

Bruno Stinghel Mattedi

**Predição de Indicadores de Qualidade de Energia
Utilizando Decomposição em Frequência e
Redes Neurais Profundas**

Brasil

Vitória, 2022

Bruno Stinghel Mattedi

**Predição de Indicadores de Qualidade de Energia
Utilizando Decomposição em Frequência e Redes Neurais
Profundas**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Universidade Federal do Espírito Santo

Programa de Pós-Graduação em Engenharia Elétrica

Orientador: Prof. Dr. Patrick Marques Ciarelli

Brasil

Vitória, 2022

Bruno Stinghel Mattedi

Predição de Indicadores de Qualidade de Energia Utilizando Decomposição em Frequência e Redes Neurais Profundas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Trabalho aprovado. Brasil, 12 de dezembro de 2022:

Prof. Dr. Patrick Marques Ciarelli
Orientador

**Prof. Dr. Domingos Sávio Lyrio
Simonetti**
Universidade Federal do Espírito Santo
Examinador Interno

Prof. Dr. Bruno Légora Souza da Silva
Instituto Federal de Educação, Ciência e
Tecnologia do Espírito Santo
Examinador Externo

Brasil
Vitória, 2022

Agradecimentos

Não poderia deixar de agradecer aos meus pais, Carmen e Francisco, e a minha irmã, Flávia, pelo companheirismo e por serem exemplos de dedicação e me ajudarem durante todo o percurso. Agradeço a minha noiva, Poliana, que me deu forças e me manteve motivado durante todas as etapas.

Agradeço a todos os professores do Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) que me instruíram nos momentos de aula, passando o conhecimento e ajudando a me tornar um profissional melhor. Além disso, gostaria de agradecer em especial ao meu orientador Patrick Marques Ciarelli e ao professor Klaus Fabian Côco por estar sempre disponível para tirar dúvidas e me ajudar ativamente com o mestrado.

Também sou grato aos meus amigos, quase irmãos, que fiz durante toda a jornada da UFES, proporcionando descontração perante os momentos difíceis, amigos sem os quais esta jornada não teria sido tão animadora.

Por fim, à banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho.

“.. the only way to survive the tides of changes is to ride them, not to stand in their path.”
(Sup. Magus, Legion)

Resumo

Com a evolução dos processos industriais, o mercado tem se tornado cada vez mais competitivo, onde qualquer vantagem pode ser o diferencial para uma empresa ser ou não bem sucedida. Este preceito também se aplica ao mercado de energia cativo, mesmo embora o consumidor não tenha a opção de escolher a sua distribuidora de energia. Isso ocorre pelo fato de que a Agência Nacional de Energia Elétrica (ANEEL), com seu poder regulatório, estabeleceu limites para alguns indicadores da qualidade do serviço, que ao serem ultrapassados pelas distribuidoras, acarretam em multas a serem pagas. Assim, o controle desses indicadores proporciona uma grande vantagem estratégica, de modo que os investimentos possam ser otimizados para evitar que os limites sejam violados.

Neste contexto, este trabalho apresenta uma abordagem baseada no uso de redes neurais para a predição de indicadores de uma distribuidora de energia elétrica. Mais especificamente, serão preditos os indicadores coletivos de continuidade da qualidade de serviço: Duração Equivalente de Interrupção por Unidade Consumidora (DEC) e Frequência Equivalente de Interrupção por Unidade Consumidora (FEC). O foco é criar predições que possuam um bom desempenho, possibilitando uma melhor gestão dos gastos com a redução dos valores pagos em multas. Neste trabalho foram utilizados três tipos de redes neurais, sendo elas: uma rede neural rasa, uma *Long Short-Term Memory* (LSTM) e uma *Convolutional Neural Network* (CNN) combinada com LSTM (CNN+LSTM). Além disso, buscando a redução da complexidade dos dados, foram utilizadas duas técnicas de pré-processamento baseadas em decomposição em frequências das séries temporais: uma baseada em Transformada Wavelet e outra baseada em decomposição empírica das séries temporais.

A abordagem proposta baseada em redes neurais e decomposição das séries temporais foi aplicada a um conjunto de dados reais composto por indicadores DEC e FEC de uma distribuidora de energia elétrica. Comparando os resultados dos métodos propostos, foi possível observar que a decomposição baseada em Transformada Wavelet combinada com as redes LSTM e CNN+LSTM apresentaram melhores desempenhos para a predição dos indicadores DEC e FEC, respectivamente.

Palavras-chave: Aprendizado de Máquina; Aprendizado Profundo; Indicadores de Qualidade; Séries Temporais; Redes Neurais Convolucionais.

Abstract

With the evolution of industrial processes, the market has become increasingly competitive, where any advantage can be the differential for a company to be successful or not. This precept also applies to the captive energy market, even though the consumer does not have the option of choosing his energy distributor. This is because the National Electric Energy Agency (ANEEL), with its regulatory power, has established limits for some service quality indicators, which, when exceeded by the distributors, result in fines being paid. Thus, the control of these indicators provides a great strategic advantage, so that investments can be optimized to avoid the limits being violated.

In this context, this work presents an approach based on the use of neural networks for the prediction of indicators of an electric energy distributor. More specifically, the collective indicators of service quality continuity will be predicted: the System Average Interruption Duration Index (SAIDI) and the System Average Interruption Frequency Index (SAIFI). The focus is on creating predictions that perform well, enabling better management of expenses by reducing amounts paid in fines. In this work, three types of neural networks were used, namely: a shallow neural network, a Long Short-Term Memory (LSTM), and a Convolutional Neural Network (CNN) combined with LSTM (CNN+LSTM). In addition, seeking to reduce the complexity of the data, two pre-processing techniques based on the frequency decomposition of the time series were used: one based on the Wavelet Transform and the other based on the empirical decomposition of the time series.

The proposed approach based on neural networks and decomposition of time series was applied to a real dataset composed of SAIDI and SAIFI indicators of an electricity distribution company. Comparing the results of the proposed methods, it was possible to observe that the decomposition based on the Wavelet Transform combined with the LSTM and CNN+LSTM networks presented better performances for the prediction of the SAIDI and SAIFI indicators, respectively.

Keywords: Machine Learning; Deep Learning; Quality Indicators; Time Series; Convolutional Neural Networks.

Lista de ilustrações

Figura 1 – Compensações mensais, trimestrais e anuais referentes a qualidade do serviço	25
Figura 2 – Indicadores coletivos de continuidade e seus respectivos limites ao longo dos anos	26
Figura 3 – Modelo de um neurônio artificial	31
Figura 4 – Função de ativação linear	32
Figura 5 – Função de ativação sigmoide	32
Figura 6 – Função de ativação sigmoide e tangente hiperbólica	33
Figura 7 – Rede <i>Feedforward</i>	34
Figura 8 – Rede TDNN	35
Figura 9 – Desdobramento de uma RNN	36
Figura 10 – Exemplo de extrações de características por camada	36
Figura 11 – Gradiente em declive acentuado vs plano	37
Figura 12 – Camada LSTM	38
Figura 13 – Bloco/Célula LSTM	38
Figura 14 – Processo de convolução 2D com <i>stride</i> e <i>padding</i>	40
Figura 15 – Processo de convolução 1D	41
Figura 16 – Janela deslizante no tempo	43
Figura 17 – EMD de um sinal. Mostrando apenas 6 das 9 IMFs extraídas.	46
Figura 18 – Exemplo de série de DEC (a) e FEC (b) para o primeiro semestre do ano de 2014.	50
Figura 19 – Amostra do <i>DEC</i> antes e depois do tratamento de <i>outliers</i>	52
Figura 20 – Subséries após Transformada Wavelet	52
Figura 21 – Subséries após CEEMDAN	53
Figura 22 – Janela Deslizante no Tempo	55
Figura 23 – Visão geral do fluxo das séries temporais no método proposto	55
Figura 24 – Estrutura da rede CNN+LSTM	56
Figura 25 – Comparação dos erros nas diferentes configurações para o indicador DEC	60
Figura 26 – Comparação dos erros nas diferentes configurações para o indicador FEC	62
Figura 27 – Predição do DEC para a rede profunda CNN+LSTM com pré-processamento MODWT	63
Figura 28 – Predição do DEC para a rede profunda CNN+LSTM com pré-processamento CEEMDAN	63
Figura 29 – Predição do DEC para a rede profunda CNN+LSTM sem decomposição (SD)	64

Figura 30 – Predição do DEC para a rede rasa TDNN com pré-processamento CEEMDAN	64
Figura 31 – Predição do DEC para a rede rasa TDNN com pré-processamento MODWT	65
Figura 32 – Predição do DEC para a rede rasa TDNN sem decomposição (SD) . . .	65
Figura 33 – Predição do DEC para a rede LSTM com pré-processamento CEEMDAN	66
Figura 34 – Predição do DEC para a rede LSTM com pré-processamento MODWT	66
Figura 35 – Predição do DEC para a rede LSTM sem decomposição (SD)	67
Figura 36 – Predição do FEC para a rede profunda CNN+LSTM com pré-processamento MODWT	67
Figura 37 – Predição do FEC para a rede profunda CNN+LSTM com pré-processamento CEEMDAN	68
Figura 38 – Predição do FEC para a rede profunda CNN+LSTM sem decomposição (SD)	68
Figura 39 – Predição do FEC para a rede rasa TDNN com pré-processamento CEEMDAN	69
Figura 40 – Predição do FEC para a rede rasa TDNN com pré-processamento MODWT	69
Figura 41 – Predição do FEC para a rede rasa TDNN sem decomposição (SD) . . .	70
Figura 42 – Predição do FEC para a rede LSTM com pré-processamento CEEMDAN	70
Figura 43 – Predição do FEC para a rede LSTM com pré-processamento MODWT	71
Figura 44 – Predição do FEC para a rede LSTM sem decomposição (SD)	71
Figura 45 – Comparação do DEC (azul) e da média mensal por dia (laranja)	73
Figura 46 – Comparação do FEC (azul) e da média mensal por dia (laranja)	74

Lista de tabelas

Tabela 1 – Compensação paga pelas concessionárias e permissionárias no Brasil . . .	25
Tabela 2 – Indicadores Coletivos de Continuidade do Conjunto de São Mateus/ES para cada ano	49
Tabela 3 – Análise estatística dos indicadores coletivos de qualidade da rede elétrica	51
Tabela 4 – Média e desvio padrão das subséries CEEMDAN	54
Tabela 5 – Média e desvio padrão das subséries MODWT	54
Tabela 6 – RMSE, MAPE e MdRAE para as variações de rede e de decomposições utilizadas no pré-processamento de cada indicador. Em negrito os menores valores de cada erro para cada indicador.	61
Tabela 7 – RMSE, MAPE e MdRAE para os indicadores em relação a média mensal por dia	73

Lista de abreviaturas e siglas

ACL	Ambiente de Contratação Livre
ANEEL	Agência Nacional de Energia Elétrica
ANN	<i>Artificial Neural Network</i>
ARIMA	<i>AutoRegressive Integrated Moving Average</i>
BDGD	Base de Dados Geográfica da Distribuidora
CEEMDAN	<i>Complete Ensemble Empirical Mode Decomposition with Adaptive Noise</i>
CGSE	Câmara de Gestão do Setor Elétrico
CNN	<i>Convolutional Neural Network</i>
CNPE	Conselho Nacional de Política Energética
DEC	Duração Equivalente de Interrupção por Unidade Consumidora
DIC	Duração de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão
DICRI	Dia Crítico por Unidade Consumidora
DICRI	Duração da Interrupção Individual ocorrida em Dia Crítico por Unidade Consumidora ou por Ponto de Conexão
DMIC	Duração Máxima de Interrupção Contínua por Unidade Consumidora ou por Ponto de Conexão
DT	<i>Decision Trees</i>
FEC	Frequência Equivalente de Interrupção por Unidade Consumidora
FIC	Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão
GCE	Câmara de Gestão da Crise de Energia Elétrica
LSTM	<i>Long Short-Term Memory</i>
MAE	Mercado Atacadista de Energia Elétrica

MARS	<i>Multivariate Adaptive Regression Splines</i>
MISO	<i>Multiple Input Single Output</i>
MODWT	<i>Maximal Overlap Discrete Wavelet Transformation</i>
NARX	<i>Nonlinear autoregressive exogenous</i>
ONS	Operador Nacional do Sistema
PND	Programa Nacional de Desestatização
PNE	Plano Nacional de Eletrificação
PPGEE	Programa de Pós-Graduação em Engenharia Elétrica
PRODIST	Procedimentos de Distribuição de Energia Elétrica
PSDB	Partido da Social Democracia Brasileira
RESEB	Reestruturação do Setor Elétrico Brasileiro
RLM	Regressão Linear Múltiplas
RNA	Rede Neural Artificial
RNN	<i>Recurrent Neural Network</i>
SAIDI	<i>System Average Interruption Duration Index</i>
SAIFI	<i>System Average Interruption Frequency Index</i>
SED	Subestação de Distribuição
TDNN	<i>Time Delay Neural Network</i>
TMAE	Tempo de Atendimento às Ocorrências Emergenciais
TMD	Tempo Médio de Deslocamento
TME	Tempo Médio de Execução
TMP	Tempo Médio de Preparação
UFES	Universidade Federal do Espírito Santo

Sumário

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.2	Estrutura da Dissertação	16
2	REGULAMENTAÇÃO DO SETOR ELÉTRICO BRASILEIRO	17
2.1	Presença do Estado	17
2.2	Regulação pré-PRODIST	18
2.3	PRODIST	19
2.4	Módulo 8 - Qualidade do Fornecimento de Energia Elétrica	21
2.5	Qualidade do Serviço	21
2.5.1	Conjuntos de unidades consumidoras	21
2.5.2	Tempos de atendimento às ocorrências emergenciais	21
2.5.3	Indicadores individuais de continuidade	22
2.5.4	Compensações por violação dos limites	23
2.5.5	Indicadores Coletivos de Continuidade	24
2.6	Trabalhos Relacionados	26
3	APRENDIZADO DE MÁQUINAS E REDES NEURAIS	28
3.1	Aprendizado de Máquinas	28
3.2	Conceitos de Redes Neurais Artificiais	29
3.2.1	Neurônio Artificial	30
3.2.2	Redes <i>Feedforward</i>	33
3.2.3	Redes Neurais com Atraso no Tempo	34
3.2.4	Redes Neurais Recorrentes	35
3.3	Aprendizado Profundo	35
3.3.1	<i>Long Short-Term Memory</i>	36
3.3.2	Convolutional Neural Network	39
4	DECOMPOSIÇÃO DE SÉRIES TEMPORAIS	42
4.1	Séries Temporais	42
4.2	Transformada Wavelet de Máxima Sobreposição	43
4.3	<i>Complete Ensemble Empirical Mode Decomposition with Adaptive Noise</i>	45
5	BASES DE DADOS E MÉTODO PROPOSTO	49
5.1	Bases de Dados e Preparação dos Dados	50

5.1.1	Bases de Dados	50
5.1.2	Pré-processamento das Bases de Dados	51
5.2	Arquiteturas	54
5.2.1	Redes com Atraso no Tempo	55
5.2.2	LSTM	56
5.2.3	CNN+LSTM	56
5.2.4	Configurações Gerais	57
6	EXPERIMENTOS E RESULTADOS	58
6.1	Métricas	58
6.2	Procedimentos Experimentais	58
6.3	Recursos	59
6.4	Resultados	59
7	CONCLUSÃO E TRABALHOS FUTUROS	75
	REFERÊNCIAS	77

1 Introdução

É bem entendido que a eletricidade tem um papel importante no desenvolvimento dos países. Em um estudo apresentado por [Ferguson et al. \(2000\)](#), os quais analisaram 100 países, foi observada uma forte correlação entre esses dois parâmetros. Apesar de correlação não indicar causalidade, [Zhang et al. \(2017\)](#) apresentam uma relação entre o consumo energético e o crescimento econômico, bem como o contrário, a depender do local.

Dessa forma, com a relevância dada a eletricidade, os setores de geração, transmissão e distribuição vem ganhando mais atenção dos agentes governamentais. No Brasil, através da Lei nº 9.427 assinada pelo então presidente Fernando Henrique Cardoso em dezembro de 1996, foi determinada a criação da Agência Nacional de Energia Elétrica (ANEEL), com o objetivo de fiscalizar e regular a geração, transmissão, distribuição e comercialização de energia elétrica.

Em relação às distribuidoras, a ANEEL através do Módulo 8 do PRODIST estabelece normas e procedimentos para que o setor, que é um monopólio natural, seja atrativo para as empresas do setor, mas também para os consumidores finais. Este Módulo diz respeito a qualidade da energia elétrica e nele foi estabelecido um conjunto de exigências que são acompanhadas através de indicadores de natureza técnica e comercial, além de penalidades, com o objetivo de assegurar a qualidade dos serviços prestados por empresas distribuidoras de energia elétrica aos seus consumidores ([DIAS, 2002](#)), de maneira semelhante ao método de penalidade e recompensa apresentado em ([SIMAB; HAGHIFAM, 2012](#)).

Assim, para medir o desempenho e a confiabilidade de sistemas de distribuição de energia a partir do aspecto da qualidade do produto, existem indicadores de qualidade estabelecidos pela ANEEL. Dentre eles, destacam-se a Duração Equivalente de Interrupção por Unidade Consumidora (DEC) e Frequência Equivalente de Interrupção por Unidade Consumidora (FEC) ([ANEEL, 2021b](#)).

Em caso de violação das metas estabelecidas no período de apuração dos indicadores, as concessionárias sofrem penalidades ([SPERANDIO et al., 2004](#)) que vão desde a elaboração de um plano de melhorias, na qual a concessionária deve elaborar um plano para atingir as metas de seus conjuntos elétricos e apresentá-lo ao órgão fiscalizador, até a realização de compensações aos consumidores na forma de descontos. Por outro lado, o bom desempenho em tais indicadores é recompensado pela ANEEL através de ajuste tarifário anual como incentivo à busca pela melhoria contínua da qualidade do serviço. Conforme dados da [ANEEL \(2017a\)](#), no Brasil as concessionárias realizaram uma quantidade de 80,265 milhões de compensações em 2021, 79,681 milhões em 2020 e 81,306 milhões em

2019, que resultaram num montante de R\$ 722,95 milhões, R\$ 637,14 milhões e R\$ 642,36 milhões, respectivamente. As multas e as compensações onerosas para as distribuidoras afetam não só o Brasil, mas também vários outros países. Por exemplo, estima-se que os custos da falta de energia nos Estados Unidos foram de aproximadamente US\$ 79 bilhões para a economia em 2006 (LACOMMARE; ETO, 2006).

O conhecimento a priori dos valores dos indicadores de continuidade DEC e FEC pode auxiliar na definição de prioridades para distribuir de forma mais eficaz os investimentos e recursos para a redução de problemas no sistema elétrico, trazendo como consequência direta a melhoria dos índices de confiabilidade Tronchoni et al. (2008). Nesse aspecto, esse trabalho busca utilizar a análise e o processamento dos dados históricos, além do uso de redes neurais, para desenvolver uma forma que permita predizer, com o menor erro possível, esses indicadores de continuidade.

Sendo assim, esse trabalho propõe o uso de redes neurais profundas com técnicas de processamento de sinais para estimar os indicadores DEC e FEC. Um conjunto de dados de uma concessionária de energia elétrica, coletado entre 2010 e 2017, será utilizado para avaliação da proposta. Nos experimentos realizados foi utilizado um tipo de redes rasa composta por uma RNA com atraso no tempo, a TDNN, e duas arquiteturas de redes profundas, sendo elas a LSTM, pelo seu bom desempenho com séries temporais, e a CNN+LSTM, rede híbrida com capacidade de melhor extrair características de dados ruidosos.

Para o tratamento dos dados, além da normalização e remoção de pontos discrepantes, foram utilizadas duas metodologia que decompõe o sinal em subséries de frequência, sendo elas a MODWT e CEEMDAN.

Os resultados obtidos sobre uma base de dados reais mostraram que a abordagem proposta de decomposição da série apresentou resultados promissores frente ao treinamento sem esta etapa. Além disso, a aplicação da rede CNN+LSTM também resultou em um melhor desempenho se comparado às outras duas, LSTM e TDNN.

1.1 Objetivos

Este trabalho explora o uso de redes neurais profundas convolucionais e recorrentes para realizar a predição dos indicadores de qualidade, DEC e FEC, de um conjunto elétrico de uma distribuidora de energia. As redes regressoras utilizadas foram duas profundas, sendo uma com arquitetura híbrida composta por uma convolucional e uma recorrente e outra profunda com apenas modelo recorrente. Além disso, foi empregada uma rede rasa desacompanada com atraso no tempo.

Busca-se também investigar a aplicação de métodos de pré-processamento na série

temporal, destacando-se as técnicas de decomposição em frequência de modo a buscar reduzir a complexidade dos dados de entrada para aumentar o desempenho da rede.

Assim, a partir a junção do pré-processamento com a rede neural, busca-se que a rede consiga prever os indicadores com o menor erro possível, bem como analisar quais técnicas proporcionaram um bom desempenho para que este trabalho possa contribuir para a aplicação das distribuidoras de energia ajudando a melhorar a qualidade do serviço prestado.

O objetivo geral pode ser alcançado por meio dos seguintes objetivos específicos:

- Analisar o conjunto de dados de DEC e FEC de forma a compreender o seu comportamento;
- Aplicar algoritmos de pré-processamento de séries temporais buscando um melhor condicionamento dos dados;
- Estudar e analisar o uso de redes neurais para a predição dos indicadores DEC e FEC;
- Realizar experimentos e comparar o desempenho de diferentes arquiteturas de aprendizado.

1.2 Estrutura da Dissertação

Este trabalho está organizado como segue. A Seção 2 apresenta o histórico da regulamentação do setor elétrico brasileiro até o presente, incluindo os indicadores de continuidade e trabalhos à eles relacionados. Na Seção 3 é descrito sobre o aprendizado de máquina e as redes utilizadas neste trabalho. Na Seção 4 são especificadas as técnicas de pré-processamento aplicadas, compostas também pelas decomposições de frequência. A Seção 5 descreve a base de dados e a metodologia desenvolvida para predição dos indicadores de continuidade. Os experimentos, resultados e discussões são apresentados na Seção 6. Por fim, é realizada a conclusão na Seção 7.

2 Regulamentação do Setor Elétrico Brasileiro

2.1 Presença do Estado

O modelo estatal predominou no setor elétrico brasileiro durante as décadas de 60 e 70, tendo a Eletrobrás como principal indutora do processo. Apesar da lei da criação da Eletrobrás ter sido promulgada em 1954, com o Plano Nacional de Eletrificação (PNE) substituindo o Departamento Nacional de Águas e Energia Elétrica, o projeto foi concretizado apenas com a Lei n.º 3.890-A, de 25 de abril de 1961.

O PNE promoveu uma reestruturação no setor brasileiro, criando um estado mais presente através de empresas federais e estaduais, como a criação da Central Elétrica de Furnas S.A. em 1957 e a Companhia Hidrelétrica do Vale do Rio Paraíba em 1961.

Com a tomada de decisões sendo feita pelo setor público, houve uma centralização das decisões, permitindo a construção de grandes usinas hidrelétricas, como Itaipu, e grandes linhas de transmissão, acarretando no aumento da interconexão do sistema brasileiro, na redução dos custos e na melhoria da qualidade do setor energético (GOLDENBERG; PRADO, 2003).

Apesar dos esforços do Governo Brasileiro em melhorar a qualidade do setor elétrico do país, vários fatores contribuíram para o surgimento de uma crise institucional no setor, provocado pelo endividamento externo do país que impedia na continuidade de investimentos nas obras de expansão. Dentre estes fatores podem ser destacados o primeiro e o segundo choque do petróleo, a desvalorização da moeda brasileira, as elevadas taxas de juros no mercado internacional, e a suspensão dos empréstimos concedidos pelo México ao Brasil (GOLDENBERG; PRADO, 2003).

Em 1995, com o início do governo de Fernando Henrique Cardoso, houve a priorização de reformas e privatização em diversos setores da economia, como o setor elétrico, através do Programa Nacional de Desestatização (PND), motivado pelo largo endividamento do estado (PRONI; LYRIO, 2005).

Buscando adotar um modelo presente em países que transferiram a responsabilidade da expansão do sistema elétrico para o capital privado e, assim, introduzir um ambiente mais competitivo num ambiente naturalmente considerado como de monopólio, o governo federal veio a criar a Agência Nacional de Energia Elétrica (ANEEL).

Em 1996, por meio do Projeto de Reestruturação do Setor Elétrico Brasileiro (RESEB), foi dado início aos estudos para a reestruturação do setor elétrico do país, aspirando atribuir ao Estado as funções de planejamento, formulação, regulamentação e

fiscalização, e transferir para o setor privado os investimentos e operações.

Atendendo a esses objetivos, a ANEEL, criada pela Lei nº 9.427 de 26 de dezembro de 1996, é uma autarquia em regime especial vinculada ao Ministério de Minas e Energia, tendo como principais atribuições a fiscalização das concessões, permissões e serviços de energia elétrica, regulação da geração, transmissão, distribuição e comércio, além de ser responsável por estabelecer tarifas, implementar políticas e diretrizes relativas à exploração da energia.

2.2 Regulação pré-PRODIST

O setor elétrico teve sua regulação inicialmente estabelecida através do Decreto nº 24.643, de 10 de julho 1934, que em seu artigo 144 define às seguintes competências para o Governo Federal:

“Art. 144. O Serviço de Aguas (*sic*) do Departamento Nacional de Produção (*sic*) Mineral do Ministerio (*sic*) da Agricultura, é o órgão (*sic*) competente do Governo Federal para:

a) proceder ao estudo e avaliação de energia hydraulica (*sic*) do territorio (*sic*) nacional;

b) examinar e instruir technica (*sic*) e administrativamente os pedidos de concessão ou autorização para a utilização da energia hydraulica (*sic*) e para produção (*sic*), transmissão, transformação e distribuição da energia hydro-electrica (*sic*);

c) regulamentar e fiscalizar de modo especial e permanente o serviço de produção (*sic*), transmissão, transformação de energia hydro-electrica (*sic*);

d) exercer todas as atribuições que lhe forem conferidas por este Codigo (*sic*) e seu regulamento.”

(DEPUTADOS, 1934)

Em 1957 ele foi editado pelo Decreto nº 41.019 que regulamentava os serviços de energia elétrica. Em 1965 houve a criação do Departamento de Águas e Energia (DNAE) seguido, em 1968, pela criação da estrutura básica do Ministério das Minas e Energia, ocorrendo a alteração do nome do DNAE para Departamento Nacional de Águas e Energia Elétrica (DNAEE), após receber as atribuições do Conselho Nacional de Águas e Energia Elétrica, este criado em 1939. Conforme regimento interno aprovado em 1977, a natureza e finalidade do DNAEE são:

“Art.1º - O Departamento Nacional de Águas e Energia Elétrica - DNAEE, instituído pela Lei nº 4.904, de 17 de dezembro de 1965, com autonomia financeira assegurada pelo art. 18 do Decreto nº 75.468 de 11 de março de 1975, é o Órgão Central de Direção Superior responsável pelo planejamento, coordenação e execução dos estudos hidrológicos em todo o território nacional; pela supervisão, fiscalização e controle dos aproveitamentos das águas que alteram o seu regime; bem como pela supervisão, fiscalização e controle dos serviços de eletricidade.” (ANEEL, 2022b)

Por fim, houve a criação da ANEEL em 1996, conforme explicado anteriormente, vinculado ao Ministério de Minas e Energia, ficando então extinto o DNAEE.

2.3 PRODIST

Por meio do projeto Reestruturação do Setor Elétrico Brasileiro, que movimentou uma consultoria que resultou em estudos que levaram a criação da ANEEL, do Operador Nacional do Sistema (ONS) responsável por gerenciar a produção e a transmissão da energia, e do Mercado Atacadista de Energia Elétrica (MAE), definindo um conjunto de regras comerciais a serem seguidas por agentes e concessionárias que compram energia por atacado.

Conforme apresentado por [Cabral et al. \(2017\)](#), apesar das mudanças citadas e outras mais, a RESEB não conseguiu acompanhar a velocidade de expansão do sistema, levando às crises elétricas de 2001 com a ocorrência de diversos apagões, chegando a ter um estado de racionamento implementado. Abaixo um relato do ocorrido na época:

Entre 2000 e 2002, o Brasil enfrentou uma série de problemas no abastecimento de energia elétrica. O Governo de Fernando Henrique Cardoso (PSDB) teve de elaborar um plano de racionamento para gerenciar a crise. Foi estabelecida uma meta de economia de 20% no consumo energético e consumidores que não atingissem este valor veriam sua conta de luz mais cara. O racionamento acabou em 2002. ([OPOVO, 2018](#))

Frente aos problemas enfrentados, por meio da Câmara de Gestão da Crise de Energia Elétrica (GCE), buscou-se elaborar estudos e propor ações de modo a ter equivalência entre demanda e oferta de energia elétrica. Os estudos operaram em metas de curto prazo, para enfrentar o racionamento, e de médio prazo, visando a reestruturação do setor elétrico brasileiro. Estes estudos apresentaram a ineficiência do governo em realizar os investimentos adequados, em tomar ações para prevenir grandes racionamentos e para corrigir as falhas do mercado.

No ano de 2002, substituindo a GCE, foi criada a Câmara de Gestão do Setor Elétrico (CGSE), com função de sugerir ao Conselho Nacional de Política Energética (CNPE) diretrizes para elaboração de políticas do setor de energia elétrica.

Em 2004 ocorreram três mudanças significativas no setor elétrico, sendo criado o programa Luz Para Todos, durante o governo de Luiz Inácio Lula da Silva, buscando ampliar o acesso a energia para populações antes marginalizadas. Além disso, com a aprovação das Leis 10.847/2004 e 10.848/2004, foi autorizada a criação da Empresa de Pesquisa Energética (EPE) para, respectivamente, elaborar estudos e pesquisas de modo a fundamentar o planejamento do setor energético e estabelecer novas regras de comercialização de energia elétrica entre as concessionárias, permissionárias e autorizadas.

Assim, conforme apresentado por Cabral et al. (2017), foi adotada uma estratégia híbrida, dividida em duas partes, o Ambiente de Contratação Regulada (ACR), contando com os agentes de geração e distribuição, e o Ambiente de Contratação Livre (ACL), com os agentes de geração, comercialização, importadores e exportadores de energia, e também consumidores livres.

Por meio desse modelo, permitiu-se a criação de contratos de venda de energia de longo prazo, definiu a oferta de menor tarifa para participação nas licitações, mudou a metodologia de cálculo de lastro para vendas e contratações de usinas e assegurou o monitoramento da continuidade e de segurança de suprimento.

Com as diversas mudanças estruturais as quais o setor elétrico brasileiro foi passando, foi mudando a forma de regulamentação com o objetivo de garantir energia a uma tarifa acessível, com um sistema seguro e disponível para todos. Assim, objetivando unificar as diversas regulamentações atribuídas a concessionárias e permissionárias do serviço de distribuição, em 2008 a ANEEL publicou o Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST).

O PRODIST são documentos regulatórios que padronizam as atividades técnicas relacionadas ao funcionamento e desempenho de sistemas de energia elétrica, referentes à distribuição de energia elétrica, sendo atualmente composto por onze módulos:

- Módulo 1 – Glossário de Termos Técnicos do PRODIST
- Módulo 2 – Planejamento da Expansão do Sistema de Distribuição
- Módulo 3 – Acesso ao Sistema de Distribuição
- Módulo 4 – Procedimentos Operativos do Sistema de Distribuição
- Módulo 5 – Sistemas de Medição
- Módulo 6 – Informações Requeridas e Obrigações
- Módulo 7 – Cálculo de Perdas na Distribuição
- Módulo 8 – Qualidade do Fornecimento de Energia Elétrica
- Módulo 9 – Ressarcimento de Danos Elétricos
- Módulo 10 – Sistema de Informação Geográfica Regulatório
- Módulo 11 – Fatura de Energia Elétrica e Informações Suplementares

Este trabalho tem como foco a qualidade do fornecimento de energia elétrica, portanto, o Módulo 8 será descrito com detalhes na seção seguinte.

2.4 Módulo 8 - Qualidade do Fornecimento de Energia Elétrica

O Módulo 8 da PRODIST estabelece parâmetros de medição e coleta de dados, cálculo dos indicadores, referências, penalidades, submissão de informações à ANEEL, etc., referentes principalmente à qualidade do produto em regime permanente ou transitório e a qualidade do serviço, referente aos atendimentos emergenciais. Ele é dividido em quatro grupos, sendo eles:

- Qualidade do Produto
- Qualidade do Serviço
- Qualidade Comercial
- Segurança do Trabalho e Instalações

Em função do trabalho realizado nesta dissertação, na próxima seção será aprofundada a explicação sobre o grupo Qualidade do Serviço, mostrando os parâmetros que compõe os indicadores coletivos de continuidade e a importância desses.

2.5 Qualidade do Serviço

2.5.1 Conjuntos de unidades consumidoras

O conjunto de unidades consumidores abrange as redes média tensão (MT) à jusante da Subestação de Distribuição (SED), que seja de propriedade da concessionária de distribuição. Aquelas com um número menor ou igual a mil devem ser agregadas, mantendo a contiguidade das áreas, formando um único conjunto. Às que possuem entre mil e dez mil unidades, podem ser agregadas a outras, sem obrigatoriedade. Ambas as agregações devem respeitar o limite máximo de dez mil unidades e contiguidade das áreas, salvo as exceções destacadas no módulo 8 ([ANEEL, 2021b](#)).

Os conjuntos têm como propriedades os dados extraídos da Base de Dados Geográfica da Distribuidora (BDGD). Dessa forma, ocorrendo alguma mudança permanente no sistema que afete a configuração dos conjuntos a distribuidora deve propor a revisão dos conjuntos perante a ANEEL.

2.5.2 Tempos de atendimento às ocorrências emergenciais

Como parte dos dados que devem ser apurados pela distribuidora de energia estão os tempos médios de atendimento às emergências (TMAE), sendo eles vinculados aos conjuntos e devendo ser avaliados e controlados. O TMAE apresenta uma visão global do tempo de atendimento de cada conjunto em um determinado período, mas a partir dos

demais indicadores que o compõe é possível avaliar pontos fortes e fracos relacionados à operação da empresa. O TMAE é composto por:

- Tempo Médio de Preparação (TMP), que mede a eficácia dos métodos de comunicação adotados, do centro de operações, bem como o dimensionamento das equipes, começando a ser cronometrado a partir do momento que a ocorrência é registrada até uma equipe ser à ele atribuída. Ele é calculado pela Equação (2.1):

$$TMP = \frac{\sum_{i=1}^n TP(i)}{n}, \quad (2.1)$$

sendo n o número total de ocorrências emergenciais registradas no conjunto durante o período e TP o tempo de preparação para cada uma delas.

- Tempo Médio de Deslocamento (TMD), que tem como objetivo avaliar a alocação geográfica das equipes de campo, manutenção e operação, contando do momento da atribuição da ordem a equipe até a chegada ao local da ocorrência. Esse indicador é definido pela Equação (2.2):

$$TMD = \frac{\sum_{i=1}^n TD(i)}{n}, \quad (2.2)$$

sendo TD o tempo de deslocamento de cada ocorrência no período e conjunto a ser calculado.

- Tempo Médio de Execução (TME), que é calculado a partir do TE , que é o tempo de execução do serviço, buscando por meio deste avaliar a eficiência do reestabelecimento do sistema de distribuição pelas equipes de manutenção e operação. O seu valor é obtido pela Equação (2.3):

$$TME = \frac{\sum_{i=1}^n TE(i)}{n}. \quad (2.3)$$

Assim, com os indicadores acima, é possível realizar o cálculo do TMAE com a Equação (2.4):

$$TMAE = TMP + TMD + TME. \quad (2.4)$$

2.5.3 Indicadores individuais de continuidade

Conforme ANEEL (2021b), os indicadores individuais de continuidade registram a frequência e a duração das interrupções, permitindo a análise da qualidade do serviço e o desempenho do sistema elétrico. Com exceção da Duração da Interrupção Individual ocorrida em Dia Crítico por Unidade Consumidora (DICRI), os indicadores devem ser apurados mensalmente para todos os clientes.

As equações abaixo discriminam esses indicadores, valendo ressaltar que o cálculo é feito por unidade consumidora.

- Duração de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão (DIC). Definida pela Equação (2.5), ela é calculada a partir da soma da duração de todas as interrupções registradas no mês:

$$DIC = \sum_{i=1}^n t(i), \quad (2.5)$$

na qual $t(i)$ é o tempo da duração da interrupção i , i é o índice de interrupções e n é o total de interrupções da unidade consumidora considerada no mês analisado.

- Frequência de Interrupção Individual por Unidade Consumidora ou por Ponto de Conexão (FIC). Apresentada pela Equação (2.6), ela é representada pela quantidade de interrupções no mês (n).

$$FIC = n. \quad (2.6)$$

- Duração Máxima de Interrupção Contínua por Unidade Consumidora ou por Ponto de Conexão (DMIC). Ela é a interrupção com o maior tempo registrado no período, conforme Equação (2.7). Por meio desse indicador, é possível dar a devida atenção a valores atípicos fora da média.

$$DMIC = \max_{i=1, \dots, n} [t(i)]. \quad (2.7)$$

- Duração da Interrupção Individual ocorrida em Dia Crítico por Unidade Consumidora ou por Ponto de Conexão (DICRI), vide a Equação (2.8), é o tempo de duração da interrupção em um dia crítico. Ele é caracterizado quando a quantidade de ocorrências emergenciais em um dia superar a média, relativa aos últimos 24 meses anteriores ao ano em curso, em três desvios padrões (ANEEL, 2021a).

$$DICRI = t_{critico}. \quad (2.8)$$

2.5.4 Compensações por violação dos limites

Conforme ANEEL (2018), as compensações não almejam ressarcir o consumidor integralmente pela interrupção, apenas atuar o transtorno causado. Elas têm como principal objetivo incentivar que as distribuidoras busquem melhorar o serviço, frente ao custo caso sigam na direção contrária.

De acordo com o ANEEL (2021b), quando há violação dos limites de continuidade individuais (DIC, FIC, DMIC e DICRI) por ele estabelecidos, a distribuidora responsável deverá efetuar o cálculo da compensação e creditar na fatura em até dois meses após o período de apuração, podendo ser mensal, trimestral ou anual, respeitando um valor mínimo de R\$ 0,01 e, quando ultrapassar o valor máximo da fatura, o valor remanescente deverá ser armazenado para crédito nas faturas posteriores.

Os cálculos da compensações são definidos pela ANEEL (2021b) conforme as equações abaixo:

- DIC:

$$Comp_{DIC} = DIC_v \frac{VRC}{730} kei_1,$$

- DMIC:

$$Comp_{DMIC} = DMIC_v \frac{VRC}{730} kei_1,$$

- FIC:

$$Comp_{FIC} = \frac{FIC_v}{FIC_p} DIC_p \frac{VRC}{730} kei_1,$$

- DICRI:

$$Comp_{DICRI} = DICRI_v \frac{VRC}{730} kei_2,$$

sendo:

$DIC_v, FIC_v, DMIC_v, DICRI_v$ → indicadores por ponto de conexão verificados no período considerado;

DIC_p, FIC_p → limite estabelecido do indicador no período da interrupção;

VRC → base monetária para o cálculo de compensação referente ao mês de apuração do indicador;

730 → número médio de horas no mês;

kei_1 → coeficiente de majoração com valor fixado em 34 para unidades de Baixa Tensão (BT), 40 para Média Tensão (MT) e 108 para Alta Tensão (AT);

kei_2 → coeficiente de majoração com valor fixado em 14 para unidades de Baixa Tensão (BT), 20 para média tensão (MT).

Assim, com o modelo atual vigente para compensação, pode-se observar que o valor médio pago por compensação em 2021 foi de R\$ 8,99, dentro de um montante de mais de 80 milhões de compensações pagas, totalizando R\$ 720 milhões de reais gastos pelas distribuidoras e permissionárias no ano passado. Além disso, conforme a Tabela 1, nota-se que nos anos de 2019 e 2020 foram gastos mais de R\$ 600 milhões cada, sendo a maior parte do valor realizada pela transgressão mensal destes, também por representar o maior volume, apresentado na Figura 1.

Dessa forma, por meio da punição pela violação dos indicadores, a ANEEL busca compensar os clientes piores servidos e evitar que a situação se mantenha com baixa qualidade, deslocando o maior volume de consumidores para uma quantidade menor de horas interrompidas por mês.

2.5.5 Indicadores Coletivos de Continuidade

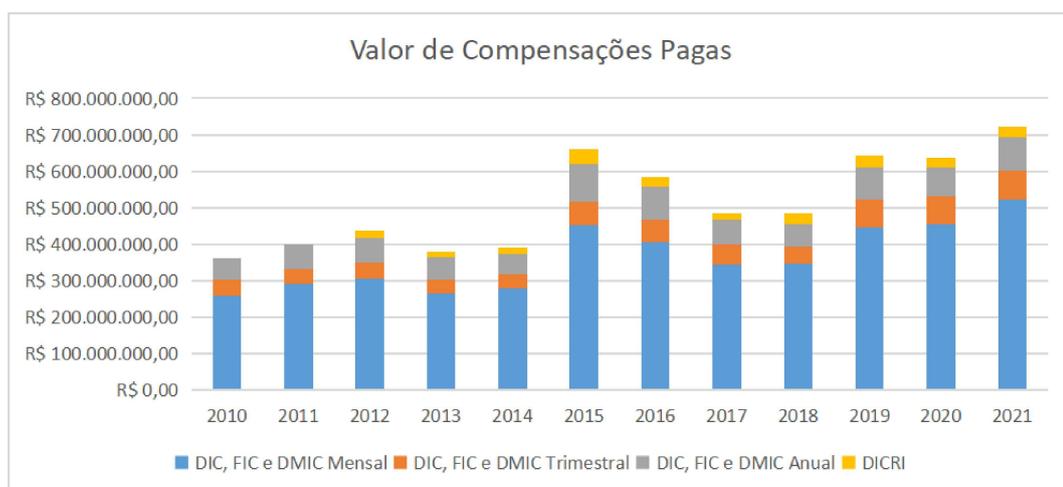
Os indicadores coletivos de continuidade são referências globais dos conjuntos elétricos e da área assistida pela distribuidora que permitem avaliar o desempenho e

Tabela 1 – Compensação paga pelas concessionárias e permissionárias no Brasil

Ano	Quantidade de Compensações (total anual)	Compensações Pagas (total anual)	Valor Médio por Compensação
2010	95.091.690	R\$ 360.797.553,60	R\$ 3,79
2011	105.139.243	R\$ 400.304.696,24	R\$ 3,81
2012	107.282.626	R\$ 438.092.874,05	R\$ 4,08
2013	104.787.128	R\$ 379.426.928,54	R\$ 3,62
2014	110.300.341	R\$ 391.887.571,81	R\$ 3,55
2015	125.454.095	R\$ 659.950.195,19	R\$ 5,26
2016	112.757.207	R\$ 583.438.543,43	R\$ 5,17
2017	99.642.998	R\$ 485.532.799,56	R\$ 4,87
2018	83.649.094	R\$ 485.099.732,88	R\$ 5,80
2019	81.306.038	R\$ 642.358.183,79	R\$ 7,90
2020	79.681.957	R\$ 637.138.643,62	R\$ 8,00
2021	80.218.644	R\$ 720.961.841,53	R\$ 8,99

Fonte: Consolidação dos dados disponibilizados pela ANEEL (2022a)

Figura 1 – Compensações mensais, trimestrais e anuais referentes a qualidade do serviço



Fonte: Consolidação dos dados disponibilizados pela ANEEL (2022a)

balizar a tomada de decisões. Tais indicadores também são regulados e acompanhados pela ANEEL, possuindo limites de forma que as empresas têm que atuar para manter os seus indicadores abaixo deles. Os limites e indicadores nacionais ao longo dos anos estão apresentados na Figura 2. Os indicadores são:

1. Duração Equivalente de Interrupção por Unidade Consumidora (DEC), definida pela Equação (2.9), é calculada a partir da soma das durações individuais (DIC) de todas as unidades consumidoras (NUC):

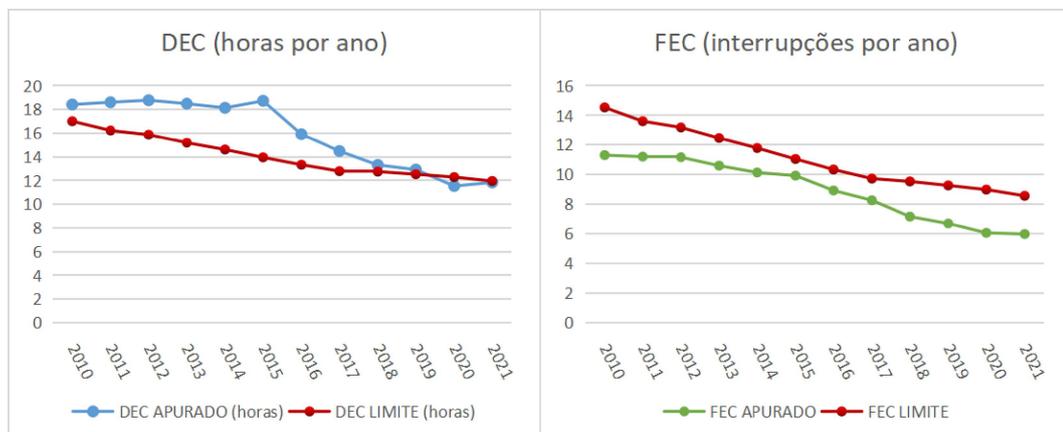
$$DEC = \frac{\sum_{i=1}^{NUC} DIC(i)}{NUC}. \quad (2.9)$$

2. Frequência Equivalente de Interrupção por Unidade Consumidora (FEC), apresentada pela Equação (2.10), é calculada a partir da soma da quantidade de interrupções (FIC) de todas as unidades consumidoras (NUC):

$$FEC = \frac{\sum_{i=1}^{NUC} FIC(i)}{NUC}. \quad (2.10)$$

Com esses indicadores, DEC e FEC, tem-se uma média de quanto tempo (DEC) e quantas vezes (FEC) cada cliente do conjunto ficou sem o fornecimento de energia elétrica.

Figura 2 – Indicadores coletivos de continuidade e seus respectivos limites ao longo dos anos



Fonte: Consolidação dos dados disponibilizados pela ANEEL (2022c)

2.6 Trabalhos Relacionados

Alguns trabalhos na literatura buscaram estimar os indicadores DEC e FEC e, assim, auxiliar no controle da qualidade do serviço prestado pelas concessionárias.

Magalhães (2017) buscou estimar os indicadores de continuidade DEC e FEC através de técnicas de Regressão Linear Múltiplas (RLM) e Redes Neurais Artificiais

(RNA) para auxiliar na priorização das tarefas de manutenção. Em seu trabalho, foram utilizados dados referentes ao período de 01/01/2013 a 31/12/2014 que, após análise estatística, foi possível observar que a maior partes da falhas no conjunto ALFA da concessionária dava-se a problemas de manutenção nos equipamentos, seguida por falha ambiental, causas desconhecidas, causada por terceiros, sistema e falha humana.

As ocorrências desse conjunto ALFA foram classificadas em 57 tipos de falhas e organizadas em 180 amostras referentes às ocorrências mensais registradas, desenvolvendo assim um modelo de múltiplas entradas e uma única saída (*Multiple Input Single Output (MISO)*), obtendo resultados positivos para seu estudo para a aplicação com RNA, após a otimização do modelo, de forma a impedir que os resultados assumissem valores menores que zero.

O trabalho de Magalhães (2017) diferencia-se em relação a organização dos dados e estrutura das redes utilizadas, efetuando a segmentação em diferentes tipos de falhas e com múltiplas entradas.

Junior (2016) analisou a predição do DEC e FEC derivados de falhas passíveis do uso de termografia para sua prevenção, buscando através deste auxiliar a distribuidora na manutenção preventiva dos equipamentos afetados. Em seu trabalho, mostrou-se que das falhas analisadas para a concessionária/conjunto, 41% delas participam do grupo de estudo, como as oriundas de alta resistência de contatos, curto circuitos iminentes, níveis de óleos incorretos, aquecimentos indutivos, dentre outras.

Nesse trabalho, Junior (2016) validou em uma de sus hipóteses que a falha gerada pelo conector, dentre as selecionadas, é a que gera mais impactos nos indicadores de continuidade. Além disso, através de uma simulação realizada, mostrou que com as técnicas apresentadas é possível obter a redução do DEC e FEC da distribuidora.

Relacionado ao campo energético, o trabalho de Jana et al. (2020) conseguiu, através da técnica de decomposição em frequência *maximal overlap discrete wavelet transformation* (MODWT) juntamente com uma rede profunda com camada *Long Short-Term Memory* (LSTM), ter resultados superiores a outros seis métodos comparados para predição do consumo energético nos EUA, sendo esses métodos ARIMA, DT, ANN, RNN e MARS. A importância de sua pesquisa dá-se devido aos fatores políticos, socio-econômicos, climáticos, etc., relacionados ao consumo energético, podendo guiar a expansão da malha de transmissão e pontos de geração.

3 Aprendizado de Máquinas e Redes Neurais

3.1 Aprendizado de Máquinas

A habilidade humana de adquirir conhecimento a partir da experiência e generalizar o aprendizado, aplicando o que foi aprendido em um contexto para uma situação diferente, tem fascinado os cientistas. Afinal, identificar como o conhecimento é adquirido, leva a formas mais eficientes de ensinar (DOUMAS et al., 2018).

Com esse fascínio, em 1943, o neurofisiologista McCulloch e o matemático Pitts publicaram um artigo (MCCULLOCH; PITTS, 1943) onde modelaram, de forma simples, o funcionamento de um neurônio utilizando circuitos elétricos. A partir de então, buscou-se criar/reproduzir a maneira como os neurônios e as redes neurais funcionam. Isto deu início a um campo que futuramente, de forma mais genérica, ficou conhecido como “Aprendizado de Máquinas”.

A área de aprendizado de máquinas busca formas de construir algoritmos que aprendem com a experiência, ou seja, com a habilidade de reconhecer padrões sem depender de codificação diretamente humana para isso (MITCHELL, 1997).

Os algoritmos “aprendem” a partir de tentativas e verificações, assim, com um objetivo definido, são feitos constantes testes de desempenho e, caso a tentativa tenha sido satisfatória, são recompensados de forma a buscarem uma solução naquele entorno. Como definido por Mitchell (1997), “Diz-se que um programa de computador aprende pela experiência **E**, com respeito a algum tipo de tarefa **T** e desempenho **P**, se seu desempenho **P** nas tarefas em **T**, na forma medida por **P**, melhoram com a experiência **E**.”

Essa metodologia é empregada em diversas áreas com obtenção de bons resultados (NAQA; MURPHY, 2015). Por exemplo, reconhecimento de fala, de imagens, pontos de falha de equipamentos, extração de informações relevantes de dados comerciais, pré-diagnósticos médicos e também para tomadas de decisões, como compra e venda de ações e carros autônomos.

Os tipos mais comuns de algoritmos para aprendizado de máquinas podem ser classificados em três categorias, sendo elas: aprendizado não supervisionado, aprendizado supervisionado e aprendizado por reforço.

O aprendizado supervisionado diz respeito ao grupo de algoritmos que precisam de dados que venham acompanhados de pares entrada-saída, ou seja, a saída esperada quando analisada a entrada. Daí a origem da nomenclatura “supervisionado”. Por exemplo, para treinar uma rede neural a classificar números em imagens, é necessário que os dados para

treinamento venham identificados com o valor esperado como resposta ([GOODFELLOW et al., 2016](#)).

Dessa forma, esses algoritmos buscam encontrar parâmetros que permitam um bom desempenho nos dados de treinamento e também nos dados de teste, para garantir que o modelo não esteja viciado, podendo, assim, classificar/estimar uma entrada desconhecida com uma boa acurácia.

Uma das subáreas que mais se destacam entre os algoritmos de aprendizado supervisionado é a de redes neurais artificiais. A aplicação delas é uma alternativa para resolver problemas que não possuem soluções matemáticas fechadas. Elas têm como inspiração o comportamento do cérebro humano, ou seja, alta complexidade, não linearidade e processamento em paralelo ([HAYKIN et al., 2009](#)).

O aprendizado não supervisionado, também conhecido por análise de *clusters*, utiliza dados para treinamento que não possuem uma saída esperada. Esses algoritmos foram desenvolvidos para extrair informações, descobrir padrões e estruturar os dados de forma a agrupá-los usando as semelhanças entre os dados como referência.

Por último, a técnica de aprendizado por reforço busca encontrar a melhor ação a ser tomada, dentre determinadas possibilidades ou regras, a partir de dados do ambiente ([BISHOP, 2006](#)). Neste caso, não são conhecidos a priori os pares entrada-saída, mas por meio de tentativa e erro, cada vez que o algoritmo toma uma decisão, ele é recompensado ou punido como “consequência” de suas ações, guiando-o na direção desejada.

3.2 Conceitos de Redes Neurais Artificiais

As redes neurais artificiais (RNA) são atualmente utilizadas para uma gama muito grande de aplicações. O menor elemento que forma uma rede neural é o neurônio. Os neurônios estão organizados em camadas, e os neurônios que estão em uma mesma camada recebem no mesmo instante uma entrada (sinal) para ser processada e enviada para a camada seguinte de neurônios ou para a saída da rede. As camadas podem ser: de entrada, que é a camada inicial da rede por onde os sinais entram na rede; de saída, que é a última camada da rede por onde sai a saída da rede; e intermediária ou oculta, que são camadas entre a camada de entrada e de saída.

A forma como as camadas estão organizadas e interconectadas gera diferentes arquiteturas de redes neurais. Dependendo da quantidade de camadas ocultas, as redes podem ser classificadas em duas categorias: redes neurais rasas ou clássicas, quando o número de camadas ocultas é menor ou igual a dois, e redes neurais profundas, quando existe um número maior de camadas ocultas, normalmente na ordem de dezenas de camadas.

Atualmente, as redes neurais profundas têm despertado o interesse da comunidade científica pela sua capacidade de resolver problemas de alta complexidade. As redes neurais profundas remetem desde os anos de 1970, como, por exemplo, [Fukushima e Miyake \(1982\)](#) que utilizaram redes profundas para reconhecer caracteres. Todavia, por muitos anos, redes profundas dependiam do uso de *clusters* de computação de alto desempenho (HPC) que eram inacessíveis para a maioria das pessoas, limitando a sua aplicação ([GREENGARD, 2016](#)). Com o desenvolvimento das placas de processamento gráfico (GPU), essas técnicas foram se tornando mais acessíveis e mais populares. Hoje, as GPU representam as principais unidades de processamento de redes profundas ([GREENGARD, 2016](#)).

As próximas subseções detalham alguns elementos das redes neurais que são importantes para este trabalho. Inicialmente é descrito o funcionamento de um neurônio artificial e depois os tipos de arquiteturas em função das conexões entre camadas. Também são descritos os funcionamentos das células LSTM e as camadas convolucionais, ambas presentes nas arquiteturas de redes neurais profundas utilizadas neste trabalho.

3.2.1 Neurônio Artificial

Redes neurais artificiais, da mesma forma que o cérebro, possuem como um elemento essencial para o seu funcionamento o neurônio. O neurônio orgânico, fonte da inspiração, possui os dendritos que recebem as várias entradas/estímulos, e o axônio que atua como saída do neurônio, comunica-se através de sinapses, e é ligado a múltiplos outros neurônios ([BISHOP, 1994](#)).

Assim, de maneira similar, segundo [Haykin et al. \(2009\)](#), o neurônio artificial consiste dos seguinte elementos, que podem ser observados na Figura 3:

- **Sinapses:** Nas sinapses as entradas (x_m) são multiplicadas por um peso (w_k), individual, podendo ser um valor tanto positivo quanto negativo.
- **Somador:** O somador, como o próprio nome indica, realiza uma adição das entradas após a etapa anterior. Uma outra entrada dessa etapa é o *bias* (b_k), ou viés, que tem como objetivo aumentar ou diminuir a entrada da função de ativação.
- **Função de ativação:** Tem por finalidade limitar a amplitude do sinal de saída.

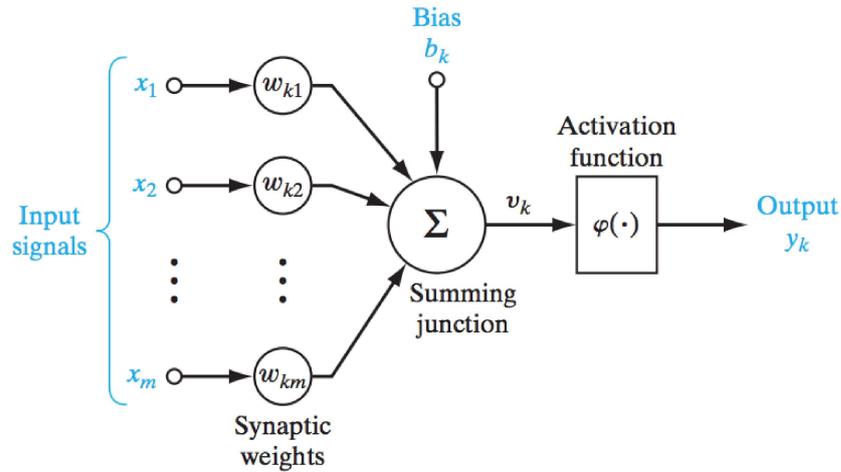
Os sinais de saída podem ser descritos como na Equação (3.1):

$$y_k = \varphi(v_k), \quad (3.1)$$

onde

$$v_k = \sum_{j=1}^m w_{kj}x_j + b_k, \quad (3.2)$$

Figura 3 – Modelo de um neurônio artificial



Fonte: Haykin et al. (2009)

sendo φ a função de ativação, b_k o *bias* e x_j o j -ésimo sinal de entrada ($j = 1, 2, \dots, m$) que é ponderado pelo peso sináptico w_{kj} do k -ésimo neurônio.

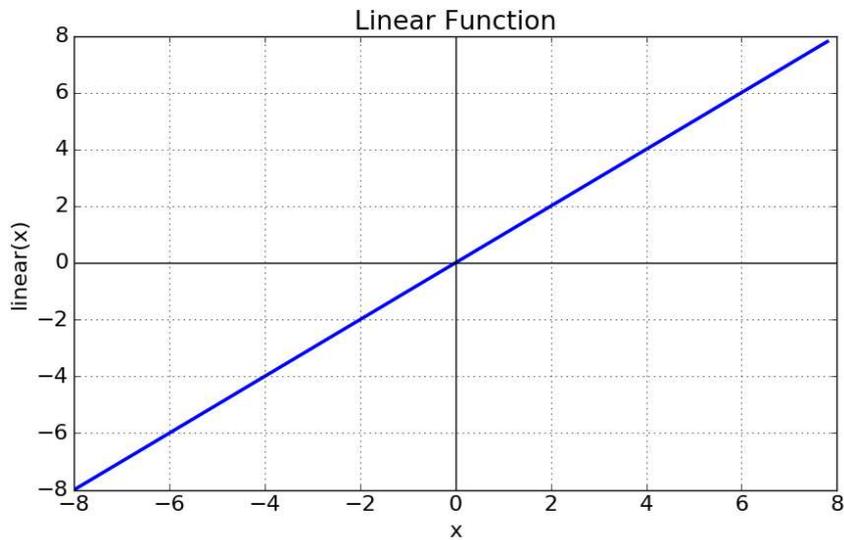
A função de ativação, atribuída anteriormente como φ , possui diversas formas, tendo a escolha de acordo com a necessidade da aplicação. Neste estudo foram utilizadas três funções de ativação, sendo elas a linear, a sigmoide e a tangente hiperbólica:

1. **Função Linear:** Nessa função a saída é diretamente proporcional a entrada, como na Figura 4, podendo ser definida pela Equação (3.3), onde v_k é a saída do somador adicionada ao sinal de *bias*.

$$\varphi_v = cv_k \quad (3.3)$$

sendo c o coeficiente de inclinação da reta.

Figura 4 – Função de ativação linear



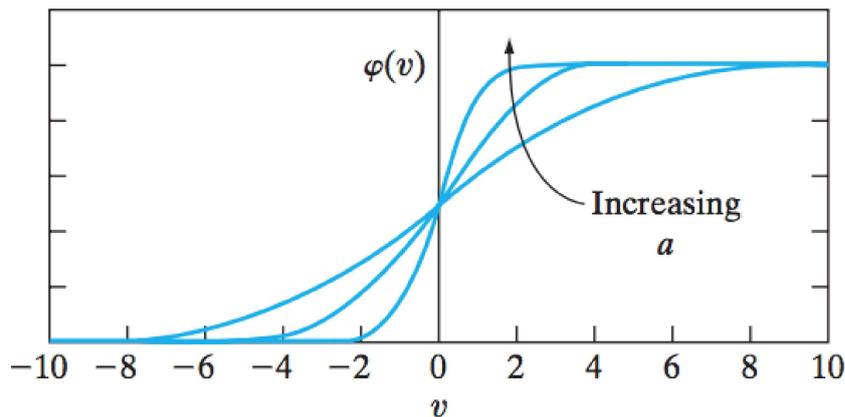
Fonte: Sharma (2017)

2. **Função sigmoide:** Conforme Haykin et al. (2009), a função sigmoide mantém o equilíbrio entre a linearidade e a não-linearidade. Esta é uma das funções de ativação mais usadas, com saída limitada ao intervalo $[0,1]$, e definida pela Equação (3.4),

$$f(x) = \frac{1}{1 + e^{-av_k}}, \quad (3.4)$$

com sinal de saída demonstrado pela Figura 5, onde a é o parâmetro de inclinação.

Figura 5 – Função de ativação sigmoide

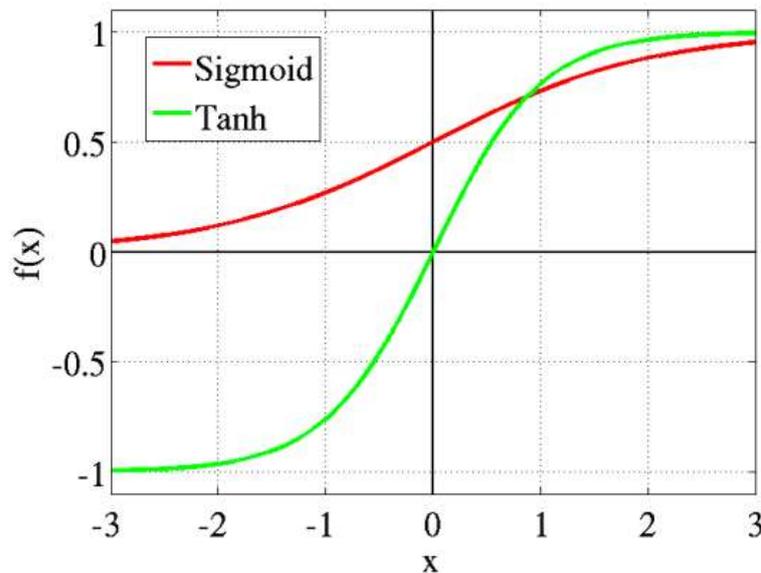


Fonte: (HAYKIN et al., 2009)

3. **Função tangente hiperbólica:** Para possibilitar que o intervalo de saída tenha valores negativos na faixa $[-1,1]$, é utilizada a função tangente hiperbólica, que possui as mesmas características, exceto a amplitude. Ela é representada pela Equação (3.5), podendo ser observada na Figura 6 pelo sinal em verde, em comparação com a função sigmoide dada em vermelho.

$$f(x) = \frac{1 - e^{-av_k}}{1 + e^{-av_k}}. \quad (3.5)$$

Figura 6 – Função de ativação sigmoide e tangente hiperbólica

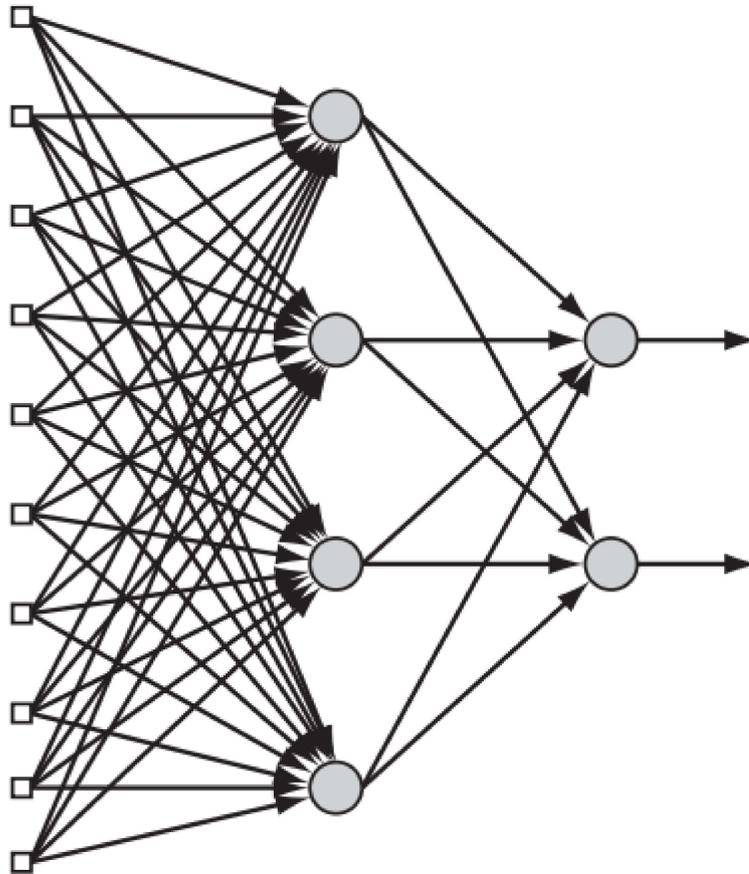
Fonte: [Sharma \(2017\)](#)

Os neurônios são organizados em camadas, tal que neurônios em uma mesma camada possuem o mesmo tipo de função de transferência. A forma como as camadas são conectadas entre si na rede dá origem a dois tipos de arquiteturas que são descritas a seguir: redes *feedforward* e redes recorrentes.

3.2.2 Redes *Feedforward*

As redes *feedforward* podem ser construídas tanto em múltiplas camadas ocultas, como em apenas uma, tendo suas conexões fluindo apenas no sentido da entrada para a saída, de forma acíclica, ([HAYKIN et al., 2009](#)). A Figura 7 ilustra uma rede *feedforward* com uma camada oculta.

Um tipo de rede *feedforward* é a perceptron multicamada, possuindo uma ou mais camadas ocultas, sendo que a camada anterior geralmente é ligada apenas a camada seguinte, podendo ser totalmente ou parcialmente conectada, dependendo se todos os neurônios anteriores são ligados aos próximos ou apenas alguns. A adição de camadas ocultas permite que mais informações possam ser extraídas da entrada ([HAYKIN et al., 2009](#)).

Figura 7 – Rede *Feedforward*

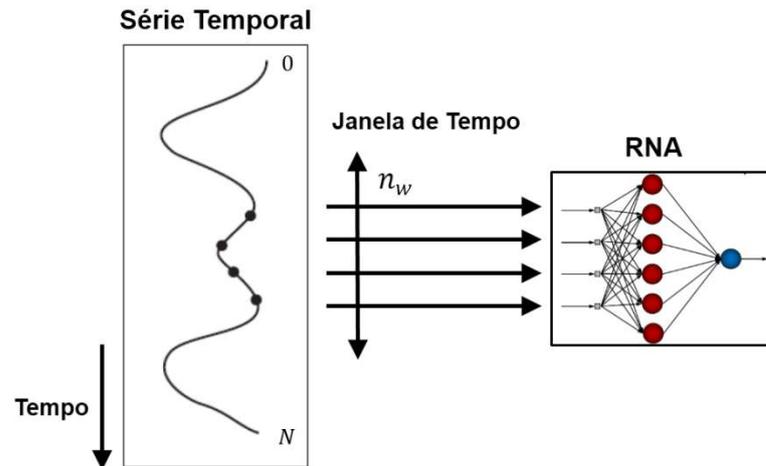
Fonte: Haykin et al. (2009)

3.2.3 Redes Neurais com Atraso no Tempo

Redes neurais também podem ser classificadas como redes estáticas ou dinâmicas. Nas redes estáticas, geralmente utilizadas na área de reconhecimento de padrões estruturais, as entradas não se relacionam. Já nas redes dinâmicas, para reconhecimentos de padrões ao longo do tempo, a saída não depende apenas do instante atual, mas também do passado. Dessa forma, uma rede é considerada dinâmica caso tenha memória de curto ou longo prazo (HAYKIN et al., 2009; JHA; SINHA, 2014).

A partir do atraso das entradas de uma série temporal é possível inserir a capacidade de memória na rede. Um modelo conhecido como *Time Delay Neural Network* (TDNN) foi criado a partir da junção entre o atraso das entradas a um perceptron multicamadas (HU; HWANG, 2002). Assim, com essa inserção, faz com que ao percorrer a janela de instantes de tempo criada, cria-se a dependência no tempo necessária. A Figura 8 traz um exemplo de TDNN, onde n_W é o número de atrasos da série temporal para estimar uma observação futura. Por exemplo, para 4 atrasos, são usadas as entradas x_{t-3} , x_{t-2} , x_{t-1} e x_t para estimar x_{t+1} .

Figura 8 – Rede TDNN



Fonte: [Diniz \(2018\)](#)

3.2.4 Redes Neurais Recorrentes

Redes neurais recorrentes (RNN) são uma variação das redes *feedforward* com a adição de ciclos, ou realimentações, fazendo com que as partes à vante se conectem com outros segmentos anteriores, como conexão entre a camada de saída e uma camada oculta, entre duas camadas ocultas, uma camada oculta e a camada de entrada, etc. As realimentações funcionam como uma entrada gerada pela própria rede. A Figura 9 apresenta um grafo de uma RNN após o seu desdobramento, permitindo melhor visualizar o fluxo de informações. Assim, a saída do neurônio depende não só da entrada atual como dos estados anteriores, funcionando como uma memória.

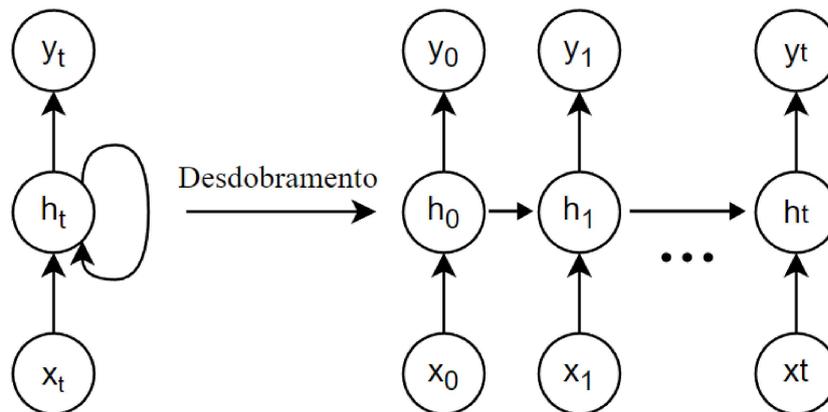
Esse tipo de rede é adequado para processamento de dados sequenciais, como séries temporais, graças a característica de compartilhar informações entre camadas. Dessa forma, permite identificar um evento mesmo ocorrendo em um diferente instante de tempo ([GOODFELLOW et al., 2016](#)). Além do mais, a introdução dos atrasos gerados pelas realimentações introduz não linearidades no sistema.

3.3 Aprendizado Profundo

Um grande problema relacionado ao aprendizado de máquinas é a quantidade de fatores que podem influenciar a entrada, como em um exemplo dado por [Goodfellow et al. \(2016\)](#), onde os *pixels* de um cartão mudam o tom conforme a exposição de luz e a forma é alterada de acordo com o ângulo da foto. Logo, é comum realizar um tratamento dos dados para buscar padronizar a entrada.

Entretanto, os algoritmos de aprendizado profundo conseguem resolver este problema a partir de representações mais simples, chegando a conceitos mais complexos, como

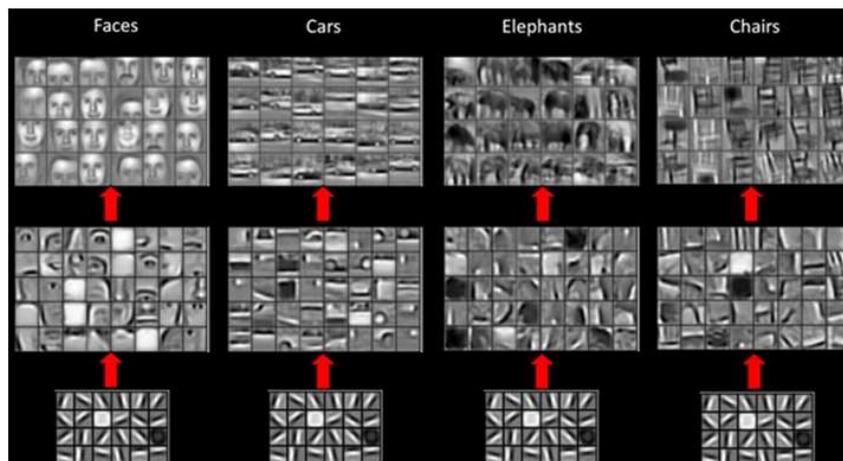
Figura 9 – Desdobramento de uma RNN



Fonte: O Autor

na Figura 10, onde primeiro são extraídas as bordas das imagens, depois as formas e, enfim, a classificação final dos objetos nas imagens.

Figura 10 – Exemplo de extrações de características por camada



Fonte: Sqalli (2016)

De acordo com a profundidade da rede, informações mais complexas podem ser aferidas, pois as instruções mais a frente podem utilizar os dados extraídos pelas camadas anteriores (LECUN et al., 2015).

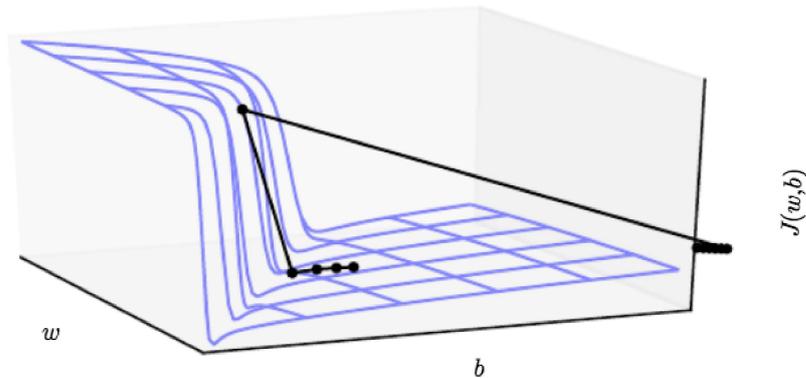
3.3.1 Long Short-Term Memory

Segundo Goodfellow et al. (2016), algumas situações que prejudicam o aprendizado da rede são comuns quando essas possuem muitas camadas. Dentre elas, existem as regiões da superfície da função de perda¹ J e das funções de ativação que podem ser muito

¹ Função de perda é uma medida que mede o quanto que a saída de uma rede neural se distancia da saída esperada. Essa função tem como parâmetros a entrada, a saída esperada e a saída da rede (que é em função dos parâmetros da rede). Quanto menor o seu valor, melhor. Existem várias tipos de função

inclinadas e levar a um gradiente grande, ou como é conhecido, gradiente explosivo. Além disso, também é comum as regiões da superfície de J e das funções de ativação quase planas em que o gradiente se torna desprezível. Os dois casos são ilustrados na Figura 11.

Figura 11 – Gradiente em declive acentuado vs plano



Fonte: Goodfellow et al. (2016)

Assim, o gradiente do erro, em redes com retropropagação (*backpropagation*) no tempo, é retroalimentado e tende a explodir ou desaparecer. Isso ocorre pois, conforme demonstrado por Hochreiter e Schmidhuber (1997), a forma como essas redes comumente aprendem faz com que o gradiente do erro seja escalado de forma a crescer ou decrescer exponencialmente. No primeiro caso, gera uma oscilação nos pesos, fazendo com que a rede tenha um aprendizado instável. Já para o decréscimo, o gradiente do erro rapidamente desaparece, assim, a rede não aprende em um intervalo de tempo aceitável.

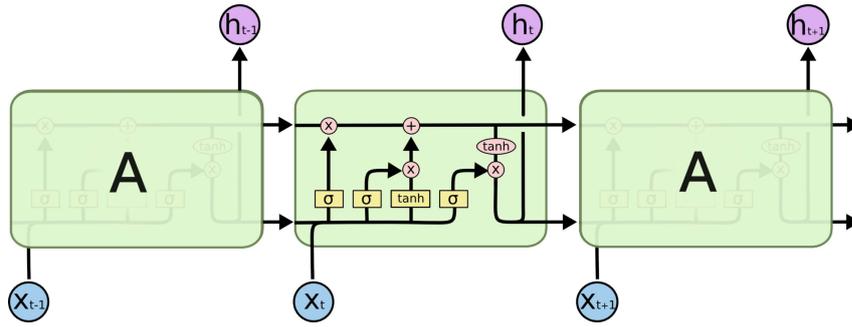
Dessa forma, por apresentar dificuldades em utilizar informações passadas no instante necessário, em 1997 foi apresentada a proposta da camada *Long Short-Term Memory* (LSTM) de modo a proporcionar a passagem do gradiente do erro sem os problemas anteriores. Nesse estudo, foi introduzido o conceito de célula de memória, permitindo às unidades conter informações fora do fluxo. Assim, dados mais antigos podem ser utilizados mesmo vários passos depois ou esquecidos, caso não sejam relevantes para o aprendizado, com estrutura exemplificada na Figura 12.

Dentro de um bloco/célula LSTM, Figura 13, existem três *gates* responsáveis por proporcionar as capacidades de uma camada LSTM, sendo eles o *gates* de entrada (*input gate*), de esquecimento (*forget gate*) e o de saída (*output gate*). *Gates* são os nomes dados para cada elemento composto por uma função de transferência (por exemplo, sigmoide) seguida por um multiplicador ponto a ponto.

Segundo Cui et al. (2018), e tomando a Figura 13 como referência visual, existem os *gates*:

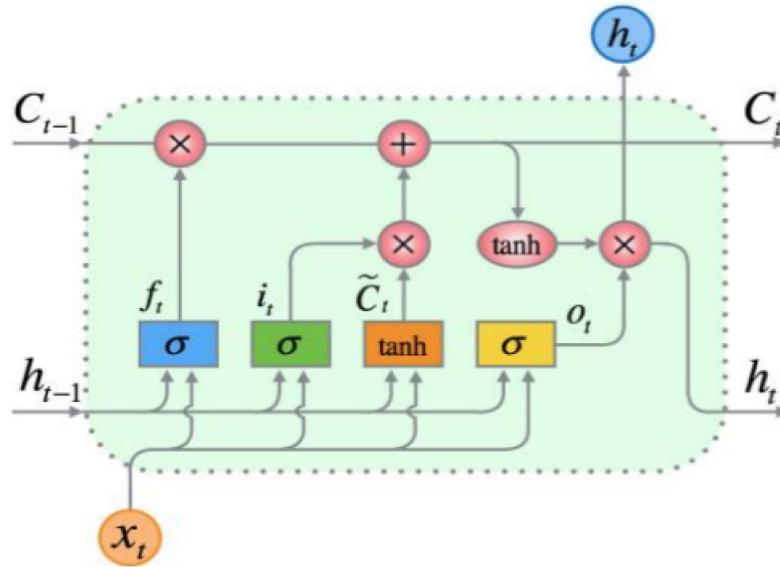
de perda, e o gradiente do erro é calculado baseado nela.

Figura 12 – Camada LSTM



Fonte: Cui et al. (2018)

Figura 13 – Bloco/Célula LSTM



Fonte: Cui et al. (2018)

1. **Forget Gate:** Representado pelo bloco azul com a sigmoide e o multiplicador seguinte, é utilizado para decidir que informação deve ser descartada, esquecida, a partir de h_{t-1} e x_t , resultando em uma saída f_t no intervalo de $[0, 1]$, onde zero representa esquecer totalmente e um indica que C_{t-1} deve ser integralmente armazenada. A saída deste *gate* é definida pela Equação (3.6),

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{3.6}$$

sendo σ a função de ativação sigmoide, W e b as matrizes de pesos e *bias*, respectivamente, x_t a entrada atual, h_{t-1} a saída definida no instante de tempo anterior e C_t representa o estado da célula, variando de 0 a 1, sendo 1 para manter em sua totalidade e 0 para apagar a informação, constituído a partir de C_{t-1} .

2. **Input Gate:** nesta etapa é definido o que será atualizado, através de i_t (Equação (3.7)), e o valor das atualizações \tilde{C}_t (Equação (3.8)). Assim, o valor a ser adicionado

é dado por $\tilde{C}_t \cdot i_t$, representando o vetor de entrada escalado.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_i] + b_i), \quad (3.7)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_i] + b_C), \quad (3.8)$$

tendo \tanh (tangente hiperbólica) como função de ativação.

3. **Output Gate:** para composição da saída, primeiro é importante definir C_t . C_t representa o estado da célula, sendo constituído a partir do estado antigo C_{t-1} , onde ao multiplicar por f_t os valores recebidos nos passos anteriores podem ou não serem esquecidos. Em seguida, os valores novos ($\tilde{C}_t \cdot i_t$) são adicionados ao estado antigo, obtendo, assim, C_t , como definido na Equação (3.9).

$$C_t = C_{t-1} \cdot f_t + i_t \cdot \tilde{C}_t \quad (3.9)$$

Após aplicar a função de ativação em h_{t-1} e x_t , para definir qual parte do novo estado (C_t) irá ser utilizado na saída, por meio de o_t (3.10), C_t é aplicado a uma função \tanh para que seu intervalo varie de $[-1, 1]$. Assim, ao multiplicar os dois é obtida a saída h_t , como definido na Equação (3.11).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (3.10)$$

$$h_t = o_t \cdot \tanh(C_t). \quad (3.11)$$

3.3.2 Convolutional Neural Network

Redes neurais convolucionais, *Convolutional Neural Networks* (CNN), são muito utilizadas para extração de características de dados como imagens e sinais. O resultado da aplicação delas leva, além de outros processos, ao uso de *kernels* de dimensões menores que a entrada, semelhante ao processo de convolução que acontece em processamento de imagens digitais. Essa aplicação faz com que redes convolucionais tenham menos parâmetros e melhoram a eficiência do modelo estatístico (GOODFELLOW et al., 2016).

Geralmente aplicada no campo de visão computacional através da convolução 2D, a CNN apresenta um alto grau de invariância frente a distorções na entrada, como mudança na escala, torções, etc. Graças às características da CNN, ela tem alcançado bons resultados em tarefas de reconhecimento visual (KHAN et al., 2018).

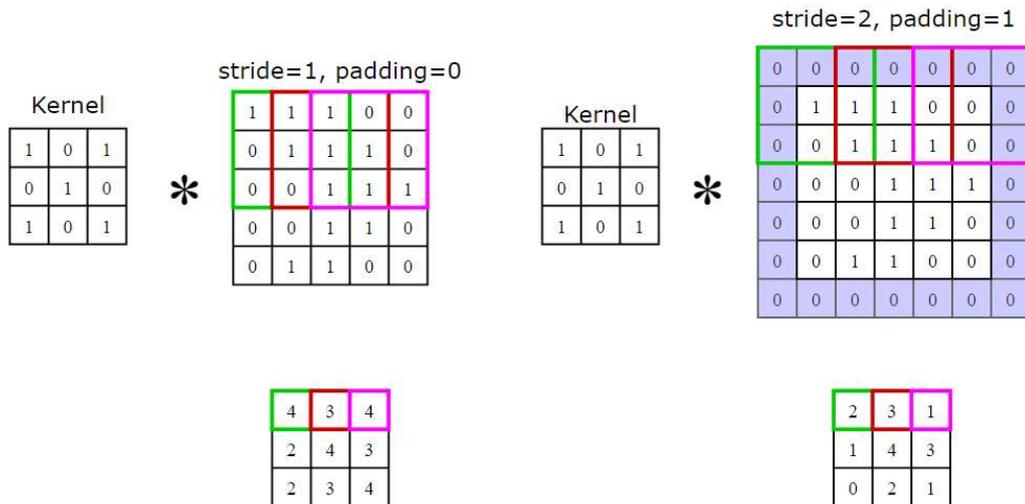
A Figura 10 exemplifica a extração de características de uma rede CNN. Com a habilidade de extração de características, esta rede elimina a necessidade de que essa seleção seja feita manualmente.

A convolução, no processamento de imagens, consiste na aplicação de uma máscara, também conhecida como *kernel*, pela imagem de entrada, fazendo com que o *kernel* percorra em passos conhecidos como *stride* a imagem base que pode ter ou não um *padding* de forma a controlar a redução das dimensões do resultado. Durante o percurso do *kernel* ao longo da entrada, é realizado o produto ponto a ponto e realizada a soma dos elementos multiplicados pelo *kernel*, conforme apresentado na Figura 14.

O *stride* indica o passo ao qual o filtro (*kernel*) é deslocado ao longo da entrada. Se ele é igual a 1, então desloca-se o centro do filtro um pixel de cada vez e caso seja 2, desloca-se de dois em dois pixels, pulando um.

Padding é o preenchimento adicionado aos extremos das bordas que permite controlar o tamanho da saída. Caso deseja-se que a saída da camada de convolução tenha o mesmo tamanho da camada anterior, faz-se necessário aplicar o *padding* com o tamanho adequado para que não haja redução das dimensões.

Figura 14 – Processo de convolução 2D com *stride* e *padding*



Fonte: O Autor

Diferente das redes tradicionais, em que cada neurônio é conectado a todos os elementos de entrada e possui seus próprios pesos e bias para produzir uma saída, nas camadas convolucionais os neurônios são parcialmente conectados a entrada e compartilham os seus parâmetros, ou seja, utilizam os mesmos pesos e bias para todas as entradas, ou seja, o mesmo *kernel* é aplicado ao longo da camada. Dessa forma, as CNN ficam mais leves e rápidas que, por exemplo, redes totalmente conectadas. Isso fica ainda mais evidente conforme a entrada cresce, já que em redes totalmente conectadas todos os neurônios da camada anterior então ligados aos da seguinte.

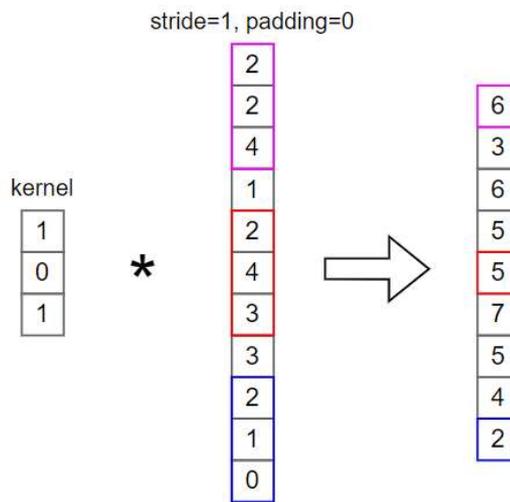
Outra etapa comumente presente com a camada convolucional é a de *Pooling*, como Média ou Máximo, por exemplo. Essa camada ajuda a reduzir o número de informações a

serem processadas, ajudando a simplificar a saída e evitar o sobreajuste.

A aplicação de CNN 1D vem apresentando sucesso em diversas séries temporais, como classificação de eletrocardiograma em tempo real (KIRANYAZ et al., 2015), detecção de falhas de motor (INCE et al., 2016) e detecção de dano estrutural baseada em vibração (ABDELJABER et al., 2017).

Para a sua aplicação em séries temporais, a ideia é semelhante ao da convolução 2D, sendo a principal diferença que neste caso o *kernel* se desloca apenas em uma dimensão, diferente da 2D que se desloca em duas dimensões. Essa técnica ajuda a reduzir a quantidade de ruídos nos dados através da extração de características mais relevantes. A Figura 15 exemplifica o processo de convolução 1D.

Figura 15 – Processo de convolução 1D



Fonte: O Autor

O processo consiste na aplicação de várias convoluções 1D aplicadas ao sinal a de tamanho n , com *stride* s , conforme definido pela Equação (3.12), com um *kernel* f de dimensão m .

$$\begin{aligned}
 Conv_{1D}(a, f) &= \left[Conv_{1D}(a[i], f), i \in \left(1, \left\lfloor \frac{n-m}{s} \right\rfloor + 1\right) \right] \\
 Conv_{1D}(a[i], f) &= \sum_{k=\lceil -m/2 \rceil}^{\lfloor m/2 \rfloor} a[i - k] \cdot f[k]
 \end{aligned}
 \tag{3.12}$$

4 Decomposição de Séries Temporais

4.1 Séries Temporais

Uma Série Temporal (ST) é um grupo de observações acompanhadas de marcações de tempo, indicando quando cada valor foi observado, podendo ser discreta ou contínua. Discreta é quando os valores são registrados em intervalos fixos de tempo, já contínua é quando os valores são gravados continuamente durante um período de tempo (BROCKWELL; DAVIS, 2002). Séries temporais possuem diversas naturezas, como atividades elétricas no cérebro, registro de chuvas, mercado de ações, número de vendas diárias do varejo, batidas do coração por minuto, consumo de energia por mês, indicadores de qualidade de energia mensal, etc. Além disso, possuem diferentes aplicações, como:

1. Investigação do mecanismo gerador da série temporal;
2. Previsões de valores futuros da série;
3. Descrição de seu comportamento;
4. Investigação de propriedades relevantes nos dados.

Predições de valores futuros de uma série é a utilização de métodos estatísticos para prever um valor futuro tendo como base o histórico de dados e, se for o caso, de entradas exógenas. Há diversas técnicas de predição de séries temporais que podem ser empregadas de acordo com a complexidade e o objetivo dos dados utilizados, como regressão para equações de primeiro grau, modelos *Autoregressive Moving Average* (ARMA), *Autoregressive Integrated Moving Average* (ARIMA), redes neurais rasas e redes neurais profundas, dentre outras técnicas.

Exemplos da aplicação de diferentes métodos de predição de séries temporais podem ser vistos na literatura, como Sahai et al. (2020) ao utilizar um modelo ARIMA para prever os números de infectados nos países com os maiores índices de COVID-19. Kouziokas (2019) apresentou em sua pesquisa a aplicação de uma rede neural *feedforward* para estimar valores futuros referentes ao número de desempregados no Reino Unido. Já Hewamalage et al. (2021) realizaram uma comparação entre a aplicação de redes neurais recorrentes, ARIMA e *exponential smoothing* (EST), de modo a avaliar os seus desempenhos com diferentes bases de séries temporais.

4.2 Transformada Wavelet de Máxima Sobreposição

A Transformada de Fourier (TF) é uma ferramenta muito utilizada para análise das componentes de frequência de um sinal (LIU, 2010). Ela, ao ser aplicada em um sinal no domínio do tempo, $g(t)$, decompõe o sinal em infinitas somas de funções senos e cossenos, sendo definida pela Equação (4.1):

$$L\{g(t)\} = G(s) = \int_{-\infty}^{+\infty} g(t)e^{-j\omega t} dt, \quad (4.1)$$

sendo ω a frequência angular, dada por $\omega = 2\pi f$, e $G(s)$ o sinal resultante após a aplicação da TF.

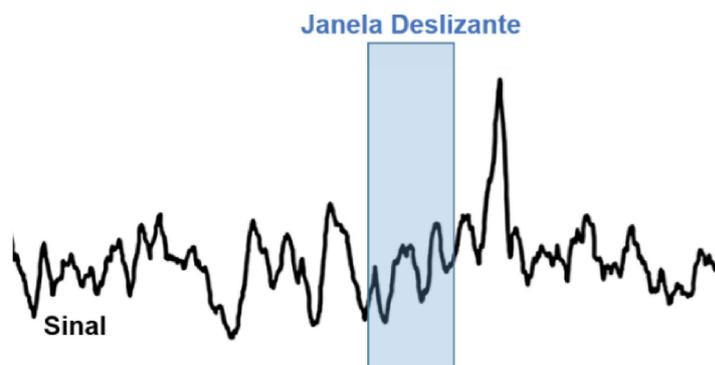
Uma forma mais explícita de observar a soma de senos e cossenos é analisando a Série de Fourier (4.2), que representa um caso da transformada aplicada a sinais periódicos, $x(t)$, onde n representa os múltiplos da frequência natural e c_n é conhecido como coeficiente de Fourier, podendo ser definido pela Equação (4.3), sendo T o período do sinal.

$$x(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(j\frac{2\pi nt}{T}\right) \quad (4.2)$$

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \exp\left(-j\frac{2\pi nt}{T}\right) dt \quad (4.3)$$

Apesar da TF apresentar uma boa forma de visualização dos dados no domínio da frequência, ela não carrega consigo dados do tempo, como o momento que determinada frequência ocorreu. Para obter informações do tempo, em conjunto com a frequência, há a Transformada de Fourier de Curto Período (*Short-time Fourier Transform - STFT*) que desloca uma janela no tempo, Figura 16, para obter conhecimento sobre ambos os domínios (LIU, 2010), porém limitado ao tamanho da janela representada por $\omega(t - \tau)$ na Equação (4.4), centrada em τ , onde $\omega(\tau)$ é comumente uma janela Gaussiana centrada em zero, ou uma janela retangular, como na Figura 16.

Figura 16 – Janela deslizante no tempo



Fonte: [Diniz \(2018\)](#)

$$[STFT]\{g(t)\} = G(\tau, s) = \int_{-\infty}^{+\infty} g(t)e^{-j\omega t}w(t - \tau) dt. \quad (4.4)$$

Um dos problemas da STFT é que é usado somente um tamanho fixo de janela, o que impõe limitações na representação do sinal no plano tempo-frequência. Ou seja, se há uma boa resolução no domínio da frequência, perde-se na precisão da localização destas frequências no domínio do tempo, por outro lado, se há uma boa localização no tempo, perde-se a precisão da representação do sinal no domínio da frequência.

Assim, conforme Liu (2010), para auxiliar na análise de sinais, buscando solucionar a limitação da STFT, foi desenvolvida a Transformada Wavelet (WT - *Wavelet Transform*), definida pela Equação (4.5), possibilitando a decomposição do sinal em diferentes frequências:

$$[WT](a, b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}(t) dt, \quad (4.5)$$

sendo $\psi_{a,b}(t)$ ¹ a função wavelet base ψ escalada e transladada pelos parâmetros a e b , respectivamente, conforme Equação (4.6):

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \quad (4.6)$$

Para sinais discretos é utilizada a Transformada Wavelet Discreta (DWT - *Discrete Wavelet Transform*), que pode ser expressa por²:

$$[DWT](a, b) = \sum_{n=0}^{N-1} f[n] \cdot \psi_{a,b}[n], \quad (4.7)$$

sendo $[DWT](a, b)$ o coeficiente da DWT para os parâmetros a e b e $f[n]$ o sinal discreto de tamanho N . A base discretizada é dada por:

$$\psi_{a,b}[n] = \frac{1}{\sqrt{|p^a|}}\psi\left(\frac{n-qp^a}{p^a}\right), \quad (4.8)$$

onde a e b são números inteiros correspondentes aos parâmetros de dilatação e translação, respectivamente, $p > 1$ é um parâmetro de dilatação fixo, que reflete a variação da dilatação, e q é o fator de translação (passo da translação), o qual depende do fator de dilatação. Normalmente são usados os valores de $p = 2$ e $q = 1$ (DINIZ, 2018).

Para utilização no processamento de sinais, o algoritmo da DWT sofre uma sensibilidade quanto à escolha do ponto inicial, ou seja, possui variância no tempo (PERCIVAL; WALDEN, 2000). Devido a essa limitação, foram formulados os algoritmos de Transformadas Wavelets não-decimadas e invariantes no tempo, como o *Maximal Overlap Discrete Wavelet Transform* (MODWT) (DINIZ, 2018).

¹ Na Equação (4.5) é assumido que ambos $x(t)$ e $\psi_{a,b}(t)$ são valores reais. Caso ambos sejam valores complexos, é usado o complexo conjugado de $\psi_{a,b}(t)$, ou seja, $\psi_{a,b}^*(t)$, na Equação (4.5).

² A mesma consideração da Equação (4.5) é feita na Equação (4.7).

A MODWT é uma transformada não ortogonal e redundante, ou seja, funcionando como um filtro, capaz de decompor uma série temporal em subséries. Essa transformada possui a capacidade de reter baixas frequências, que outrora seriam descartadas, como no caso da DWT, de modo que a partir da soma das subséries consegue-se reconstruir o sinal original (CORNISH et al., 2006).

Dessa forma, conforme exposto por Nason e Sachs (1999), a MODWT é muito utilizada para análise e previsão de séries temporais.

4.3 Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

Outra limitação da Transformada de Fourier, assim como de grande parte das técnicas de análise, é que ela requer que os dados sejam estacionários no tempo e lineares, ou estritamente periódicos. Uma série temporal pode ser tida como estacionária caso ela atenda às Equações (4.9), (4.10) e (4.11):

$$E(|X(t)|^2) < \infty, \quad (4.9)$$

$$E(X(t)) = m, \quad (4.10)$$

$$C(X(t_1), X(t_2)) = C(X(t_1 + \tau), X(t_2 + \tau)) = C(t_1 + t_2), \quad (4.11)$$

sendo $E(\cdot)$ o valor esperado, $X(t)$ a série temporal e $C(\cdot)$ a função de covariância.

Como explicado por Huang et al. (1998), uma vez que sempre serão usados dados com duração finita ($N < \infty$), aproximações terão que ser usadas desde o começo. Além disso, poucas séries temporais satisfazem os requisitos de estacionariedade. Dessa forma, Huang et al. (1998) propuseram a técnica *Empirical Model Decomposition* (EMD).

EMD é um método iterativo frequentemente aplicado para extração de ruído em dados de voz (HADHAMI; BOUZID, 2013; JUN-CHANG; LI, 2011), que decompõe sinais complexos em diversos elementos de componentes oscilatórios, conhecidos como *intrinsic mode functions* (IMF). IMF é uma função que atende a duas condições, a primeira é que o número de máximos e mínimos locais não pode ser maior que um e a segunda diz que a média entre os “envelopes” de máximo e de mínimo tem que ser zero.

Através da aplicação do algoritmo EMD, funções IMF podem ser extraídas de forma que as frequências de cada IMF irão diminuir até que a última componente seja monoatômica, não mais que um extremo, também chamada de resíduo. A Figura 17 apresenta a decomposição de um sinal a partir da aplicação da EMD.

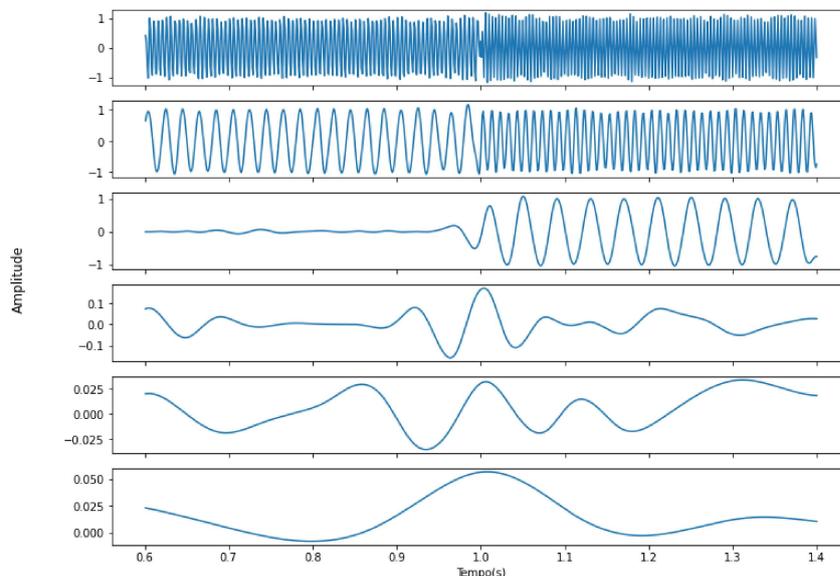
O algoritmo consiste em decompor interativamente um sinal em séries de oscilação rápida sobrepostas em componentes de sinal mais lento. O algoritmo para realizar a EMD pode ser dividido em oito passos:

1. Obter os extremos do sinal $x(t)$;
2. Conectar, separadamente os máximo e mínimo, através da interpolação cúbica, obtendo $e_{min}(t)$ e $e_{max}(t)$;
3. Encontrar a média $m(t)$ dos envelopes obtidos após a interpolação: $m(t) = (e_{min}(t) + e_{max}(t))/2$;
4. Subtrair a média do sinal, $h(t) = x(t) - m(t)$;
5. Verificar se atende as condições de IMF;
6. Se não atender, repetir os passos 1 à 4 até que seja uma IMF;
7. Caso seja válido, $h(t)$ é extraído como uma componente IMF;
8. Repetir os passos de 1 à 7 até que o critério de parada seja atendido, ou seja, D_k menor que um valor pré-determinado.

O valor de D_k é definido pela Equação (4.12):

$$D_k = \sum_{t=0}^T \frac{[h_{k-1}(t) - h_k(t)]^2}{[h_{k-1}(t)]^2}. \quad (4.12)$$

Figura 17 – EMD de um sinal. Mostrando apenas 6 das 9 IMFs extraídas.



Fonte: O Autor

Todavia, a técnica tem problemas devido a interpolação cúbica que pode vir a introduzir, como a propagação de oscilações para outras IMF, criação de picos e vales (*overshoots and the undershoots*) (DELÉCHELLE et al., 2005). Além disso, Hadiyoso et al. (2020) mostraram que oscilações na mesma escala podem ser armazenadas em diferentes IMFs e em uma única IMF pode ter sinais de variadas escalas, podendo levar a perda do significado físico individual das componentes.

Buscando superar esses problemas, Wu e Huang (2009) propuseram o desenvolvimento do método *Ensemble Empirical Mode Decomposition* (EEMD). O EEMD, através da inserção de ruído branco que preenche o espaço de tempo e frequência de maneira distribuída, permitindo que as IMF não tenham conexões entre elas, conforme mencionado anteriormente em relação à distribuição de um sinal em várias IMF. A eficiência do método dá-se em função da amplitude do ruído (A) e do número de rodadas utilizadas para gerar cada IMF (N).

O algoritmo para realizar o EEMD pode ser dividido em seis passos, tendo $y_n(t) = x(t) + u_n(t)$ como a n -ésima observação:

1. Inicializar o número de rodadas N e a amplitude do ruído A ;
2. Adicionar o ruído branco $u_n(t)$ ao sinal original $x(t)$, resultando em $y_n(t)$;
3. Decompor $y_n(t)$ em várias IMF utilizando o método de EMD;
4. Repetir os passos 2 e 3 enquanto a quantidade de rodadas for menor que N , mas adicionando um diferente sinal de ruído branco a cada rodada;
5. Calcular a média do conjunto para N tentativas de cada IMF; A média das IMF do passo 3 resultará nas IMF finais.

O efeito da decomposição usando EEMD é que os ruídos brancos adicionados irão cancelar uns aos outros na média final das correspondentes IMF, ficando as médias das IMF dentro da dinâmica natural do sinal (WU; HUANG, 2009), desde que sejam usadas um número suficiente de componentes para o cálculo da média.

O EEMD, contudo, é um processo computacionalmente custoso e pode apresentar um número variado de componentes e resíduo do ruído, além de não ser possível eliminar completamente a sua influência (TEJA et al., 2020; DAS; BHUIYAN, 2016).

Torres et al. (2011) apresentaram uma versão aprimorada do EEMD, conhecida como *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN). Por meio do CEEMDAN, que possui menor custo computacional quando comparado à EEMD, é possível obter a reconstrução exata do sinal. Abaixo o algoritmo da CEEMDAN, considerando $E_j(\cdot)$ um operador que resulta na j -ésima componente da decomposição EMD.

1. Obter $x_n^i(t)$, onde $x_n^i(t) = x_n(t) + \sigma_0 u_n(t)$, sendo σ_0 o parâmetro que permite o ajuste da relação sinal-ruído;
2. Calcular a primeira IMF e o primeiro resíduo $u_1(t)$ usando as Equações 4.13 e 4.14, sendo L o número de realizações para cada IMF;

$$IMF_1(t) = \frac{1}{L} \sum_{i=1}^L IMF_1^L, \quad (4.13)$$

$$u_1(t) = x(t) - IMF_1(t). \quad (4.14)$$

3. Tendo σ_1 como o desvio padrão do ruído branco neste ponto, determinar a IMF_2 pela Equação 4.15;

$$IMF_2(t) = \frac{1}{L} \sum_{i=1}^L E_1\{u_1(t) + \sigma_1 E_1[r^i(t)]\}. \quad (4.15)$$

4. Continuar o processo obtendo a k -ésima componente e resíduo usando as Equações 4.16 e 4.17;

$$IMF_{k+1}(t) = \frac{1}{L} \sum_{i=1}^L E_1\{u_k(t) + \sigma_k E_k[r^i(t)]\}, \quad (4.16)$$

$$r_k(t) = r_{k-1}(t) - IMF_k(t). \quad (4.17)$$

5. Repetir o passo anterior até que o resíduo seja uma função monotônica, sendo possível reconstruir o sinal por meio das k s IMF obtidas (Equação 4.18).

$$x(t) = \sum_{i=1}^R IMF_i + r_i(t). \quad (4.18)$$

5 Bases de Dados e Método Proposto

Este trabalho trata-se da predição dos indicadores coletivos de qualidade do fornecimento de energia elétrica, DEC e FEC, a partir do cálculo dos valores diários de interrupção. Para chegar aos indicadores diários, é necessário saber qual a base de clientes no mês, quanto tempo durou a ordem de serviço para o atendimento de religação emergencial e quantas unidades consumidoras foram afetadas por ela.

A obtenção desses dados é difícil, já que a ANEEL fornece apenas a estratificação mensal e para toda a base da empresa. Além disso, dado que os conjuntos elétricos variam com o tempo, é necessário projetar a base de clientes que compõe o conjunto atual às ordens anteriores a ele de modo a tornar possível a composição da série temporal, dado que os conjuntos elétricos são mutáveis com o tempo e, por isso, o grupo de unidades consumidoras abrangido irá variar.

Assim, isso explica a utilização dos dados limitados ao período de 2010 a 2017, visto que não foi possível obter informações mais recentes.

Dentre os conjuntos existentes na base, foi escolhido um do município de São Mateus, Espírito Santo (ES), por sua característica do DEC, 10,59, estar próximo ao limite estabelecido pela ANEEL, de 11, conforme apresentado na Tabela 2 para o ano de 2017, além de ao longo dos anos anteriores estar margeando o limite, ora abaixo e ora acima. A concessionária e os conjuntos disponíveis não foram revelados por questão de proteção aos dados.

Tabela 2 – Indicadores Coletivos de Continuidade do Conjunto de São Mateus/ES para cada ano

Ano	Clientes	DEC Apurado	DEC Limite	FEC Apurado	FEC Limite
2010	48.930	15,69	14,00	10,04	13,00
2011	35.984	12,00	13,00	6,60	13,00
2012	37.787	13,44	13,00	8,98	12,00
2013	39.653	11,87	13,00	7,71	12,00
2014	41.313	11,56	13,00	10,57	12,00
2015	42.842	12,28	12,00	8,90	11,00
2016	44.315	11,43	11,00	6,46	10,00
2017	44.879	10,59	11,00	7,40	10,00

Fonte: ANEEL (2017b)

Dada a grande distribuição geográfica do conjunto analisado, notou-se a complexidade dos dados, pois não há um comportamento uniforme e nem variação lenta. Dessa forma, neste trabalho é proposta a aplicação de técnicas de decomposição de séries temporais baseadas em frequência, sem a perda da informação temporal, permitindo a

aplicação de cada subsérie em uma rede diferente, de modo que ela consiga se especializar na frequência da componente fornecida. Além disso, este trabalho utiliza redes híbridas, LSTM e CNN, em comparação com a aplicação de redes neurais rasas.

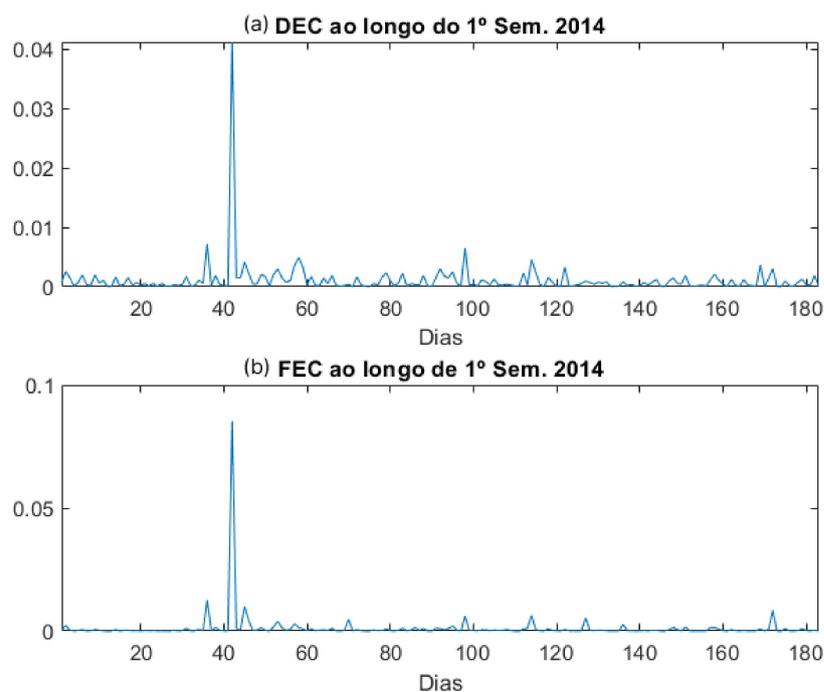
Neste capítulo são expostas as arquiteturas utilizadas, as técnicas de decomposição, o processo de preparação das séries temporais para treinamento e teste, e como foram realizados os experimentos.

5.1 Bases de Dados e Preparação dos Dados

5.1.1 Bases de Dados

Para o desenvolvimento deste estudo, foram utilizados os dados diários de DEC e FEC oriundos de um Conjunto Elétrico que compõe a área da concessionária estudada, referentes ao período de 01/01/2010 a 29/01/2017, formando assim duas séries temporais (DEC e FEC). A Figura 18 mostra um exemplo da disposição dos dados de DEC e FEC para o primeiro semestre de 2014. Pode-se observar que existem vários picos de valores, inclusive alguns bem proeminentes. De maneira complementar, buscando uma melhor análise dos dados, a Tabela 3 apresenta uma breve estatística da base de dados em relação aos indicadores coletivos diários.

Figura 18 – Exemplo de série de DEC (a) e FEC (b) para o primeiro semestre do ano de 2014.



Fonte: O Autor

Tabela 3 – Análise estatística dos indicadores coletivos de qualidade da rede elétrica

Indicador	Qnt	$\bar{x}(10^3)$	Min(10^3)	Max (10^3)	$\sigma(10^3)$
DEC	2220	0,4547	1,0483e-03	41,1690	2,1
FEC	2220	0,2144	7,9328e-04	85,1780	2,7

Por meio da Tabela 3 é possível observar que as séries apresentam um alto desvio padrão em relação a média, de 4,6 e 12,6 vezes maior para o DEC e FEC, respectivamente.

5.1.2 Pré-processamento das Bases de Dados

Antes da aplicação das técnicas de decomposição das séries temporais (DEC e FEC), cada uma das séries recebeu tratamento de *outliers*, onde foram mantidos inalterados os dados no intervalo de valor de $[0, \sigma]$, sendo σ o desvio padrão calculado a partir de toda a série temporal até $t - 1$, sendo t o instante que da início aos dados de teste, garantindo assim apenas os dados de treinamento sejam utilizados para cálculo do desvio padrão. Os valores acima desse intervalo foram retirados da série e preenchidos com um valor por meio do algoritmo *Piecewise Cubic Hermite Data Interpolation* (PCHIP) (FRITSCH; CARLSON, 1980; FRITSCH, 1985), conforme Figura 19, onde é possível observar a média (linha preta inferior), essa somada ao desvio padrão (linha preta superior) e a linha de corte realizada, com valor de um desvio padrão, destacada pela linha verde. O PCHIP não produz *outlier* para dados com muitas oscilações e utiliza polinômios cúbicos monotônicos para encontrar os novos pontos. Para aplicação deste algoritmo, é passado como parâmetro a quantidade de pontos que deseja-se inserir entre os existentes, nesse caso, apenas os pontos removidos serão preenchidos.

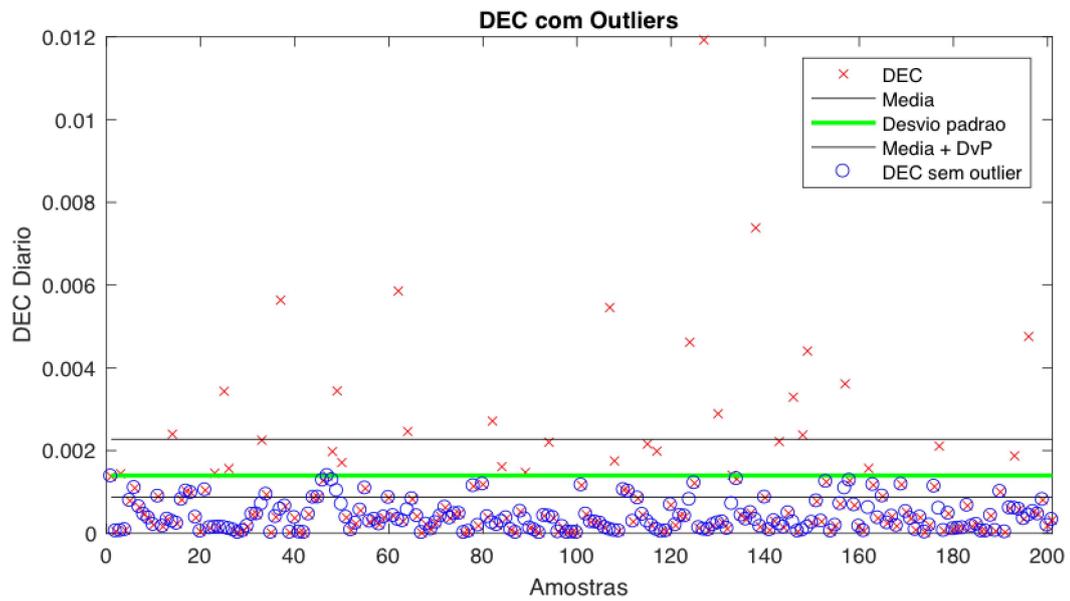
Em seguida, foram aplicados os métodos de decomposição de séries temporais, sendo eles a Transformada Wavelet de Máxima Sobreposição e o CEEMDAN.

Através da Transformada Wavelet, a série foi dividida em 8 subséries utilizando a Wavelet Mãe Daubechies 10 (*db10*), conforme Figura 20. Pela imagem, é notável a diferença entre os diferentes níveis, sendo que quanto maior o nível da subsérie, maiores são os valores das frequências que as compõem, contendo, assim, oscilações mais rápidas.

A decomposição usando a MODWT tem a finalidade de dividir as séries originais em subséries aditivas e somáveis que possuem menor complexidade. Aditivas, pois não apresentam efeito de borda como na Transformada Wavelet Discreta, sendo possível efetuar a adição de uma amostra ao final de cada subsérie antes da transformada sem a necessidade de modificar as demais amostras que compõem as respectivas subséries. Somáveis, pois a série original é recomposta pela soma das subséries da decomposição MODWT.

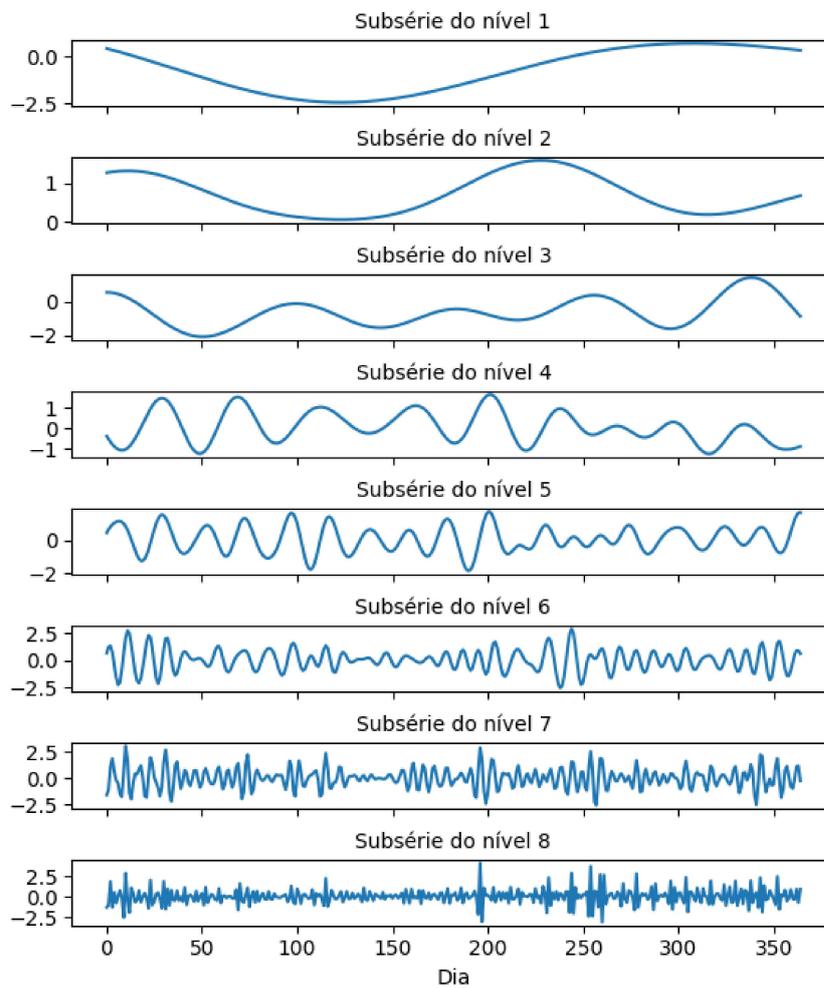
Outra técnica utilizada para a decomposição foi a *Complete Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN) que através das funções IMF

Figura 19 – Amostra do *DEC* antes e depois do tratamento de *outliers*



Fonte: O Autor

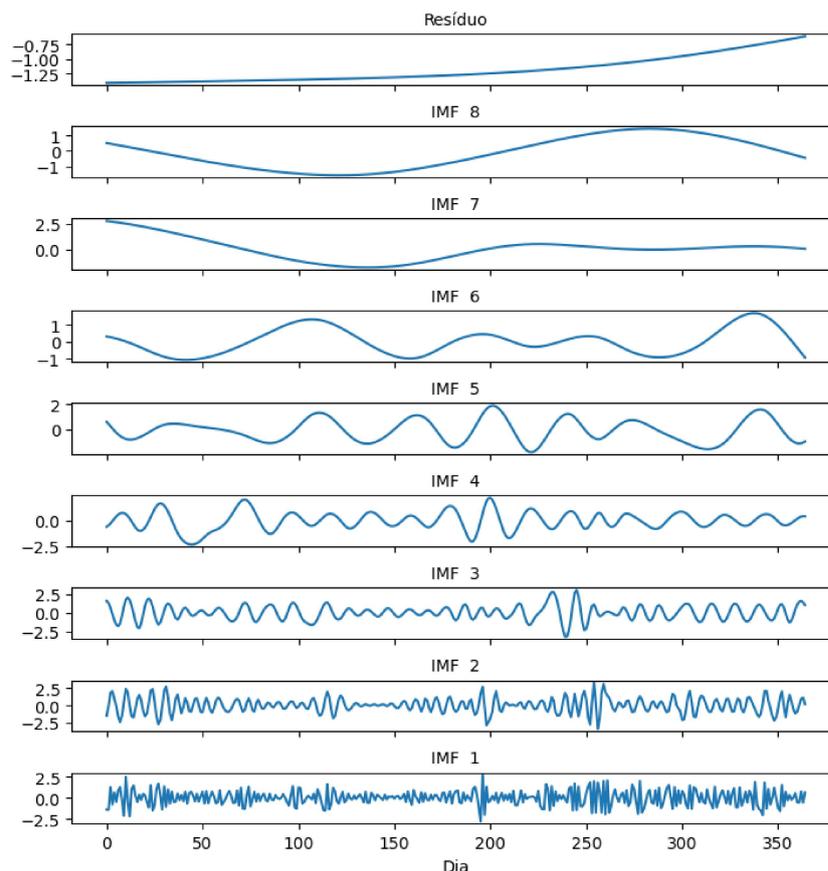
Figura 20 – Subséries após Transformada Wavelet



Fonte: O Autor

consegue decompor de modo interativo o sinal de entrada, buscando solucionar os problemas da EMD e EEMD, proporcionando o processo com menor custo computacional, maior cancelamento do ruído aplicado e permitir a recomposição do sinal com a soma de suas componentes. A Figura 21 ilustra a decomposição de um sinal em 8 subséries pelo CEEMDAN e o resíduo da decomposição.

Figura 21 – Subséries após CEEMDAN



Fonte: O Autor

Após as decomposições dos indicadores em 8 subséries pela aplicação da Transformada Wavelet e 8 subséries mais resíduo pela utilização da CEEMDAN, os dados foram separados em 75% para treino e 25% para teste, para então efetuar a normalização (*standardized*) de forma a ter-se média zero e desvio padrão unitário, conforme apresentado nas Figuras 20 e 21, buscando evitar distorções causadas pela escala durante o aprendizado. Além disso, vale ressaltar que os dados de teste foram normalizados com a média e desvio padrão do treino para que assim não houvesse contaminação do conjunto de testes.

As Tabelas 4 e 5 apresentam a média e desvio padrão de cada subsérie obtida pelo respectivo método, indicadores esses que serão utilizados no processo de recomposição da série após a predição realizada pela rede neural.

Após a normalização das séries, elas foram agrupadas em janelas de 250 dias conforme Rezaei et al. (2021), deslocando dia-a-dia até o final dos dados de treino/teste,

para formar os dados de entrada. Assim, $w_1 = [x_1, \dots, x_{250}]$, $w_2 = [x_2, \dots, x_{251}]$, $w_{n-1-249} = [x_{n-1-249}, \dots, x_{n-1}]$, sendo n a quantidade de elementos na série. Já a saída será o dia seguinte ao último valor da janela deslizante, $y_1 = [x_{251}]$, $y_2 = [x_{252}]$, $y_{n-1-249} = [x_n]$, conforme apresentado na Figura 22.

5.2 Arquiteturas

Para a realização da tarefa de predição de indicadores de qualidade, foram analisados dois tipos de redes neurais. Como um regressor mais simples, foi aplicada uma rede neural rasa com atraso de tempo, e como arquitetura profunda, foi aplicada uma rede neural híbrida, que utiliza uma camada CNN seguida por uma camada LSTM.

A Figura 23 apresenta de maneira genérica a estrutura do método utilizado, o sentido do fluxo de informações e a predição da saída. Destaca-se que os dados de treinamento de cada uma das subséries obtidas são utilizados para treinar todas as redes com o mesmo pré-processamento e decomposição. Entretanto, as redes são otimizadas individualmente de modo que há uma variação do tamanho da rede e da taxa de aprendizado entre elas. Observe que, após o treinamento dos modelos, em uma etapa de predição dos dados de teste, a série é decomposta, cada subsérie é passada por uma das redes e, no final, as

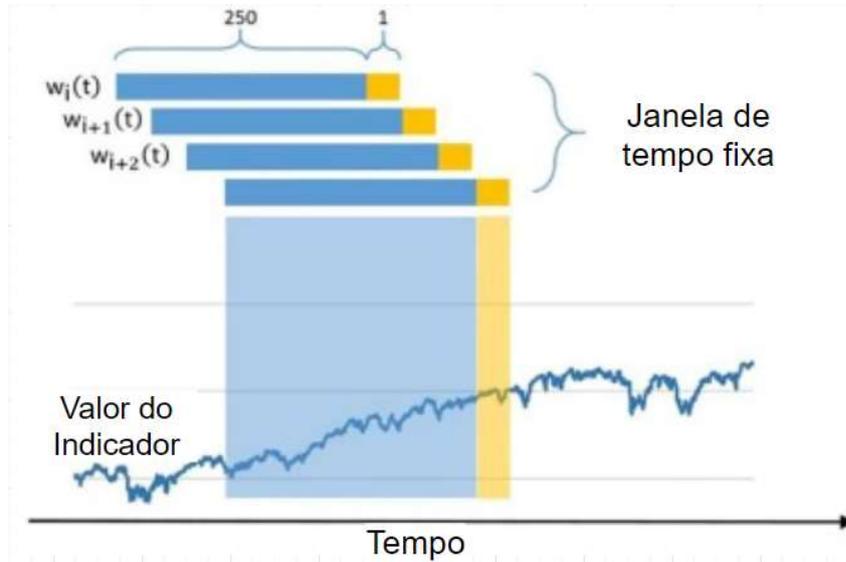
Tabela 4 – Média e desvio padrão das subséries CEEMDAN

CEEMDAN		
Componente	média (10^{-6})	desvio padrão (10^{-6})
IMF 8	-47,52	618,42
IMF 7	-9,60	426,52
IMF 6	-8,97	339,90
IMF 5	-3,02	234,41
IMF 4	-2,57	186,52
IMF 3	3,28	122,74
IMF 2	-5,34	126,18
IMF 1	7,53	54,60
Resíduo	815,07	55,64

Tabela 5 – Média e desvio padrão das subséries MODWT

MODWT		
Componente	média(10^{-6})	desvio padrão (10^{-6})
Subsérie do Nível 1	$-1,52 \times 10^{-13}$	147,06
Subsérie do Nível 2	$-6,65 \times 10^{-14}$	118,40
Subsérie do Nível 3	$-6,29 \times 10^{-14}$	167,59
Subsérie do Nível 4	$-6,17 \times 10^{-14}$	179,99
Subsérie do Nível 5	$-6,23 \times 10^{-14}$	206,96
Subsérie do Nível 6	$-5,73 \times 10^{-14}$	281,37
Subsérie do Nível 7	$-6,60 \times 10^{-14}$	366,44
Subsérie do Nível 8	748,85	497,09

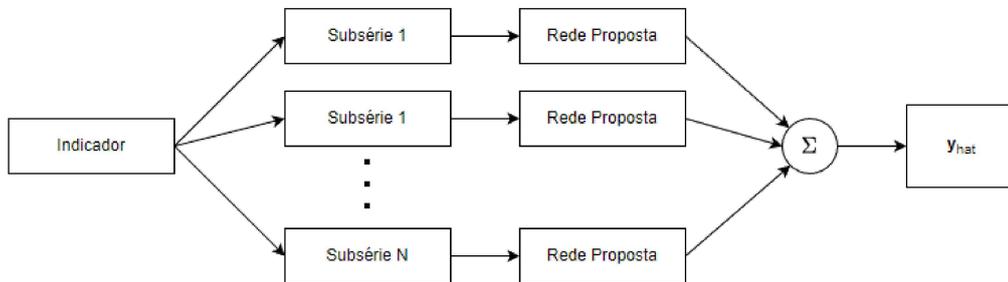
Figura 22 – Janela Deslizante no Tempo



Fonte: Rezaei et al. (2021) com adaptações.

saídas das redes são somadas para obter o valor predito.

Figura 23 – Visão geral do fluxo das séries temporais no método proposto



Fonte: O Autor.

5.2.1 Redes com Atraso no Tempo

A primeira arquitetura utilizada como base de comparação, frente a rede CNN+LSTM proposta, é uma rede rasa com atraso no tempo. Esta consiste na utilização de uma camada totalmente conectada de tamanho não unitário, seguida por outra do mesmo tipo, de tamanho um, que leva à saída, estrutura representada pela Figura 8.

Hiperparâmetros Específicos da Arquitetura:

- Número de Unidades: 10 – 100.

5.2.2 LSTM

Outra rede utilizada para comparação é a rede LSTM, amplamente aplicada para predição de séries temporais devido a sua capacidade de memória de longo e curto prazo.

Para esta rede foi utilizada apenas uma camada LSTM, cuja entrada é a janela de 250 dias, seguida por uma densa composta por único neurônio que leva a saída.

Hiperparâmetros Específicos da Arquitetura:

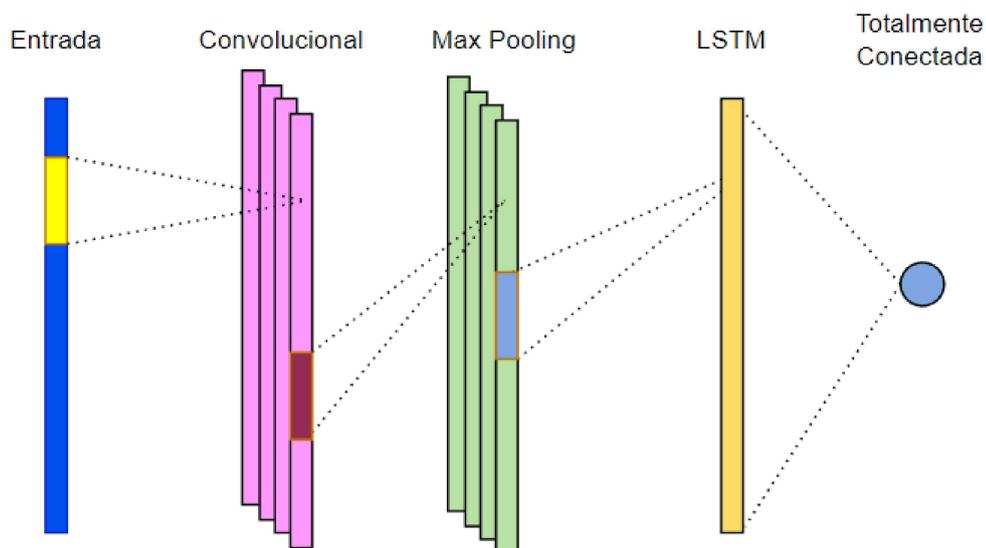
- Número de Unidades: 100 – 250.

5.2.3 CNN+LSTM

A rede profunda proposta para a predição é a CNN+LSTM. Ela é composta por dois blocos principais, sendo o primeiro a camada de convolução com *max pooling*, e o segundo a camada LSTM.

Dessa forma, a janela de 250 dias é inserida na camada de entrada que tem suas informações “filtradas” pela camada CNN e só então passa para a camada LSTM que é responsável pelas relações temporais das informações, cuja saída é encaminhada para uma camada totalmente conectada, seguida por uma função de ativação sigmoideal. A Figura 24 apresenta a organização das camadas.

Figura 24 – Estrutura da rede CNN+LSTM



Fonte: O Autor.

Hiperparâmetros Específicos da Arquitetura

- Canais (Conv): 256 – 512;

- Kernel Pool Size: 2 – 10;
- Unidades LSTM: 200 – 350;

5.2.4 Configurações Gerais

Os hiperparâmetros das redes, TDNN e CNN+LSTM, foram ajustados através de otimização bayesiana (SNOEK et al., 2012) por meio do módulo KerasTuner (O'MALLEY et al., 2019) para Python. Dado que cada rede foi treinada individualmente para cada indicador e subsérie/IMF deste, os hiperparâmetros também foram otimizados dessa forma.

Parâmetros gerais:

- Taxa de Aprendizado: 10^{-3} – 10^{-5} ;
- Função de ativação: sigmoid
- Épocas: 200
- Função de Perda: Huber

A função Huber, definida pela Equação (5.2.4), foi utilizada como função de perda pela sua larga aplicação no campo de regressão por combinar a função MSE e a MAE (*Mean Absolute Error*), fazendo com que tenha um comportamento quadrático para erros pequenos e linear para os maiores (ESMAEILI; MARVASTI, 2019).

$$L_{\delta}(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \leq \delta, \\ \delta(|a| - \frac{1}{2}\delta) & \text{for } |a| > \delta, \end{cases} \quad (5.1)$$

sendo a o resíduo e δ o ponto onde a função deixa de ser linear e passa a ser quadrática.

6 Experimentos e Resultados

6.1 Métricas

Tendo em vista que a tarefa em questão trata-se de uma predição, cuja saída são valores contínuos, foi utilizado o *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) e *Median Relative Absolute Error* (MdRAE) como métricas de avaliação:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (6.1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (6.2)$$

$$MdRAE = \text{Mediana}_{i=1,n} \left(\frac{|y_i - \hat{y}_i|}{|y_i - b_i|} \right), \quad (6.3)$$

onde n é o número de amostras avaliadas, \hat{y}_i é o valor estimado para a i -ésima amostra, y_i é o valor esperado como saída, b_i é uma referência calculada a partir de uma previsão ingênua.

O erro foi calculado após os dados passarem por um processo reverso ao de pré-processamento, retornando-os a sua forma padrão. Ou seja, após as saídas de cada rede para cada subsérie, elas são restauradas do processo de normalização, ao multiplicar pelo desvio padrão e somar a média e, por fim, a soma das subséries tratadas é a saída estimada.

O RMSE é um erro com escala dependente, ou seja, é relativo a unidade dos dados utilizados. Por isso, conforme apresentado na próxima seção, terá uma pequena semelhança com os indicadores de continuidade.

O MAPE é um erro percentual, independente da escala, e muito utilizado para comparação entre diferentes séries. Nesse caso, aplicado para permitir uma melhor comparação entre o DEC e o FEC.

Já o MdRAE é um erro relativo, que compara os erros obtidos com os de um modelo de referência, geralmente a partir de uma previsão ingênua. Nesse caso, as predições iguais aos dias anteriores.

6.2 Procedimentos Experimentais

Cada base de dados foi dividida em duas partes: um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento, que corresponde a 75% da base de dados,

foi utilizado para treinar e ajustar os hiperparâmetros da rede usada em cada componente a partir da função Huber, sendo salvos os pesos da rede para utilização posterior na etapa de predição com os dados de teste. Durante os testes executados, foi possível observar que o treinamento dos modelos convergia com cerca de 30 épocas, sendo configurado um limite máximo de 200 épocas com *Early Stop*, ou parada antecipada, de modo a evitar *overfitting*.

6.3 Recursos

Para os experimentos realizados neste trabalho foi utilizado um computador com as seguintes configurações:

- Windows 10;
- CPU Intel(R) Core(TM) i7-10750H 2.59GHz;
- 16 GB de memória RAM;
- GPU Nvidia GTX 1660 Ti.

Todo o código desenvolvido foi implementado em Python 3 (ROSSUM; DRAKE, 2009), desde a etapa de leitura dos dados, decomposição em frequência, modelos de redes neurais rasas e profundas e a otimização de hiperparâmetros. Sendo que para as redes neurais foi utilizado o *framework* Keras (CHOLLET et al., 2015), baseado em TensorFlow. Além disso, outras duas principais bibliotecas utilizadas para o desenvolvimento do código foram às Matplotlib (HUNTER, 2007) e NumPy (HARRIS et al., 2020).

6.4 Resultados

Foram feitos experimentos com as redes profundas e rasas, com e sem o pré-processamento baseado em decomposição das séries temporais. Assim, foram feitos nove testes para cada indicador, sendo eles:

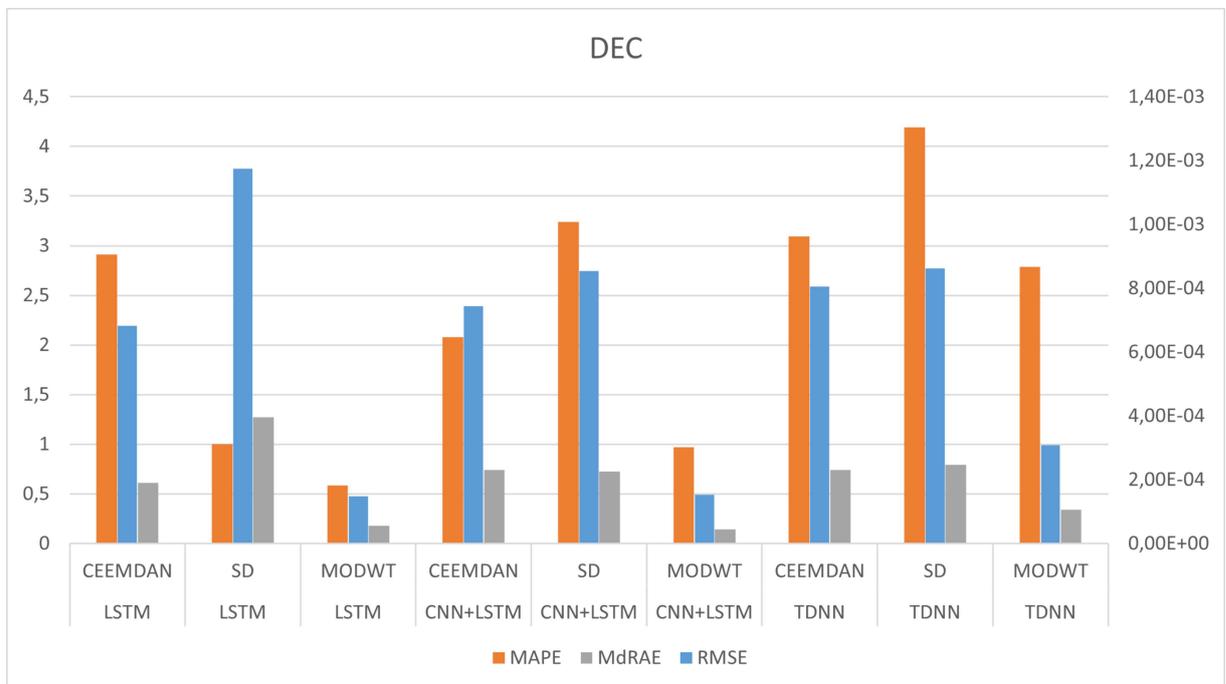
1. TDNN e SD;
2. TDNN e CEEMDAN;
3. TDNN e MODWT;
4. LSTM e SD;
5. LSTM e CEEMDAN;

- 6. LSTM e MODWT;
- 7. CNN+LSTM e SD;
- 8. CNN+LSTM e CEEMDAN;
- 9. CNN+LSTM e MODWT;

onde a sigla SD significa “sem decomposição”.

A Tabela 6 apresenta os valores dos três cálculos de erros utilizados para os diferentes tipos de redes utilizadas, bem como para as técnicas de pré-processamento aplicadas. De maneira complementar, as Figuras 25 e 26 mostram de forma gráfica os valores da tabela. Através dessas, pode-se observar que a aplicação do pré-processamento MODWT obteve o melhor desempenho em todos os casos, em segundo lugar tem-se o método CEEMDAN, e, por último, estão os resultados sem decomposição. Em relação às redes, é possível notar uma diferença entre qual obteve o melhor desempenho para o DEC e FEC. Para o primeiro, a rede LSTM com MODWT destacou-se com duas das três métricas de erro utilizadas, já para o FEC, a rede CNN+LSTM com MODWT obteve melhores resultados.

Figura 25 – Comparação dos erros nas diferentes configurações para o indicador DEC

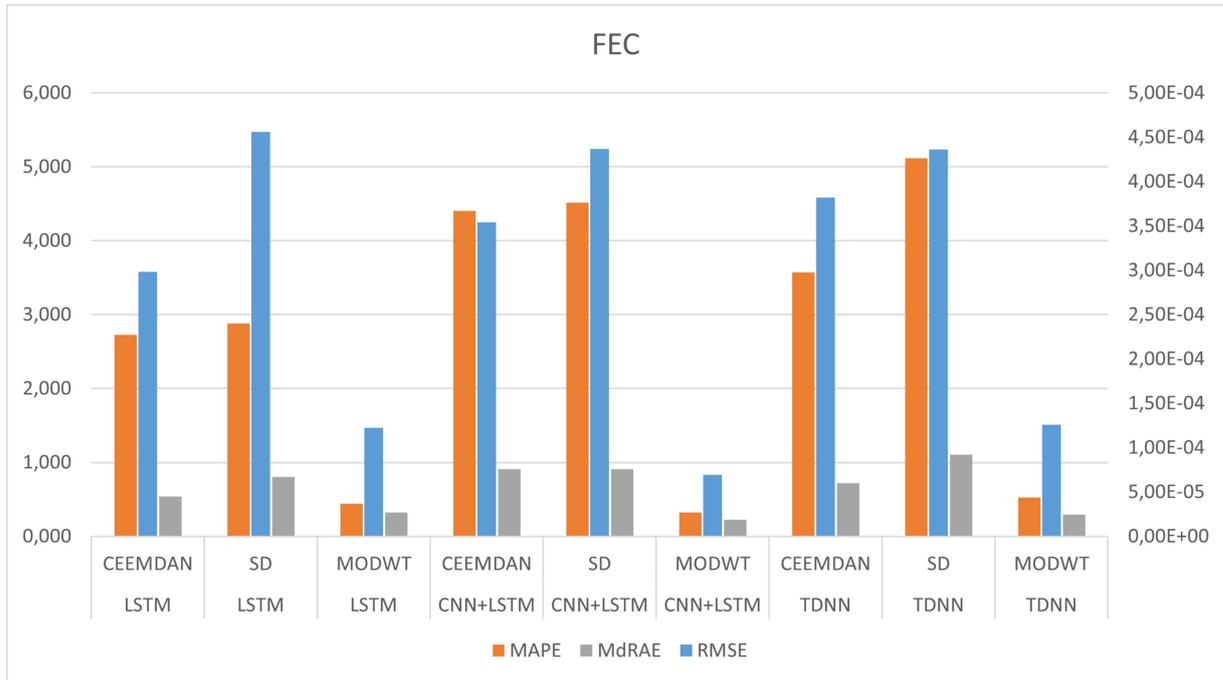


Fonte: O Autor.

Tabela 6 – RMSE, MAPE e MdRAE para as variações de rede e de decomposições utilizadas no pré-processamento de cada indicador. Em negrito os menores valores de cada erro para cada indicador.

			FEC	DEC
RMSE	TDNN	SD	$4,36 \times 10^{-4}$	$8,62 \times 10^{-4}$
		CEEMDAN	$3,82 \times 10^{-4}$	$8,05 \times 10^{-4}$
		MODWT	$1,26 \times 10^{-4}$	$3,09 \times 10^{-4}$
	LSTM	SD	$4,56 \times 10^{-4}$	$1,18 \times 10^{-3}$
		CEEMDAN	$2,98 \times 10^{-4}$	$6,83 \times 10^{-4}$
		MODWT	$1,22 \times 10^{-4}$	$1,49 \times 10^{-4}$
	CNN+LSTM	SD	$4,37 \times 10^{-4}$	$8,54 \times 10^{-4}$
		CEEMDAN	$3,54 \times 10^{-4}$	$7,44 \times 10^{-4}$
		MODWT	$6,91 \times 10^{-5}$	$1,53 \times 10^{-4}$
MAPE	TDNN	SD	5,119	4,193
		CEEMDAN	3,571	3,096
		MODWT	0,525	2,786
	LSTM	SD	2,879	1,000
		CEEMDAN	2,723	2,911
		MODWT	0,444	0,587
	CNN+LSTM	SD	4,517	3,238
		CEEMDAN	4,406	2,079
		MODWT	0,324	0,969
MdRAE	TDNN	SD	1,103	0,796
		CEEMDAN	0,724	0,740
		MODWT	0,295	0,342
	LSTM	SD	0,803	1,270
		CEEMDAN	0,540	0,609
		MODWT	0,326	0,180
	CNN+LSTM	SD	0,909	0,728
		CEEMDAN	0,910	0,741
		MODWT	0,224	0,144

Figura 26 – Comparação dos erros nas diferentes configurações para o indicador FEC



Fonte: O Autor.

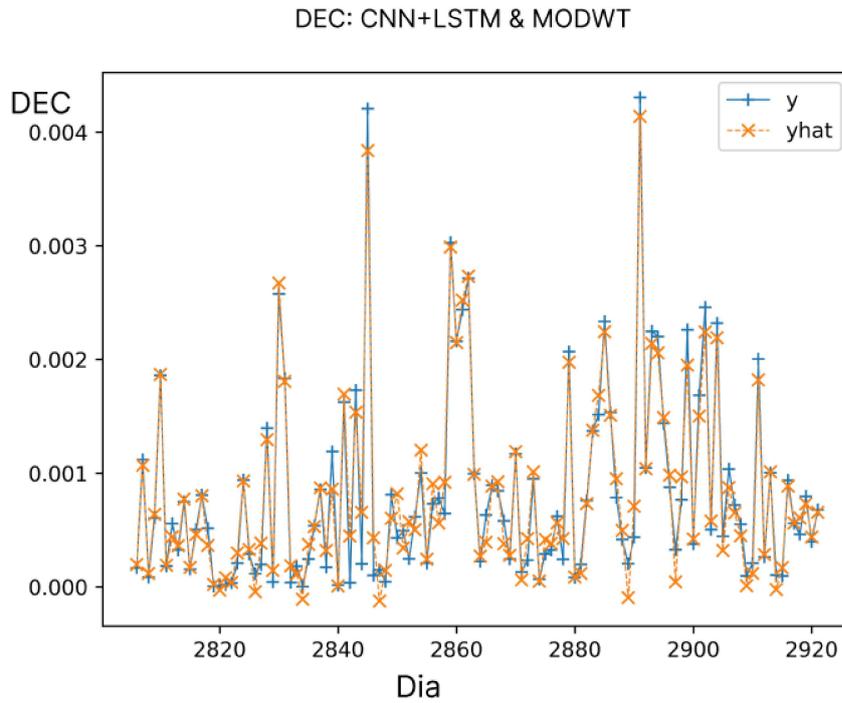
As Figuras 27 a 44 mostram de forma gráfica os valores das predições do DEC e FEC para os diferentes tipos de pré-processamento analisados, sendo os valores estimados \hat{y} , em laranja tracejado, e os valores reais do indicador y , em azul contínuo.

A primeira análise é em relação ao pré-processamento de dados, observando qual decomposição em frequência apresentou o melhor resultado, e se houve realmente um ganho com a aplicação dessas técnicas. Observando os resultados apresentados nos gráficos com as predições do DEC e FEC, nota-se que apesar dos erros (RMSE, MAPE e MdRAE) para os testes sem decomposição estarem dentro da mesma ordem de grandeza que os demais, não houve um bom desempenho deste método, conforme apresentado nas Figuras 29, 32 e 35 para o DEC, e Figuras 38, 41 e 44 para o FEC.

Essa afirmação pode também ser destacada pelos valores apresentados na Tabela 6 e Figuras 25 e 26, onde os erros quando não aplicado a decomposição foram maiores em todos os casos. Destaca-se que essas redes, nas condições de hiperparâmetros configuradas, não conseguiram aprender o comportamento dos dados para o indicador DEC e apresentaram um aprendizado incipiente para o FEC. Dessa forma, a não utilização da decomposição indica um prejuízo no desempenho do método.

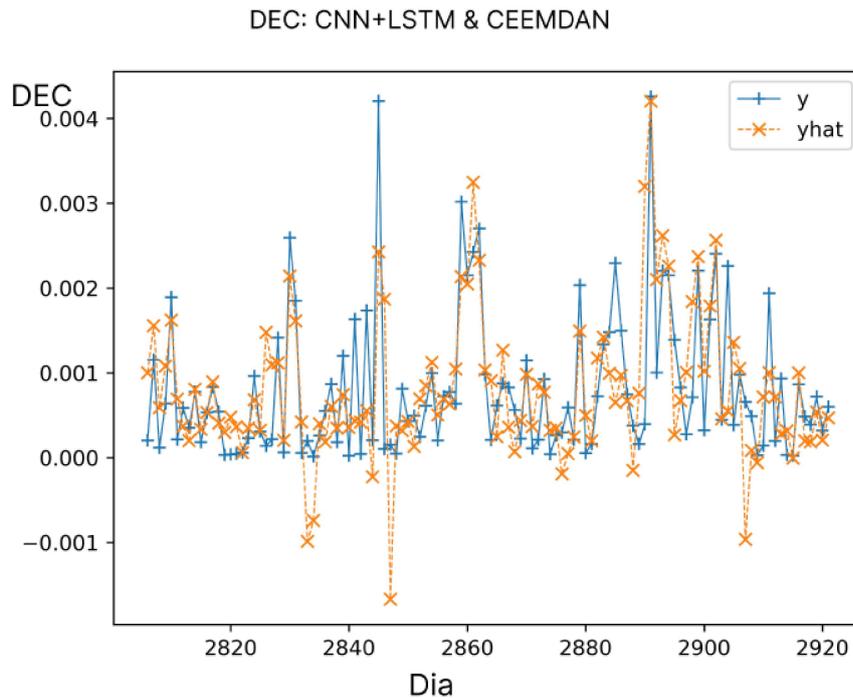
Já entre os outros dois métodos de decomposição utilizados, CEEMDAN e MODWT, é possível observar que o último apresenta um resultado superior, tanto para a rede profunda quanto para a rasa, e que a aplicação do CEEMDAN resultou em predições mais discrepantes (*outliers*). Nota-se também que a aplicação do método CEEMDAN resultou

Figura 27 – Predição do DEC para a rede profunda CNN+LSTM com pré-processamento MODWT



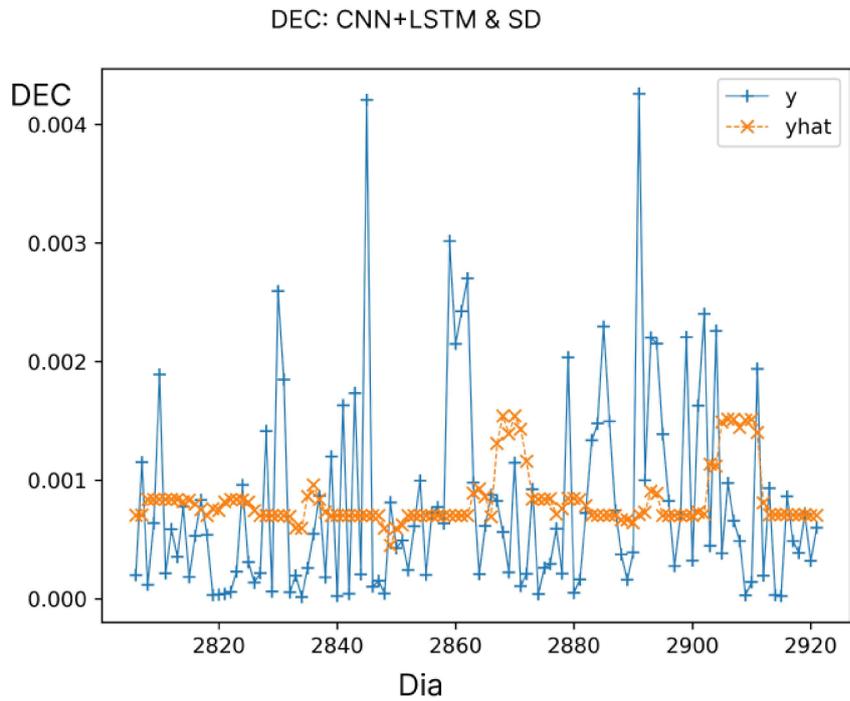
Fonte: O Autor.

Figura 28 – Predição do DEC para a rede profunda CNN+LSTM com pré-processamento CEEMDAN



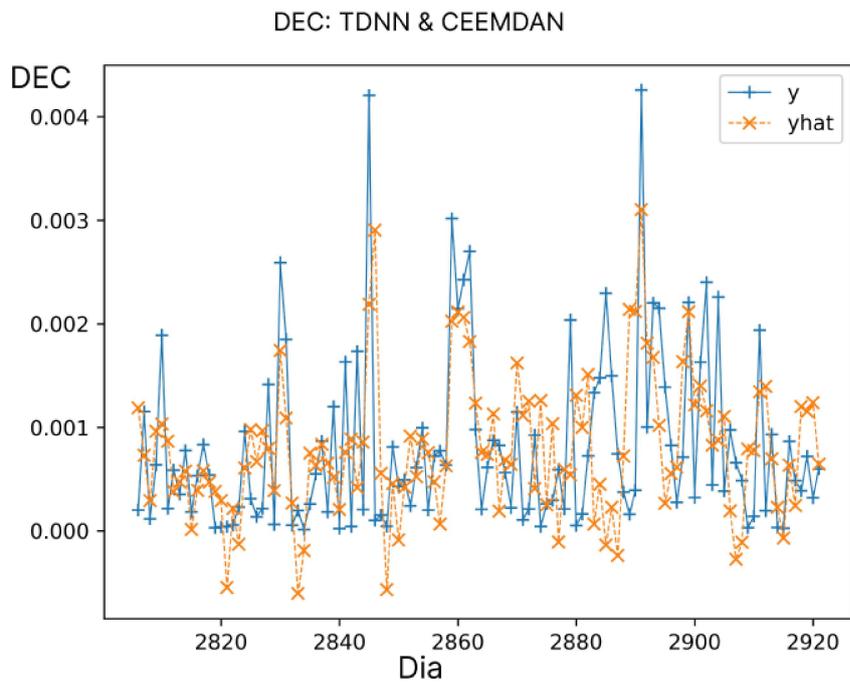
Fonte: O Autor.

Figura 29 – Predição do DEC para a rede profunda CNN+LSTM sem decomposição (SD)



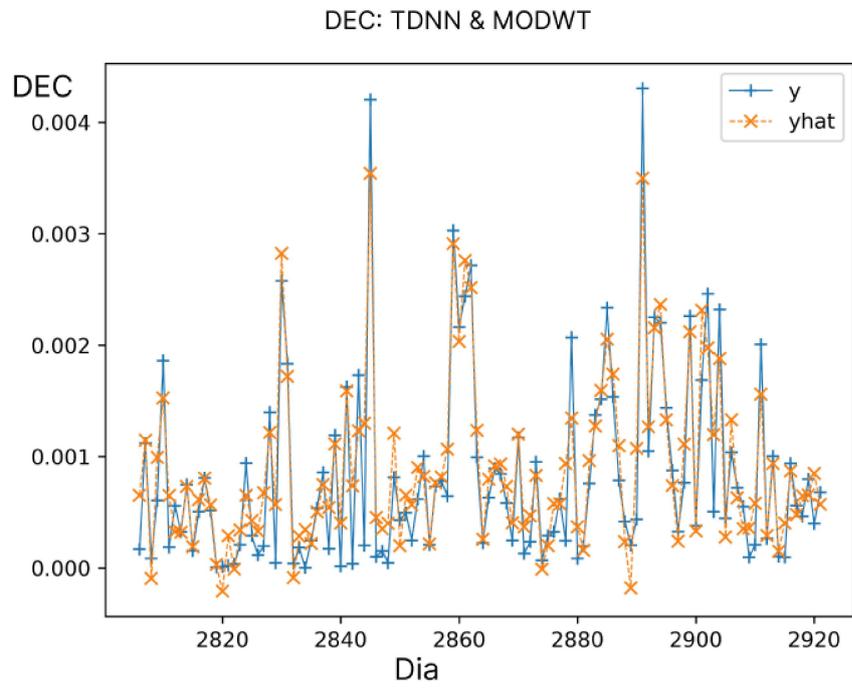
Fonte: O Autor.

Figura 30 – Predição do DEC para a rede rasa TDNN com pré-processamento CEEMDAN



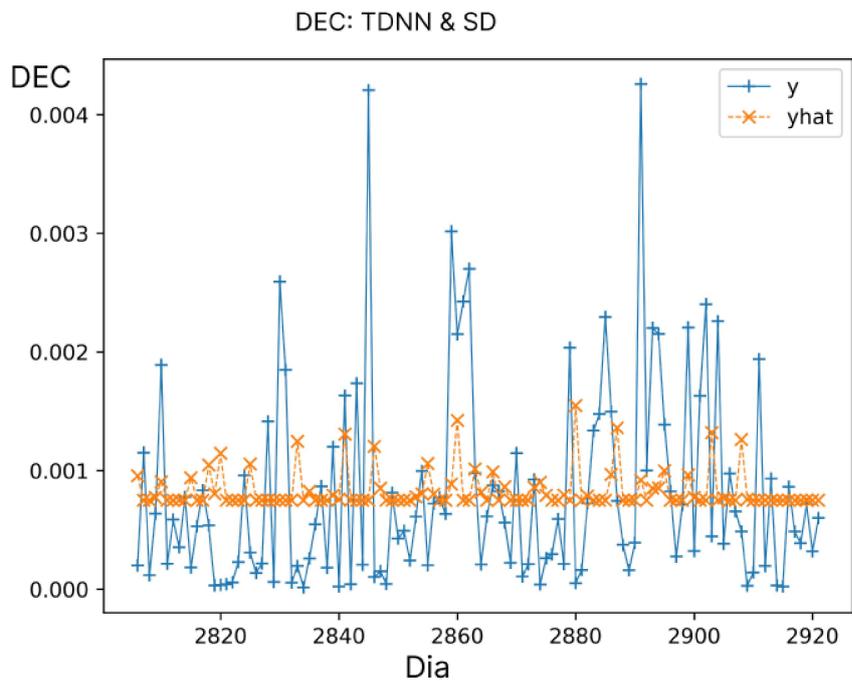
Fonte: O Autor.

Figura 31 – Predição do DEC para a rede rasa TDNN com pré-processamento MODWT



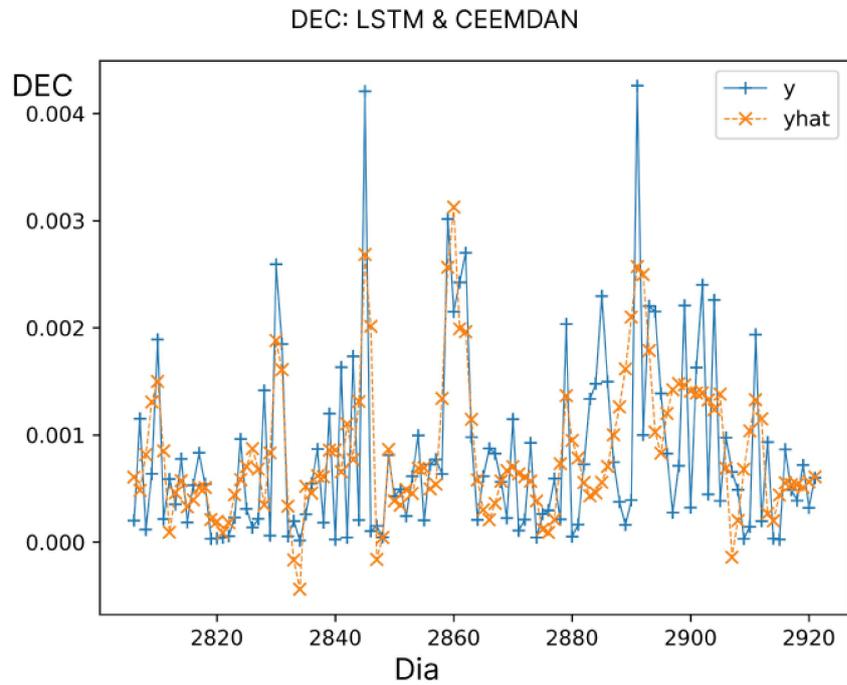
Fonte: O Autor.

Figura 32 – Predição do DEC para a rede rasa TDNN sem decomposição (SD)



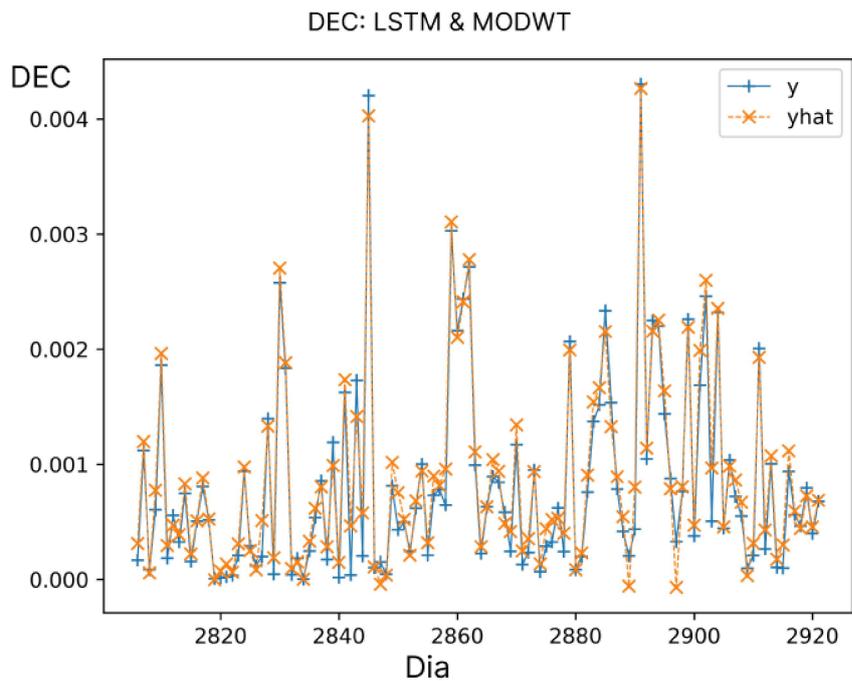
Fonte: O Autor.

Figura 33 – Predição do DEC para a rede LSTM com pré-processamento CEEMDAN



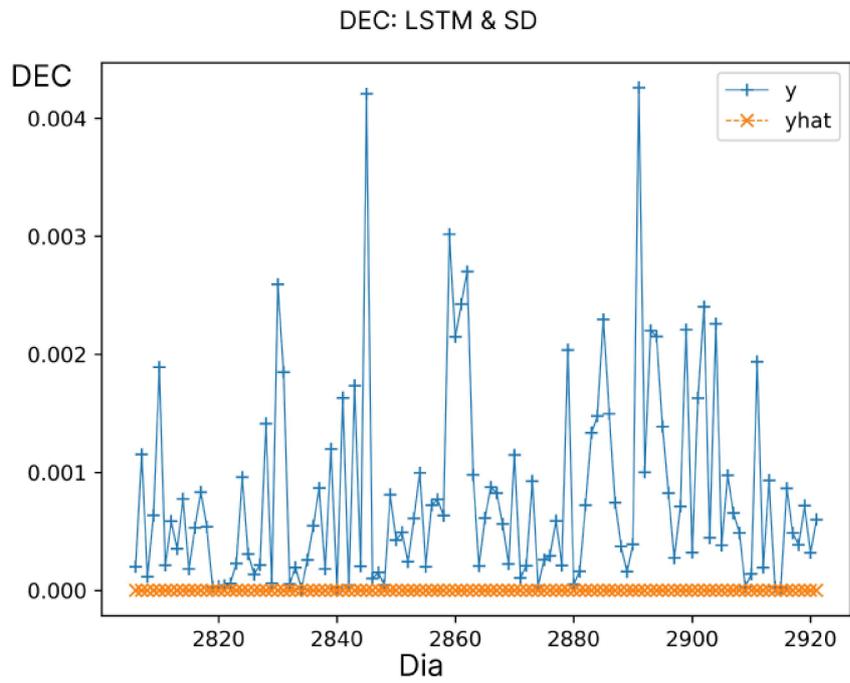
Fonte: O Autor.

Figura 34 – Predição do DEC para a rede LSTM com pré-processamento MODWT



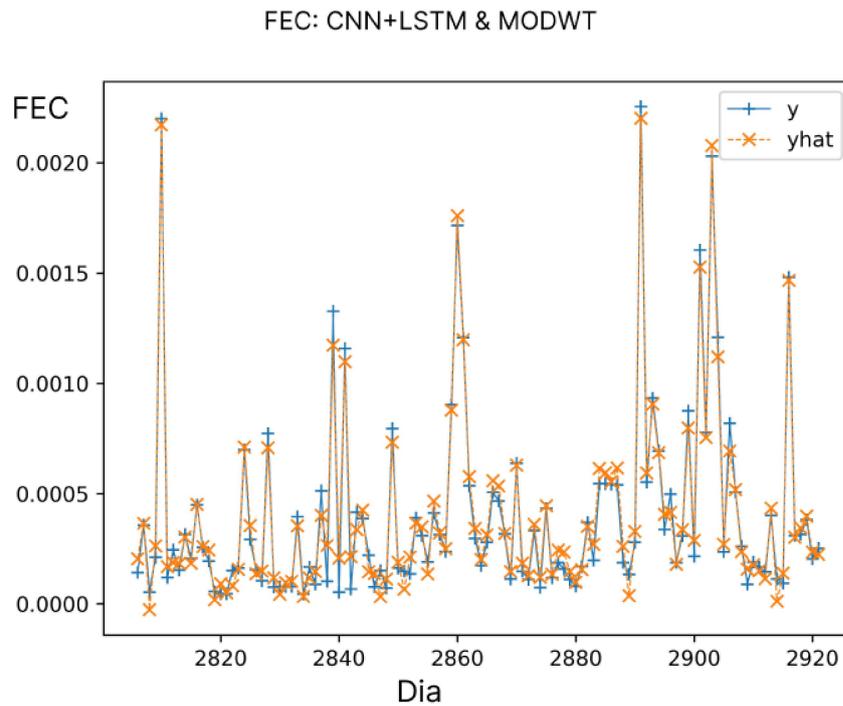
Fonte: O Autor.

Figura 35 – Predição do DEC para a rede LSTM sem decomposição (SD)



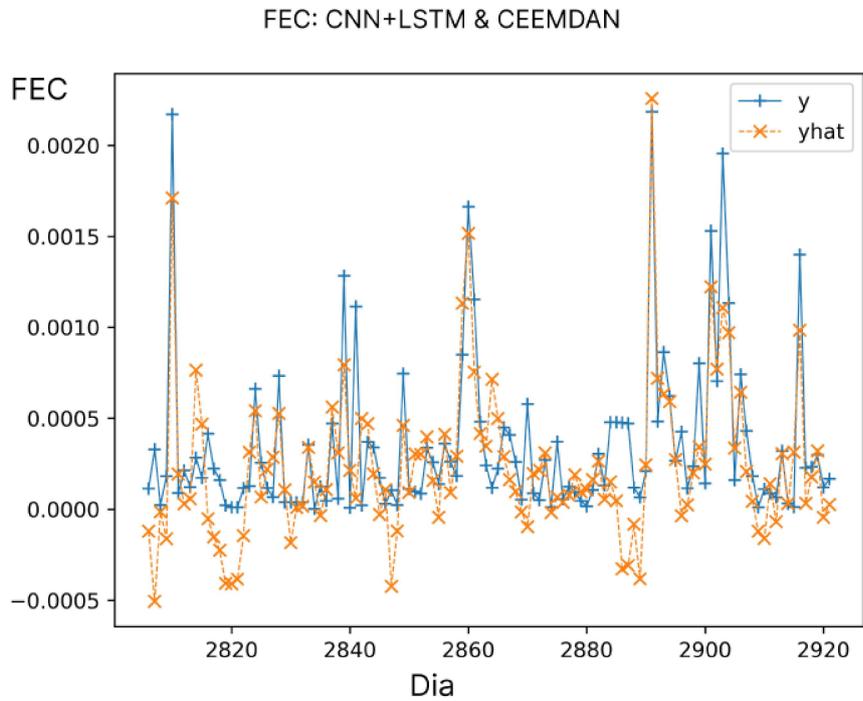
Fonte: O Autor.

Figura 36 – Predição do FEC para a rede profunda CNN+LSTM com pré-processamento MODWT



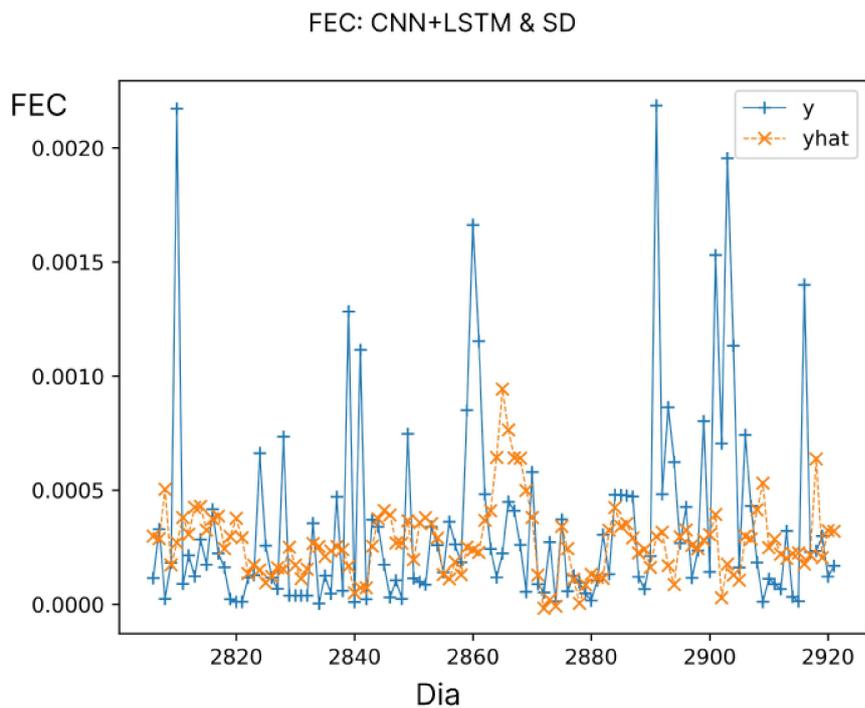
Fonte: O Autor.

Figura 37 – Predição do FEC para a rede profunda CNN+LSTM com pré-processamento CEEMDAN



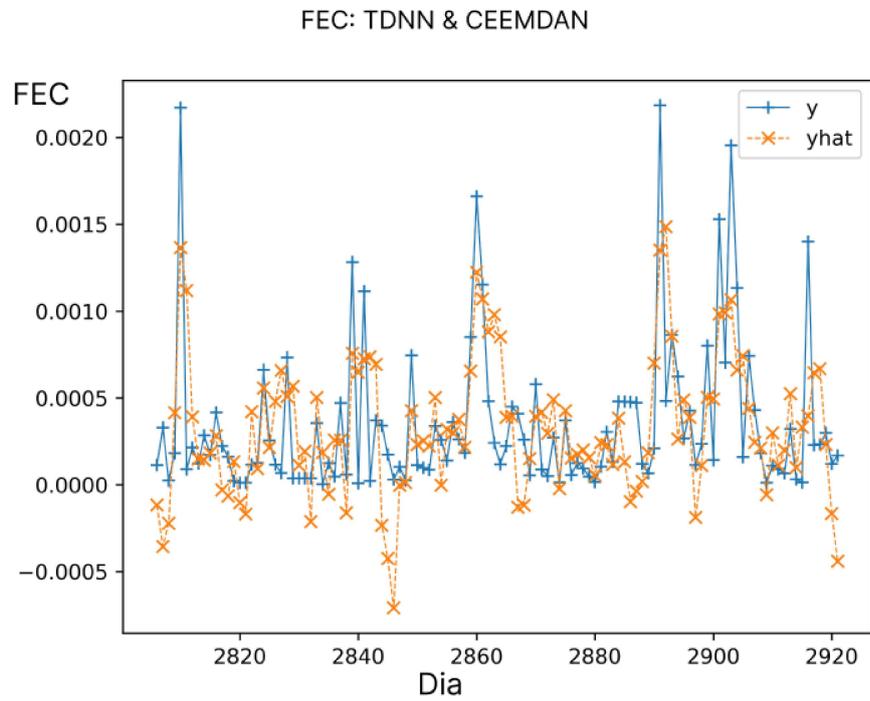
Fonte: O Autor.

Figura 38 – Predição do FEC para a rede profunda CNN+LSTM sem decomposição (SD)



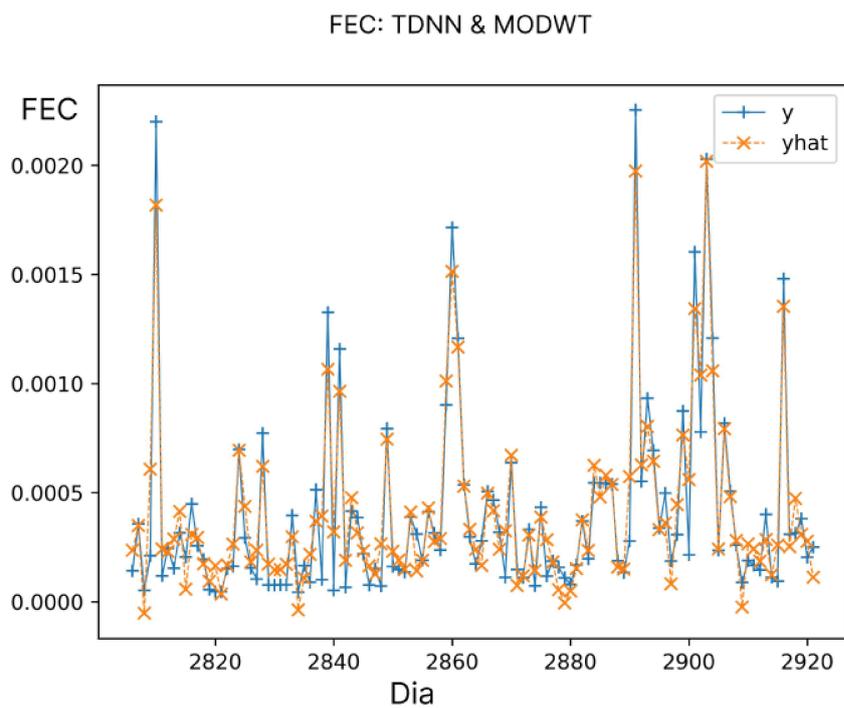
Fonte: O Autor.

Figura 39 – Predição do FEC para a rede rasa TDNN com pré-processamento CEEMDAN



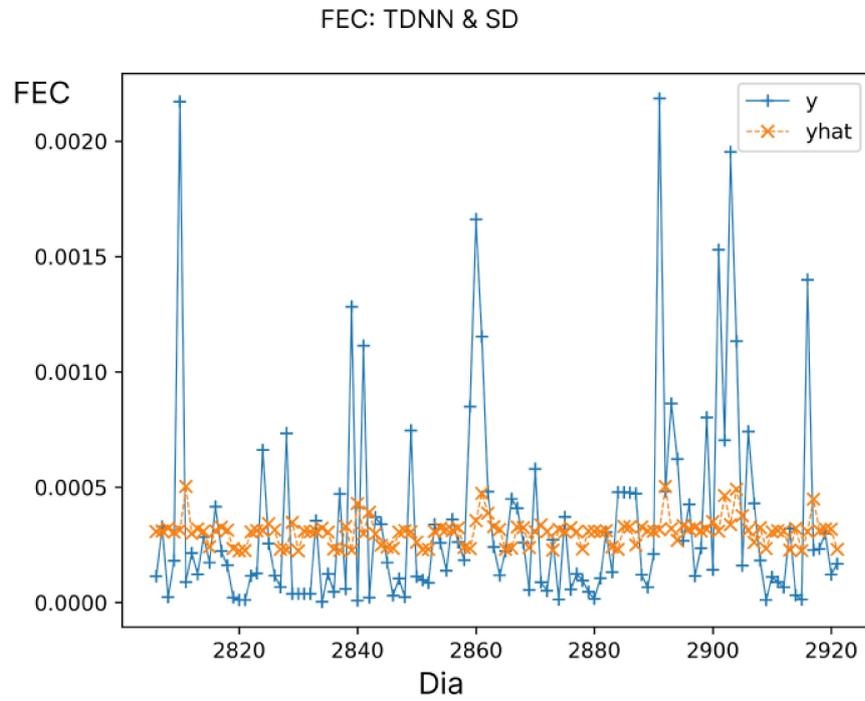
Fonte: O Autor.

Figura 40 – Predição do FEC para a rede rasa TDNN com pré-processamento MODWT



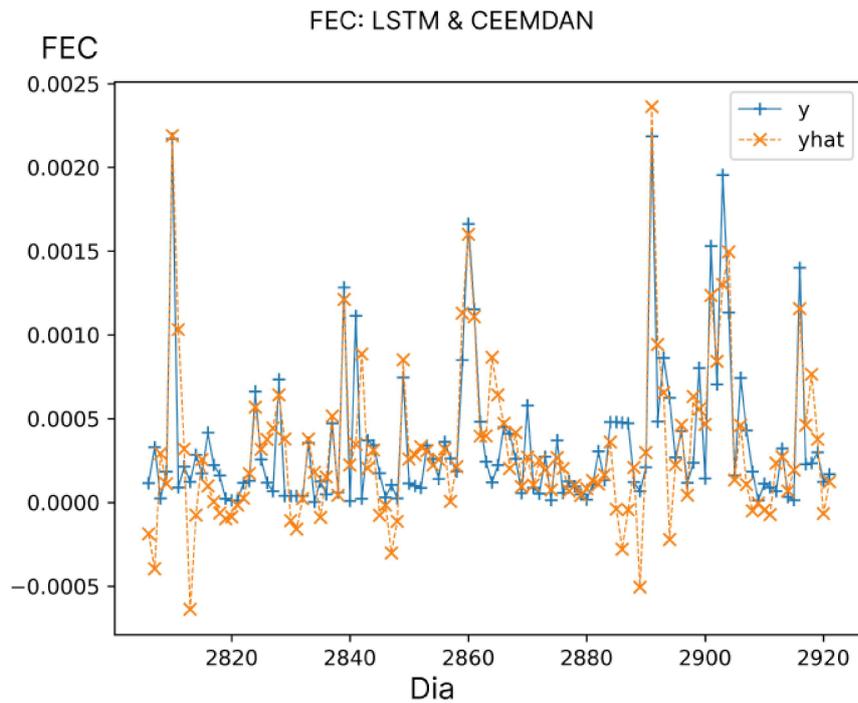
Fonte: O Autor.

Figura 41 – Predição do FEC para a rede rasa TDNN sem decomposição (SD)



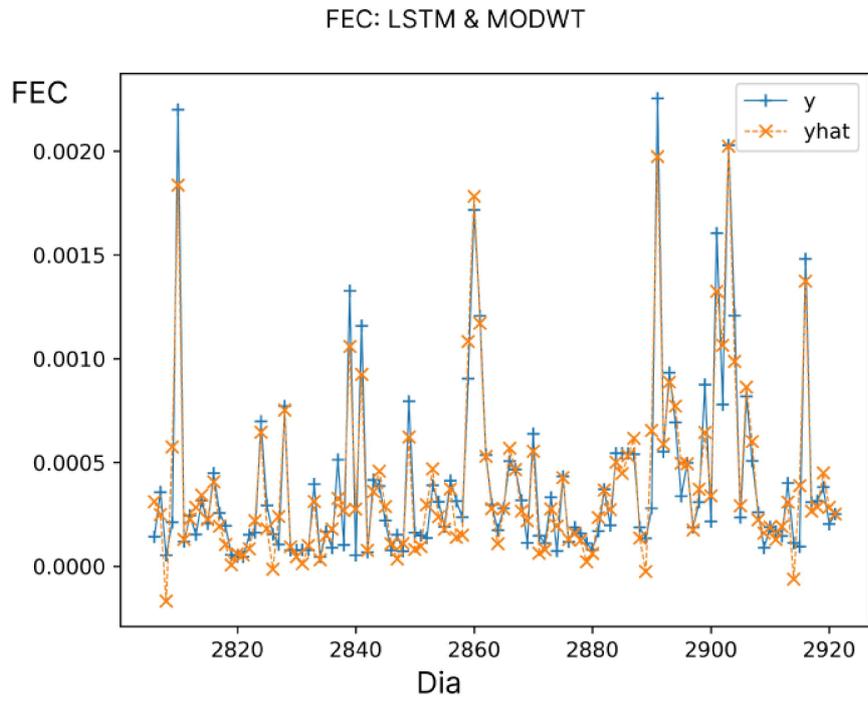
Fonte: O Autor.

Figura 42 – Predição do FEC para a rede LSTM com pré-processamento CEEMDAN



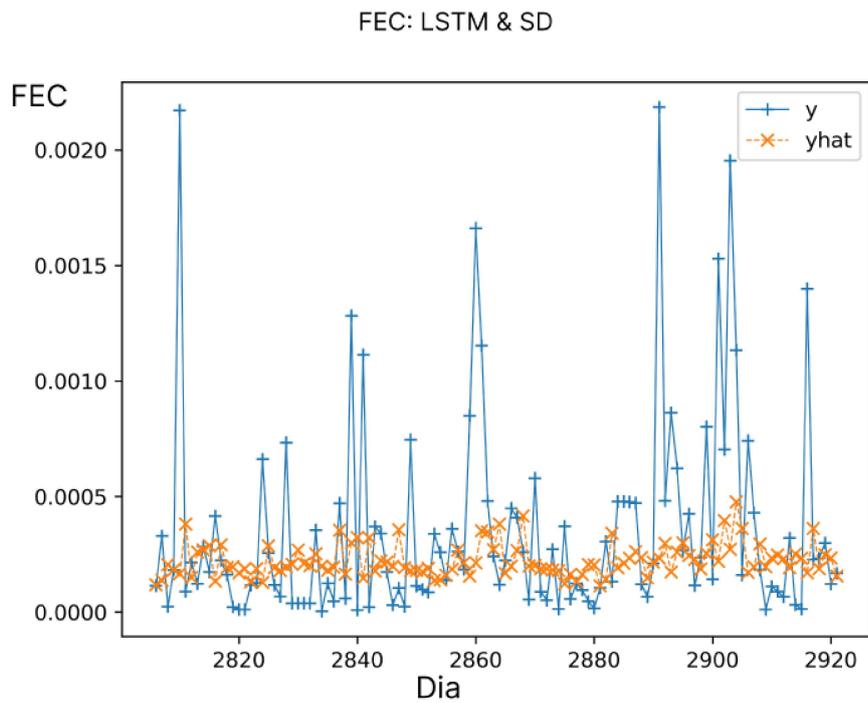
Fonte: O Autor.

Figura 43 – Predição do FEC para a rede LSTM com pré-processamento MODWT



Fonte: O Autor.

Figura 44 – Predição do FEC para a rede LSTM sem decomposição (SD)



Fonte: O Autor.

num maior erro das redes (entre os dois métodos de decomposição), com os pontos preditos mais distantes dos reais. Outro parâmetro relevante é o tempo levado para a execução do algoritmo de decomposição. O CEEMDAN levou em média 28 segundos, enquanto o MODWT precisou de somente 6 segundos, possuindo assim um menor tempo de execução.

A próxima análise realizada é em relação as redes aplicadas, observando a rede profunda híbrida, CNN+LSTM, em relação a rede rasa TDNN, utilizada como parâmetro de comparação para avaliar a viabilidade da aplicação de uma rede mais complexa. Comparando as duas redes através das figuras apresentadas e da Tabela 6, nota-se que para um mesmo pré-processamento e para o mesmo indicador, a rede CNN+LSTM teve melhores resultados, como para o DEC e decomposição MODWT, as métricas de erro foram TDNN: $RMSE = 1,26 \times 10^{-4}$, $MAPE = 0,525$ e $MdRAE = 0,295$, já para a CNN+LSTM: $RMSE = 6,91 \times 10^{-5}$, $MAPE = 0,324$ e $MdRAE = 0,224$. Dessa forma, apesar da TDNN ter tido um bom desempenho, a CNN+LSTM desempenhou melhor.

Em relação aos erros, a Tabela 6 mostra que a decomposição MODWT foi a técnica que proporcionou os melhores resultados, e nota-se que a CNN+LSTM com a decomposição MODWT obteve o melhor desempenho para o FEC, já para o DEC em primeiro lugar ficou a rede LSTM com decomposição MODWT. Também de interesse observar que a aplicação dos dados sem decomposição, apresentados pela sigla SD, apresentaram resultados que mostram que as redes não foram capazes de generalizar bem. Estima-se que a razão para tal deve-se a natureza demasiadamente ruidosa dos dados com transições muito abruptas de valores.

Com os dados obtidos, é possível observar que as técnicas de decomposição de frequência apresentam ganhos de desempenho em ambas as redes, tendo uma vantagem clara na utilização do MODWT juntamente com a rede profunda híbrida, sendo esse conjunto o mais eficaz dentre as técnicas aplicadas.

Outra análise possível é referente ao método atual utilizado pela distribuidora de energia. Método este que consiste basicamente na média dos meses anteriores. Dessa forma, para efeito de comparação, foi calculado o valor médio por dia dos indicadores dos meses anteriores para estimar os dos dias do mês atual. Por exemplo, para estimar os valores dos dias de Janeiro, foi calculada a média diária dos meses de Janeiro dos anos anteriores, e o mesmo foi feito para os demais meses. A Figura 45 apresenta os valores estimados para o DEC por este método e os seus valores reais, enquanto a Figura 46 apresenta os dados para o FEC.

Além disso, ao comparar os valores dos erros da Tabela 7, que mostra os erros para a métrica de média mensal por dia, e as métricas obtidas anteriormente pelas redes e pré-processamento utilizados (Tabela 6), nota-se que todos os erros foram maiores que as opções que se destacaram anteriormente. Principalmente ao comparar as métricas de MAPE e MdRAE. Dessa forma, comparando os valores e observando as Figuras 45 e 46,

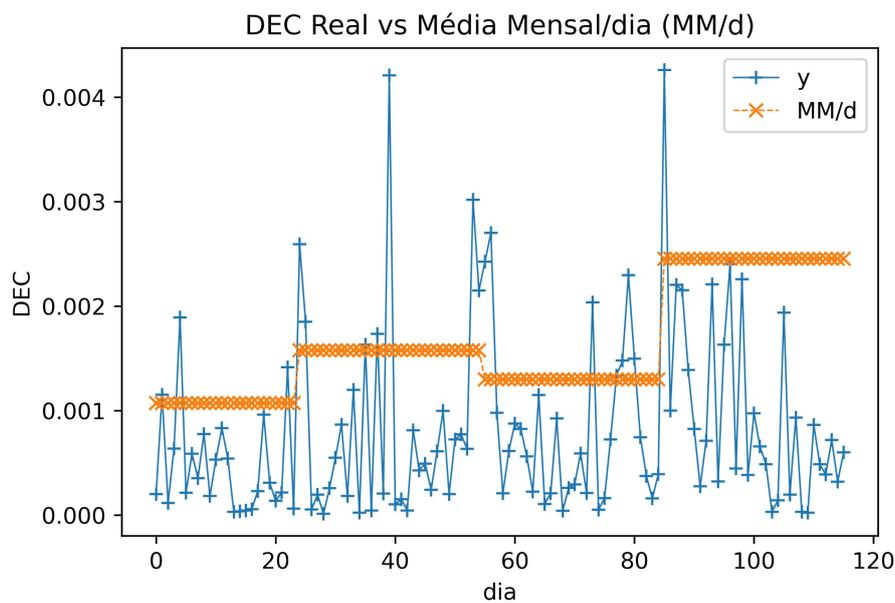
fica evidente que a nível diário o método atual não permite uma aplicação com qualidade adequada.

Tabela 7 – RMSE, MAPE e MdRAE para os indicadores em relação a média mensal por dia

	DEC	FEC
RMSE	$1,22 \times 10^{-3}$	$6,4 \times 10^{-4}$
MAPE	8,582	12,226
MdRAE	1,636	2,862

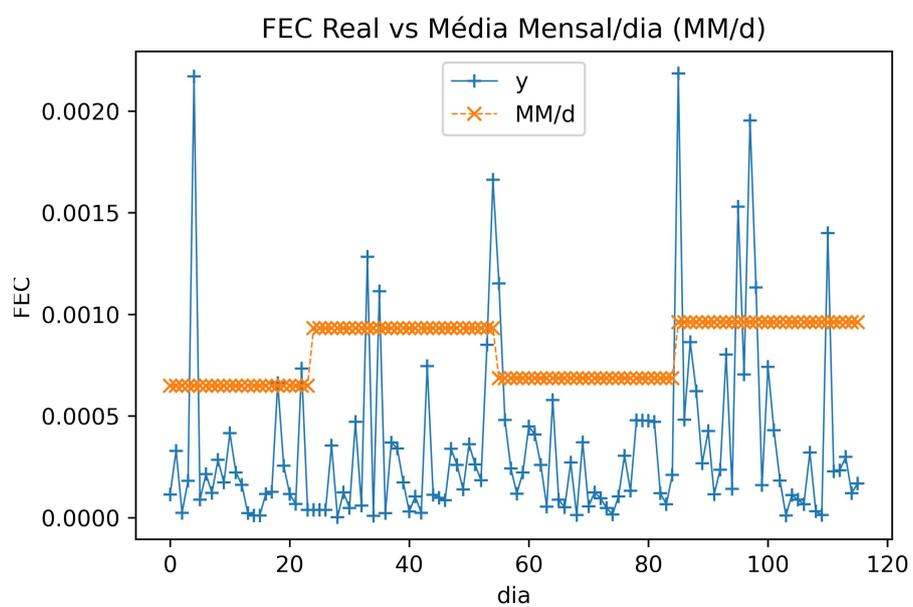
Assim, caracteriza-se como não tão eficiente para tomada de ações quando comparado com o método proposto neste trabalho. Embora não seja uma representação exata do método utilizado atualmente na distribuidora de energia, ele dá uma ideia que os resultados obtidos neste trabalho conseguem alcançar resultados melhores.

Figura 45 – Comparação do DEC (azul) e da média mensal por dia (laranja)



Fonte: O Autor.

Figura 46 – Comparação do FEC (azul) e da média mensal por dia (laranja)



Fonte: O Autor.

7 Conclusão e Trabalhos Futuros

Este trabalho teve como objetivo propor o desenvolvimento de um modelo matemático para realizar a predição dos indicadores de continuidade DEC e FEC de modo a permitir que as concessionárias de energia possam efetuar um melhor planejamento através do direcionamento de suas equipes de atendimento e alocação de recursos, buscando a redução das penalidades aplicadas pela ANEEL referente a esses indicadores. Para tal, durante o estudo deste trabalho, foram analisadas técnicas para predição de séries temporais e propostas possíveis soluções.

Primeiramente foi elaborada a etapa de pré-processamento dos dados. Essa etapa consistiu na normalização da série, seguido pela decomposição em frequência e, por último, divisão em uma janela móvel a ser utilizada como entrada da rede. Para a etapa de decomposição, os métodos ficaram limitados ao MODWT e CEEMDAN, sendo também utilizado um pré-processamento sem decomposição.

Em relação às redes neurais, foram analisadas as que apresentaram um bom desempenho com séries temporais, limitando o escopo às redes CNN+LSTM e LSTM, e a rede TDNN como rede rasa.

A aplicação em conjunto das técnicas de pré-processamento, juntamente com as redes selecionadas, trouxeram resultados significativos para ambos os indicadores. Para o FEC, a rede híbrida CNN+LSTM com a decomposição MODWT teve o melhor desempenho, com $RMSE = 6,91 \times 10^{-5}$, $MAPE = 0,324$ e $MdRAE = 0,224$. Já para o DEC, a rede LSTM com decomposição MODWT obteve o melhor desempenho em duas das três métricas, $RMSE = 1,49 \times 10^{-4}$ e $MAPE = 0,587$.

É possível concluir que a aplicação de técnicas de decomposição apresentou um ganho de desempenho significativo tendo em vista os resultados obtidos quando estas técnicas não foram aplicadas, que resultaram com que as redes não aprendessem nas condições apresentadas.

Também foi observado neste trabalho um melhor desempenho da rede CNN+LSTM para predição de dados com grande variações de valores, como os indicadores de continuidade diário utilizados. Um motivo para este melhor resultado pode ter sido porque a camada LSTM aprendeu com as características extraídas pela camada convolucional, que conseguiu suavizar o sinal e extrair informações mais relevantes.

Este trabalho focou na aplicação de técnicas de decomposição de frequência e na aplicação da rede híbrida CNN+LSTM, permitindo que outros trabalhos explorem demais técnicas de pré-processamento ou tipos de redes neurais.

Durante a execução deste trabalho também foi avaliada a aplicação de dados climáticos, disponibilizados pelo Instituto Nacional de Meteorologia (INMET), do município pertencente ao conjunto elétrico como dados exógenos em uma rede neural NARX (*Nonlinear autoregressive exogenous*). Os dados obtidos tiveram que passar por uma etapa de pré-processamento, removendo os dados naturalmente não realizáveis ou fisicamente impossíveis, como por exemplo temperaturas negativas e pressões muito maiores que 1 atm. Tais fatores, juntamente com a presença de dados faltantes, levaram a aplicação de técnicas de preenchimento dos valores incoerentes e dos faltantes. Todavia, apesar dos métodos aplicados, os resultados obtidos não foram satisfatórios dado que não aumentaram de maneira significativa o desempenho da rede sem estes dados e não apresentaram correlação relevante.

Para um trabalho futuro é possível utilizar a segmentação por tipo de falha, conforme Magalhães (2017), criando um modelo com entradas exógenas que podem facilitar o aprendizado da rede, semelhante a uma decomposição de frequência.

Em complemento, outras técnicas de tratamento de dados podem ser empregadas, buscando reduzir a retirada dos *outliers*, por exemplo, a suavização através da média móvel exponencial, que atribui um peso aos valores que decaem conforme se afastam do ponto em questão (KALEKAR et al., 2004).

Outra possibilidade, além de treinar individualmente redes do mesmo tipo para cada subsérie, é analisar o ganho frente a aplicação de redes diferentes de acordo com a componente de frequência utilizada. Aplicando assim uma rede com melhor desempenho para séries lentas em uma subsérie menor e outra para subséries com componentes de maior frequência.

Referências

- ABDELJABER, O. et al. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. Journal of Sound and Vibration, Elsevier, v. 388, p. 154–170, 2017. Citado na página 41.
- ANEEL. Compensação pela Transgressão dos Limites de Continuidade. [S.l.]: Agência Nacional de Energia Elétrica, 2017. <<http://www.ANEEL.gov.br/indicadores-de-compensacao-de-continuidade>>. Accessed: 2018-04-19. Citado na página 14.
- ANEEL. Indicadores de Continuidade por Conjunto. [S.l.]: Agência Nacional de Energia Elétrica, 2017. <<http://www.aneel.gov.br/indicadores-coletivos-de-continuidade>>. Accessed: 2018-10-25. Citado na página 49.
- ANEEL. Relatório de Análise de Impacto Regulatório - Avaliar Aprimoramentos na Regulação de Continuidade do Fornecimento de Energia Elétrica. [S.l.]: Agência Nacional de Energia Elétrica, 2018. <<https://antigo.aneel.gov.br/web/guest/impacto-regulatorio>>. Accessed: 2021-06-19. Citado na página 23.
- ANEEL. Módulo 1. [S.l.]: Agência Nacional de Energia Elétrica, 2021. <<https://www.gov.br/aneel/pt-br/centrais-de-conteudos/procedimentos-regulatorios/prodist>>. Accessed: 2021-06-19. Citado na página 23.
- ANEEL. Módulo 8. [S.l.]: Agência Nacional de Energia Elétrica, 2021. <<https://www.gov.br/aneel/pt-br/centrais-de-conteudos/procedimentos-regulatorios/prodist>>. Accessed: 2021-06-19. Citado 4 vezes nas páginas 14, 21, 22 e 23.
- ANEEL. Compensação pela Transgressão dos Limites de Continuidade. [S.l.]: Agência Nacional de Energia Elétrica, 2022. <<https://antigo.aneel.gov.br/web/guest/indicadores-de-compensacao-de-continuidade>>. Accessed: 2021-06-19. Citado na página 25.
- ANEEL. Histórico da Governança. 2022. <<https://www.gov.br/aneel/pt-br/aceso-a-informacao/institucional/historico>>. Acessado em 01/07/2022. Citado na página 18.
- ANEEL. Indicadores Coletivos de Continuidade. [S.l.]: Agência Nacional de Energia Elétrica, 2022. <<https://antigo.aneel.gov.br/indicadores-coletivos-de-continuidade>>. Accessed: 2021-06-19. Citado na página 26.
- BISHOP, C. M. Neural networks and their applications. Review of scientific instruments, American Institute of Physics, v. 65, n. 6, p. 1803–1832, 1994. Citado na página 30.
- BISHOP, C. M. Pattern recognition and machine learning. [S.l.]: springer, 2006. Citado na página 29.
- BROCKWELL, P. J.; DAVIS, R. A. Introduction to time series and forecasting. [S.l.]: Springer, 2002. Citado na página 42.

- CABRAL, J. de A.; LEGEY, L. F. L.; CABRAL, M. V. de F. Electricity consumption forecasting in brazil: A spatial econometrics approach. Energy, Elsevier, v. 126, p. 124–131, 2017. Citado 2 vezes nas páginas 19 e 20.
- CHOLLET, F. et al. Keras. GitHub, 2015. Disponível em: <<https://github.com/fchollet/keras>>. Citado na página 59.
- CORNISH, C. R.; BRETHERTON, C. S.; PERCIVAL, D. B. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. Boundary-Layer Meteorology, Springer, v. 119, n. 2, p. 339–374, 2006. Citado na página 45.
- CUI, Z.; KE, R.; WANG, Y. Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. arXiv preprint arXiv:1801.02143, 2018. Citado 2 vezes nas páginas 37 e 38.
- DAS, A. B.; BHUIYAN, M. I. H. Discrimination of focal and non-focal eeg signals using entropy-based features in eemd and ceemdan domains. In: IEEE. 2016 9th International Conference on Electrical and Computer Engineering (ICECE). [S.l.], 2016. p. 435–438. Citado na página 47.
- DELÉCHELLE, E.; LEMOINE, J.; NIANG, O. Empirical mode decomposition: an analytical approach for sifting process. IEEE Signal Processing Letters, IEEE, v. 12, n. 11, p. 764–767, 2005. Citado na página 47.
- DEPUTADOS, C. d. DECRETO Nº 24.643, DE 10 DE JULHO DE 1934. 1934. <<https://www2.camara.leg.br/legin/fed/decret/1930-1939/decreto-24643-10-julho-1934-498122-publicacaooriginal-1-pe.html>>. Acessado em 01/07/2022. Citado na página 18.
- DIAS, E. B. Avaliação de indicadores de continuidade e seu impacto no planejamento de sistemas de distribuição. Tese (Doutorado) — Universidade de São Paulo, 2002. Citado na página 14.
- DINIZ, A. P. M. Modelos de Previsão do Conteúdo de Silício no Ferro-Gusa Usando Redes Neurais Artificiais. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, Vitória, 2018. Citado 3 vezes nas páginas 35, 43 e 44.
- DOUMAS, L. A.; PUEBLA, G.; MARTIN, A. E. Human-like generalization in a machine through predicate learning. arXiv preprint arXiv:1806.01709, 2018. Citado na página 28.
- ESMAEILI, A.; MARVASTI, F. A novel approach to quantized matrix completion using huber loss measure. IEEE Signal Processing Letters, IEEE, v. 26, n. 2, p. 337–341, 2019. Citado na página 57.
- FERGUSON, R.; WILKINSON, W.; HILL, R. Electricity use and economic development. Energy policy, Elsevier, v. 28, n. 13, p. 923–934, 2000. Citado na página 14.
- FRITSCH, F. Pchip, piecewise cubic hermite data interpolation. 1985. Citado na página 51.
- FRITSCH, F. N.; CARLSON, R. E. Monotone piecewise cubic interpolation. SIAM Journal on Numerical Analysis, SIAM, v. 17, n. 2, p. 238–246, 1980. Citado na página 51.

FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. [S.l.]: Springer, 1982. p. 267–285. Citado na página 30.

GOLDENBERG, J.; PRADO, L. T. S. Reforma e crise do setor elétrico no período fhc. Tempo social, SciELO Brasil, v. 15, p. 219–235, 2003. Citado na página 17.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado 5 vezes nas páginas 29, 35, 36, 37 e 39.

GREENGARD, S. Gpus reshape computing. Communications of the ACM, ACM New York, NY, USA, v. 59, n. 9, p. 14–16, 2016. Citado na página 30.

HADHAMI, I.; BOUZID, A. Speech denoising based on empirical mode decomposition and improved thresholding. In: SPRINGER. International Conference on Nonlinear Speech Processing. [S.l.], 2013. p. 200–207. Citado na página 45.

HADIYOSO, S.; DEWI, E.; WIJAYANTO, I. Comparison of emd, vmd and eemd methods in respiration wave extraction based on ppg waves. In: IOP PUBLISHING. Journal of Physics: Conference Series. [S.l.], 2020. v. 1577, n. 1, p. 012040. Citado na página 47.

HARRIS, C. R. et al. Array programming with NumPy. Nature, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <<https://doi.org/10.1038/s41586-020-2649-2>>. Citado na página 59.

HAYKIN, S. S. et al. Neural networks and learning machines. [S.l.]: Pearson Upper Saddle River, 2009. v. 3. Citado 6 vezes nas páginas 29, 30, 31, 32, 33 e 34.

HEWAMALAGE, H.; BERGMEIR, C.; BANDARA, K. Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, Elsevier, v. 37, n. 1, p. 388–427, 2021. Citado na página 42.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. Neural computation, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 37.

HU, Y. H.; HWANG, J.-N. Handbook of neural network signal processing. [S.l.]: Acoustical Society of America, 2002. Citado na página 34.

HUANG, N. E. et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, The Royal Society, v. 454, n. 1971, p. 903–995, 1998. Citado na página 45.

HUNTER, J. D. Matplotlib: A 2d graphics environment. Computing in Science & Engineering, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007. Citado na página 59.

INCE, T. et al. Real-time motor fault detection by 1-d convolutional neural networks. IEEE Transactions on Industrial Electronics, IEEE, v. 63, n. 11, p. 7067–7075, 2016. Citado na página 41.

- JANA, R. K.; GHOSH, I.; SANYAL, M. K. A granular deep learning approach for predicting energy consumption. Applied Soft Computing, Elsevier, v. 89, p. 106091, 2020. Citado na página 27.
- JHA, G. K.; SINHA, K. Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in india. Neural Computing and Applications, Springer, v. 24, n. 3, p. 563–571, 2014. Citado na página 34.
- JUN-CHANG, Z.; LI, Z. A speech denoising method based on improved emd. In: IEEE. 2011 International Conference on Multimedia and Signal Processing. [S.l.], 2011. v. 2, p. 305–309. Citado na página 45.
- JUNIOR, J. A. A. Modelo para predição de indicadores de continuidade em um sistema de distribuição de energia elétrica, uma aplicação à gestão de manutenção com a perspectiva do uso da termografia. Escola Politécnica, 2016. Citado na página 27.
- KALEKAR, P. S. et al. Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi school of information Technology, v. 4329008, n. 13, p. 1–13, 2004. Citado na página 76.
- KHAN, S. et al. A guide to convolutional neural networks for computer vision. Synthesis lectures on computer vision, Morgan & Claypool Publishers, v. 8, n. 1, p. 1–207, 2018. Citado na página 39.
- KIRANYAZ, S.; INCE, T.; GABBOUJ, M. Real-time patient-specific ecg classification by 1-d convolutional neural networks. IEEE Transactions on Biomedical Engineering, IEEE, v. 63, n. 3, p. 664–675, 2015. Citado na página 41.
- KOUZIOKAS, G. N. Unemployment prediction in uk by using a feedforward multilayer perceptron. In: Operational Research in the Digital Era–ICT Challenges. [S.l.]: Springer, 2019. p. 65–74. Citado na página 42.
- LACOMMARE, K. H.; ETO, J. H. Cost of power interruptions to electricity consumers in the united states (us). Energy, Elsevier, v. 31, n. 12, p. 1845–1855, 2006. Citado na página 15.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. nature, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado na página 36.
- LIU, C.-L. A tutorial of the wavelet transform. NTUEE, Taiwan, 2010. Citado 2 vezes nas páginas 43 e 44.
- MAGALHÃES, E. F. A. Modelagem e simulação de indicadores de continuidade: ferramenta auxiliar para a manutenção em redes de distribuição de energia elétrica. Escola Politécnica, 2017. Citado 3 vezes nas páginas 26, 27 e 76.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 28.
- MITCHELL, T. M. Machine Learning. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997. Citado na página 28.

- NAQA, I. E.; MURPHY, M. J. What is machine learning? In: machine learning in radiation oncology. [S.l.]: Springer, 2015. p. 3–11. Citado na página 28.
- NASON, G. P.; SACHS, R. v. Wavelets in time-series analysis. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, The Royal Society, v. 357, n. 1760, p. 2511–2526, 1999. Citado na página 45.
- O'MALLEY, T. et al. KerasTuner. 2019. <<https://github.com/keras-team/keras-tuner>>. Citado na página 57.
- OPOVO. os maiores apagões da história do Brasil. 2018. <<https://www.opovo.com.br/noticias/brasil/2018/03/confira-os-maiores-apagoes-da-historia-do-brasil.html>>. Acessado em 01/07/2022. Citado na página 19.
- PERCIVAL, D. B.; WALDEN, A. T. The maximal overlap discrete wavelet transform. In: _____. Wavelet Methods for Time Series Analysis. [S.l.]: Cambridge University Press, 2000. (Cambridge Series in Statistical and Probabilistic Mathematics), p. 159–205. Citado na página 44.
- PRONI, M. W.; LYRIO, P. M. A privatização do setor elétrico e seus impactos sobre o trabalho. Revista Gestão Industrial, v. 1, n. 1, 2005. Citado na página 17.
- REZAEI, H.; FAALJOU, H.; MANSOURFAR, G. Stock price prediction using deep learning and frequency decomposition. Expert Systems with Applications, Elsevier, v. 169, p. 114332, 2021. Citado 2 vezes nas páginas 53 e 55.
- ROSSUM, G. V.; DRAKE, F. L. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009. ISBN 1441412697. Citado na página 59.
- SAHAI, A. K. et al. Arima modelling & forecasting of covid-19 in top five affected countries. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, Elsevier, v. 14, n. 5, p. 1419–1427, 2020. Citado na página 42.
- SHARMA, S. Activation Functions: Neural Networks. [S.l.]: Towards Data Science, 2017. <<https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>>. Accessed: 2018-10-25. Citado 2 vezes nas páginas 32 e 33.
- SIMAB, M.; HAGHIFAM, M. R. Quality performance based regulation through designing reward and penalty scheme for electric distribution companies. International Journal of Electrical Power & Energy Systems, Elsevier, v. 43, n. 1, p. 539–545, 2012. Citado na página 14.
- SNOEK, J.; LAROCHELLE, H.; ADAMS, R. P. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems, v. 25, 2012. Citado na página 57.
- SPERANDIO, M. et al. Classificação de conjuntos consumidores de energia elétrica via mapas auto-organizáveis e estatística multivariada. Florianópolis, SC, 2004. Citado na página 14.
- SQALLI, M. Traffic signs classification with Deep Learning. [S.l.]: Hacker Noon, 2016. <<https://hackernoon.com/traffic-signs-classification-with-deep-learning-b0cb03e23efb>>. Accessed: 2018-11-07. Citado na página 36.

- TEJA, K.; TIWARI, R.; MOHANTY, S. Adaptive denoising of ecg using emd, eemd and ceemdan signal processing techniques. In: IOP PUBLISHING. Journal of Physics: Conference Series. [S.l.], 2020. v. 1706, n. 1, p. 012077. Citado na página [47](#).
- TORRES, M. E. et al. A complete ensemble empirical mode decomposition with adaptive noise. In: IEEE. 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). [S.l.], 2011. p. 4144–4147. Citado na página [47](#).
- TRONCHONI, A. B. et al. Identificação de causas de desligamentos não programados em redes de distribuição. Pontifícia Universidade Católica do Rio Grande do Sul, 2008. Citado na página [15](#).
- WU, Z.; HUANG, N. E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in adaptive data analysis, World Scientific, v. 1, n. 01, p. 1–41, 2009. Citado na página [47](#).
- ZHANG, C. et al. On electricity consumption and economic growth in china. Renewable and Sustainable Energy Reviews, Elsevier, v. 76, p. 353–368, 2017. Citado na página [14](#).