dos dados. Já Assunção (2023), investigou o uso de CNNs em sistemas energéticos, reforçando a aplicabilidade de arquiteturas convolucionais em problemas práticos de engenharia.

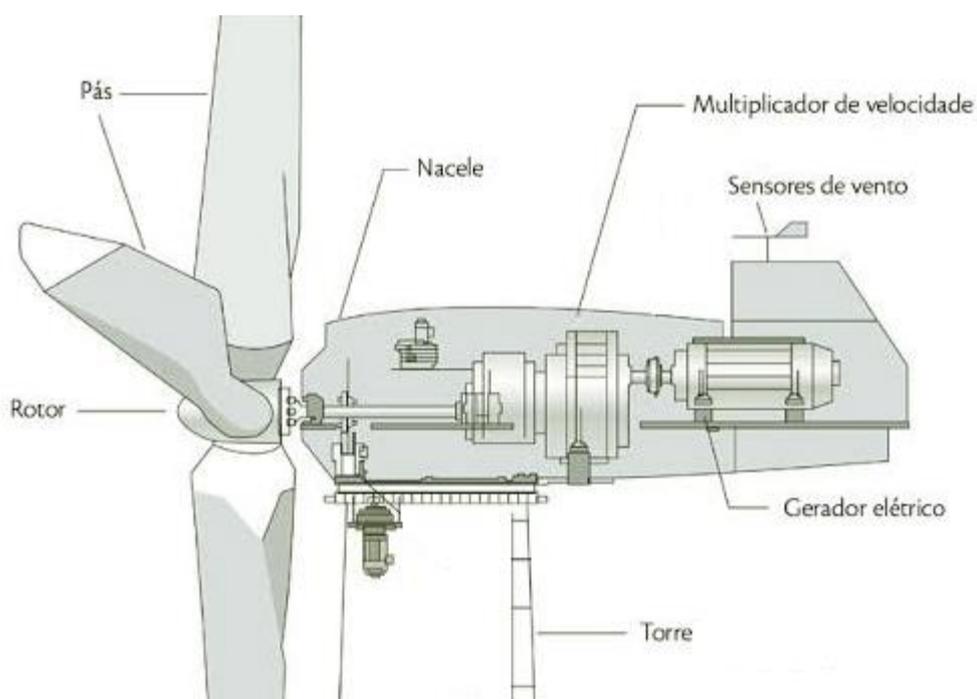
Dessa forma, este trabalho se justifica por sua relevância acadêmica e prática. Do ponto de vista científico, contribui para o avanço da integração entre processamento de sinais acústicos e imagens com aprendizado profundo, área em constante expansão. Do ponto de vista aplicado, a pesquisa propõe um método para identificar falhas em máquinas rotativas de ventilação, um equipamento de uso comum em ambientes industriais e domésticos, abrindo espaço para a expansão da abordagem em sistemas mecânicos mais complexos.

1.3 REVISÃO LITERÁRIA

1.3.1 Manutenção preditiva e falhas mecânicas

Os aerogeradores, também chamados de turbinas eólicas, são máquinas rotativas projetadas para converter a energia cinética do vento em energia elétrica. Seu princípio de funcionamento baseia-se no movimento das pás do rotor, que giram devido à ação do vento. Esse movimento é transferido a um eixo principal, conectado a uma caixa multiplicadora de velocidade, responsável por aumentar a rotação e transmiti-la ao gerador elétrico (BURTON, 2011).

Figura 1 - Componentes do Aerogerador



Fonte: WINDBOX BRASIL (2023).

Além desses elementos presentes na figura 1, os aerogeradores modernos contam com sistemas auxiliares, como:

- Sistema de controle: regula a orientação das pás (*pitch control*) e da nacele (*yaw control*) em relação à direção do vento.
- Sistema hidráulico: auxilia na frenagem e na mudança do ângulo das pás.

- Sistema de monitoramento: coleta dados de vibração, temperatura, torque e velocidade, fundamentais para manutenção preditiva.
- Estrutura de suporte (torre e fundação): garante estabilidade diante de cargas dinâmicas do vento e vibrações.

Os aerogeradores assumem papel estratégico na matriz energética mundial por serem fontes de energia limpa e renovável. Segundo a Agência Internacional de Energia (IEA, 2022), a energia eólica é responsável por uma parcela crescente da eletricidade global, sendo essencial para a transição energética e a mitigação das mudanças climáticas. No entanto, os altos custos de operação e manutenção, especialmente em parques *offshore*, tornam a confiabilidade desses equipamentos um fator crítico para a competitividade do setor.

Os aerogeradores operam em condições ambientais severas, sujeitas a cargas variáveis e impactos de turbulências. Estudos apontam que aproximadamente 60% das falhas estão associadas a componentes mecânicos (FENG, 2013). Entre as mais recorrentes destacam-se:

- Desgaste e falhas em rolamentos, resultando em vibrações excessivas.
- Falhas na caixa multiplicadora (*gearbox*), devido a cargas cíclicas e falta de lubrificação adequada.
- Trincas ou desalinhamentos nas pás do rotor, ocasionados por fadiga do material e impactos externos.
- Superaquecimento no gerador elétrico, decorrente de sobrecarga ou falhas no sistema de refrigeração.

Esses problemas podem levar a paradas não programadas, perda de produção e altos custos de reparo, que chegam a representar até 25% do custo total de geração em turbinas eólicas *offshore* (ZAHER; MCARTHUR; INFIELD, 2009).

Para reduzir os riscos de falhas e otimizar a disponibilidade dos aerogeradores, a manutenção preditiva tem sido amplamente adotada. Essa estratégia baseia-se na análise contínua de dados coletados por sensores embarcados — vibração, acústica, temperatura e corrente elétrica — que permitem prever falhas antes de sua ocorrência.

O uso de técnicas tradicionais, como análise de vibração e termografia infravermelha, já se mostrou eficaz para identificar desgaste em rolamentos e superaquecimento em geradores (RANDALL, 2011). Contudo, com a digitalização industrial, o emprego de algoritmos avançados de aprendizado de máquina e redes

neurais convolucionais tem ampliado a capacidade de detecção de falhas de forma mais precisa e automática.

Estudos como o de (RANDALL, 2016) evidenciam que CNNs aplicadas a sinais de vibração e espectrogramas acústicos são capazes de identificar falhas em rolamentos com maior acurácia que os métodos convencionais. Além disso, a integração entre sensores IoT, sistemas de monitoramento remoto e análise em nuvem possibilita o desenvolvimento da chamada manutenção inteligente (*smart maintenance*), reduzindo custos de O&M e aumentando a vida útil dos aerogeradores.

1.3.2 Processamento de sinais e imagens

O processamento de sinais é uma parte fundamental para a análise, interpretação e transformação de dados obtidos por sensores, convertendo informações brutas em representações mais significativas. No caso de sinais acústicos, as técnicas de análise espectral assumem papel central, permitindo identificar padrões de frequência que podem estar associados a falhas ou condições normais de operação de sistemas mecânicos.

A análise de sinais de áudio pode ser realizada por diferentes representações matemáticas, sendo o espectrograma uma das mais utilizadas em aplicações de processamento digital de sinais e aprendizado de máquina. Para compreender seu funcionamento, é necessário partir do conceito de sinal sonoro, entendido como uma variação da pressão do ar ao longo do tempo, que pode ser digitalizada por meio da amostragem em intervalos regulares, como ocorre, por exemplo, em uma taxa de 44,1 kHz (ROBERTS, 2020).

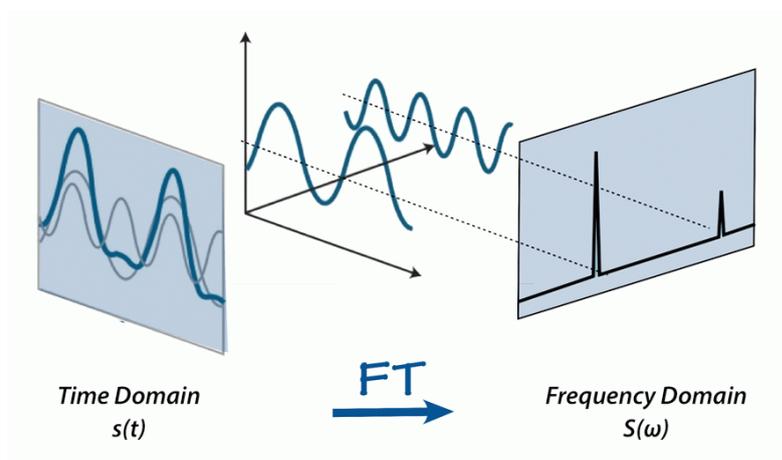
A partir dessa digitalização, aplica-se a Transformada de Fourier, que tem como objetivo decompor o sinal no domínio do tempo em suas componentes de frequência, revelando quais frequências estão presentes e com que intensidade. O algoritmo mais empregado nesse processo é a Transformada Rápida de Fourier (FFT), que viabiliza a análise computacional eficiente do sinal. Quando o interesse está na variação espectral ao longo do tempo, como em sinais de fala ou música, utiliza-se a aplicação da Transformada de Fourier de Curto Prazo (STFT) em janelas sobrepostas do sinal, gerando o chamado espectrograma, uma representação bidimensional em que o eixo

vertical corresponde à frequência e o horizontal ao tempo, enquanto a intensidade é indicada por variações de cor ou amplitude.

Entretanto, a escala linear de frequências não corresponde à forma como o ouvido humano percebe o som. Os indivíduos são mais sensíveis a diferenças em frequências baixas do que em altas, o que motivou a criação da escala Mel, uma escala perceptiva que traduz a frequência em valores mais próximos da experiência auditiva humana. Dessa forma, ao converter as frequências do espectrograma tradicional para a escala Mel, obtém-se o espectrograma mel, que constitui uma representação mais alinhada à percepção auditiva e, portanto, mais adequada para aplicações em reconhecimento automático de fala, classificação de áudio e análise de padrões sonoros.

Além disso, essa representação costuma ser ajustada para trabalhar em decibéis, a fim de normalizar os níveis de intensidade, tornando a análise mais robusta a variações na amplitude do sinal. Assim, a EM pode ser entendido como um processo em etapas: o sinal é amostrado, transformado para o domínio da frequência com a STFT, representado no tempo por meio do espectrograma, convertido em escala perceptiva por meio da escala Mel e, por fim, normalizado. Tal abordagem tem se mostrado essencial para sistemas de aprendizado profundo voltados ao processamento de sinais de áudio, uma vez que fornece uma representação compacta e perceptualmente significativa (ROBERTS, 2020).

Figura 2 - Representação esquemática da Transformada de Fourier (tempo → frequência)



Fonte: RANGEL (2021).

No campo do processamento de imagens, técnicas clássicas descritas por Gonzalez e Woods (2018) incluem operações de filtragem espacial, equalização de histograma e normalização, utilizadas para remover ruídos, melhorar o contraste e padronizar a escala de intensidades. Esses pré-processamentos são fundamentais quando imagens ou espectrogramas são utilizados como entrada em modelos de aprendizado de máquina, pois reduzem variações indesejadas e aumentam a robustez do sistema.

O processo de cálculo do TME segue, em geral, as seguintes etapas (GONZALEZ; WOODS 2018):

- 1) Da STFT ao espectrograma de potência.

Primeiro, obtém-se o espectrograma de potência a partir do sinal de áudio $x[n]$:

$$P(k, m) = |X(k, m)|^2 \quad (1)$$

Onde o $X(k, m)$ é a STFT do sinal, k é o índice de frequência discreta, m é o índice temporal (janela). A STFT é definida por:

$$X(k, m) = \sum_{n=0}^{N-1} x[n] \cdot w[n - m] \cdot e^{-j\frac{2\pi}{N}kn} \quad (2)$$

Onde, $w[n]$ é a janela e N o tamanho da FFT.

- 2) Aplicação de um banco de filtros triangulares espaçados na EM.

$$E_m = \sum_{k=0}^{K-1} P(k) \cdot H(k), \quad m = 1, 2, \dots, M \quad (3)$$

Onde E_m é a energia no filtro m , $H_m(k)$ é a resposta em frequência do filtro triangular na EM, M é o número total de filtros (ex.: 40, 64, 128).

- 3) Transformada Mel-Espectral

Assim, a TME é dada por:

$$M(m, t) = \sum_{k=0}^{K-1} |X(k, t)|^2 \cdot H_m(k) \quad (4)$$

Onde, $M(m, t)$ representa a energia espectral no filtro m , no instante t .

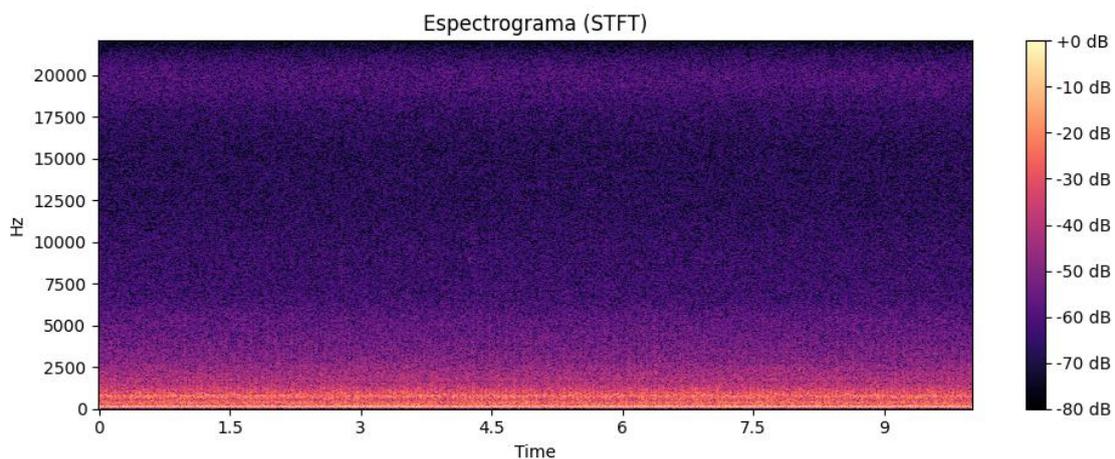
4) Conversão para dB (escala logarítmica).

A conversão entre frequência em Hertz f e frequência em Mel m é dada por (STEVENS, 1937):

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (5)$$

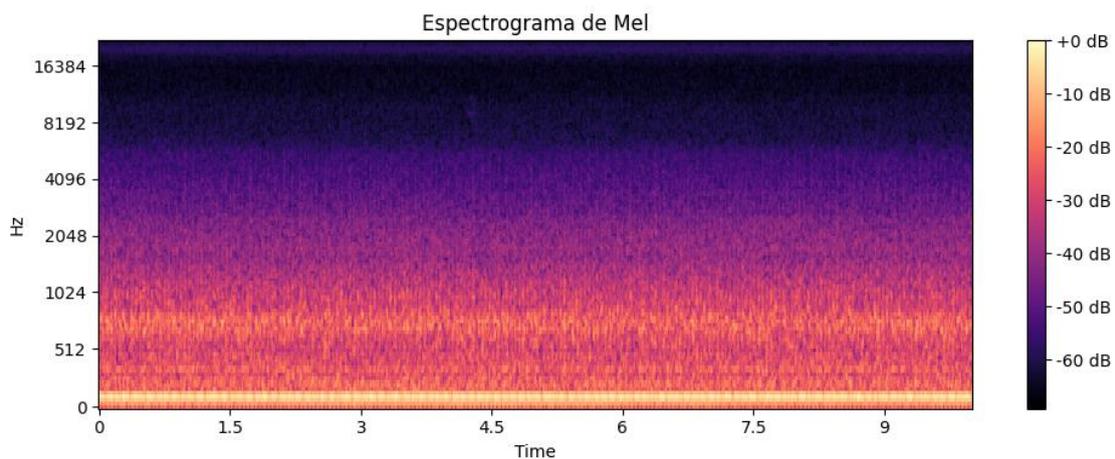
Essa transformação permite que padrões acústicos como harmônicos, ruídos e ressonâncias sejam representados de modo mais próximo da percepção humana, tornando o Mel-espectrograma uma entrada altamente eficaz para CNNs em tarefas como reconhecimento de fala, classificação de sons ambientais e diagnóstico de falhas mecânicas (PICZAK, 2015).

Figura 3 - Espectrograma STFT, áudio na condição normal de funcionamento



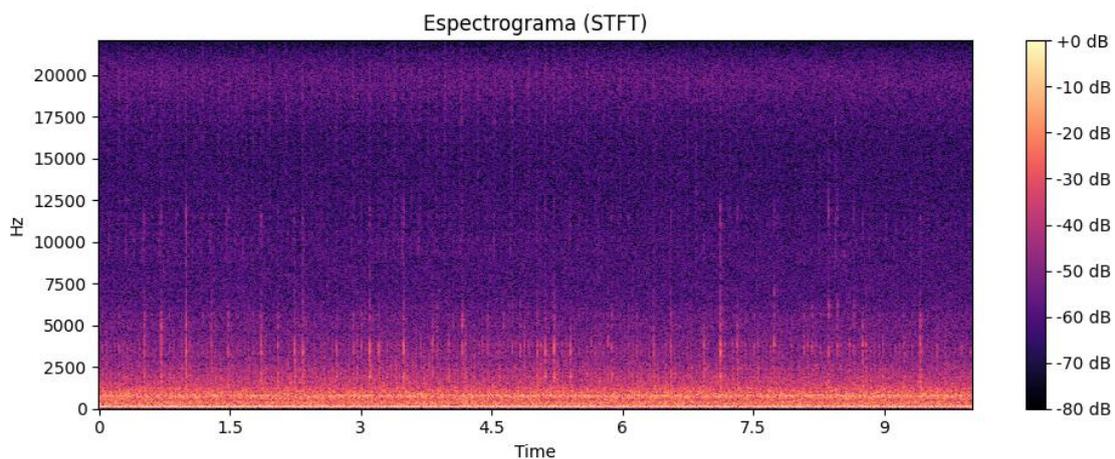
Fonte: Autor (2025).

Figura 4 - Espectrograma TME, áudio na condição normal de funcionamento



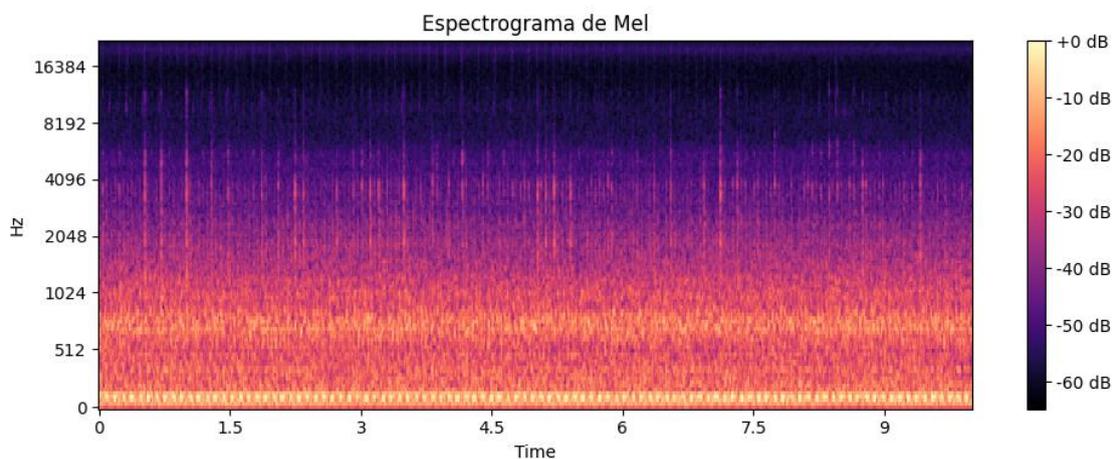
Fonte: Autor (2025).

Figura 5 - Espectrograma STFT, áudio na condição de defeito de funcionamento



Fonte: Autor (2025).

Figura 6 - Espectrograma TME, áudio na condição de defeito de funcionamento



Fonte: Autor (2025).

Na primeira figura observa-se maior concentração de energia nas baixas frequências (até aproximadamente $2,5\text{ kHz}$), representadas pelas cores mais claras (amarelo/laranja). Isso é esperado, já que o som gerado pela máquina em funcionamento normal é composto principalmente por ruídos de baixa frequência relacionados à rotação das pás. O restante do espectro apresenta baixa intensidade (tons escuros), indicando ausência de harmônicos fortes ou ruído significativo.

Na segunda figura o mapeamento na EM torna o espectro mais próximo da percepção auditiva humana, dando maior ênfase às faixas graves e médias. Nota-se claramente a predominância de componentes até 1 kHz , com gradativa atenuação nas regiões acima. Isso reforça que a máquina em operação normal mantém um padrão estável de baixas frequências, sem grandes picos anômalos.

Na terceira figura diferente da condição normal, aparecem faixas verticais mais intensas e descontínuas ao longo do tempo. Essas linhas indicam a presença de ruídos harmônicos adicionais, provavelmente causados por desbalanceamento ou vibração irregular. A energia nas frequências acima de 2 kHz é mais perceptível, sugerindo que o defeito introduziu ruídos mais agudos.

Na quarta figura fica ainda mais clara a diferença em relação ao funcionamento normal: há maior densidade de energia distribuída entre 500 Hz e 2 kHz , com regiões de concentração não observadas na condição normal. O mapa de cores revela maior variabilidade temporal, evidenciando instabilidade na assinatura acústica da máquina defeituoso. Essa diferença é fundamental para a CNN, pois os espectrogramas fornecem padrões visuais distintos que permitem a classificação automática das condições de operação.

1.3.3 Redes neurais

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano, compostos por camadas de neurônios interconectados que processam informações de maneira distribuída. Elas podem ser classificadas em diferentes tipos, de acordo com sua arquitetura e aplicação. As redes

feedforward, também conhecidas como perceptrons multicamadas (MLPs), apresentam fluxo unidirecional da informação, da camada de entrada para a de saída, sendo amplamente utilizadas em tarefas de classificação e regressão. Já as redes recorrentes (RNNs) possuem conexões cíclicas, o que permite que informações de estados anteriores influenciem as decisões atuais, tornando-as indicadas para o processamento de dados sequenciais, como séries temporais e linguagem natural. Por fim, as redes convolucionais aplicam operações de convolução para extrair características espaciais de dados, mostrando-se especialmente eficazes no reconhecimento de imagens e vídeos (DEEP LEARNING BOOK, 2025).

O desenvolvimento das RNAs ocorreu ao longo de décadas, passando por períodos de avanço e estagnação, até atingir o patamar atual conhecido como *Deep Learning (DL)*, no qual arquiteturas profundas são capazes de resolver problemas de alta complexidade, como reconhecimento facial, tradução automática e diagnósticos médicos baseados em imagens (GOODFELLOW, 2016).

As origens das RNAs remontam ao ano de 1943, quando Warren McCulloch e Walter Pitts publicaram o artigo “*A Logical Calculus of the Ideas Immanent in Nervous Activity*”. Nesse trabalho pioneiro, os autores propuseram o primeiro modelo matemático de neurônio artificial, baseado em operações lógicas binárias (MCCULLOCH; PITTS, 1943).

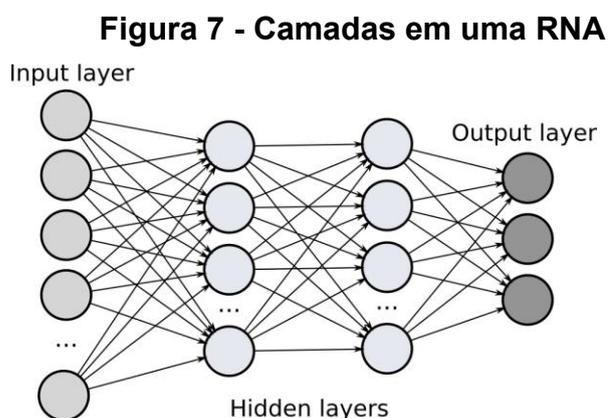
Posteriormente, em 1958, Frank Rosenblatt introduziu o Perceptron, considerado a primeira rede neural treinável, capaz de ajustar seus parâmetros para aprender padrões simples (ROSENBLATT, 1958). Esse modelo marcou um avanço importante, mas apresentava limitações: era incapaz de resolver problemas não linearmente separáveis, como o famoso problema da função lógica XOR.

Na década de 1980, ocorreu um marco decisivo: o ressurgimento das RNAs com o desenvolvimento do algoritmo de retropropagação do erro (*backpropagation*), difundido por David Rumelhart, Geoffrey Hinton e Ronald Williams (1986). Esse algoritmo permitiu o treinamento eficiente de redes multicamadas, tornando possível a modelagem de problemas mais complexos. Se tornando uma parte fundamental para evolução das RNAs.

Nos anos 2000, com o aumento da capacidade computacional e a popularização das unidades de processamento gráfico (GPUs), além da disponibilidade de grandes bases de dados, as RNAs ganharam novo impulso. Esse período ficou conhecido como a era do DL, em que surgiram arquiteturas como as

Redes Neurais Convolucionais (CNNs) e as Redes Recorrentes (RNNs), aplicadas em diversas áreas do conhecimento (LECUN, 2015).

Uma RNA é composta por unidades chamadas neurônios artificiais, organizados em camadas. A camada de entrada (*input layer*) é responsável por receber os dados do ambiente externo. Por exemplo, em uma rede para reconhecimento de imagens, cada pixel da imagem pode corresponder a um neurônio de entrada. As camadas ocultas (*hidden layers*) realizam o processamento intermediário, extraindo padrões e características dos dados. Quanto maior o número de camadas ocultas, maior a profundidade da rede. A camada de saída (*output layer*) fornece o resultado final do processamento, que pode ser uma classificação, previsão ou valor numérico (HAYKIN, 2001).



Fonte NORDBY, J (2019).

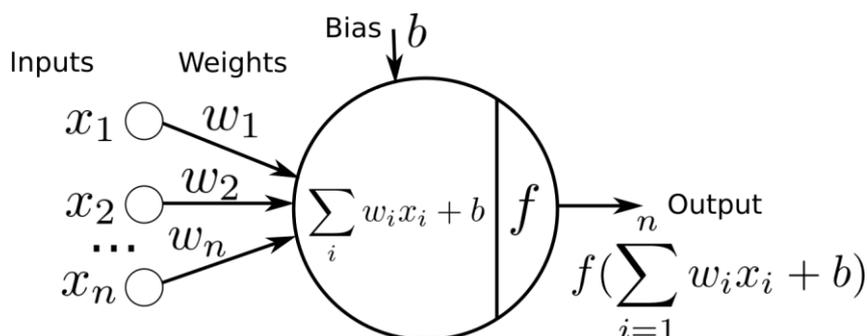
A figura acima representa um conjunto de neurônios formando uma camada (coluna) e se comunicando com os demais.

Matematicamente, o funcionamento de cada neurônio pode ser representado da seguinte forma:

$$z = \sum_{i=1}^n (w_i * x_i) + b \quad (6)$$

Nessa expressão, z representa a saída do neurônio, x_i são as entradas, w_i os pesos, b o *bias* (viés) e n o número de conexões para cada neurônio.

Figura 8 - Princípio computacional de um neurônio artificial



Fonte NORDBY, J (2019).

Posteriormente, o valor z é transformado por uma função de ativação $f(z)$, gerando a saída do neurônio: $a = f(z)$. Esse processo é repetido em cada camada da rede, até a geração da saída final.

Os pesos w são os parâmetros que determinam a importância de cada entrada no neurônio. Durante o treinamento da rede, os pesos são ajustados para que a rede aprenda os padrões do conjunto de dados (GOODFELLOW, 2016).

O *bias* tem a função de deslocar a função de ativação, permitindo maior flexibilidade no ajuste da rede. Sem o *bias*, as funções ficariam restritas a passar pela origem do gráfico, o que limitaria a capacidade de modelagem.

De forma mais didática, para aprender como funciona os cálculos por trás da rede neural mais simples, o perceptron, supondo que alguém deseja criar uma rede neural para lhe dizer se alguém deve ou não comprar um computador. Considerando três pontos de análise: existem computadores bons no mostruário? Está com promoção? É uma compra de urgência? Cada ponto representa uma variável x_1 , x_2 , x_3 , respectivamente. E Y representa a resposta real. O sistema pode adquirir valores verdadeiros (1) ou falsos (0). Assim, se cria uma tabela de possibilidades.

Tabela 1 - Análise de situações

Situação	x_1	x_2	x_3	Y
1°	1	1	1	1
2°	0	0	0	0
3°	0	0	1	1
4°	0	1	1	1
5°	1	0	0	0
6°	1	1	0	1
7°	0	1	0	0

Fonte: Autor (2025).

Observando a 4ª situação como exemplo: não tem computadores bons no mostruário ($x_1 = 0$), mas está em promoção ($x_2 = 1$) e é uma compra urgente ($x_3 = 1$), logo, deve-se comprar o computador ($Y = 1$).

Agora considerando um valor aleatório para os pesos e para o bias igual a zero (de preferência diferente de zero, considerou-se zero para simplificação didática), os primeiros cálculos já podem ser feitos. Destrinchando a equação (6):

$$z = x_1 \cdot w_1 + x_2 \cdot w_2 + x_3 \cdot w_3 + b_0 \quad (7)$$

Os valores obtidos de z serão colocados na função de ativação ($f(z)$):

$$f(z) = \begin{cases} 1, & \text{se } z > 0 \\ 0, & \text{se } z \leq 0 \end{cases} \quad (8)$$

Calculando a primeira situação, temos:

$$z = 1 \cdot 0 + 1 \cdot 0 + 1 \cdot 0 + 0 = 0 \quad (9)$$

Como $z = 0$, então $f(z) = 0$ e $f(z) \neq Y$, logo a rede previu errado. Então o ajuste é feito através de duas equações:

$$w_i = w_i + \Delta w_i \quad (10)$$

O Δw_i é dado por:

$$\Delta w_i = \eta \cdot (Y_i - \hat{Y}_i) \cdot x_i \quad (11)$$

Onde η é a taxa de aprendizado (constante), Y_i é o valor real e o \hat{Y}_i é o valor fornecido pela rede. Para o bias a equação acima é a mesma, fazendo apenas uma consideração: o x_i é sempre 1.

Assim, considerando o η igual a 0,1. O ajuste do w_1 pode ser calculado:

$$\Delta w_1 = 0,1 \cdot (1 - 0) \cdot 1 = 0,1 \quad (12)$$

Logo:

$$w_1 = 0 + 0,1 = 0,1 \quad (13)$$

Dessa forma, temos um novo valor para o peso w_1 . O ajuste dos demais pesos e o bias seguem o mesmo raciocínio. Após feito para a primeira situação, segue o cálculo para as outras.

As funções de ativação são fundamentais para introduzir não linearidade no modelo. Entre as mais utilizadas estão: Sigmoide, Tangente Hiperbólica (\tanh), ReLU e *Softmax*. Cada uma possui vantagens e desvantagens, sendo escolhidas conforme a natureza do problema.

A função sigmoide, também conhecida como função logística, é amplamente utilizada em redes neurais artificiais como mecanismo de ativação, especialmente em problemas de classificação binária. Sua formulação matemática é dada por:

$$f(x) = \text{sigmoide}(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

Onde x representa a entrada do neurônio. Essa função apresenta um comportamento não linear e produz valores contínuos no intervalo $(0, 1)$, o que possibilita a interpretação direta de suas saídas como probabilidades (HAYKIN, 2001).

Uma característica importante da sigmoide é a existência de uma região de saturação: para valores de entrada muito positivos ou muito negativos, a derivada da função torna-se próxima de zero. Essa propriedade, embora contribua para limitar a saída em um intervalo finito, pode ocasionar o chamado problema do gradiente desvanecente (*vanishing gradient*), dificultando a atualização dos pesos durante o processo de retropropagação em redes profundas (GOODFELLOW, 2016).

Apesar dessa limitação, a função sigmoide permanece relevante em situações que requerem a obtenção de probabilidades normalizadas, como na camada de saída de classificadores binários, onde sua interpretação probabilística e sua suavidade são vantajosas para a modelagem estatística de decisões.

A função tangente hiperbólica é expressa como:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (15)$$

Seus valores estão no intervalo $[-1, 1]$, o que a torna centrada em zero, diferentemente da sigmoide. Isso ajuda a acelerar a convergência durante o treinamento. Melhor que a sigmoide em muitos casos, pois os valores negativos ajudam no balanceamento dos gradientes, apropriada em camadas intermediárias. Também sofre com o problema do desvanecimento do gradiente (diminuição da capacidade de atualização dos pesos), não é a melhor escolha em redes muito profundas.

A *Rectified Linear Unit* (ReLU) é atualmente uma das funções mais utilizadas em redes convolucionais. É definida como:

$$f(x) = \max(0, x) \quad (16)$$

Ou seja, valores negativos são transformados em zero, enquanto valores positivos permanecem inalterados (NAIR, 2010). Simples e computacionalmente eficiente, evita o problema do desvanecimento do gradiente em grande parte dos casos, permite treinamento mais rápido e eficaz em redes profundas. Pode sofrer com o problema dos neurônios mortos, quando muitas ativações se tornam zero e deixam de atualizar seus pesos.

A função *softmax* é amplamente utilizada em camadas de saída de redes neurais para classificação multiclasse. É definida como:

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (17)$$

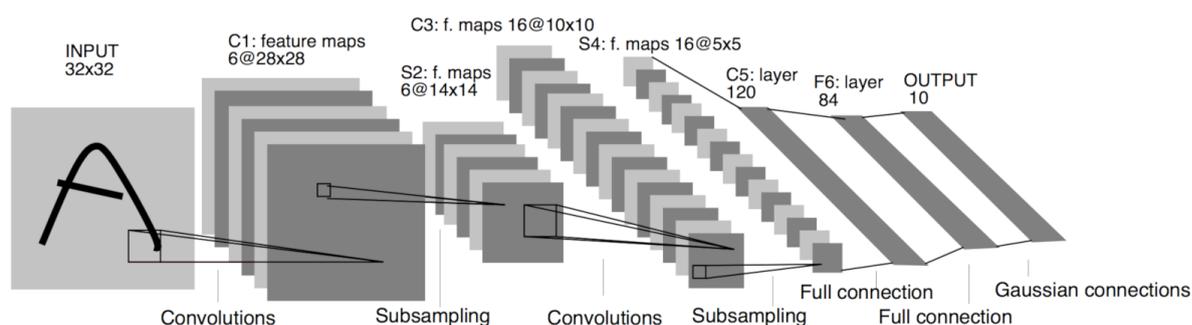
Onde x_i representa a saída do neurônio i e n é o número de classes. Essa função transforma os valores em probabilidades normalizadas, cujo somatório é igual a 1. Essencial para classificação multiclasse, fornece interpretação probabilística clara, porém, mais sensível a valores extremos na entrada e computacionalmente mais custosa que a ReLU.

As funções de ativação são um dos principais elementos que determinam a capacidade de generalização de uma rede neural. Enquanto a sigmoide e a tanh foram

amplamente usadas em arquiteturas tradicionais, a ReLU tornou-se padrão em redes profundas devido à sua simplicidade e desempenho. Já a softmax é indispensável para problemas de classificação multiclasse. A escolha adequada depende do tipo de rede, profundidade do modelo e natureza dos dados analisados.

Diversos modelos de redes neurais foram desenvolvidos ao longo da história, cada um com finalidades específicas. O Perceptron é o modelo simples de camada única, eficiente apenas para problemas linearmente separáveis (ROSENBLATT, 1958). O Multilayer Perceptron (MLP) é uma rede de múltiplas camadas, treinada por backpropagation, capaz de modelar problemas não lineares. Já as CNNs são especializadas em imagens, utilizando filtros convolucionais para detecção de padrões (LECUN, 1998).

Figura 9 - Representação da arquitetura de uma CNN clássica (LeNet-5).



Fonte LeCun (1998).

A imagem mostra o fluxo de uma CNN (LeNet-5) desde a entrada da imagem até a classificação final. As primeiras camadas aprendem padrões locais (bordas, texturas), as camadas intermediárias combinam esses padrões para formar representações mais complexas, as últimas camadas tomam decisões de classificação.

O processo de treinamento de uma RNA envolve duas etapas principais: propagação direta (*feedforward*) e retropropagação do erro (*backpropagation*). Na fase de propagação direta, os dados percorrem a rede da entrada até a saída, realizando os cálculos em cada camada. Já na retropropagação do erro, os erros entre saída prevista e desejada são propagados de volta pela rede, ajustando os pesos de acordo com a regra do gradiente descendente (RUMELHART, 1986). Esse processo é repetido até que a rede atinja um nível satisfatório de desempenho.

As RNAs representam um dos pilares da inteligência artificial moderna, com aplicações em diversas áreas do conhecimento, como saúde, engenharia, economia e processamento de linguagem natural. Seu desenvolvimento histórico demonstra a evolução contínua da área, desde os primeiros modelos simplificados até arquiteturas complexas de aprendizado profundo. A compreensão de seus elementos fundamentais — camadas, neurônios, pesos, bias, funções de ativação e algoritmos de treinamento — é essencial para o estudo e aplicação prática dessa tecnologia.

2 PROCEDIMENTO METODOLÓGICO

O desenvolvimento do estudo envolveu três etapas principais: coleta de dados, construção do modelo de CNN e por último o treinamento e teste do sistema.

A pesquisa foi desenvolvida de forma experimental e aplicada, com enfoque em manutenção preditiva de sistemas mecânicos rotativos. O objeto de estudo foi uma máquina rotativa para ventilação de pequeno porte, submetido a duas condições de operação: normal – funcionamento sem alterações estruturais e desbalanceado – adicionando uma massa em uma das pás da hélice para simular falha mecânica.

2.1 COLETA DOS DADOS

Os áudios foram coletados em ambiente controlado, utilizando microfones do celular a uma taxa de amostragem de 44,1 kHz, com duração de 10 a 11 segundos. Foram realizadas 240 gravações de áudio, sendo 120 referentes ao funcionamento normal e 120 ao funcionamento com falha. A extração dos dados foi feita em duas partes: primeiro após ligar a máquina com a hélice em perfeito estado, e esperado a estabilização da velocidade, foram feitas as gravações; em seguida foi adicionado um pequeno peso em uma das pás da hélice, após a ligar o motor novamente e esperar alguns segundos para estabilizar a velocidade foi feita a segunda parte das gravações.

2.2 ARQUITETURA DA REDE NEURAL CONVOLUCIONAL

A arquitetura de rede neural desenvolvida neste trabalho foi baseada em uma CNN implementada em Python por meio da biblioteca TensorFlow/Keras pela Plataforma do Google Colab. A arquitetura da CNN é composta por três blocos convolucionais sucessivos. O primeiro bloco contém uma camada Conv2D com 64 filtros de tamanho 3×3 , seguida por uma camada de *Batch Normalization* e uma

operação de *MaxPooling2D* com janela 2×2 . Esse padrão é repetido no segundo bloco, com 128 filtros, e no terceiro bloco, com 256 filtros, permitindo a extração progressivamente mais abstrata de padrões locais a partir do espectrograma. A convolução é responsável por detectar estruturas espectrais, como variações de energia em determinadas frequências, enquanto a normalização contribui para estabilizar o treinamento e acelerar a convergência. O *pooling*, por sua vez, reduz a dimensionalidade dos mapas de ativação, preservando as informações mais relevantes e aumentando a robustez contra pequenas variações no sinal.

Após a fase convolucional, o modelo utiliza uma camada *Flatten*, que transforma os mapas bidimensionais em um vetor unidimensional, o qual é então conectado a uma camada densa (*fully connected*) com 256 neurônios e função de ativação ReLU. Essa etapa é responsável por combinar as características extraídas e construir representações mais discriminativas do sinal. Para mitigar o risco de *overfitting*, foi aplicada uma camada de *Dropout* com taxa de 50%, desativando aleatoriamente metade das ativações durante o treinamento.

A camada de saída é composta por um único neurônio com função de ativação sigmóide, que produz um valor entre 0 e 1, representando a probabilidade de a amostra pertencer à classe “defeituosa”. Essa configuração é apropriada para problemas de classificação binária, sendo treinada com a função de perda *binary cross-entropy* e o otimizador Adam com taxa de aprendizado de 10^{-4} , é um dos otimizadores mais utilizados em redes neurais.

A *Sample Rate* (taxa de amostragem) corresponde ao número de amostras de áudio capturadas por segundo. Essa taxa utilizada (44,1 kHz) é a mesma utilizada em gravações de qualidade de CD e garante fidelidade suficiente para representar os harmônicos característicos de falhas mecânicas. Apesar de taxas mais altas fornecerem maior resolução sonora, elas também aumentam o custo computacional do processamento.

A duração dos sinais foi padronizada. Com isso, obteve-se consistência no banco de dados, garantindo que todos os espectrogramas possuíssem o mesmo tamanho. Embora gravações mais longas capturem maior variação do comportamento dinâmico do sistema, também tornam o treinamento da rede mais pesado.

O número de filtros Mel foi estabelecido como parâmetro para a conversão dos sinais em espectrogramas. Esse valor define a resolução espectral utilizada para representar o áudio. Um número maior de filtros permite identificar detalhes de

frequência mais sutis, o que pode favorecer a classificação de padrões relacionados a falhas. Entretanto, o aumento da dimensionalidade também eleva o risco de sobreajuste (*Overfitting*) do modelo e o tempo de processamento.

Outro parâmetro importante é o *Batch Size* (tamanho do lote), que determina quantas amostras são processadas antes da atualização dos pesos da rede. Lotes menores tendem a gerar aprendizado mais estável, embora demandem mais tempo de processamento. Já lotes maiores aceleram o treinamento, mas podem comprometer a capacidade de generalização do modelo.

O *Epochs* (número de épocas), que corresponde às vezes em que a rede passa por todo o conjunto de treinamento, também deve ser controlado. Poucas épocas podem resultar em *underfitting*, ou seja, quando a rede não aprende adequadamente os padrões dos dados. Por outro lado, muitas épocas aumentam o risco de *overfitting*, situação em que o modelo memoriza os exemplos de treino e perde capacidade de generalizar. Uma prática recomendada nesses casos é o uso do método *early stopping*, que interrompe o treinamento quando não há mais melhoria no desempenho da validação.

A definição de uma semente aleatória (*Seed*) foi utilizada para garantir reprodutibilidade dos resultados. Isso significa que, ao repetir o treinamento com a mesma configuração, os resultados obtidos permanecem consistentes, o que é fundamental em trabalhos científicos.

Por fim, aplicou-se a técnica de aumento de dados (*Data Augmentation*), obtidas por meio da adição de ruído, alteração de pitch (percepção da altura de um som – grave ou agudo), deslocamento temporal (*time shift*), estiramento. Esse procedimento aumenta o número de amostras disponíveis para treinamento e melhora a robustez do modelo frente a ruídos externos. No entanto, também aumenta o tempo necessário para processamento e treinamento da rede.

2.3 MÉTRICAS DE ANÁLISES

No contexto da avaliação de modelos de aprendizado de máquina, especialmente em redes neurais, diversas métricas são empregadas para mensurar o desempenho preditivo. Dentre as mais utilizadas, destacam-se acurácia, precisão

(*precision*), revocação (*recall*), medida-F1 (*f1-score*) e suporte (*support*).

Acurácia é uma métrica amplamente utilizada em problemas de classificação para avaliar o desempenho de modelos de aprendizado de máquina, incluindo redes neurais. Ela indica a proporção de previsões corretas em relação ao total de observações avaliadas, fornecendo uma visão geral da capacidade do modelo em classificar corretamente os dados. Matematicamente, a acurácia é expressa como a razão entre o número de previsões corretas (verdadeiros positivos e verdadeiros negativos) e o número total de amostras:

$$Acurácia = \frac{VP + VN}{T} \quad (18)$$

Onde, *VP* (Verdadeiros Positivos): número de instâncias positivas corretamente classificadas; *VN* (Verdadeiros Negativos): número de instâncias negativas corretamente classificadas e *T* são as observações totais.

Embora seja intuitiva e de fácil interpretação, a acurácia apresenta limitações em conjuntos de dados desbalanceados, nos quais algumas classes possuem muito mais exemplos que outras. Nesses casos, um modelo pode apresentar alta acurácia mesmo sem identificar corretamente as classes minoritárias, tornando necessário o uso de métricas complementares, como precisão, revocação e F1-score, para uma avaliação mais completa do desempenho do modelo.

Precisão consiste na proporção de exemplos classificados corretamente como positivos em relação ao total de exemplos previstos como positivos. Em outras palavras, representa a capacidade do modelo em evitar falsos positivos. Essa métrica é particularmente relevante em situações em que o custo de classificar incorretamente uma amostra negativa como positiva é elevado (TAN, 2019).

Essa métrica é definida como:

$$Precisão = \frac{VP}{VP + FP} \quad (19)$$

Onde, *FP* (Falsos Positivos): número de instâncias negativas incorretamente classificadas como positivas.

A revocação, também conhecida como sensibilidade ou taxa de verdadeiros

positivos, corresponde à proporção de exemplos corretamente identificados como positivos em relação ao total real de positivos no conjunto de dados. Essa métrica é essencial em cenários nos quais a prioridade é minimizar a ocorrência de falsos negativos, como no diagnóstico de falhas em sistemas ou em contextos médicos (GOODFELLOW, 2016).

$$Revocação = \frac{VP}{VP + FN} \quad (20)$$

Onde FN (Falsos Negativos): número de instâncias positivas incorretamente classificadas como negativas.

A $F1$ -score é definida como a média harmônica entre precisão e revocação. Ela busca equilibrar ambas as métricas, sendo apropriada quando se deseja uma avaliação global do modelo em contextos nos quais tanto os falsos positivos quanto os falsos negativos apresentam impactos significativos. Assim, a medida-F1 fornece uma visão mais equilibrada do desempenho do modelo (JURAFSKY, 2021).

$$F1 = 2 \cdot \frac{Precisão \cdot Revocação}{Precisão + Revocação} \quad (21)$$

Métricas, quando analisadas em conjunto, permitem uma avaliação robusta e detalhada do desempenho de modelos de classificação supervisionada, sendo fundamentais para validar a eficácia de redes neurais aplicadas à identificação de falhas em sistemas mecânicos, como ventiladores industriais.

O *loss*, também denominado função de perda ou função de custo, corresponde a uma métrica que quantifica a diferença entre a saída prevista por um modelo e o valor esperado, servindo como guia para o processo de otimização durante o treinamento da rede neural. Em termos práticos, a função de perda mede o grau de erro cometido pela rede a cada iteração, sendo o parâmetro fundamental que orienta o algoritmo de *backpropagation* no ajuste dos pesos sinápticos (GOODFELLOW, 2016).

3 RESULTADOS E DISCUSSÃO

Para obtenção dos resultados da pesquisa certos parâmetros no código foram modificados afim de otimizar a acurácia enquanto reduz o tempo de treinamento e o *overfitting*. Além disso, essas modificações tem como objetivo de entender o comportamento da rede diante de diversas situações.

A avaliação do desempenho de redes neurais convolucionais requer o uso de métricas e representações gráficas que permitam analisar tanto a capacidade de aprendizado do modelo quanto sua generalização para dados não vistos. Nesse contexto, três ferramentas se destacam: a matriz de confusão, as curvas de acurácia e as curvas de *loss*.

A matriz de confusão apresenta de forma estruturada o número de acertos e erros do classificador em cada classe, possibilitando identificar não apenas o desempenho global, mas também os tipos de erros cometidos.

As curvas de acurácia de treino e validação permitem observar a evolução da taxa de acertos ao longo das épocas de treinamento.

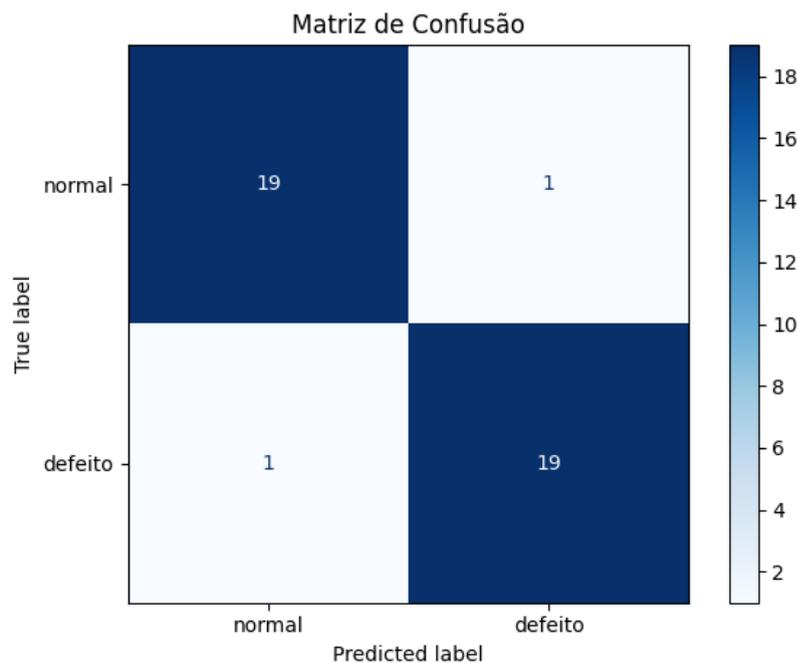
Por fim, a análise das curvas de *loss* fornece informações sobre o erro cometido pelo modelo em cada época.

Tabela 2 - Parâmetro e resultados, primeira análise

Parâmetros	Valores	Resultados	Valores
Duração (s)	5	Precisão – Normal	0,95
Filtros Mel	32	Recall – Normal	0,95
Batch Size	8	F1-Score – Normal	0,95
Epochs	20	Precisão – Defeito	0,95
Dados para Teste	40	Recall – Defeito	0,95
Data Augmentation	2	F1-Score – Defeito	0,95
		Acurácia Geral	0,95

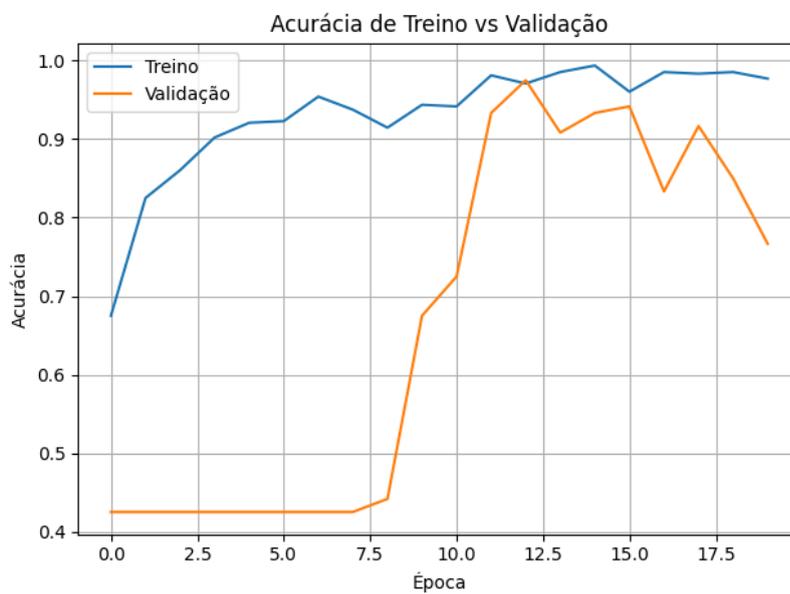
Fonte: Autor (2025).

Figura 10 - Matriz de confusão, primeira análise



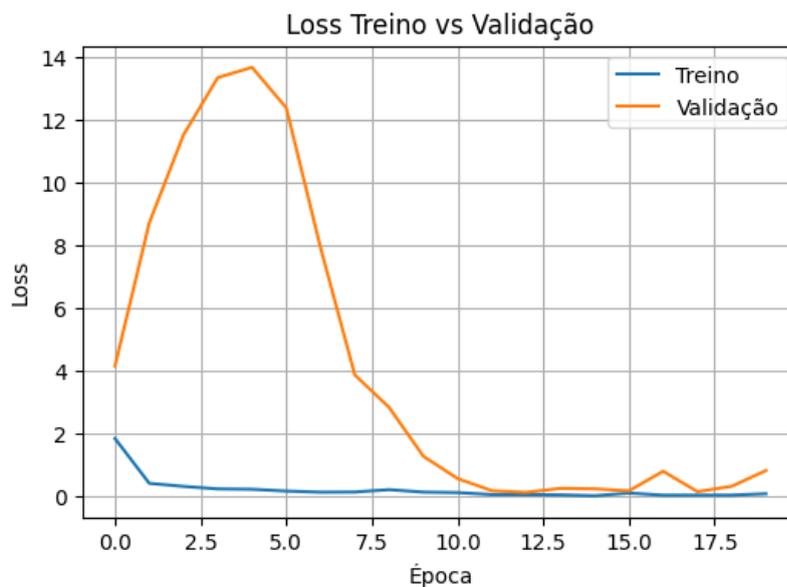
Fonte: Autor (2025).

Figura 11 - Acurácia de treino e de validação nas épocas, primeira análise



Fonte: Autor (2025).

Figura 12 - Loss do treino e da validação nas épocas, primeira análise



Fonte: Autor (2025).

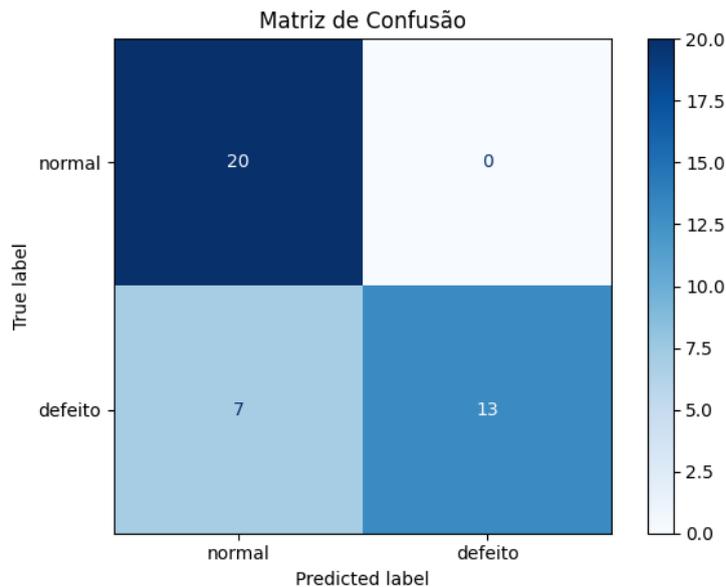
O teste inicial mesmo com apenas 5 segundos de áudio, o modelo já separa bem as classes, acurácia geral em 95%. Entretanto, a perda de 5 segundos de conteúdo temporal provavelmente privou a rede de pistas acústicas importantes que aparecem em janelas mais longas. A acurácia da validação apresentou uma inércia inicial e diminuiu em aproximadamente 12 épocas, devidas as limitações impostas, o modelo deve estar sofrendo *overfitting*.

Tabela 3 - Parâmetro e resultados, segunda análise

Parâmetros	Valores	Resultados	Valores
Duração (s)	10	Precisão – Normal	0,71
Filtros Mel	64	Recall – Normal	1,00
Batch Size	64	F1-Score – Normal	0,83
Epochs	40	Precisão – Defeito	1,00
Dados para Teste	40	Recall – Defeito	0,60
Data Augmentation	4	F1-Score – Defeito	0,75
		Acurácia Geral	0,80

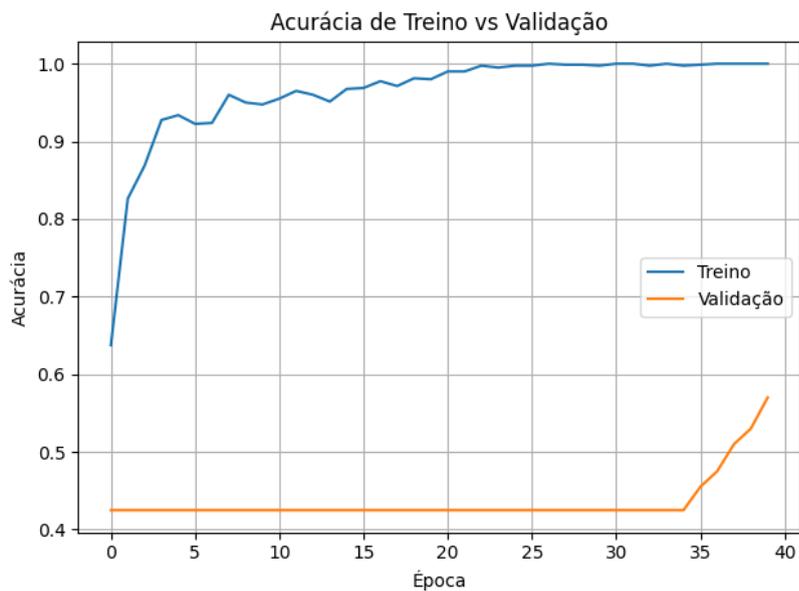
Fonte: Autor (2025).

Figura 13 - Matriz de confusão, segunda análise



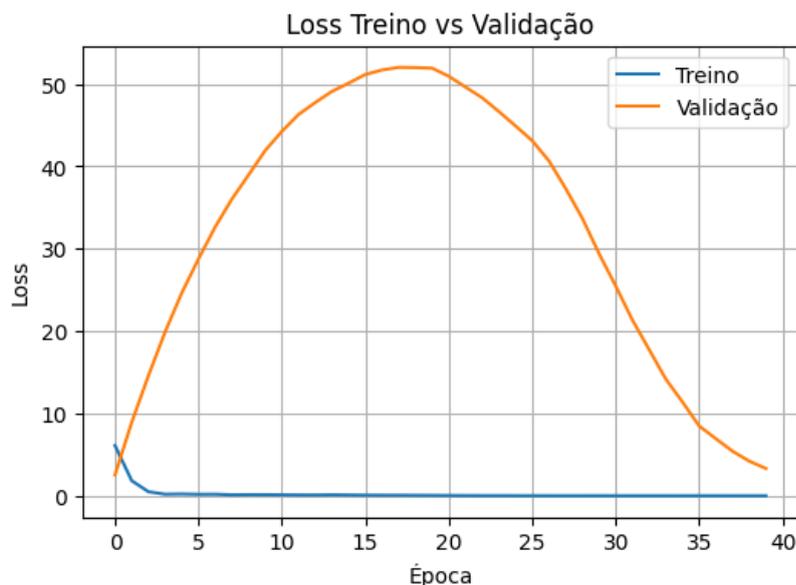
Fonte: Autor (2025).

Figura 14 - Acurácia de treino e de validação nas épocas, segunda análise



Fonte: Autor (2025).

Figura 15 - Loss do treino e da validação nas épocas, segunda análise



Fonte: Autor (2025).

A segunda combinação de valores mostrou resultados piores do que o primeiro teste. O tamanho do lote foi o fator preponderante para um baixo nível de treinamento da rede. Acurácia de treino sobe rápido e chega próxima de 1,0, mostrando que a rede aprendeu bem os padrões nos dados de treino. Já a acurácia de validação fica estagnada em torno de 0,40 por quase 30 épocas e só começa a subir timidamente no final, chegando a aproximadamente 0,55. Isso indica forte sobreajuste (*overfitting*): a rede decorou o conjunto de treino, mas não conseguiu generalizar para a validação.

O *loss* de treino cai rapidamente e se mantém baixo, sinal de aprendizado nos dados de treino. O *loss* de validação sobe progressivamente até atingir valores muito altos e só depois começa a cair, mas permanece bem acima do treino. Esse padrão confirma que o modelo não conseguiu generalizar, ficando preso em soluções que funcionam apenas para o treino.

Batch size muito alto (64): reduziu a variabilidade dos gradientes, levando a um aprendizado menos robusto e pior generalização. Data augmentation insuficiente (4): não gerou variabilidade suficiente para compensar o efeito do lote grande. Desequilíbrio no ajuste da rede: pode ter criado um viés em favor da classe “normal”, já que todos normais foram reconhecidos, mas defeitos foram parcialmente ignorados.

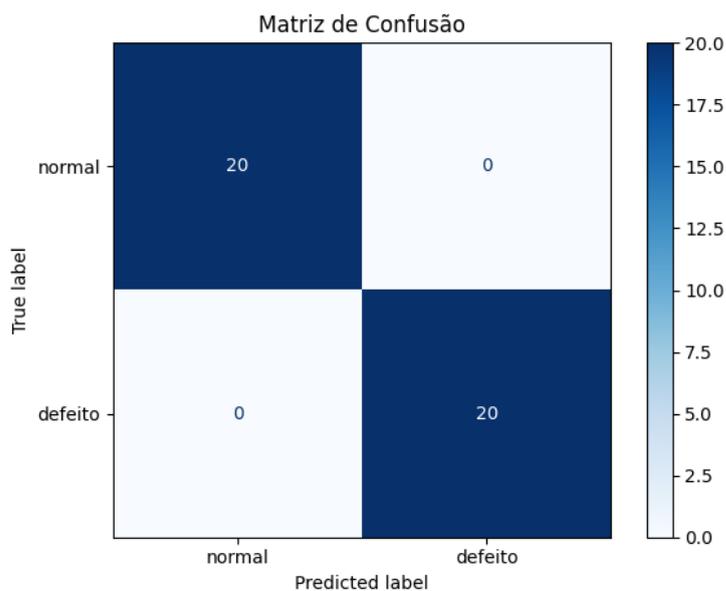
Tabela 4 - Parâmetro e resultados, terceira análise

Parâmetros	Valores	Resultados	Valores
Duração (s)	10	Precisão – Normal	1,00

Filtros Mel	64	Recall – Normal	1,00
Batch Size	64	F1-Score – Normal	1,00
Epochs	40	Precisão – Defeito	1,00
Dados para Teste	40	Recall – Defeito	1,00
Data Augmentation	6	F1-Score – Defeito	1,00
		Acurácia Geral	1,00

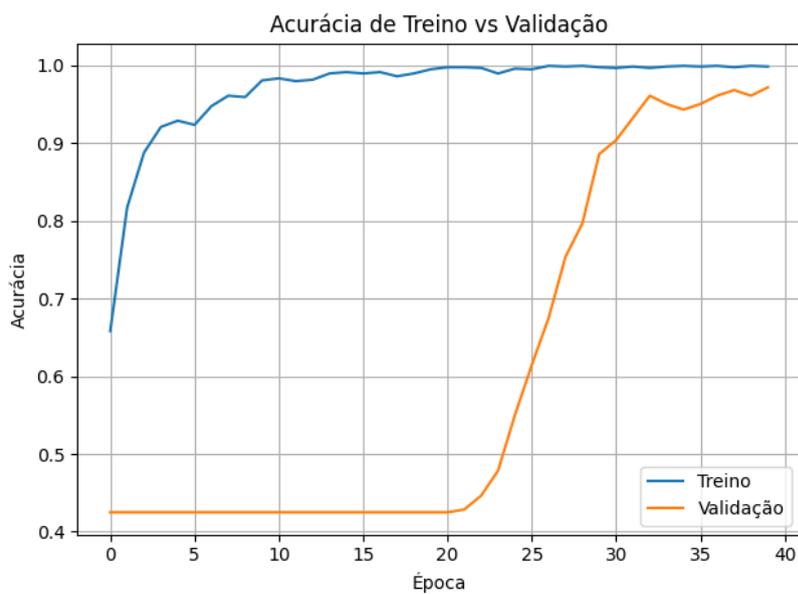
Fonte: Autor (2025).

Figura 16 - Matriz de confusão, terceira análise.



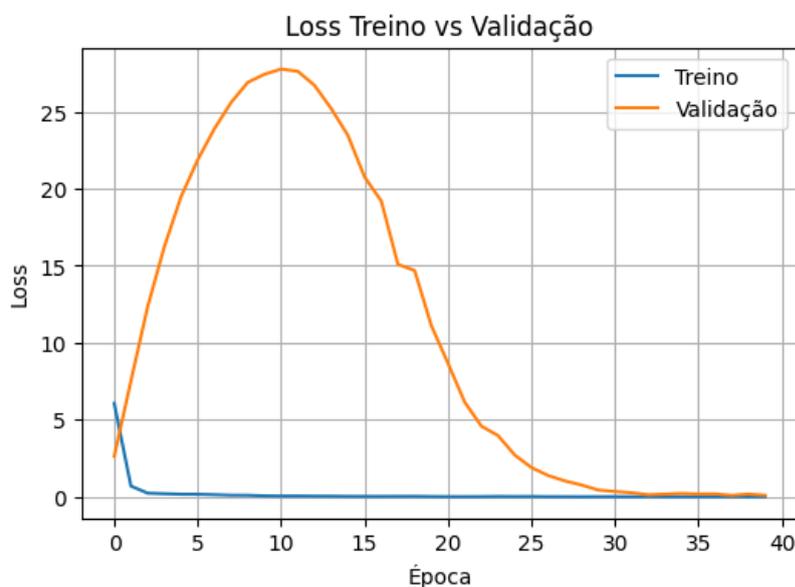
Fonte: Autor (2025).

Figura 17 - Acurácia de treino e de validação nas épocas, terceira análise



Fonte: Autor (2025).

Figura 18 - Loss do treino e da validação nas épocas, terceira análise



Fonte: Autor (2025).

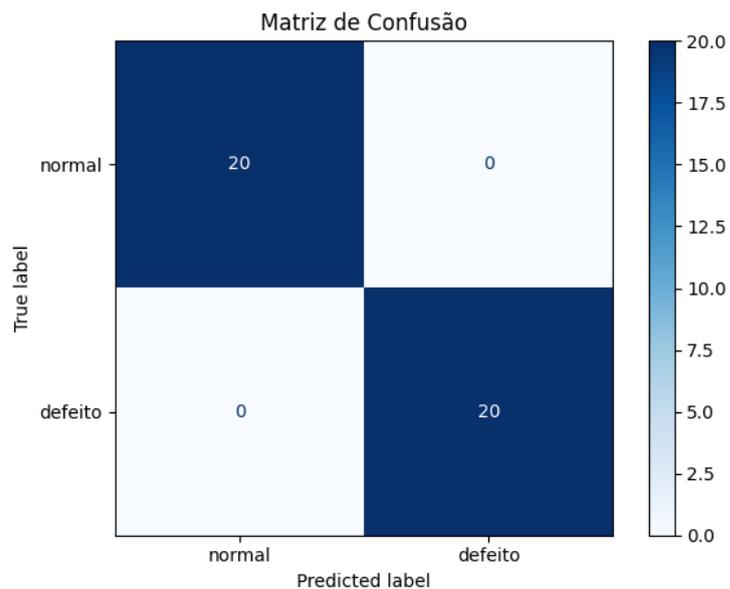
O aumento do *data argmentation* foi crucial para a estabilidade da rede e a possibilidade de um treinamento efetivo. Isto, demonstra a importância do grande número de dados para treinamento.

Tabela 5 - Parâmetro e resultados, quarta análise

Parâmetros	Valores	Resultados	Valores
Duração (s)	10	Precisão – Normal	1,00
Filtros Mel	64	Recall – Normal	1,00
Batch Size	8	F1-Score – Normal	1,00
Epochs	40	Precisão – Defeito	1,00
Dados para Teste	40	Recall – Defeito	1,00
Data Augmentation	7	F1-Score – Defeito	1,00
		Acurácia Geral	1,00

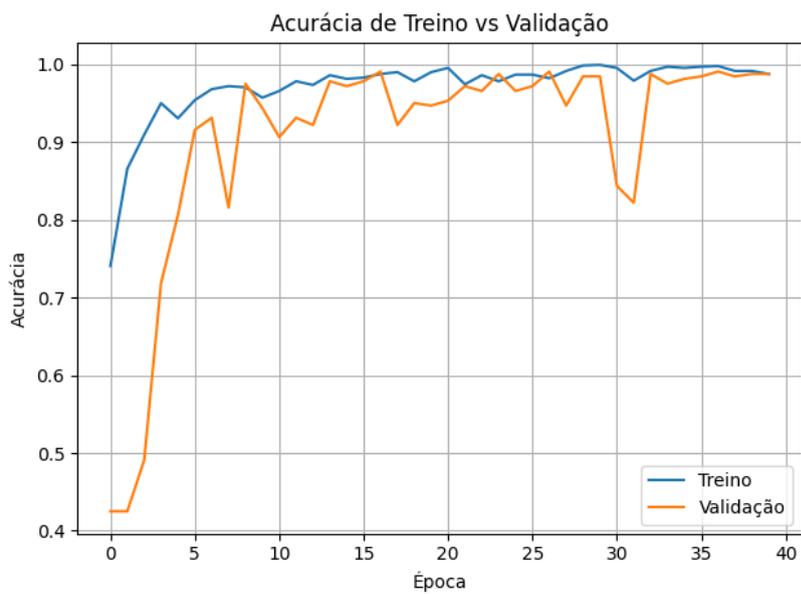
Fonte: Autor (2025).

Figura 19 - Matriz de confusão, quarta análise



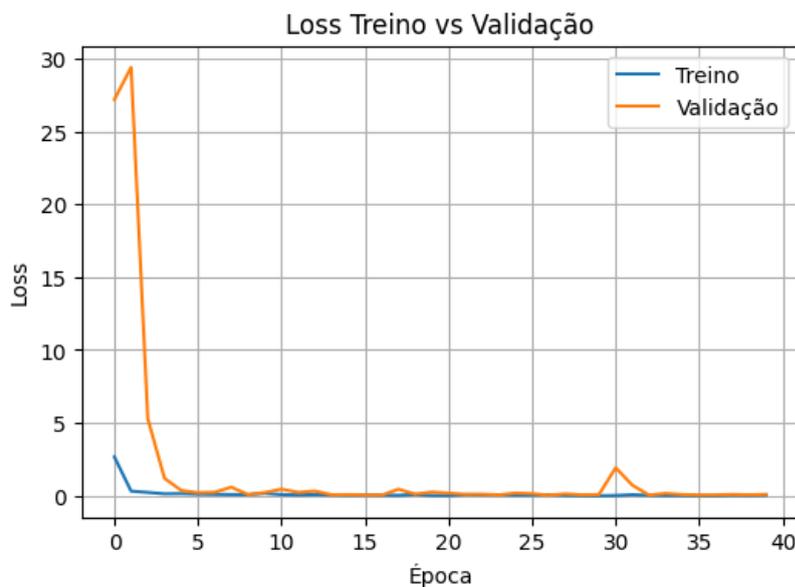
Fonte: Autor (2025).

Figura 20 - Acurácia de treino e de validação nas épocas, quarta análise



Fonte: Autor (2025).

Figura 21 - Loss do treino e da validação nas épocas, quarta análise



Fonte: Autor (2025).

A matriz de confusão apresenta zero erros, ou seja, nenhum falso positivo e nenhum falso negativo. Isso é o melhor cenário possível, mostrando que o modelo conseguiu distinguir de forma clara os dois padrões sonoros.

Na curva de curácia (acurácia vs validação), a curva azul (treino) cresce rápido e estabiliza próxima de 1,0. A curva laranja (validação) apresenta mais oscilação nas primeiras épocas, mas converge para aproximadamente 0,98 a 1,00 após 15 épocas. Essa convergência sem divergência entre treino e validação indica que o modelo não sofreu *overfitting* significativo, mesmo com 40 épocas.

Na curva de *loss* (Treino vs Validação), a *loss* (perda) de treino cai rapidamente já nas primeiras 5 épocas. O *loss* de validação começa alto, mas desaba logo no início e se estabiliza próximo de zero.

Pequenos picos ocasionais no *loss* de validação (ex.: perto da época 30) são normais e indicam flutuações na generalização, mas não comprometem o resultado final.

Batch pequeno (8): Forneceu gradientes mais ruidosos, mas esse “ruído” ajudou na generalização, evitando que a rede ficasse presa em mínimos locais.

Duração (10 segundos): Fundamental para capturar assinaturas acústicas completas do ciclo de funcionamento dos motores.

Filtros Mel (64): Garantiu boa resolução espectral, aumentando a riqueza de detalhes nos espectrogramas sem sobrecarregar demais o modelo.

Augmentation (7): Adicionou variabilidade suficiente para que a rede não “decorasse” os exemplos originais, melhorando robustez.

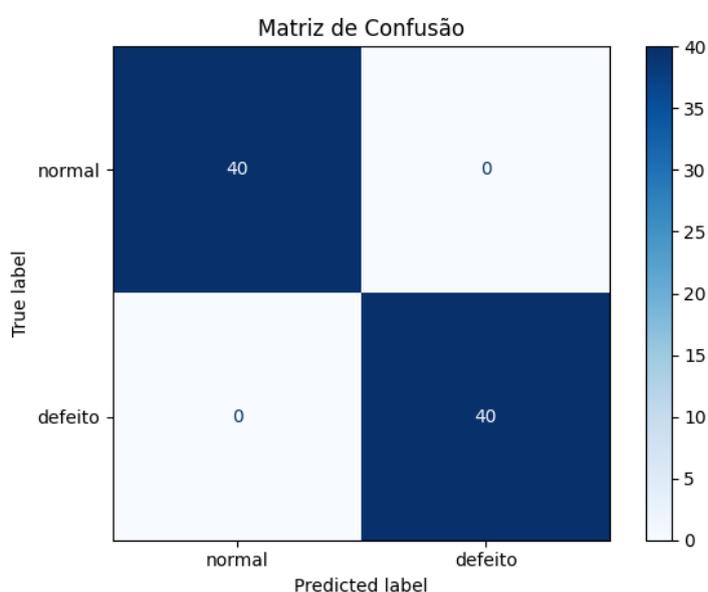
O modelo não apenas alcançou 100% no conjunto de teste, como também apresentou estabilidade entre treino e validação, indicando forte capacidade de generalização.

Tabela 6 - Parâmetro e resultados, quinta análise

Parâmetros	Valores	Resultados	Valores
Duração (s)	10	Precisão – Normal	1,00
Filtros Mel	64	Recall – Normal	1,00
Batch Size	8	F1-Score – Normal	1,00
Epochs	80	Precisão – Defeito	1,00
Dados para Teste	80	Recall – Defeito	1,00
Data Augmentation	7	F1-Score – Defeito	1,00
		Acurácia Geral	1,00

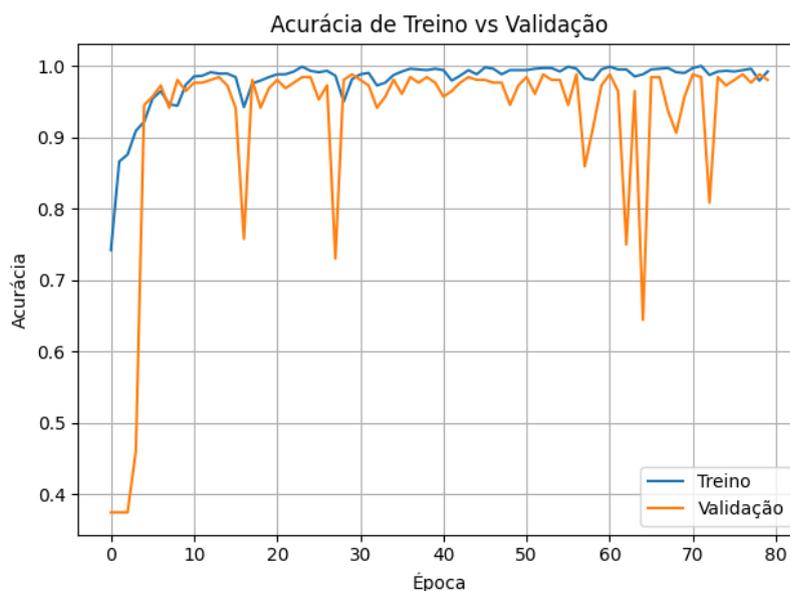
Fonte: Autor (2025).

Figura 22 - Matriz de confusão, quinta análise



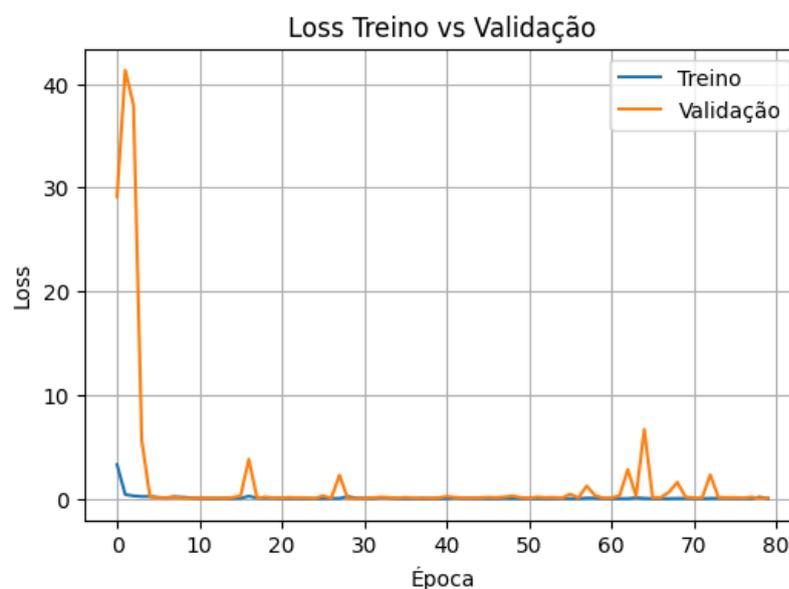
Fonte: Autor (2025).

Figura 23 - Acurácia de treino e de validação nas épocas, quinta análise



Fonte: Autor (2025).

Figura 24 - Loss do treino e da validação nas épocas, quinta análise



Fonte: Autor (2025).

A última análise realizada teve como principal objetivo a validação final da rede neural desenvolvida, permitindo avaliar de maneira mais aprofundada o seu desempenho e a consistência dos resultados obtidos. Esse procedimento foi essencial para verificar a capacidade do modelo em generalizar para novos dados, confirmando sua eficácia e evidenciando todo o potencial da abordagem proposta no contexto estudado.

3.1 INTERPRETAÇÃO DOS RESULTADOS

O presente trabalho demonstrou a viabilidade da utilização das CNNs para a identificação automática de falhas em máquinas rotativas a partir de sinais acústicos. Os experimentos realizados evidenciaram que a combinação de parâmetros adequados de pré-processamento, como a transformação em espectrogramas Mel, juntamente com técnicas de *data augmentation*, é capaz de proporcionar uma classificação altamente precisa entre equipamentos normais e defeituosos.

Os resultados obtidos mostraram acurácia superior a 95% na maioria das configurações testadas, alcançando até 100%. Observou-se também que lotes excessivamente grandes (*batch size* = 64) sem a devida compensação em *augmentation* resultaram em queda significativa no desempenho, reforçando a importância do balanceamento entre complexidade do modelo, tamanho do lote e variabilidade dos dados.

Dessa forma, conclui-se que a abordagem proposta apresenta elevado potencial de aplicação prática, podendo ser empregada como ferramenta auxiliar na manutenção preditiva e no monitoramento em tempo real de sistemas giratórios.

4 CONCLUSÃO

Este trabalho apresentou o desenvolvimento e aplicação de uma Rede Neural Convolucional para detecção de falhas em motores rotativos, a partir da análise de áudios. O experimento comprovou que a abordagem é capaz de identificar anomalias com alto índice de acerto, validando sua utilização em contextos de manutenção preditiva.

As principais contribuições do estudo foram:

- Desenvolvimento de um banco de dados experimental de áudios em condições normais e desbalanceadas.
- Avaliação de métricas de desempenho que evidenciaram a robustez do modelo.
- Demonstração da viabilidade de CNNs para análise de sinais acústicos e espectrogramas para manutenção preditiva.

Como limitações, destaca-se o tamanho reduzido da base de dados e a restrição do estudo a apenas um tipo de falha (desbalanceamento). Para trabalhos futuros, sugere-se a ampliação do banco de dados, o uso de diferentes tipos de falhas (desgaste de rolamentos, desalinhamento, entre outros) e a integração de técnicas híbridas envolvendo sinais de vibração e imagens, também, a exploração de arquiteturas mais avançadas de aprendizado profundo e a validação em ambientes industriais, a fim de consolidar e generalizar os resultados obtidos.

Dessa forma, o estudo contribui para o avanço da aplicação de inteligência artificial na Engenharia Mecânica, reforçando o papel das CNNs como ferramentas promissoras no contexto da Indústria 4.0 e da manutenção inteligente.

REFERÊNCIAS

BURTON, T. et al. **Wind energy handbook**. 2. ed. Chichester: John Wiley & Sons, 2011.

DEEP LEARNING BOOK. **Capítulo 9 – A Arquitetura das Redes Neurais**. Disponível em: <https://www.deeplearningbook.com.br/a-arquitetura-das-redes-neurais/>. Acesso em: 01 set. 2025.

FENG, Y.; TAVNER, P. J.; CHEN, B. **Condition monitoring for wind turbine planetary gearbox**. *Wind Energy*, v. 16, n. 5, p. 728–740, 2013.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 4. ed. New York: Pearson, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016.

GRIEVES, M.; VICKERS, J. **Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems**. In: TRANS-DISCIPLINARY PERSPECTIVES ON COMPLEX SYSTEMS. Springer, Cham, 2017. p. 85–113.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.

HE, K.; ZHANG, X.; REN, S.; SUN, J. **Deep Residual Learning for Image Recognition**. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. p. 770–778.

HENG, A. et al. **Rotating machinery prognostics: State of the art, challenges and opportunities**. *Mechanical Systems and Signal Processing*, v. 23, n. 3, p. 724–739, 2009.

IEA – International Energy Agency. **World Energy Outlook 2022**. Paris: OECD/IEA, 2022.

JANSSENS, O.; SLUIJS, B.; VERVLIET, S.; VANHOECKE, S.; LUCHTENBERG, K.; VANRUMSTE, B. **Convolutional neural network based fault detection for rotating machinery**. *Journal of Sound and Vibration*, v. 377, p. 331-345, 2016.

JARDINE, A. K. S.; LIN, D.; BANJEVIC, D. **A review on machinery diagnostics and prognostics implementing condition-based maintenance**. *Mechanical Systems and Signal Processing*, v. 20, n. 7, p. 1483–1510, 2006.

JURAFSKY, Daniel; MARTIN, James H. **Speech and Language Processing**. 3. ed. Stanford: Draft, 2021.

KHEDDAR, Hamza; HEMIS, Mustapha; HIMEUR, Yassine. **Automatic Speech**

Recognition using Advanced Deep Learning Approaches: A survey. Information Fusion, 2024.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **Imagenet classification with deep convolutional neural networks.** In: Advances in Neural Information Processing Systems (NIPS), 2012. p. 1097–1105.

KROKHMAL, Pavlo; ZHANG, Hao. **On the Evaluation Metrics for Classification Problems.** Journal of Applied Statistics, v. 47, n. 4, p. 689-706, 2020.

LECUN, Yann; BOTTOU, Leon; BENGIO, Yoshua; HAFFNER, Patrick. **Gradient-based learning applied to document recognition.** *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning.** Nature, v. 521, p. 436–444, 2015.

LEE, J.; KAO, H.-A.; YANG, S. **Service innovation and smart analytics for Industry 4.0 and big data environment.** Procedia CIRP, v. 16, p. 3–8, 2014.

LEE, Jay. **Industrial big data analytics and cyber-physical systems for future maintenance & service innovation.** Procedia CIRP, v. 38, p. 3-7, 2015.

LI, X.; DING, Q.; SUN, J.-Q. **Remaining useful life estimation in prognostics using deep convolution neural networks.** Reliability Engineering & System Safety, v. 172, p. 1–11, 2020.

LIU, R. et al. **Machine health monitoring using adaptive deep transfer learning.** IEEE Transactions on Industrial Electronics, v. 68, n. 12, p. 12361-12371, 2021.

MALLAT, S. **A Wavelet Tour of Signal Processing.** 2. ed. San Diego: Academic Press, 1999.

MCCULLOCH, W. S.; PITTS, W. **A logical calculus of the ideas immanent in nervous activity.** Bulletin of Mathematical Biophysics, v. 5, p. 115–133, 1943.

MUNDIM FILHO, Nelson Alves. **Classificação da maturação de frutos de café utilizando redes neurais convolucionais.** 2021. 61 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) – Universidade Federal de Uberlândia, Uberlândia, 2021.

NAIR, V.; HINTON, G. E. **Rectified linear units improve restricted boltzmann machines.** In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010. p. 807-814.

NORDBY, J. **Environmental Sound Classification Using Convolutional Neural Networks for Resource Constrained Devices.** Master's Thesis – Norwegian University of Science and Technology, Trondheim, 2019.

PICZAK, K. J. **Environmental sound classification with convolutional neural**

networks. In: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2015. p. 1–6.

RANDALL, R. B. **Vibration-based condition monitoring: industrial, aerospace and automotive applications**. Chichester: John Wiley & Sons, 2011.

RANGEL, Rodrigo Fill. **Transformada de Fourier em Visão Computacional**. Medium (Turing Talks), 21 mar. 2021. Disponível em: <https://medium.com/turing-talks/transformada-de-fourier-b1775e891cc5>. Acesso em: [06 set. 2025].

ROBERTS, Leland. **Understanding the Mel Spectrogram**. Analytics Vidhya, 6 mar. 2020. Disponível em: <https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53>. Acesso em: 01 set. 2025.

ROSENBLATT, F. **The perceptron: a probabilistic model for information storage and organization in the brain**. Psychological Review, v. 65, n. 6, p. 386–408, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. **Learning representations by back-propagating errors**. Nature, v. 323, p. 533–536, 1986.

SANVEZZO JUNIOR, J. **Redes neurais convolucionais aplicadas à detecção de não-conformidades em equipamentos industriais**. Monografia (Engenharia de Computação) – Universidade Federal de Uberlândia, 2022.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. In: International Conference on Learning Representations (ICLR), 2014.

STEVENS, S. S.; VOLKMAN, J.; NEWMAN, E. B. **A scale for the measurement of the psychological magnitude pitch**. The Journal of the Acoustical Society of America, v. 8, n. 3, p. 185–190, 1937.

SUN, R. et al. **Convolutional neural networks for fault diagnosis of wind turbine bearings**. Energies, v. 9, n. 4, p. 1-15, 2016.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. 2. ed. Boston: Pearson, 2019.

WINDBOX BRASIL. **Componentes dos aerogeradores**. Blog Windbox, [S.l.], 2023. Disponível em: <https://windbox.com.br/blog/componentes-dos-aerogeradores/>. Acesso em: 6 set. 2025.

ZAHER, A.; MCARTHUR, S. D. J.; INFELD, D. **Application of data-mining techniques for wind turbine condition monitoring**. Wind Energy, v. 12, n. 6, p. 574–593, 2009.