



UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

PREDIÇÃO DE DEMANDA DE ENERGIA ELÉTRICA PARA O DIA SEGUINTE
APLICANDO REDES NEURAS CONVOLUTIVAS E REDES RECORRENTES

NEILSON LUNIERE VILAÇA

Manaus – Amazonas

2022

NEILSON LUNIERE VILACA

PREDIÇÃO DE DEMANDA DE ENERGIA ELÉTRICA PARA O DIA SEGUINTE
APLICANDO REDES NEURASIS CONVOLUTIVAS E REDES RECORRENTES

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Orientadora: Prof^a. Dr^a. Marly Guimarães Fernandes Costa
Coorientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

V695p Vilaça, Neilson Luniere
Predição de demanda de energia elétrica para o dia seguinte
aplicando redes neurais convolutivas e redes recorrentes / Neilson
Luniere Vilaça . 2022
86 f.: il. color; 31 cm.

Orientadora: Marly Guimarães Fernandes Costa
Coorientador: Cícero Ferreira Fernandes Costa Filho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade
Federal do Amazonas.

1. Predição de demanda. 2. Redes neurais convolutivas. 3. Redes
neurais recorrentes. 4. Redes neurais híbridas. 5. Covid-19. I.
Costa, Marly Guimarães Fernandes. II. Universidade Federal do
Amazonas III. Título

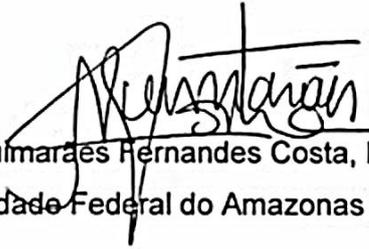
NEILSON LUNIERE VILAÇA

**PREDIÇÃO DE DEMANDA DE ENERGIA ELÉTRICA PARA O DIA
SEGUINTE APLICANDO REDES NEURAIAS CONVOLUTIVAS E
REDES RECORRENTES**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 14 de dezembro de 2022.

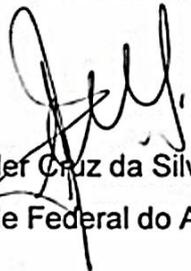
BANCA EXAMINADORA



Profª Dra. Marly Guimarães Fernandes Costa, Presidente
Universidade Federal do Amazonas



Prof. Dr. Ozenir Farah da Rocha Dias, Membro
Universidade Federal do Amazonas



Prof. Dr. Helder Cruz da Silva, Membro
Universidade Federal do Amazonas

AGRADECIMENTOS

Ao Senhor Deus, por tantas bênçãos derramadas em minha vida. Por ser o criador de todas as coisas. Por ser meu guia, meu auxiliador, meu consolador, meu Senhor e meu Salvador. Graças a Ele que concluo mais essa etapa na minha carreira.

A minha família, por tanto amor, carinho, dedicação e apoio. A mamãe e papai, por sempre estarem comigo e por serem pais maravilhosos. A minha irmã, pela sua companhia e ajuda em todos os momentos.

Aos meus orientadores, Prof. Dr. Cícero Ferreira Fernandes Costa Filho e Prof^ª. Dr^ª. Marly Guimarães Fernandes Costa, por toda a contribuição e orientação neste trabalho.

Aos amigos Gustavo Aquino, Lennon Nascimento, Ruan Teixeira, Rubens Fernandes, Samuel Rego e a todos amigos que de forma direta ou indireta me ajudaram e contribuíram para a minha formação.

A todos os colaboradores do instituto Hub Tecnologia e Inovação, pelo apoio e companheirismo durante essa jornada.

Pois qual de vós, pretendendo construir uma torre, não se assenta primeiro para calcular a despesa e verificar se tem os meios para a concluir?

(Bíblia Sagrada, Lucas 14:28)

RESUMO

A predição da demanda de energia em cenários adversos, como a pandemia da COVID-19, é fundamental para assegurar o fornecimento de eletricidade e o funcionamento de serviços básicos de uma região metropolitana. Um outro fator importante é a autossustentabilidade do sistema energético, frente a índices históricos de emissão de CO₂ ocasionado pelo alto consumo de energia. Na literatura, é possível observar uma tendência na utilização de inteligência artificial para prever demanda de energia e mitigar esses efeitos. Contudo, foram desenvolvidos poucos trabalhos envolvendo cenários anômalos. Nessa dissertação, é proposto um modelo de aprendizado profundo para prever a demanda para o próximo dia, utilizando a base de dados da competição “*IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*”. Três arquiteturas de redes neurais profundas são propostas, sendo uma rede convolucional, uma rede recorrente e uma rede híbrida (convolucional-recorrente). O melhor modelo apresentado é caracterizado pela extração de características espaço-temporais dos dados de entrada, por meio da rede híbrida. É realizado uma análise preliminar dos dados de entrada, com a exclusão de variáveis anômalas e a aplicação de uma janela deslizante, com intervalo de 24 horas, que define a quantidade de horas necessárias para prever a demanda da próxima hora. Também foi implementada a normalização dos dados de entrada, com uma ponderação com fator de 10 para a variável demanda. O desempenho dos modelos propostos foi comparado com os modelos desenvolvidos na competição, por meio de uma análise *benchmark*. A arquitetura híbrida proposta neste trabalho apresentou um erro absoluto médio de 2361,84 kW (78,22% inferior ao modelo de melhor desempenho da competição) e uma similaridade da curva de demanda real e a curva de demanda predita também superior aos referidos modelos. Tais fatos evidenciam a eficiência da aplicação de redes profundas, frente aos métodos clássicos aplicados por outros autores.

Palavras chave: Predição de demanda, redes neurais convolutivas, redes neurais recorrentes, redes neurais híbridas, COVID-19.

ABSTRACT

Predicting energy demand in adverse scenarios, such as the COVID-19 pandemic, is critical to ensure the electricity supply and operation of basic services in a metropolitan region. Another important factor is the self-sustainability of the energy system, compared to historical levels of CO₂ emissions caused by high energy consumption. In the literature, it is possible to observe a trend in the use of artificial intelligence to predict energy demand and mitigate these effects. However, few works involving anomalous scenarios have been developed. In this dissertation, a deep learning model is proposed to predict the demand for the next day using the “IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm” database. Three deep neural network architectures are proposed, being a convolutional network, a recurrent network, and a hybrid (convolutional-recurrent) network. The best model presented is characterized by extracting spatio-temporal features from the input data by means of the hybrid network. A preliminary analysis of the input data is performed, with the exclusion of anomalous variables and the application of a sliding window, with a 24-hour interval, which defines the number of hours needed to predict the demand of the next hour. The normalization of the input data was also implemented, with a weighting with a factor of 10 for the demand variable. The performance of the proposed models was compared with the models developed in the competition, by means of a benchmark analysis. The hybrid architecture proposed in this work presented an average absolute error of 2361.84 kW (78.22% lower than the model with the best performance in the competition) and a similarity between the real demand curve and the predicted demand curve, also superior to the referred models. Such facts demonstrate the efficiency of the application of deep networks, compared to the classical methods applied by other authors.

Keywords: Load forecasting, convolutional neural network, recurrent neural network, hybrid neural network, COVID-19.

LISTA DE FIGURAS

Figura 1 - Diagrama unifilar do sistema elétrico de potência. Fonte: (KAGAN; OLIVEIRA; ROBBA, 2005).	37
Figura 2 - Curva de demanda do Sistema Interligado Nacional. Fonte: Operador Nacional do Sistema Elétrico (2022).	38
Figura 3 - Variação percentual do consumo setorial de eletricidade. Fonte: Balanço Energético Nacional (2021).	40
Figura 4 - Subconjuntos de estudos da Inteligência Artificial. Fonte: Adaptado de Alzubaidi et al. (2021).	41
Figura 5 - Diferença entre o método clássico e o aprendizado profundo. Fonte: Adaptado de Alzubaidi et al. (2021).	42
Figura 6 - Estrutura do neurônio: (a) neurônio biológico; (b) neurônio artificial. Fonte: Adaptado de Neapolitan e Jiang, 2018.	42
Figura 7 - Principais funções de ativação aplicadas nas redes neurais. Fonte: (AGGARWAL, 2018a).	43
Figura 8 - Rede tipo perceptron multicamadas. Fonte: (HAYKIN, 2009).	44
Figura 9 - Arquitetura da rede LeNet-5. Fonte: (LECUN et al., 1998).	45
Figura 10 – Abstração do processamento das redes convolutivas. (a) Cérebro humano. (b) Processamento de imagem pela rede neural convolutiva. Fonte:(AHMAD; JOE; HAN, 2018; XU; VAZIRI-PASHKAM, 2021).	45
Figura 11 - Desempenho das redes de aprendizado profundo frente aos métodos clássicos de aprendizado de máquina. Fonte: Adaptado de Aggarwal, 2018a.	46
Figura 12 - Aplicação da operação de convolução bidimensional. Fonte: Adaptado de Raschka e Mirjalili, 2019.	47
Figura 13 - Função max pooling 3x3 com stride de 1 e 2. Fonte: Adaptado de Aggarwal, 2018a.	49
Figura 14 - Processo de dropout em uma rede neural: (a) rede inteiramente conectada; (b) neurônios e conexão em vermelho que serão desativados; (c) neurônios selecionados foram desativados. Fonte: (THEODORIDIS, 2020).	50
Figura 15 - Arquitetura de uma rede recorrente.	53
Figura 16 - Formas de sequenciamento dos dados em uma RNN: (a) Many-to-one; (b) One-to-Many; (c) Many-to-many sincronizado; (d) Many-to-many não sincronizado.	53
Figura 17 - Arquitetura da rede LSTM.	54

Figura 18 - Comparação do método SGD convencional (a) e SGDM (b). Fonte: (MARSLAND, 2015).....	57
Figura 19 - Etapas de desenvolvimento do sistema proposto.....	60
Figura 20 - Curva de demanda de eletricidade disponibilizada. Fonte: (FARROKHABADI et al., 2022)	62
Figura 21 - Curvas de pressão disponibilizadas. (a) Pressão do ar real. (b) Pressão do ar pela previsão meteorológica. Fonte: (FARROKHABADI et al., 2022).	63
Figura 22 - Curvas de cobertura de nuvens do banco de dados. (a) Cobertura de nuvens predita. (b) Cobertura de nuvens real. Fonte: (FARROKHABADI et al., 2022).	64
Figura 23 - Curvas de temperatura do banco de dados. (a) Temperatura predita. (b) Temperatura medida. Fonte: (FARROKHABADI et al., 2022).	65
Figura 24 - Curvas de direção do vento do banco de dados. (a) Direção do vento prevista. (b) Direção do vento medida. Fonte: (FARROKHABADI et al., 2022).	66
Figura 25 - Curvas de velocidade do vento do banco de dados. (a) Velocidade do vento prevista. (b) Velocidade do vento real. Fonte: (FARROKHABADI et al., 2022).	67
Figura 26 - Método de janela deslizante aplicado nos dados de entrada.....	68
Figura 27 - Primeira arquitetura de rede neural profunda proposta.	69
Figura 28 - Segunda arquitetura de rede neural profunda proposta.	70
Figura 29 - Terceira arquitetura de rede neural profunda proposta.	71
Figura 30 - Parâmetros variados durante o treinamento das redes de aprendizado profundo. Os blocos verdes foram os métodos e valores adotados.	72
Figura 31 - Sistemática de funcionamento da função auxiliar implementada, que supervisiona o desempenho da arquitetura e retorna a rede com o melhor desempenho.	73
Figura 32 - Gráfico de convergência da arquitetura 1 proposta.	74
Figura 33 - Gráfico de convergência da arquitetura 2 proposta.	75
Figura 34 - Gráfico de convergência da arquitetura 3 proposta.	75
Figura 35 - Curva de predição de demanda da arquitetura 3. A linha azul é o valor da demanda real e a linha vermelha tracejada é o valor da demanda predita.	77

LISTA DE TABELAS

Tabela 1 - Variáveis disponibilizadas pela competição.....	61
Tabela 2 - Desempenho dos modelos da competição e do modelo proposto.....	76
Tabela 3 - Resumo dos valores reais do banco de dados e valores preditos pela arquitetura 3.	76

LISTA DE QUADROS

Quadro 1 - Resumo dos trabalhos relacionados a predição de demanda em regiões metropolitanas.	28
Quadro 2 - Resumo dos trabalhos relacionados a predição de demanda no cenário da COVID-19.	31
Quadro 3 - Resumo dos trabalhos relacionados a competição IEEE.	32

LISTA DE SIGLAS

AdaGrad	<i>Adaptive Gradient Algorithm</i>
ADALINE	<i>Adaptive Linear Element</i>
ADAM	<i>Adaptive Moment Estimation</i>
AM	Aprendizado de Máquina
ANEEL	Agência Nacional de Energia Elétrica
AP	Aprendizado Profundo
BEN	Balanço Energético Nacional
BPTT	<i>Backpropagation Through Time</i>
CNN	<i>Convolutional Neural Network</i>
CPU	<i>Central Processing Unit</i>
CO ₂	Dióxido de Carbono
DBN	<i>Deep Belief Network</i>
EMD	<i>Empirical Mode Decomposition</i>
EPE	Empresa de Pesquisa Energética
GPU	<i>Graphics Processing Unit</i>
GRU	<i>Gated Recurrent Unit</i>
GWO	<i>Grey Wolf Optimize</i>
IA	Inteligência Artificial
IEA	<i>International Energy Agency</i>
IESO	<i>Independent Electricity System Operator</i>
IEEE	<i>Institute of Electrical and Electronic Engineers</i>
LOC	<i>Lateral Occipital Cortex</i>
LSTM	<i>Long Short-Term Memory</i>
LTLF	<i>Long-Term Load Forecasting</i>
MAE	<i>Mean Absolut Error</i>
MAPD	<i>Mean Absolute Percentage Deviation</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MLP	<i>Multilayer Perceptron</i>

MSE	<i>Mean Squared Error</i>
MSLE	<i>Mean Squared Logarithmic Error</i>
MTLF	<i>Medium-Term Load Forecasting</i>
NARX	<i>Nonlinear Autoregressive Exogenous</i>
ONS	Operador Nacional do Sistema Elétrico
PCA	<i>Principal Component Analysis</i>
R ²	R-quadrado
RAM	<i>Random Access Memory</i>
RGB	<i>Red-Blue-Green</i>
RF	<i>Random Forest</i>
RNA	Rede Neural Artificial
RNC	Redes Neurais Convolutivas
RNN	<i>Recurrent Neural Network</i>
RMSE	<i>Root Mean Squared Error</i>
RMSProp	<i>Root Mean Square Propagation</i>
SGD	<i>Stochastic Gradient Descent</i>
SGDM	<i>Stochastic Gradient Descent With Momentum</i>
SIN	Sistema Interligado Nacional
STLF	<i>Short-Term Load Forecasting</i>
SVM	<i>Support Vector Machine</i>
SVR	<i>Support Vector Regression</i>
VOT	<i>Ventral Occipitotemporal Cortex</i>
XGBoost	<i>Extreme Gradient Boosting</i>

Sumário

1	INTRODUÇÃO	17
1.1	Objetivo Geral.....	19
1.2	Objetivos Específicos	19
1.3	Organização do Trabalho	20
2	REVISÃO DA LITERATURA.....	21
2.1	Análise dos trabalhos - previsão de demanda em regiões metropolitanas.....	21
2.2	Análise dos trabalhos - previsão de demanda durante a COVID-19	25
2.3	Análise dos trabalhos – IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm	26
2.4	Discussão dos trabalhos	34
3	REFERENCIAL TEÓRICO	37
3.1	Sistema elétrico de potência	37
3.2	Aprendizagem profunda	40
3.3	Redes Neurais Profundas - Convolutivas	44
3.3.1	Camada de convolução	46
3.3.2	Camada de ativação	47
3.3.3	Camada de <i>pooling</i>	48
3.3.4	Camada de <i>Dropout</i>	49
3.3.5	Treinamento de uma RNC	50
3.4	Redes Neurais Recorrentes	52
3.4.1	Redes <i>long short-term memory</i> (LSTM).....	54
3.4.2	Treinamento das redes LSTM.....	55
3.5	Métodos de Otimização	56
3.5.1	<i>Stochastic Gradient Descent</i> (SGD)	56
3.5.2	<i>Stochastic Gradient Descent With Momentum</i> (SGDM).....	57
3.5.3	<i>Adaptive Moment Estimation</i> (ADAM).....	58

4	MATERIAIS E MÉTODOS	60
4.1	Introdução	60
4.2	Modelagem do banco de dados	61
4.3	Definição das arquiteturas	69
4.3.1	Arquitetura 1 (CNN)	69
4.3.2	Arquitetura 2 (LSTM)	70
4.3.3	Arquitetura 3 (CNN + LSTM)	70
4.4	Definição do ambiente de simulação	71
4.5	Experimentos	72
4.6	Métricas de desempenho	73
5	RESULTADOS E DISCUSSÕES	74
5.1	Resultados da arquitetura 1	74
5.2	Resultados da arquitetura 2	74
5.3	Resultados da arquitetura 3	75
5.4	Discussão dos resultados	77
6	CONCLUSÕES	80
7	REFERÊNCIAS	82

1 INTRODUÇÃO

A preocupação global relacionada a sistemas energéticos ecologicamente corretos, econômicos e confiáveis tem sido o foco de políticas atuais, frente a preocupação de países na redução da emissão de CO₂ (dióxido de carbono). O relatório anual da IEA (*International Energy Agency*) aponta que em 2021 as emissões globais de CO₂, decorrentes do uso de combustíveis fósseis na geração de eletricidade e em processos industriais, atingiu recorde histórico de 36,3 Gigatoneladas (ENERGY AGENCY, 2021).

No entanto, em 2020, a pandemia de COVID-19 trouxe impactos globais na demanda de energia, reduzindo a emissão de CO₂ em 5,2%. A IEA enfatiza que essa rápida transição aliada a condições climáticas adversas e as condições do mercado de energia, ocasionou em uma maior queima do carvão em 2021, apesar da energia renovável registrar o maior crescimento anual (ENERGY AGENCY, 2021).

Assim, a aplicação de técnicas de aprendizado de máquina em sistemas energéticos é uma das ferramentas para o gerenciamento de recursos e planejamento estratégico em todas as suas etapas. Ações como ajuste de tarifa, avaliação de contratos, cronogramas, manutenção e qualidade de serviço podem ser aprimorados por meio da previsão da demanda elétrica (MD SALLEH; SULIMAN; JORGENSEN, 2020; NTI et al., 2020; RAZA; KHOSRAVI, 2015; VERWIEBE et al., 2021).

Diversos métodos e cenários têm sido explorados com o objetivo de generalizar modelos preditivos. Algoritmos para predição de demanda de energia elétrica, por exemplo, têm sido desenvolvidos visando determinar o valor de consumo de energia elétrica, em kW, para as próximas horas, meses ou anos. Tais períodos são definidos como previsão de carga de curto prazo (STLF) referente a predições com até 1 dia de antecedência, previsão de carga de médio prazo (MTLF) referente ao período de um 1 dia a 1 ano à frente e a previsão de carga de longo prazo (LTLF) referente a 1 a 10 anos à frente (ROMÁN-PORTABALES; LÓPEZ-NORES; PAZOS-ARIAS, 2021).

Na literatura, para previsão da demanda, observa-se que 90% dos métodos propostos utilizam técnicas de inteligência artificial, enquanto 10% utilizam métodos estatísticos tradicionais, sendo necessário mais estudos que avaliem esses dois paradigmas (NTI et al., 2020). Destaca-se também que a descarbonização do setor demanda esforços adicionais para obtenção de uma solução adequada para o problema de predição da demanda (VERWIEBE et al., 2021). Um ingrediente fundamental para o desenvolvimento de métodos robustos para previsão da demanda energética é a disponibilidade de bancos de dados que, na maioria dos

casos, não estão acessíveis à comunidade para a produção de novos experimentos. Além da disponibilidade de bancos de dados, com o objetivo de se chegar a uma avaliação imparcial do desempenho dos diferentes algoritmos, os documentos de demanda de carga devem usar um *benchmark* de referência comum.

A avaliação do desempenho de tais modelos em cenários adversos, como o caso da COVID-19, também requer estudos mais aprofundados. Por conta de períodos de *lockdown*, associados a pandemia, a curva de demanda resultante desse período foi completamente fora dos padrões dos anos anteriores, ficando abaixo da média histórica. A descontinuidade no padrão de consumo causada pela pandemia de COVID-19 engendrou estudos sobre novos métodos de previsão de demanda de energia, em adição aos métodos tradicionais já desenvolvidos (LU; MA; MA, 2021; OBST; DE VILMAREST; GOUDE, 2021; TUDOSE et al., 2021).

Nesse contexto, foi criada a competição “*IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*” que teve como objetivo promover uma análise detalhada dos impactos causados pelas medidas relacionadas a COVID-19 na demanda de energia elétrica. A competição buscou o desenvolvimento de novas estratégias de mitigação de impacto da previsão de demanda para o dia seguinte, em cenários anômalos. Para isso, foi disponibilizado um banco de dados, contendo informações de demanda de energia e variáveis climáticas de uma concessionária de energia.

Os trabalhos mais bem sucedidos nessa competição aplicaram técnicas clássicas de aprendizado de máquina e modelos estatísticos para prever a demanda, como modelos de espaço de estados e combinação de previsões por Agregação Online de Bernstein (BOA) suavizada (DE VILMAREST; GOUDE, 2022; ZIEL, 2022).

No entanto, técnicas de inteligência artificial, que incluem extração de características e dependência temporal, têm apresentado resultados promissores em trabalhos de previsão de demanda, utilizando outras bases de dados (AL-SAUDI; DEGELER; MEDEMA, 2021; LE et al., 2019).

Assim, essa dissertação se justifica devido ao impacto da previsão da demanda para a sustentabilidade de sistemas energéticos, em decorrência da alta relação entre a redução de emissões de CO₂ e o gerenciamento adequado de recursos. Ações como ajuste de tarifa, avaliação de contratos, cronogramas, manutenção e qualidade de serviço também podem ser aprimorados por meio da previsão da demanda. Esse trabalho também se justifica devido ao preenchimento de uma lacuna na literatura relacionada a previsão de demanda, em que há a presença de dados anômalos no banco de dados.

Essa dissertação tem o objetivo de contribuir com o avanço do estado da arte do tema predição de demanda de energia elétrica para o próximo dia, em um contexto pandêmico. Para isso, utiliza-se a base de dados da competição IEEE, referida anteriormente, e são desenvolvidas três arquiteturas de redes profundas, com destaque para a rede híbrida que obteve o melhor resultado. Para viabilizar o *benchmark* dos modelos desenvolvidos neste trabalho com os modelos preditivos desenvolvidos na competição, as mesmas métricas de desempenho são utilizadas. Além disso, uma análise prévia para seleção das características mais apropriadas à previsão e um tratamento adequado dos dados são realizados.

1.1 Objetivo Geral

Implementar um método baseado em redes neurais profundas para a predição da demanda de energia elétrica para o dia seguinte em condições adversas, como o cenário pandêmico da COVID-19.

1.2 Objetivos Específicos

1. Verificar o estado da arte relacionado a predição de demanda de energia em regiões metropolitanas e em cenários envolvendo COVID-19.
2. Investigar preliminarmente as variáveis de entrada, propondo métodos para seleção e reescalonamento de tais variáveis.
3. Avaliar o desempenho de arquiteturas de redes profundas, sendo uma rede convolutiva, uma rede recorrente e uma rede híbrida (convolucional-recorrente).
4. Comparar o melhor modelo de aprendizado profundo obtido com os modelos desenvolvidos na competição IEEE, empregando a mesma métrica para análise de desempenho.

1.3 Organização do Trabalho

Este trabalho está organizado conforme a divisão descrita a seguir:

- Capítulo 1: Introdução;
- Capítulo 2: Revisão da Literatura;
- Capítulo 3: Referencial Teórico;
- Capítulo 4: Materiais e Métodos;
- Capítulo 5: Resultados e Discussão;

O Capítulo 1 contextualiza a importância da previsão da demanda, relacionando-a com a preocupação global na redução dos níveis de CO₂. Apresenta também o impacto da COVID-19 na demanda de energia e os objetivos gerais e específicos deste trabalho.

O Capítulo 2 apresenta trabalhos publicados na literatura na área de previsão de demanda de energia em regiões metropolitanas. Também apresenta trabalhos que buscaram prever a demanda no cenário da COVID-19 e na competição IEEE. Três quadros comparativos são apresentados, destacando a origem do banco de dados, o pré-processamento, o método e os resultados obtidos.

No Capítulo 3, são apresentados os fundamentos teóricos necessários para o desenvolvimento deste trabalho. Os conceitos de sistemas elétricos de potência, aprendizagem profunda, redes neurais convolutivas e redes neurais recorrentes são descritos. As definições de treinamento e otimização dessas redes também são apresentadas.

No Capítulo 4 constam os materiais e os métodos utilizados neste trabalho. São apresentadas as características e a modelagem do banco de dados e as arquiteturas propostas. Também são apresentados o ambiente de simulação, os experimentos realizados e descrição da métrica para análise de desempenho.

O Capítulo 5 apresenta os resultados decorrentes do treinamento, validação e teste das arquiteturas desenvolvidas. Os resultados são discutidos e comparados com outros trabalhos relacionados.

Por fim, as conclusões são apresentadas no Capítulo 6.

2 REVISÃO DA LITERATURA

A revisão de literatura consistiu na pesquisa de artigos, teses e dissertações acerca do tema “predição de demanda de energia elétrica”, com objetivo de se obter o estado da arte. Tal pesquisa, foi realizada nas plataformas *Engineering Village*, *IEEE Xplore* e *MDPI*. As palavras-chave utilizadas para a sistematização da busca foram definidas como “*Machine learning*”, “*Electricity*”, “*Energy demand*”, “*Forecast*” e “*Prediction*”, e seus respectivos sinônimos e/ou termos similares. Esses termos foram associados com os operadores Booleanos AND e OR. Termos similares ou sinônimos agrupados com o operador OR e não similares com o operador AND, conforme exemplo a seguir: (“*Machine learning*” OR “*Neural Network*” OR “*Deep learning*”) AND (“*Electricity*” OR “*Energy demand*”) AND (“*Forecast*” OR “*Prediction*”).

Além disso, os trabalhos obtidos na busca preliminar passaram por um processo de triagem, com o objetivo de selecionar artigos mais recentes e relacionados diretamente com o tema. Ainda nessa etapa, os artigos publicados em revistas tiveram prioridade no processo de seleção, haja vista o alto nível de qualidade, confiabilidade e fator de impacto. Também foi possível obter novas referências literárias, por meio da citação de outros trabalhos nos artigos anteriormente selecionados, utilizando a técnica de mapas de citação.

Após a leitura e análise dos trabalhos selecionados, foi identificado que os principais métodos aplicados consistiam na aplicação de redes *perceptron* multicamadas, redes neurais convolutivas, redes recorrentes e técnicas estatísticas clássicas. Alguns trabalhos aplicaram a combinação de duas ou mais técnicas supracitadas, construindo modelos híbridos. Também foram analisados os artigos relacionados a competição “*IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*” que buscou promover a predição da demanda de energia elétrica para o próximo dia, no cenário pandêmico da COVID-19.

O Quadro 1 apresenta as técnicas aplicadas pelos trabalhos relacionados a predição de demanda em regiões metropolitanas. O Quadro 2 apresenta os trabalhos de predição de demanda no cenário da COVID-19. E o Quadro 3 se concentra no resumo dos trabalhos desenvolvidos na competição *IEEE*, evidenciando a complexidade da predição da demanda em cenários atípicos.

2.1 Análise dos trabalhos - predição de demanda em regiões metropolitanas

O trabalho desenvolvido por Peng *et al.* (2021) apresentou um método para predição diária da demanda total de 10 mil residências da cidade de Xangai utilizando redes *long short-term memory* (LSTM) e algoritmos de *clusterização*. Na etapa de pré-processamento, valores

ausentes foram preenchidos por meio de interpolação linear e valores discrepantes foram substituídos pela média dos valores adjacentes. Por fim, os autores ainda apresentaram um método de fluxo potencial espaço-temporal, indicando o deslocamento do consumo de eletricidade para diferentes regiões e horas do dia. Para avaliar o desempenho da rede, foram aplicados o erro quadrático médio (RMSE), o erro absoluto médio (MAE), o erro logarítmico quadrado médio (MSLE) e o R-quadrado (R^2). Os melhores resultados alcançados para essas métricas foram: 9446,15 kW, 6741,18 kW, 1,4% e 0,83 kW, respectivamente.

Zhu e colaboradores (2021) desenvolveram um método para predição de cargas em períodos de feriados, aplicando o algoritmo XGBoost (*Extreme Gradient Boosting*) em um conjunto de dados reais, de 2015 a 2018, obtidos em uma província no leste da China. Foram empregados a normalização dos dados, seleção de características e identificação dos feriados, segundo o tipo (Dia de Ano Novo, Festival Qingming, Primeiro de Maio, Festival do Barco do Dragão, Festival do Meio Outono, Dia da Primavera e Dia Nacional) e o tempo de duração (feriado de 3 dias e feriado de 7 dias). A rede com o melhor desempenho obteve um valor para o erro absoluto médio percentual (MAPE) de 2,98%, na capital da província.

Hafeez *et al.* (2020) apresentaram um modelo híbrido para prever a demanda de energia elétrica mensal, utilizando máquina Boltzmann restrita, em uma *Smart Grid* dos EUA. No pré-processamento, os autores aplicaram normalização, filtro de relevância e filtro de redundância nos dados de entrada. Os dados são compostos do histórico de quatro anos de demanda e de variáveis climáticas (temperatura, ponto de orvalho, umidade e velocidade do vento). O modelo foi avaliado por meio do desvio percentual absoluto médio (MAPD), variância e coeficiente de correlação, alcançando os resultados 1,10%, 1,03 e 0,79, respectivamente.

O artigo de Barman e Dev Choudhury (2020) aplicou um método baseado em máquina de vetores de suporte (SVM) híbrido, com otimização dos parâmetros por meio do algoritmo *grey wolf optimizer* (GWO) para a predição de demanda de eletricidade dos eventos regionais Rongali Bihu, Durga Puja e Diwali, na Índia. Foram utilizadas as variáveis de demanda e temperatura, do período de 2012 a 2017. A métrica utilizada para analisar o desempenho do modelo foi o MAPE, que obteve os valores de 1,79%, 1,72% e 1,68% para cada um dos feriados supracitados, respectivamente.

Um código esparso para o treinamento de uma rede neural para prever a demanda de energia de 5 mil residências na cidade de Chattanooga, a partir de dados de demanda, temperatura real e temperatura predita pela meteorologia foi proposto por Yu; Mirowski e Ho (2017). Foram empregados dados da semana anterior para prever a demanda total do próximo

dia. O desempenho do modelo foi avaliado utilizando-se a temperatura real ou a temperatura predita, nos conjuntos de treinamento e teste. O modelo com melhor resultado utilizou valores reais de temperatura nos dados de treinamento e teste, obtendo valores de MAE, RMSE e MAPE de 36,85 kWh, 46,76 kWh e 19,79%, respectivamente.

Ayub *et al.* (2020) propuseram um *framework* composto por uma SVM, uma rede *gated recurrent unit* (GRU) e uma rede neural convolucional (CNN) para a predição de demanda, utilizando dados da região de Nova Inglaterra, nos EUA. No pré-processamento, para calcular a importância das características, os autores empregaram um híbrido dos métodos *random forest* (RF) e *extreme gradiente boosting* (XGBoost). Foram utilizadas catorze características, dentre as quais destacamos a demanda em tempo real, temperatura do ponto de orvalho e energia em tempo real. Os autores aplicaram também o método *recursive feature elimination* (RFE) para retirar variáveis com pouca relevância para o modelo. Foram utilizados 8 anos de informações presentes no banco de dados. Com essa abordagem, os autores obtiveram os seguintes resultados: F1-Score – 95,23%, Acurácia – 96,33%, Precisão – 94,00%, Recall – 94,62%, MAPE – 6,00%, RMSE – 7,00 kW, MAE – 10,00 kW e erro quadrático médio (MSE) -13,00 kW².

Bedi e Toshniwal (2018) aplicaram uma decomposição empírica (EMD), seguido de uma rede LSTM, com o objetivo de estimar a demanda para o próximo dia, utilizando os dados da cidade de Chandigar. Primeiramente, foram aplicados uma limpeza e transformação dos dados de entrada, por meio da média dos valores vizinhos e da subtração da média pelo valor atual de cada série temporal, respectivamente. Foram utilizadas as variáveis demanda, dia da semana e estação do ano, disponíveis em um período de cinco anos. O valor MAPE obtido foi de 7,56%.

O trabalho de Senjyu *et al.* (2002) propôs uma arquitetura de rede neural rasa para prever a demanda para a próxima hora. Foram utilizados dados de temperatura e de demanda de uma concessionária do Japão, de um período de 4 anos. Para prever a demanda da próxima hora, foram utilizados dados de 30 dias anteriores. O melhor resultado alcançado para a métrica MAPE foi 1,18%.

Para predição de demanda, Jiang *et al.* (2022) utilizaram uma rede neural empilhada e um método de similaridade dos dias. O desempenho do modelo foi avaliado utilizando-se cinco banco de dados de diferentes regiões dos Estados Unidos. Após a normalização dos dados de entrada, uma arquitetura da rede *deep belief network* (DBN) é empregada. Os menores valores de RMSE e MAPE alcançados para a predição de demanda foram na primavera, 18,48 MW e 1,44%, respectivamente.

O trabalho de Jawad *et al.* (2020) propõe um modelo de predição baseado em *support vector regression* (SVR) relacionando características meteorológicas correlatas, como temperatura e ponto de orvalho. Para as simulações, os dados de eletricidade utilizados são de uma subestação de energia, na cidade de Muzaffarabad, de 1º de janeiro de 2014 a 31 de dezembro de 2015. Dados meteorológicos foram obtidos do Departamento de Meteorologia do Paquistão da mesma região. No pré-processamento, os valores ausentes são preenchidos por meio da interpolação de valores vizinhos. Os melhores valores das métricas MAPE, RMSE e MSE foram, respectivamente, iguais a 5,62%, 2746,58 kW e 7543721 kW².

Kondaiah e Saravanan (2022) definiram um método para aprimorar a predição da demanda de energia de redes *short-term load forecasting* (STLF), por meio da transformada de Wavelets. Para isso, foram utilizados dados de energia do Operador Independente do Sistema Elétrico (IESO) de Canadá, da cidade de Ontario, no período de 1 de janeiro de 2017 a 30 de setembro de 2019, sem a aplicação de dados meteorológicos. Após a agregação dos dados por estações do ano, o melhor valor obtido de MAPE, MAE e RMSE, para um dos modelos testados, foi de 0,59%, 51 MW e 139 MW, respectivamente.

O trabalho desenvolvido por Veeramsetty *et al.* (2022) aplicou a combinação de redes recorrentes e análise de componentes principais (PCA) para prever a demanda de uma subestação de 33/11 kV na Índia. Os dados compreendem o período de setembro a novembro de 2018. Os autores realizaram predições para a próxima hora e predições diárias. Para este último, os melhores valores de MSE e RMSE foram de 0,02 kW² e 0,141 kW.

O trabalho de Oliveira (2016) empregou modelos baseados em redes neurais tipo *perceptron* multicamadas e um modelo de análise por série de Fourier para prever a demanda anual do Brasil, utilizando a base de dados do Sistema Interligado Nacional (SIN), no período de 1970 a 2016, e outras 32 variáveis relacionadas ao clima, a indicadores socioeconômicos e valores tarifários. No pré-processamento, foi aplicado o algoritmo *Relief* para seleção de características mais importantes e normalização das séries de demanda entre -1 e 1. Os melhores resultados alcançados pela rede neural obtiveram o RMSE e MAPE iguais a 2663,63 MW e 2,75%.

Rosa *et al.* (2020) utilizaram a base de dados de uma concessionária de energia elétrica brasileira, no período de 2003 a 2017, para prever a demanda mensal estratificados por setores comerciais, industriais, residenciais e rurais. Uma matriz de correlação das variáveis foi aplicada no pré-processamento e um modelo de previsão linear foi desenvolvido para prever a demanda em cada setor. O valor do R-quadrado (R²) para o setor industrial foi de 0,5283 e

para os setores residencial, comercial e rural os valores foram respectivamente iguais a 0,9218, 0,9723, e 0,9118.

Ferreira (2016) utilizou a base de dados do Sistema Interligado Nacional (SIN), no período de 1981 a 2004, e variáveis socioeconômicas, como crescimento populacional, produto interno bruto e tarifa de energia, para prever a demanda anual de energia. No pré-processamento, foram empregadas normalização, análise de componentes independentes e análise de correlação das variáveis. Uma rede neural *adaptive linear element* (ADALINE) e uma função de aproximação da variável demanda foram implementadas para prever a demanda do ano seguinte, obtendo o valor de MSE igual a $2,79 \times 10^{-4}$.

2.2 Análise dos trabalhos - predição de demanda durante a COVID-19

Huang *et al.* (2021) avaliaram o caso de *lockdown* na China, predizendo a demanda mensal de eletricidade por meio do método de varredura *grey model* modificado (*Rolling MSGM*), em que os coeficientes de treinamento do modelo são controlados pela série de Maclaurin. Como dados de entrada, foram empregados valores de carga no período de 2012 a 2018. A variável predita foi o consumo para os três primeiros meses de 2019. Os valores de MAPE e RMSE obtidos nesse intervalo foram de 0,34% e 19,61 TWh, respectivamente.

Alasali *et al.* (2021) aplicaram o modelo autorregressivo ARIMAX para prever a demanda mensal, diária e a cada trinta minutos, no período da COVID-19, na Jordânia. Dados correspondentes a um período de cinco anos foram utilizados. Três regiões foram selecionadas para a predição (centro da cidade, Ashrafiah e Rashadieh). As variáveis demanda e temperatura foram aplicadas para o treinamento do modelo. O melhor desempenho para a métrica MAPE, na predição diária, foi de 4,1%.

Para predição do consumo diário de energia elétrica na Turquia, Özbay e Dalcali (2021) aplicaram dois métodos preditivos, uma rede LSTM e uma rede exógena não linear autorregressiva (NARX). Foram utilizados dados normalizados referentes às variáveis demanda, temperatura. Para o treinamento do modelo, foram utilizados tanto dados de períodos normais quanto dados referentes ao período da pandemia. Os melhores resultados foram obtidos com a rede LSTM. Para as métricas RMSE, MAE e MAPE foram obtidos os valores 8717,878 kW, 8004,736 kW e 1,1%, respectivamente.

Para prever a demanda para o próximo dia, no Norte da Itália, o trabalho de Scarabaggio *et al.* (2020) aplicou redes neurais tipo *perceptron* multicamadas. O banco de dados utilizado contém dados de demanda, temperatura, precipitação, hora e outras sete características obtidas no período de 2015 a 2020, com intervalos de amostras de 15 minutos.

Aplicando a normalização z-score, o melhor resultado, MAE e MSE de 0,53 e 1,02%, respectivamente, foi obtido com uma rede com 4 camadas escondidas.

2.3 Análise dos trabalhos – IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm

Os autores De Vilmarrest e Goude (2022) aplicaram uma correção estatística autorregressiva nas variáveis de previsão meteorológicas, com o objetivo de corrigir valores residuais de previsões meteorológicas feitas com 48 horas de antecedência. Além disso, estabeleceram modelos independentes para diferentes horas do dia. Os seguintes modelos foram utilizados: modelo autorregressivo, regressão linear, modelo aditivo Gaussiano, *random forest* e rede *perceptron* multicamadas. Por fim, foram desenvolvidos modelos lineares de espaço de estados gaussianos para sincronizar os modelos preditivos no tempo, seguido da aplicação do filtro de Kalman como estimador bayesiano do modelo de espaço de estados. Como resultado, o trabalho obteve, para a métrica MAE, um valor de 10844 MW, alcançando a primeira colocação da competição.

O trabalho de Ziel (2022) aplica uma limpeza nos dados, por meio de interpolação linear, seguido da transformação dos dados da direção do vento, em graus, para pontos cardeais norte-sul (NS) e leste-oeste (EW), por meio da análise de valores trigonométricos da direção do vento. As médias diárias das variáveis meteorológicas foram calculadas e adicionadas como novas variáveis no banco de dados. Ainda, no pré-processamento, seis processos foram aplicados e definidos, como carregamento de registro atrasado, transformada ReLU das variáveis meteorológicas, aproximação de Taylor de segunda ordem das 14 variáveis de entrada, efeitos determinísticos diários e semanais, efeitos determinísticos anuais e procedimento de ajuste de feriados. O modelo que obteve melhor desempenho foi baseado em regressão linear de alta dimensão, com parâmetros estimados através do método Lasso, combinado com o algoritmo de previsão por agregação online suavizada de Bernstein (BOA). Com um resultado para a métrica MAE de 11890 MW, o trabalho alcançou a terceira colocação na competição.

Os trabalhos dos demais competidores é sumarizado por Farrokhhabadi *et al.* (2022) e na plataforma da competição (“Day-ahead electricity demand forecasting: Post-COVID paradigm”, 2021). O time 14 aplicou uma normalização nos dados de entrada por meio do pico diário. Também combinou dias similares na etapa de treinamento, determinando o tipo do dia e o pico de temperatura. Esse time alcançou o segundo lugar na competição, com MAE de 11849 MW.

O time 36, por sua vez, aplicou a normalização nos dados de entrada, estabelecendo valores entre 0 e 1. Para prever a demanda, foi aplicado um conjunto de modelos incluindo os modelos *random forest*, *gradient boost machine* e XGBoost. Aplicaram também uma configuração recursiva de múltiplos passos usando carga defasada, hora de dia e de dia da semana. Adicionalmente, foram aplicados modelos para cada hora do dia, com previsões de temperatura e combinação média ponderada com base no desempenho recente. Essa equipe finalizou a competição em quarto lugar, com MAE igual a 12280 MW.

O time 19 aplicou redes neurais residuais (*Deep Residual Networks*), sem aplicação de pré-processamento nos dados de entrada, alcançando a quinta colocação, com MAE de 12317 MW. O time 25 apresentou no pré-processamento o seno e cosseno da hora do dia e do dia do ano. A técnica de previsão consistiu no *random forest* com aumento de gradiente e perda de quantil. Os modelos foram separados para 0–8h e 9–24h devido à disponibilidade de 24h dos dados. Os hiperparâmetros foram ajustados via *grid search*. Essa equipe alcançou o oitavo lugar com MAE igual a 14263 MW.

O time 13 não aplicou pré-processamento nos dados de entrada e o método de predição se baseou no cálculo da carga do mesmo dia nas últimas três semanas, sendo depois multiplicada por 1,02. Essa equipe alcançou o nono lugar com MAE igual a 14579 MW. Por fim, o time 6 se destacou pelo método aplicado baseado em 400 conjunto de modelos. Foi aplicado o algoritmo Facebook *prophet* para prever a temperatura, cobertura de nuvens. O XGBoost foi aplicado para prever temperatura e a demanda, tendo como dado de entrada a saída do algoritmo *prophet*. Os 400 modelos foram combinados utilizando regressão linear simples. Essa equipe terminou na vigésima segunda colocação com MAE igual a 17432 MW.

Quadro 1 - Resumo dos trabalhos relacionados a previsão de demanda em regiões metropolitanas.

Citação	Origem dos dados	Pré-processamento	Método	Resultados
(PENG et al., 2021)	10 mil residências em Xangai.	Média dos valores vizinhos.	Redes recorrente LSTM e algoritmos de <i>clusterização</i> .	RMSE: 9446,15 kW MAE: 6741,18 kW MSLE: 1,4% R²: 0,83 kW
(ZHU; GENG; WANG, 2021)	Capital de uma província da China	Normalização dos dados, seleção de características e identificação dos feriados, segundo o tipo e duração.	Algoritmo XGBoost.	MAPE: 2,98%
(HAFEEZ et al., 2020)	<i>Smart Grid</i> dos EUA.	Normalização, filtro de relevância e filtro de redundância.	Máquina Boltzmann restrita.	MAPD: 1,10% Variância: 1,03 Coefficiente de correlação: 0,79
(BARMAN; DEV CHOUDHURY, 2020)	Rongali Bihu, Durga Puja e Diwali, na Índia.	Similaridade dos dados de entrada.	SVM híbrido, com otimização dos parâmetros por meio do algoritmo <i>grey wolf optimizer</i> (GWO).	MAPE 1: 1,79% MAPE 2: 1,72% MAPE 3: 1,68%
(YU; MIROWSKI; HO, 2017)	5 mil residências de Chattanooga.	Não aplicado.	Código Esparso para treinamento de uma rede neural.	MAE: 36,85 kWh RMSE: 46,76 kWh MAPE: 19,79%
(AYUB et al., 2020)	Nova Inglaterra, EUA.	Cálculo da importância das características.	Modelo híbrido: SVM, GRU e CNN.	F1-Score: 95,23% Acurácia: 96,33% Precisão: 94,00% Recall: 94,62% MAPE: 6,00% RMSE: 7,00 kW MAE: 10,00 kW MSE: 13,00 kW ²

Citação	Banco de dados	Pré-processamento	Método	Resultados
(BEDI; TOSHNIWAL, 2018)	Cidade de Chandigar.	Limpeza, média dos valores vizinhos ausentes e subtração da média pelo valor atual de cada série temporal.	Rede recorrente LSTM.	MAPE: 7,56%.
(SENJYU et al., 2002)	Concessionária do Japão.	Não aplicado.	Rede neural rasa.	MAPE: 1,18%.
(JIANG et al., 2022)	Diferentes regiões dos Estados Unidos	Normalização dos dados de entrada.	Rede <i>deep belief network</i> e similaridade dos dias.	RMSE: 18,48 MW MAPE: 1,44%
(JAWAD et al., 2020)	Subestação de energia, de Muzaffarabad.	Preenchimento de valores ausentes por meio da interpolação de valores vizinhos.	Máquina de vetor de suporte regressora.	MAPE: 5,62%, RMSE: 2746,58 kW MSE: 7543721 kW ²
(KONDAIAH; SARAVANAN, 2022)	Ontario (IESO-Canadá).	Agregação dos dados por estações do ano	Redes <i>short-term load forecasting</i> (STLF) e transformada de Wavelets.	MAPE: 0,59% MAE: 51 MW RMSE: 139 MW
(VEERAMSETTY et al., 2022)	Subestação de 33/11 kV, na Índia.	Análise de componentes principais (PCA).	Combinação de redes recorrentes.	MSE: 0,02 kW ² RMSE: 0,141 kW

Citação	Banco de dados	Pré-processamento	Método	Resultados
(LOPES DE OLIVEIRA, 2016)	Sistema interligado nacional brasileiro	Normalização dos dados de entrada e algoritmo <i>Relief</i> .	Rede <i>perceptron</i> multicamadas.	RMSE: 2663,63 MW MAPE: 2,75%
(DE FIGUEIREDO ROSA et al., 2020)	Concessionária de energia brasileira	Análise de correlação das variáveis.	Modelo de previsão linear.	R² industrial: 0,5283 R² residencial: 0,9218 R² comercial: 0,9723 R² rural: 0,9118
(FERREIRA, 2016)	Sistema interligado nacional brasileiro	Normalização dos dados de entrada, análise de componentes independentes e análise de correlação das variáveis.	Rede ADALINE e aproximação de função.	MSE (normalizado): $2,79 \times 10^{-4}$

Quadro 2 - Resumo dos trabalhos relacionados a predição de demanda no cenário da COVID-19.

Citação	Origem dos dados	Pré-processamento	Método	Resultados
(HUANG et al., 2021)	Consumo energético da China.	Correção estatística autorregressiva nas variáveis de previsão meteorológicas.	Rolling MSGM.	MAPE: 0,34% RMSE: 19,61 TWh
(ALASALI et al., 2021)	Consumo energético da Jordânia.	Análise sazonal e de correlação das variáveis.	Modelo autorregressivo ARIMAX.	MAPE: 4,1%.
(ÖZBAY; DALCALI, 2021)	Consumo energético da Turquia.	Normalização das variáveis.	Rede recorrente LSTM.	RMSE: 8717,878 kW MAE: 8004,736 kW MAPE: 1,1%
(SCARABAGGIO et al., 2020)	Consumo do Norte da Itália.	Normalização z-score.	Rede <i>perceptron</i> multicamadas.	MAE: 0,53 MSE: 1,02%

Quadro 3 - Resumo dos trabalhos relacionados a competição IEEE.

Citação	Rank	Pré-processamento	Método	Resultado MAE (kW)
(DE VILMAREST; GOUDE, 2022)	1	Aplicação de um modelo de previsão linear sobre erros de previsão meteorológica e suavização da temperatura com um fator de 0,95 e 0,99.	Conjunto de modelos: Auto-regressão, regressão linear, modelos aditivos generalizados, <i>random forest</i> , <i>random forest</i> para GAM residuais, <i>perceptron</i> multicamadas e adaptação de filtro de Kalman.	10844
(“Day-ahead electricity demand forecasting: Post-COVID paradigm”, 2021; FARROKHABADI et al., 2022)	2	Normalização da carga por pico diário.	Combinação de dias semelhantes por meio do tipo do dia e temperatura de pico. Ajuste baseado em padrões recentes de previsão de pico de carga de dias do mesmo tipo.	11849
(ZIEL, 2022)	3	Treinamento ajustado para feriados. Definição dos componentes cardinais de direção do vento, características climáticas diárias, transformações múltiplas de ReLU. Detecção visual de outliers, preenchidos com interpolação.	Conjunto de modelos: Suavização exponencial descomposta STL, AR(p), modelos aditivos generalizados e modelos lineares de alta dimensão lasso. Definição de modelos para cada tempo de execução. Combinação pela agregação on-line suavizada de Bernstein.	11890
(“Day-ahead electricity demand forecasting: Post-COVID paradigm”, 2021; FARROKHABADI et al., 2022)	4	Normalização dos dados entre 0 e 1.	Conjunto de modelos random forest, gradient boosted machine e XGBoost. Combinação média ponderada com base no desempenho recente.	12280
	5	Não aplicado.	<i>Residual deep networks</i> .	12317
	6	Informações não disponíveis.	Informações não disponíveis.	13947
	7	Informações não disponíveis.	Informações não disponíveis.	14203
	8	Sin/Cos da hora do dia e do dia do ano, médias de carga recente, médias de erro de previsão do tempo recente.	Árvore de regressão com aumento de gradiente com perda de quantil, modelos separados para 0–8h e 9–24h devido à disponibilidade de 24h de atraso, ajuste via <i>grid search</i> .	14263

Citação	Rank	Pré-processamento	Método	Resultado MAE (KW)
("Day-ahead electricity demand forecasting: Post-COVID paradigm", 2021; FARROKHABADI et al., 2022)	9	Não aplicado.	A carga do mesmo dia nas últimas três semanas é calculada e depois multiplicada por 1,02.	14579
	22	Não aplicado.	Conjunto de modelos baseado no Facebook prophet para prever temperatura, cobertura de nuvens e carga. XGBoost para prever temperatura e carga usando previsões do prophet como entrada. Modelos e previsões combinados usando regressão linear simples.	17432

2.4 Discussão dos trabalhos

Os trabalhos apresentados evidenciam a importância de pesquisa na área, devido a necessidade de garantir o pleno funcionamento do serviço de energia elétrica que é essencial à sociedade. Nesse sentido, ações como ajuste de tarifa, avaliação de contratos, cronogramas, manutenção e qualidade de serviço podem ser aprimorados por meio da previsão de demanda. Além disso, manter o serviço de fornecimento de energia é fundamental em situações críticas.

Os Quadro 1 e 2 resumem as diversas técnicas aplicadas em trabalhos que envolvem a predição da demanda. Na etapa de pré-processamento, verificou-se que os autores Jiang *et al.* (2022), Özbay e Dalcali (2021) e Scarabaggio *et al.* (2020) aplicaram exclusivamente a normalização nos dados de entrada para o treinamento dos modelos. Além disso, Scarabaggio *et al.* (2020) especificou que a normalização aplicada foi do tipo z-score, o que leva em consideração o desvio padrão e a média dos dados. Tal técnica é interessante quando os dados seguem um padrão de distribuição normal.

Os autores Zhu e colaboradores (2021), Hafeez *et al.* (2020), Ferreira (2016) e Oliveira (2016) além de aplicarem a normalização, buscaram implementar outros métodos para o pré-processamento, como a seleção de características, análise de componentes independentes, identificação dos feriados, segundo o tipo e duração. Além disso, aplicaram filtro de relevância e filtro de redundância. Essas técnicas colaboram de forma significativa para uma melhor apresentação dos dados ao modelo de inteligência a ser treinado.

Outras técnicas de pré-processamento apresentadas por Peng *et al.* (2021), Bedi e Toshniwal (2018), Jawad *et al.* (2020), Kondaiah e Saravanan (2022) e Alasali *et al.* (2021) consistiram na média ou interpolação dos valores vizinhos para preenchimento de valores ausentes, agregação dos dados por estações do ano e análise de correlação das variáveis. Tais técnicas são populares quando se trata de predição de demanda e os resultados apresentados têm sido satisfatórios.

Já os autores Veeramsetty *et al.* (2022) e Huang *et al.* (2021) apresentaram técnicas inovadoras no pré-processamento. Elas se baseiam na correção estatística autorregressiva das variáveis de previsão meteorológicas e análise de componentes principais (PCA). Essa última permite reduzir a dimensionalidade dos dados de entrada, resultando em tempo menor de treinamento do modelo e redução do custo computacional. No caso em que há uma quantidade excessiva de parâmetros, essa técnica tem se mostrado promissora.

Contudo, observa-se que uma técnica interessante que poderia ter sido aplicada pelos autores seria a atribuição de um nível de importância maior para a variável demanda. Esse método permite que o modelo de inteligência artificial interprete a variável com uma

ponderação maior durante a etapa de treinamento. Além disso, a análise gráfica das variáveis de entrada é uma ferramenta importante para remoção de valores discrepantes e preenchimento de valores ausentes, podendo ser aplicada para o cenário de predição de demanda.

Quanto as técnicas aplicadas pelos trabalhos presentes no Quadro 1 e 2, observa-se a aplicação de redes recorrentes pelos autores Peng *et al.* (2021), Kondaiah e Saravanan (2022), Veeramsetty *et al.* (2022) e Özbay e Dalcali (2021). Essa técnica permite o estabelecimento de dependência temporal dos dados de entrada, o que é conveniente para dados baseados em séries temporais.

Ayub e colaboradores (2020) aplicaram modelos híbridos envolvendo máquina de vetor de suporte, redes neurais convolutivas e *gated recurrent unit* (GRU). Essa abordagem é promissora devido a capacidade de aprendizagem de diferentes características dos dados de entrada, como a relação espaço-temporal e similaridade dos dados de demanda e dados climáticos.

Alguns autores como Alasali *et al.* (2021), Senjyu *et al.* (2002), Scarabaggio *et al.* (2020) e Rosa *et al.* (2020) propuseram técnicas de aprendizado de máquina clássicas como redes neurais rasas, modelo de previsão linear, *perceptron* multicamadas e o modelo autorregressivo ARIMAX. Tais técnicas foram pioneiras nos trabalhos de predição de demanda e algumas costumam compor modelos híbridos ou parte de um conjunto de modelos.

Quanto às métricas para análise de desempenho aplicadas nos trabalhos analisados do Quadro 1 e 2, observa-se que as mais comuns são MAE, MAPE, MSE e RMSE escolhidas particularmente por cada um dos autores, não havendo uma métrica estabelecida como padrão. Tal abordagem, em alguns casos, dificulta uma análise comparativa dos modelos para o estabelecimento de uma arquitetura mais adequada.

Os trabalhos apresentados no Quadro 3, por sua vez, estão relacionados com a predição de demanda para a competição “IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm”. Em uma análise geral, foi verificado um número limitado de publicações relacionadas a predição de demanda de energia elétrica em cenários adversos, como o causado pela pandemia de COVID-19. Nesse sentido, a competição proporcionou um avanço nessa área, com o estabelecimento de um banco de dados público, uma métrica de desempenho padrão e um valor alvo a ser superado.

Em sua maioria, os métodos de pré-processamento aplicados consistiram na normalização dos dados de entrada, ajustes nos dados de predição meteorológica e detecção de valores discrepantes. Contudo, é possível aplicar um identificador do período de *lockdown*

existente no conjunto de dados. Esse ajuste pode aprimorar o desempenho dos modelos, já que nesse período o consumo de energia é atípico, se comparado aos períodos normais.

É interessante destacar o trabalho de De Vilmarest e Goude (2022) que foi inovador ao aplicar modelos lineares de espaço de estados gaussianos a fim de sincronizar os modelos no tempo. Contudo, o elevado número de modelos utilizados para prever a demanda, exige muito recurso computacional e complexidade na etapa de treinamento. O trabalho de Ziel (2022) também se destacou ao aplicar um método de tratamento de feriados, ainda que de forma analítica, ao observar picos e vales atípicos na curva de demanda. Porém, também apresentou 674 conjuntos de modelos para prever a demanda, exigindo muitos recursos para processamento.

Observa-se ainda que os métodos empregados pelos demais competidores não correspondiam ao uso de técnicas de redes neurais convolutivas e recorrentes. Estes utilizaram modelos clássicos de aprendizado de máquina ou modelos estatísticos, apesar de que a aplicação de redes profundas ter se mostrado promissora em alguns trabalhos da área. Também, a análise gráfica e tratamento de variáveis com valores discrepantes, bem como a identificação da característica mais importante para a realização do treinamento pode ser uma boa estratégia para o aprimoramento da previsão da demanda.

3 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os conceitos relacionados ao desenvolvimento deste trabalho. A Seção 3.1 apresenta conceitos do sistema elétrico de potência. A Seção 3.2 discorre acerca das definições de aprendizagem profunda. A Seção 3.3 apresenta os conceitos de redes neurais convolutivas. A Seção 3.4 expõe as definições de redes recorrentes. Por fim, a Seção 3.4 apresenta o conceito dos métodos de otimização empregados.

3.1 Sistema elétrico de potência

O desenvolvimento de fontes de energia para a execução de tarefas úteis é a base para o progresso industrial e tem como maior desafio a conversão de energia, de uma forma para outra, sem gerar poluição e degradação à biosfera. Nesse sentido, o sistema elétrico de potência constitui uma das ferramentas para converter e transportar energia, sendo subdividida em três áreas principais: geração, transmissão e distribuição (STEVENSON, 1986).

O sistema de geração tem a função de converter formas de energia presentes na natureza em energia elétrica. A transmissão transporta a energia das centrais geradoras para os centros de consumo. A distribuição recebe a energia da linha de transmissão e distribui para grandes, médios e pequenos consumidores (KAGAN; OLIVEIRA; ROBBA, 2005). A Figura 1 apresenta um diagrama unifilar do sistema elétrico de potência.

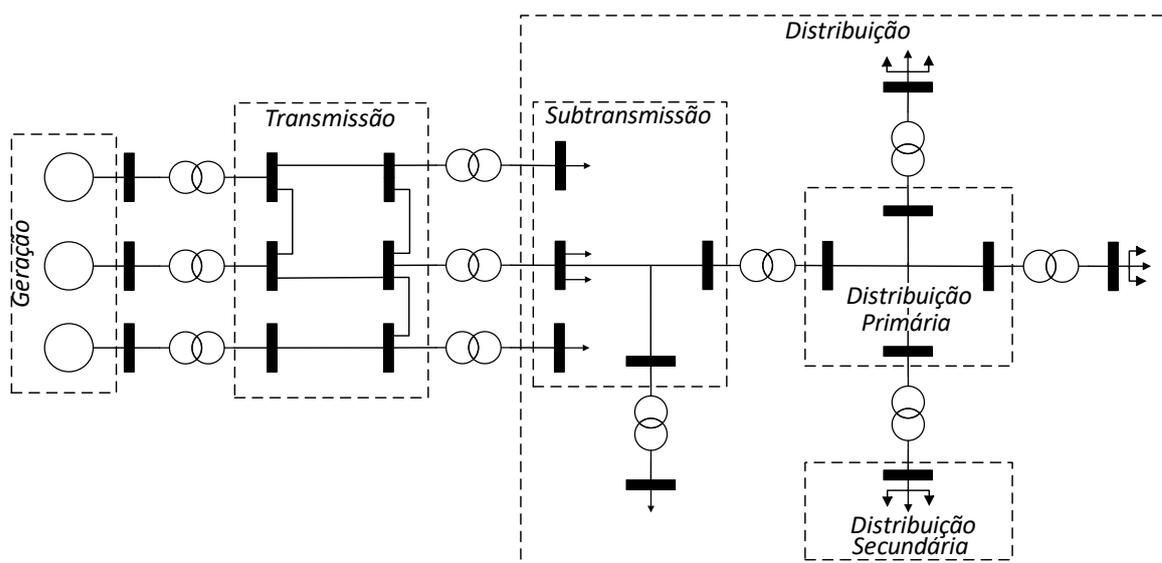


Figura 1 - Diagrama unifilar do sistema elétrico de potência. Fonte: (KAGAN; OLIVEIRA; ROBBA, 2005).

Na Figura 1 destaca-se a existência de três usinas de geração, um conjunto de linhas de transmissão, uma rede de subtransmissão que opera com um nível de tensão mais baixo, uma

rede de distribuição primária e três de distribuição secundária que o nível de tensão é denominado como “tensão secundária” ou “baixa tensão” (KAGAN; OLIVEIRA; ROBBIA, 2005).

3.2 Demanda de energia

Nesse trabalho, a previsão da demanda de energia constitui o principal objeto de estudo, possuindo uma relação direta com o sistema de geração de energia. Kagan *et al.* (2005) define a demanda como a carga instalada nos terminais receptores de energia, em valor médio, durante um determinado período. A carga é medida em termos de potência aparente, ativa ou reativa. O período no qual o valor é amostrado é denominado como intervalo de demanda. Assim, conforme a aplicação, pode-se levantar uma curva de demanda com intervalos de minutos ou horas. A Figura 2 apresenta a curva de demanda horária do Sistema Interligado Nacional (SIN), no período de janeiro a novembro de 2022, disponibilizada pelo Operador Nacional do Sistema Elétrico (ONS).

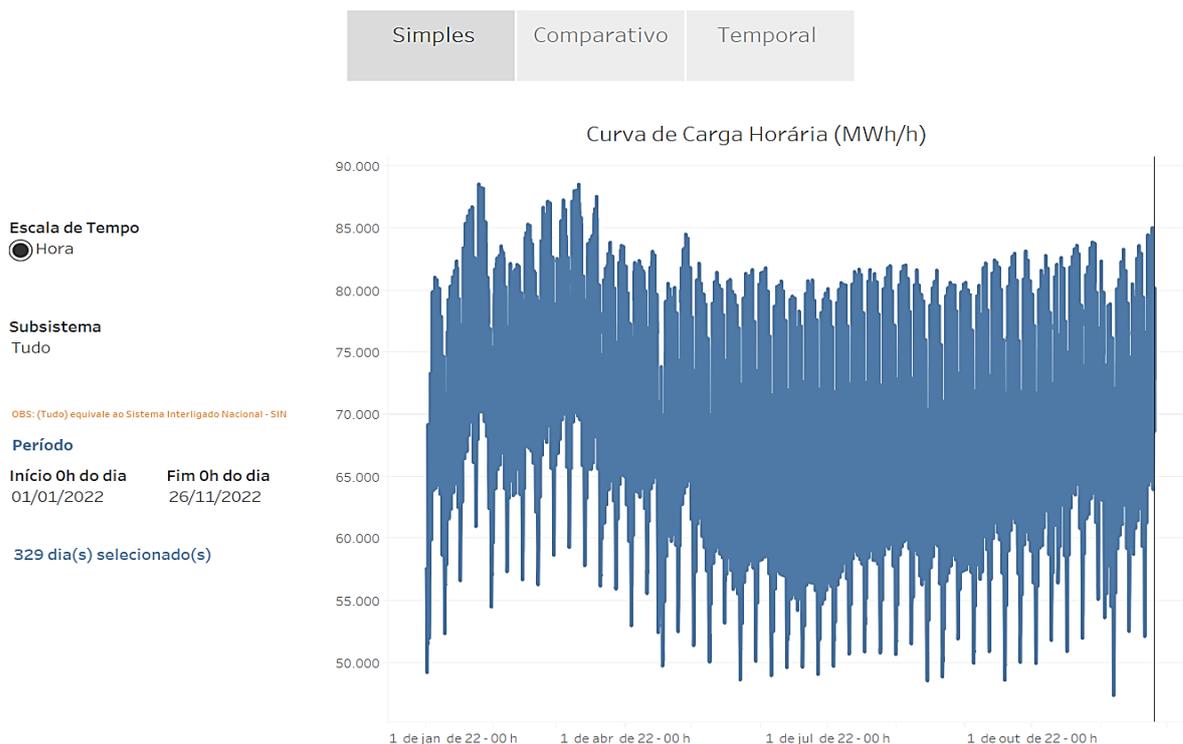


Figura 2 - Curva de demanda do Sistema Interligado Nacional. Fonte: (Operador Nacional do Sistema Elétrico, 2022)

A ANEEL (Agência Nacional Reguladora de Energia Elétrica) apresenta ainda alguns conceitos específicos relacionados a demanda de energia. A demanda registrada, por exemplo, refere-se a maior demanda de potência ativa injetada ou requerida do sistema elétrico de distribuição, verificada por instrumentos de medição, integralizada a cada 15 minutos durante

o intervalo de faturamento. A demanda contratada é a demanda de potência ativa a ser disponibilizada obrigatoriamente pela distribuidora no ponto de conexão, conforme o período contratual acordado. A demanda máxima corresponde a máxima potência que a concessionária de energia deve fornecer à instalação. A demanda percentual equivale a 85% da demanda máxima registrada nos últimos 11 meses (ANEEL, 2021). Todas essas métricas são impactadas por fatores demográficos, sociais e econômicos de uma região metropolitana. Além disso, são fundamentais para o gerenciamento do sistema energético e projeção de novas fontes geradoras.

Com base nesses conceitos, pode ser calculado o fator de demanda (f_{dem}), definido por Kagan *et al.* (2005) como a relação entre a demanda máxima ($D_{máx}$) e carga nominal ou instalada total ($D_{nom,i}$) de um elemento na rede elétrica. O cálculo do fator de demanda é dado pela equação a seguir, sendo imprescindível que as grandezas utilizadas tenham as mesmas unidades.

$$f_{dem} = \frac{D_{máx}}{\sum_{i=1,n} D_{nom,i}} \quad (1)$$

Um outro coeficiente importante é o fator de carga (f_{carga}) de um sistema, conceituado por Kagan *et al.* (2005) como a relação entre as demandas média ($D_{média}$) e máxima do sistema ($D_{máx}$), dado um período de tempo (τ) e tensão instantânea ($d(t)$), conforme a equação a seguir.

$$f_{carga} = \frac{D_{média}}{D_{máx}} = \frac{\int d(t) \cdot dt}{D_{máx} \cdot \tau} \quad (2)$$

O fator de utilização (f_{util}) também é um fundamental para observar a capacidade do sistema (C_{sis}) de atender a demanda máxima ($D_{máx}$) da instalação, em um intervalo de tempo (τ), definido por Kagan *et al.* (2005) segundo o equacionamento a seguir. O valor resultante normalmente é menor que 1. Contudo, quando há sobrecarga no sistema, esse fator ultrapassa valores unitários e inviabiliza a correta operação do sistema.

$$f_{util} = \frac{D_{máx}}{C_{sis}} \quad (3)$$

Com relação a anomalias no sistema energético, o Balanço Energético Nacional (BEN), fornecido pela Empresa de Pesquisa Energética (EPE), apontou no relatório de 2021 os impactos causados pela pandemia nos setores comerciais, públicos e energéticos. A Figura 3 apresenta a variação do consumo de energia setorial de energia elétrica do ano de 2020, em relação ao ano anterior, registrando quedas acentuadas em alguns setores e aumento em outros segmentos.

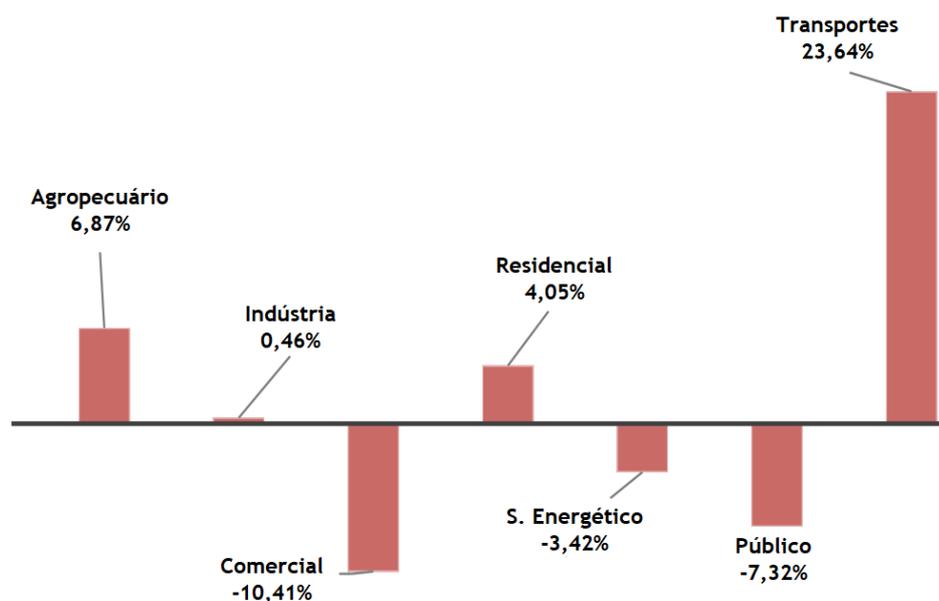


Figura 3 - Variação percentual do consumo setorial de eletricidade. Fonte: (EMPRESA DE PESQUISA ENERGÉTICA, 2021).

A preocupação dos efeitos pós pandemia também foi elencado pela Agência Nacional de Energia (ANEEL) que apontou queda nos rendimentos das distribuidoras e elevação do número da inadimplência dos consumidores (INSTITUTO ACENDE BRASIL, 2020). Esses comportamentos demonstram a dificuldade de predição da demanda de energia em um contexto pandêmico.

Nesse sentido, a predição de demanda proporciona o desenvolvimento de um sistema autossustentável, pois viabiliza a identificação de instantes de maior consumo de energia, possibilitando o uso de energia renovável, como energia solar fotovoltaica, para suprir a demanda, ao invés do uso de combustíveis fósseis. Também permite o aprimoramento na qualidade dos serviços, pois a manutenção pode ser realizada em momentos oportunos, em que há um menor consumo de energia, suprimindo a demanda com outras fontes geradoras. Além disso, oportuniza o ajuste adequado de tarifas, que tem como base de cálculo os custos de compra, transmissão e distribuição de energia. Com a predição da demanda, tais custos podem ser estimados, o que possibilita um valor de tarifa apropriado.

3.3 Aprendizagem profunda

A inteligência artificial (IA) pode ser definida como o desenvolvimento de máquinas de aprendizagem de modo que se comportem como se fossem inteligentes (ERTEL, 2017). Nesse sentido, o avanço das pesquisas ao longo dos anos subdividiu essa área em diversos campos de atuação, como por exemplo, o aprendizado de máquina clássico (AM) e o aprendizado de

máquina profundo (AP), como mostra a Figura 4. O aprendizado de máquina clássico constitui-se de um conjunto de técnicas clássicas utilizadas para classificação, regressão, transcrição, estimativa de densidade de probabilidades, etc. Dentre essas técnicas destacam-se os classificadores probabilísticos, árvores de decisão, máquinas de vetores de suporte, as redes neurais clássicas, também conhecidas como *perceptron* multicamadas (*perceptron multilayer – MLP*), k-vizinhos mais próximos, etc. O aprendizado de máquina profundo realiza as mesmas tarefas do aprendizado de máquina clássico, porém utiliza as redes neurais profundas, que são as redes convolutivas e as redes recorrentes. As redes neurais artificiais, principal ferramenta utilizada nesse trabalho, são inspiradas nos neurônios biológicos do cérebro humano.



Figura 4 - Subconjuntos de estudos da Inteligência Artificial. Fonte: Adaptado de Alzubaidi *et al.* (2021).

As redes neurais profundas têm a capacidade de automatizar o aprendizado de conjuntos de características para a realização de uma tarefa. Ao contrário, as redes neurais clássicas, MLP, requerem que as características sejam extraídas pelos usuários (ALZUBAIDI *et al.*, 2021). Dessa forma, as redes neurais profundas permitem que o computador gere conceitos, ou características complexas a partir dos próprios dados de entrada, em um processo hierárquico de geração: características mais simples gerando características mais complexas. (GOODFELLOW *et al.*, 2016). A Figura 5 apresenta o modelo de um aprendizado de máquina

profundo comparado a um modelo de aprendizado de máquina clássico. Enquanto o modelo clássico necessita de operações independentes para pré-processamento, extração e seleção de características, o modelo de aprendizado profundo é capaz de integrar todas essas funcionalidades por meio de suas camadas internas.

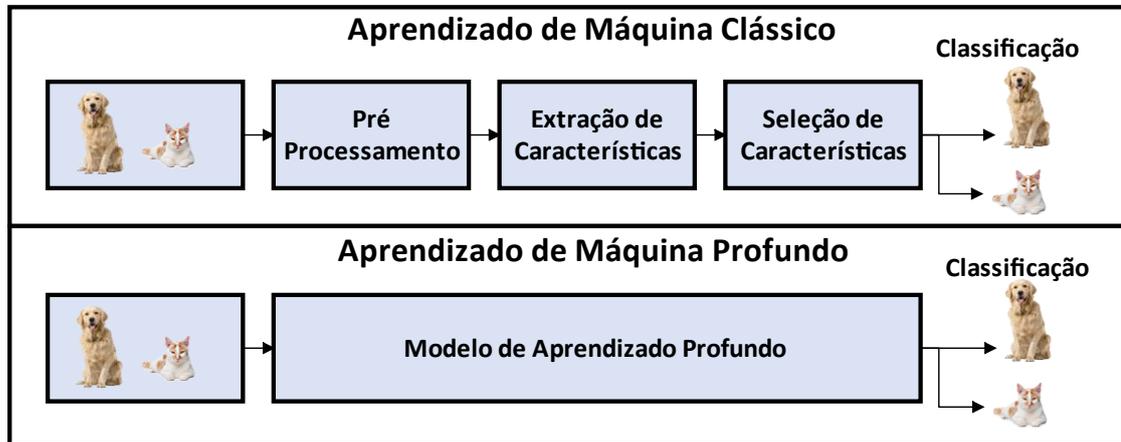


Figura 5 - Diferença entre o método clássico e o aprendizado profundo. Fonte: Adaptado de Alzubaidi *et al.* (2021).

Uma rede neural artificial consiste de muitos neurônios artificiais, que se comunicam uns com os outros. Na biologia, um neurônio contém algumas partes básicas denominadas de dendritos, corpo da célula ou soma e axônio. Os dendritos transmitem sinais para o corpo celular, que processa o sinal recebido. O axônio envia os sinais processados pelo corpo da célula para outros neurônios. Os sinais de entrada são somados no corpo da célula. Quando essa soma ultrapassa um certo limite, é gerado um sinal de saída que é passado adiante pelo axônio (NEAPOLITAN; JIANG, 2018). A Figura 6(a) representa essa estrutura de neurônio biológico. De forma análoga, o neurônio artificial proposto tem como entrada um vetor (x_1, x_2, \dots, x_k) , que é multiplicado pelos pesos (w_0, w_1, \dots, w_k) , representando os dendritos. Em seguida, há a aplicação de uma função de ativação f associado a soma da multiplicação dos pesos pela entrada, representado o corpo da célula. O valor Y é gerado na saída, representado o axônio.

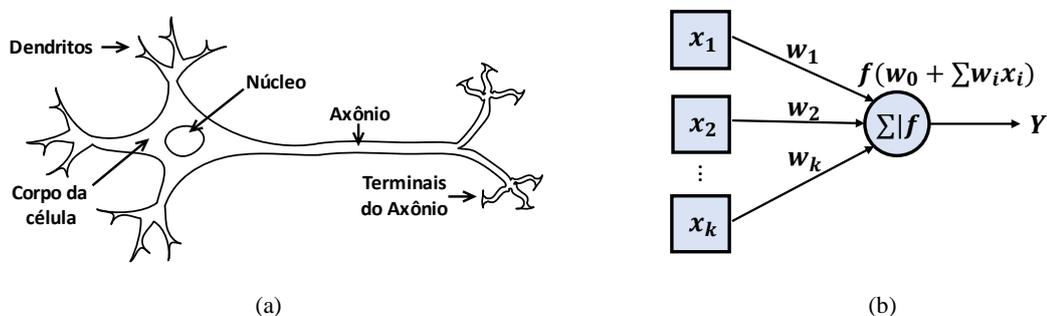


Figura 6 - Estrutura do neurônio: (a) neurônio biológico; (b) neurônio artificial. Fonte: Adaptado de Neapolitan e Jiang, 2018.

Assim, a expressão matemática do neurônio artificial é definida conforme através da Equação 4. A Equação 5 apresenta essa mesma definição, no formato matricial, em que $W =$

$$[w_1 \quad w_2 \quad \dots \quad w_n] \text{ e } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

$$Y = f\left(\sum_{i=1}^k w_i x_i + w_0\right) \quad (4)$$

$$Y = f(W \cdot X + w_0) \quad (5)$$

A matriz X é denominada de vetor de características e a seleção desse conjunto de forma apropriada é uma tarefa fundamental, compreendendo uma das áreas mais importantes dentro do campo do reconhecimento de padrões. A matriz W é denominado de vetor de pesos e w_0 de polarização da rede neural. Tais variáveis são ajustadas no processo de treinamento da rede, com o objetivo de produzir uma saída Y mais próxima ou igual ao valor esperado. Além disso, a função de ativação desempenha um papel importante pois é a partir dela que o neurônio é ativado. A Figura 7 apresenta as principais funções de ativação aplicadas.

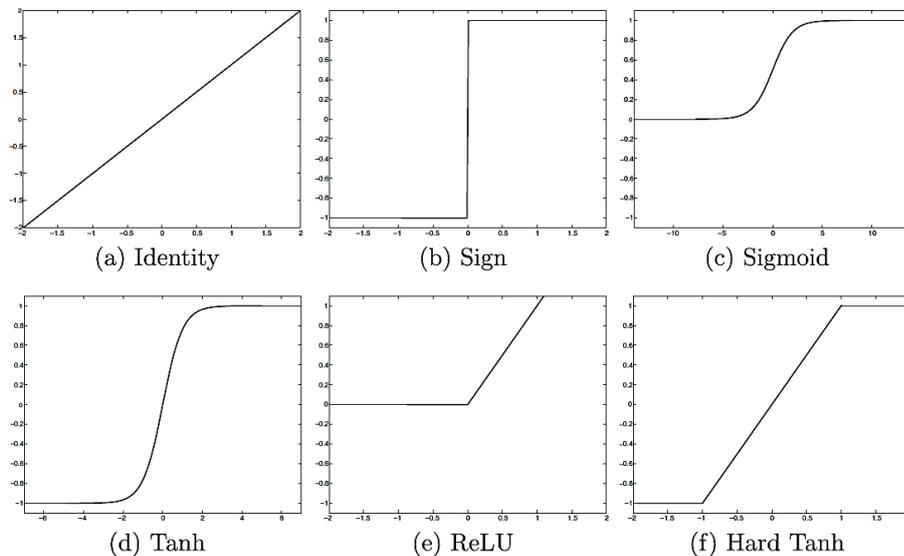


Figura 7 - Principais funções de ativação aplicadas nas redes neurais. Fonte: (AGGARWAL, 2018a).

Uma rede neural pode ser estruturada em camadas. Na forma mais simples, temos uma camada de entrada de nós fonte que projeta diretamente sobre uma camada de saída de neurônios. Conforme mostrado na Figura 8, uma MLP é formada por mais de uma camada de neurônios artificiais (HAYKIN, 2009). Observa-se que essa rede tem a característica de ser inteiramente conectada, pois todos os neurônios da camada anterior se conectam aos neurônios da camada adjacente.

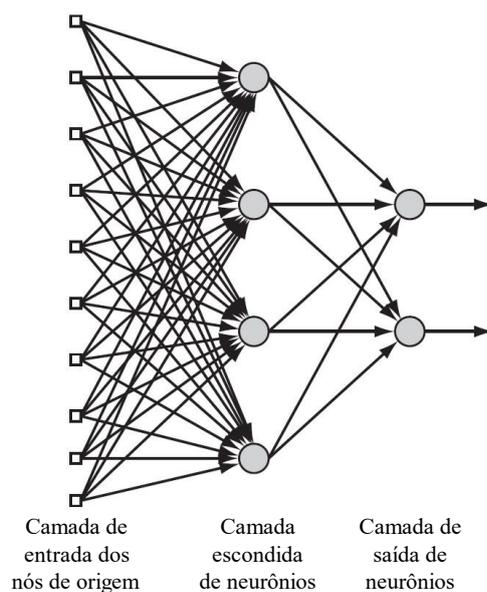


Figura 8 - Rede tipo *perceptron* multicamadas. Fonte: (HAYKIN, 2009).

3.4 Redes Neurais Profundas - Convolutivas

Embora as redes neurais clássicas tenham se tornado uma ferramenta de aprendizado de máquina dominante na segunda metade do século passado, elas passaram a ter menos popularidade no advento da virada do século, devido às limitações de desempenho. Em cenários envolvendo imagens, por exemplo, o processamento exigia um elevado número de neurônios, que implicava em um elevado número de parâmetros, como pesos e polarizações. Além disso, nos casos em que são utilizadas imagens RGB (*red-green-blue*) na entrada da rede, onde a dimensionalidade da entrada é multiplicada por três, o número de parâmetros da rede aumenta mais ainda (THEODORIDIS, 2020).

Na primeira década do século atual, as redes neurais convolutivas (RNC) vieram apresentar um novo método de treinamento de sistemas inteligentes. Na sua origem, tais redes foram propostas para resolver problemas de visão computacional, em aplicações de classificação de imagens e detecção de objetos (AGGARWAL, 2018a). Na arquitetura de rede neural convolucional, cada camada da rede é tridimensional, com extensão espacial (largura e altura) e profundidade. Cada plano de uma camada representa um mapa de características. A primeira arquitetura básica baseada nesta inspiração biológica foi o *neocognitron*, que foi então generalizado para a arquitetura LeNet-5 (LECUN, 1989). A Figura 9 representa essa rede e lista a denominação de cada uma das camadas.

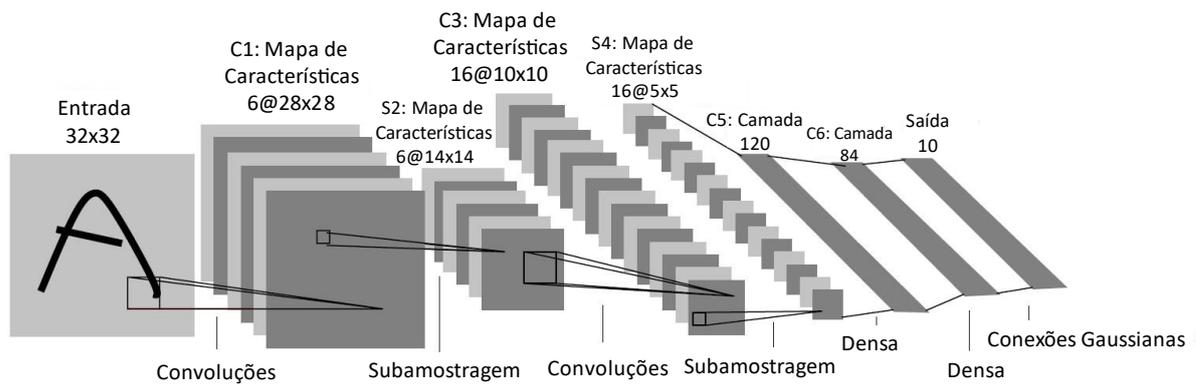


Figura 9 - Arquitetura da rede LeNet-5. Fonte: (LECUN et al., 1998).

A RNC tem uma inspiração biológica no córtex visual dos mamíferos, que está subdividido em áreas visuais inferiores, denominadas de V1, V2, V3 e V4, e regiões superiores de processamento de objetos, localizados no córtex occipito-temporal lateral (LOC) e occipito-temporal ventral (VOT), cada uma com funções específicas no processamento de imagem (XU; VAZIRI-PASHKAM, 2021). A Figura 10 apresenta uma abstração do processamento realizado pelo cérebro humano que é realizado pelas RNCs.

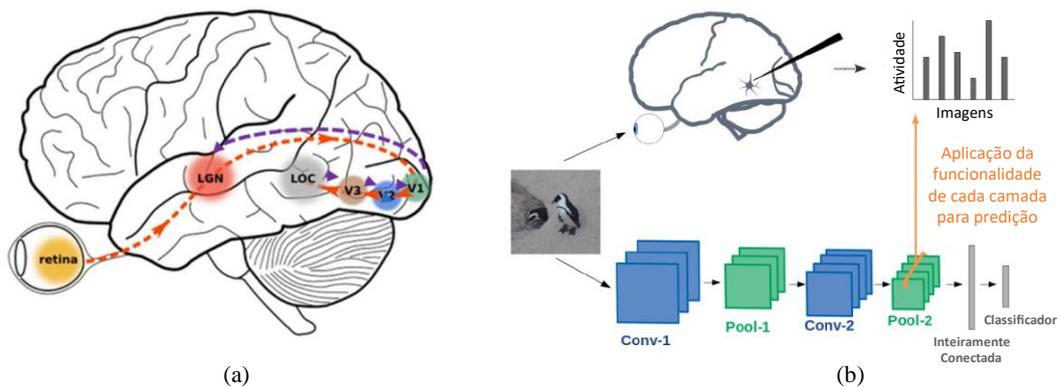


Figura 10 – Abstração do processamento das redes convolutivas. (a) Cérebro humano. (b) Processamento de imagem pela rede neural convolutiva. Fonte:(AHMAD; JOE; HAN, 2018; XU; VAZIRI-PASHKAM, 2021).

A aplicação de redes neurais convolutivas são mais atraentes quando há dados suficientes e poder computacional disponível. Nos últimos anos, esses dois atrativos têm sido acessíveis à comunidade acadêmica o que impulsionou a aplicação dessas redes para a solução de diversos problemas. Na Figura 11 observa-se que os métodos de aprendizagem de máquina profundos têm se mostrado mais eficiente frente aos métodos clássicos de aprendizado de máquina, quando a quantidade de dados aumenta. Tal fato, incentiva a aplicação dessa técnica neste trabalho.

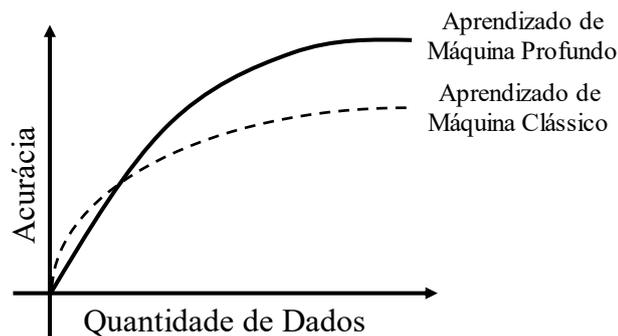


Figura 11 - Desempenho das redes de aprendizado profundo frente aos métodos clássicos de aprendizado de máquina. Fonte: Adaptado de Aggarwal, 2018a.

Como mostrado nas Figuras 9 e 10 (b), as redes neurais convolucionais possuem várias camadas e cada uma delas têm funções específicas para identificação de padrões dos dados de entrada. Assim, a quantidade e o tipo de camadas aplicadas na rede convolucional definem a arquitetura da rede e são delimitadas de forma experimental. Algumas dessas camadas são denominadas como *Convolução*, *Ativação*, *Pooling* e *Dropout*.

3.4.1 Camada de convolução

Em uma arquitetura de rede neural convolucional, a camada de convolução é a mais significativa. Ela consiste na aplicação de uma série de filtros convolucionais, denominados de *kernels*. Os filtros são aplicados aos dados de entrada para a geração dos dados de saída, denominados de mapa de características, (ALZUBAIDI et al., 2021). A equação matemática que descreve essa operação no formato bidimensional é definida como:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (6)$$

Na Equação 6, i e j definem o tamanho dos dados de entrada, enquanto que m e n correspondem ao dimensionamento do *kernel* aplicado. Assim, a operação de convolução possibilita a aplicação do filtro em cada posição dos dados, por meio da soma dos produtos dos coeficientes do filtro de forma deslizante sobre os dados de entrada. A Figura 9 apresenta a dinâmica dessa operação.

Com o objetivo de manter o dimensionamento dos dados de entrada igual aos dados de saída, é necessário a aplicação de *padding*s. Essa técnica visa preencher as bordas da matriz de entrada com novos valores, sendo comum o preenchimento com zeros, como mostra a Figura 12. Além disso, pode-se definir o passo (*stride*) que o *kernel* irá percorrer durante a aplicação da convolução.

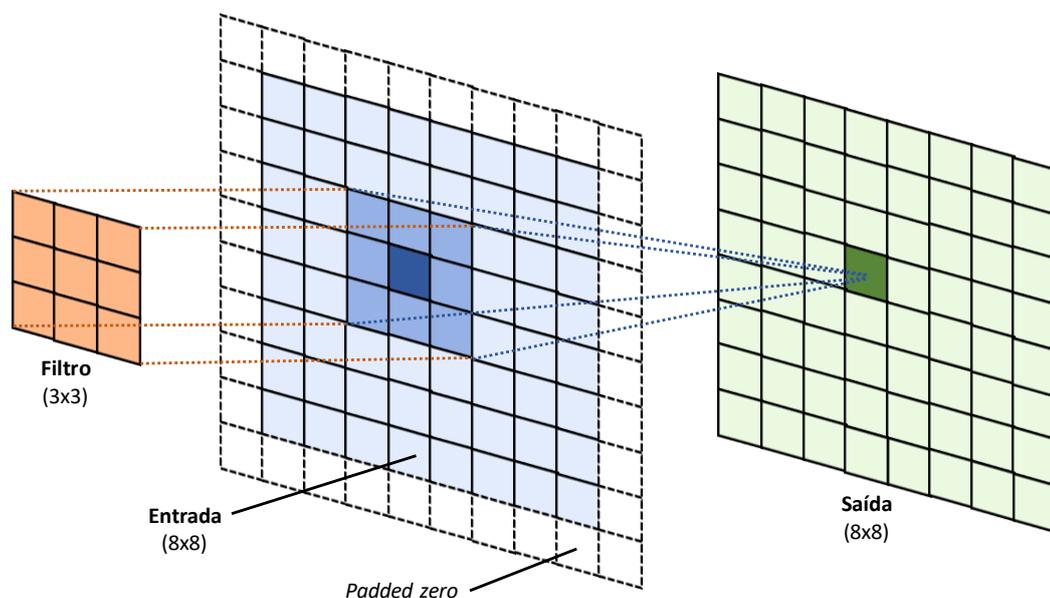


Figura 12 - Aplicação da operação de convolução bidimensional. Fonte: Adaptado de Raschka e Mirjalili, 2019.

Em uma operação de convolução, um mesmo filtro é aplicado aos dados de todos os planos da entrada, gerando um plano de características na saída da camada convolutiva. Esse compartilhamento de parâmetros diminui muito o número de parâmetros de cada camada da rede. Observa-se também que a operação de convolução correlaciona as informações de uma vizinhança com os pesos do filtro. O mapa de características gerado por uma operação de convolução será, na arquitetura da rede, a entrada da próxima camada, possibilitando a formação de uma arquitetura com várias camadas de extração de características. (THEODORIDIS, 2020).

3.4.2 Camada de ativação

A escolha de uma função de ativação é fundamental no projeto de sistemas inteligentes. No caso de uma rede neural artificial, por exemplo, uma função de ativação do tipo sinal (*sign*) permite a classificação de uma classe binária a ser predita. Contudo, em outros cenários em que mais classes devem ser preditas ou problemas que envolvam regressão, a escolha de outras funções de ativação pode ser mais adequada (AGGARWAL, 2018a).

A função de ativação ReLU (*Rectified Linear Activation Function*) tem sido amplamente utilizada em redes neurais convolucionais devido à sua não linearidade e facilidade de uso no treinamento de redes neurais com o método de *backpropagation*, pois a sua derivada ou é 0 ou é 1 (AGGARWAL, 2018a). Matematicamente, ela assume valor de saída igual a 1

para valores de entrada positivos e valor 0 para valores de entrada negativos, como expressa a Equação 7.

$$f(x) = \max(0, x) \quad (7)$$

Assim, selecionando a ReLU como função de ativação, é possível contornar problemas relacionados ao desvanecimento do gradiente, quando as derivadas obtêm valores pequenos. Esse efeito ocorre devido a derivada da função de ativação entrar de forma multiplicativa no algoritmo de retropropagação, quando os gradientes em relação aos parâmetros das camadas escondidas são calculados (THEODORIDIS, 2020). Contudo muitos neurônios podem ter valores negativos durante a etapa de treinamento, resultando em um valor de ativação igual a zero, após a multiplicação do somatório efetuado pelo neurônio pela função ReLU. Esse efeito é denominado como “*Dying ReLU*” (ALZUBAIDI et al., 2021).

3.4.3 Camada de *pooling*

Tipicamente, uma rede neural convolucional é composta por três estágios. O primeiro está relacionado com a camada convolucional, em que a operação da convolução é aplicada. Em seguida, há aplicação de uma função de ativação, como a ReLU citada anteriormente. Em um terceiro estágio ocorre a aplicação de uma camada chamada de amostragem (*pooling*).

O objetivo dessa etapa é reduzir a dimensionalidade do mapa de características. Para isso, é definido uma janela bidimensional para deslizar sobre a matriz correspondente, definindo-se um tamanho de passo (*stride*). Assim, a função consiste em escolher um único valor para representar todos os valores que estão dentro da janela. A operação mais utilizada é *max pooling* que define o valor máximo, contido dentro da janela, para ser o valor da matriz de saída (THEODORIDIS, 2020). A Figura 13 apresenta essa operação, em que uma janela 3×3 é definida, e é apresentado o resultado quando aplicado o *stride* de valor 1 e 2, respectivamente.

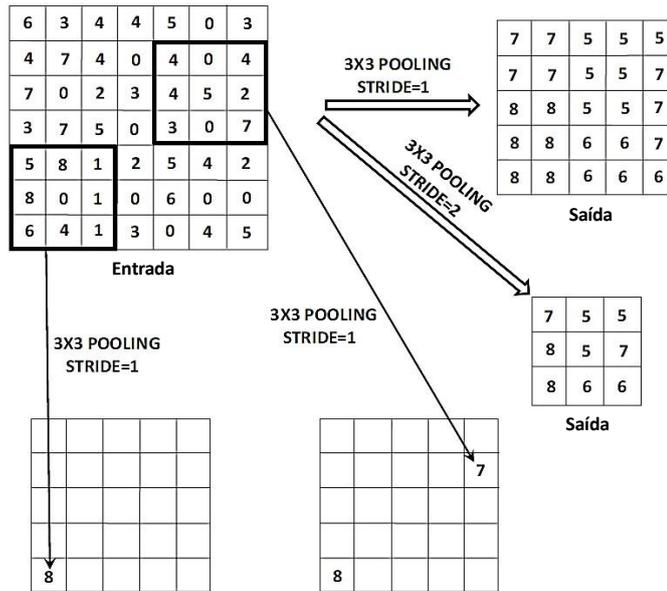


Figura 13 - Função *max pooling* 3x3 com *stride* de 1 e 2. Fonte: Adaptado de Aggarwal, 2018a.

É possível observar na Figura 10 a mudança de dimensionalidade da matriz de saída, dependendo do *stride* definido. Assim, a aplicação de *pooling*, com dimensão 2×2 e *stride* 2 é outra prática comum utilizada para redução do mapa de características. No caso de imagens, por exemplo, esse agrupamento de valores resulta em uma invariância à translação, pois, com a redução de dimensionalidade causado pela camada *pooling*, a operação de translação de uma imagem de entrada não altera significativamente o mapa de características gerado pela camada de *pooling*. (THEODORIDIS, 2020). A Equação 8 descreve o tamanho da matriz de saída (M_o), dado a matriz de entrada (M_i), o tamanho da janela *max pooling* (M_{pool}) e o tamanho do *stride* (S).

$$M_o = \frac{M_i - M_{pool}}{S} + 1 \quad (8)$$

3.4.4 Camada de *Dropout*

A camada de *dropout* é amplamente utilizada para a generalização de modelos de inteligência artificial. Durante a etapa de treinamento, neurônios arbitrários são descartados multiplicando seu valor de saída por zero. Ao fazer isso, a seleção de características é distribuída igualmente por todo o grupo de neurônios, de modo que o modelo aprenda características diferentes de forma independente (ALZUBAIDI et al., 2021). A Figura 14(a)-(c) apresenta o processo de *dropout* em uma rede neural.

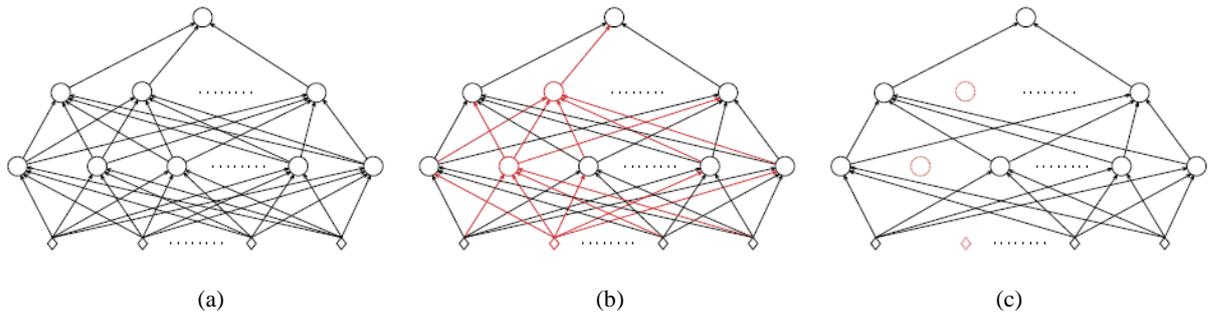


Figura 14 - Processo de *dropout* em uma rede neural: (a) rede inteiramente conectada; (b) neurônios e conexão em vermelho que serão desativados; (c) neurônios selecionados foram desativados. Fonte: (THEODORIDIS, 2020).

Primeiramente, uma amostragem de neurônios é aplicada na rede neural completa. Os nós de entrada são amostrados com uma probabilidade p_i , e os nós ocultos são amostrados com probabilidade p_h . Todas as amostras são independentes umas das outras e quando um nó é removido da rede, todas as conexões incidentes no mesmo também são removidas. Em seguida, o treinamento de um lote é realizado. Por fim, os pesos de todos os neurônios são atualizados, incluindo os que foram retirados, por meio do algoritmo de retropropagação (*backpropagation*) (AGGARWAL, 2018a).

Assim, o principal efeito do *dropout* é incorporar a regularização no processo de aprendizagem, combinando amostragem de nós e compartilhamento de pesos. Se o conjunto de dados de treinamento for grande o suficiente para reduzir as probabilidades de *overfitting*, as vantagens computacionais do *dropout* podem ser pequenas, mas ainda perceptíveis. Na literatura a porcentagem de *dropout* entre 20% e 50% é bastante utilizada.

3.4.5 Treinamento de uma RNC

Em uma rede neural artificial as informações fluem a partir de uma entrada (x) em direção a saída (y) da rede, passando por camadas escondidas. Esse fluxo é denominado como propagação direta (*forward propagation*). Durante o treinamento supervisionado da rede, é necessário a minimização de uma função de custo $J(w)$. Nessa etapa, o algoritmo de retropropagação permite que as informações da derivada da função de custo fluam em direção a entrada da rede, com o objetivo de atualizar os parâmetros da rede através do algoritmo do gradiente descendente.

O algoritmo de retropropagação consiste em aplicar repetidamente ajustes nos pesos das conexões da rede de modo a minimizar a diferença entre o valor real na saída da rede e o valor desejado (RUMELHART et al., 1986). Como resultado, os pesos e polarizações das camadas escondidas são atualizados a cada iteração do algoritmo de retropropagação.

Para minimizar o erro por meio do gradiente descendente, é necessário primeiramente calcular as derivadas parciais da função de custo $J(w)$, referente a cada parâmetro da rede. Na demonstração a seguir, aplica-se a função do erro médio absoluto (MAE) para aferir o desempenho na saída da rede, em que $y_i(x_k; w)$ é o valor predito pela rede, \hat{y}_i é o valor esperado e n a quantidade de dados observados:

$$J(w) = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n |y_i(x_k; w) - \hat{y}_i| \quad (9)$$

Na versão online desse algoritmo, em que, após a apresentação de cada entrada ocorre a atualização dos parâmetros, Bishop (2006) sintetiza o algoritmo de retropropagação em quatro etapas principais. A primeira consiste na aplicação do vetor de entrada x_k na rede neural de propagação direta para encontrar as ativações das camadas escondidas e da camada de saída. Em seguida, o valor do erro absoluto δ_k deve ser calculado:

$$\delta_k = |y_i(x_k; w) - \hat{y}_i| \quad (10)$$

Então, na terceira etapa, o valor de δ_j de cada neurônio das camadas escondidas é obtido por meio do retropropagação, em que $h(a_j)$ é a função de ativação não linear, conforme a equação:

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k \quad (11)$$

Em seguida, devem ser calculadas as derivadas parciais da função de custo:

$$\frac{\partial J_n}{\partial w_{ji}} = \delta_j z_i \quad (12)$$

Por fim, através do algoritmo do gradiente descendente, atualiza-se o valor de cada parâmetro:

$$w_{ji} = w_{ji} - \eta \frac{\partial J_n}{\partial w_{ji}} \quad (13)$$

Assim, o método do gradiente atualiza cada parâmetro tendo como objetivo reduzir o erro da classificação de um dado objeto x_k . Se essa derivada for positiva, um aumento no valor do parâmetro está causando um aumento da função de custo, então o valor do mesmo é diminuído. Caso o resultado da derivada seja negativo, um aumento do parâmetro está contribuindo para uma diminuição da função de custo, então o valor do mesmo é aumentado.

Na literatura esse método, devido ao seu desempenho, é o mais aplicado no treinamento das redes neurais. Contudo, com o aumento da complexidade das RNA's, o tempo de processamento torna-se elevado, devido a quantidade de pesos e polarizações que devem ser

atualizados. Nesse contexto, uma elevada capacidade computacional é fundamental para um treinamento mais rápido dessas redes, por meio da utilização de CPU's e unidades gráficas de processamento (GPU) de alto desempenho.

3.5 Redes Neurais Recorrentes

As redes neurais apresentadas nos capítulos anteriores têm como premissa a relação de independência na maioria dos atributos dos dados de entrada. Contudo, a natureza de alguns dados possui características de dependência sequencial, como séries temporais, textos e dados biológicos. Nas séries temporais, os registros de data e hora (*timestamp*) estão diretamente relacionados com as demais variáveis do conjunto de dados. Quando não se aplica uma relação de dependência temporal, informações valiosas relacionadas aos registros de data e hora são perdidas (AGGARWAL, 2018a).

As redes neurais recorrentes (RNN) constituem uma família das redes neurais profundas projetadas para lidar com dados sequenciais. Enquanto uma rede neural inteiramente conectada possui parâmetros separados para cada característica de entrada, a rede neural recorrente compartilha os mesmos pesos, em instantes de tempo distintos. Observa-se ainda que nas redes convolucionais há o compartilhamento de pesos, porém sem a dependência temporal de uma rede neural recorrente.

A RNN é denominada de uma máquina de Turing completa, pois é capaz de simular estados, memórias e transições, desde que haja entrada suficiente de dados e recursos computacionais. Na arquitetura da RNN, a informação percorre camada a camada com ciclos internos (*loops*) na rede, de modo que cada estado é influenciado por seus estados anteriores. Esses *loops* permitem o armazenamento de valores passados, evidenciando um comportamento dinâmico temporal. A Figura 15 apresenta a arquitetura da rede, e as variáveis envolvidas em uma RNN que são denominadas como:

- a) um vetor de estado em um tempo t , denominado como h_t . Esse vetor constitui a memória do sistema;
- b) um vetor de entrada no tempo t , denominado como x_t ;
- c) um vetor de saída no tempo t , denominado como y_t ;
- d) as matrizes W_{hy} , W_{hh} e W_{xh} que são as matrizes de pesos.

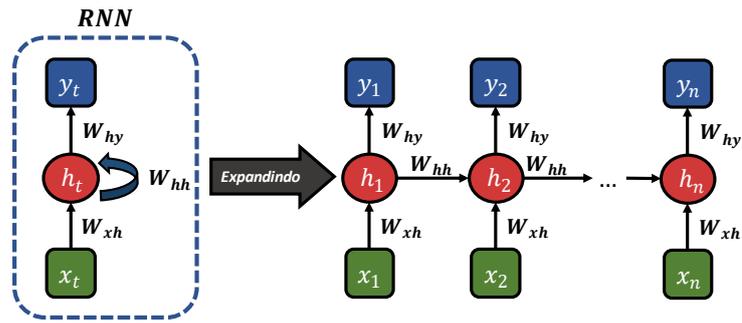


Figura 15 - Arquitetura de uma rede recorrente.

Em termos práticos, é possível modelar o sequenciamento da informação que será transmitida pela rede de aprendizado em algumas categorias, conforme mostrado na Figura 16. As aplicações de rede recorrentes podem ser feitas utilizando-se uma dentre várias configurações, dependendo do tipo da aplicação. Na configuração muitos-para-um (*many-to-one*), os dados de entrada são sequenciais, mas a saída é um escalar ou vetor. Na configuração um-para muitos (*one-to-many*), o dado de entrada é escalar ou vetor e a saída é sequencial. Por fim, na configuração muitos-para-muitos (*many-to-many*), tanto a entrada quanto a saída são sequenciais, podendo ser com entrada e saída sincronizadas ou não sincronizadas.

Frente aos desafios do aprendizado dos parâmetros de uma rede neural recorrente, existem alguns problemas relacionados ao desvanecimento e explosão dos valores do gradiente. Esse problema é frequente no contexto das redes neurais recorrentes. Para contornar esses problemas, foram propostas diversas variantes das RNNs, como a arquitetura *long short-term memory* (LSTM) e *gated recurrent unit* (GRU).

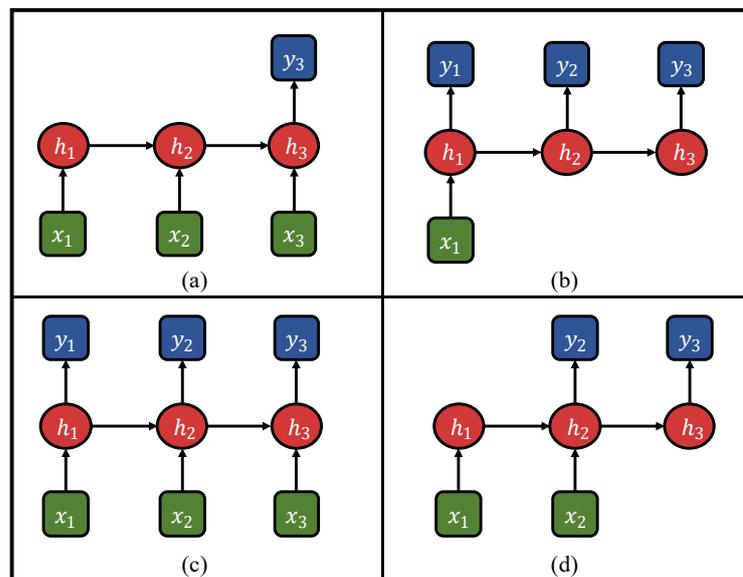


Figura 16 - Formas de sequenciamento dos dados em uma RNN: (a) *Many-to-one*; (b) *One-to-Many*; (c) *Many-to-many* sincronizado; (d) *Many-to-many* não sincronizado.

As RNNs e suas variantes têm sido utilizadas em aplicações como aprendizagem de legenda de imagens, tradução automática e análise dos sentimentos, dentre outras. A seguir, a rede LSTM e sua estrutura é apresentada, sendo uma das técnicas mais eficientes no estabelecimento de dependência temporal dos modelos de inteligência artificial.

3.5.1 Redes long short-term memory (LSTM)

A principal característica de uma rede LSTM é a estrutura interna da sua célula de memória que essencialmente representa ou substitui a camada oculta de uma rede neural recorrente padrão. Hochreiter e Schmidhuber (1997) propuseram esse modelo de rede, introduzindo técnicas para solucionar o problema do gradiente. A arte da técnica consiste em controlar o fluxo de informações para dentro e para fora da célula de memória do sistema, por meio de elementos não lineares conhecidos como portas (*gates*).

Por meio da função logística e das matrizes de peso multiplicadoras, essas portas são implementados e são denominados de *input gate* (i_t), *forget gate* (f_t) e *output gate* (o_t). Também há o vetor que indica o estado interno da célula LSTM, denominado como estado da célula (*cell state* C_t) e o valor candidato (\tilde{C}_t). Algebricamente, as portas aplicam uma ponderação no fluxo de informação, em um intervalo de 0 a 1, dependendo dos valores das variáveis que ativam a função logística. A Figura 17 apresenta a arquitetura da rede LSTM e suas conexões com as células de memória existentes na rede.

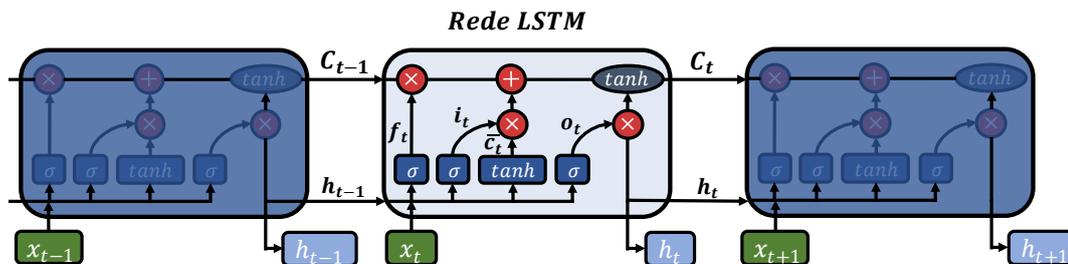


Figura 17 - Arquitetura da rede LSTM.

O *forget gate* permite que a célula de memória reinicie seu estado, evitando um crescimento exponencial. Assim, essa porta decide quais informações podem ser repassadas e quais serão descartadas. O *input gate* e o valor candidato são responsáveis por atualizar o *cell state*. O *output gate* transmite o estado interno da célula para outra. O equacionamento das portas, do estado da célula e do valor candidato da rede LSTM é apresentado a seguir, incluindo as respectivas polarizações b_c , b_i , b_f e b_o .

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (14)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (15)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (16)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (17)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (18)$$

Uma das desvantagens das redes LSTM é o seu alto custo computacional, devido a sua arquitetura complexa. Comparando com uma rede recorrente convencional, o custo da rede LSTM é quatro vezes maior. Assim, redes com estruturas modificadas foram desenvolvidas para contornar esse problema como a rede *gated recurrent unit* (GRU), que possui uma menor quantidade de portas. Outra abordagem é comprimir o vetor de saída da LSTM, por meio de uma projeção linear, denominada como projeção recorrente.

3.5.2 Treinamento das redes LSTM

Na RNN adota-se o algoritmo *Backpropagation Through Time* (BPTT), proposto por Werbos (1990), para atualização dos pesos das células. Para trabalhar com pesos compartilhados é necessário interpretar os parâmetros como sendo independentes uns dos outros nas diferentes camadas temporais. Assim, são introduzidas as variáveis W_{xh}^t , W_{hh}^t e W_{hy}^t , em relação ao tempo t . Dessa forma, primeiramente aplica-se o algoritmo de retropropagação convencional, onde as variáveis são distintas umas das outras. Posteriormente, essas contribuições individuais para o gradiente são somadas, a fim de unificar a atualização do peso, para cada parâmetro. Aggarwal (2018b) sumariza o algoritmo BPTT da seguinte forma:

1. Aplicar sequencialmente as entradas, realizar as propagações diretas e calcular os erros (perda da camada *softmax*) em cada instante de tempo.

2. Calcular o gradiente dos pesos na propagação reversa, considerando que os pesos em diferentes camadas de tempo não são compartilhados, ou seja, assumindo que os pesos W_{xh}^t , W_{hh}^t , W_{hy}^t e o tempo t são distintos de outros registros de data e hora. Assim, utiliza-se o

algoritmo de retropropagação convencional para calcular $\frac{\partial J}{\partial W_{xh}^{(t)}}$, $\frac{\partial J}{\partial W_{hh}^{(t)}}$ e $\frac{\partial J}{\partial W_{hy}^{(t)}}$.

3. Por fim, soma-se as derivadas relativas das diferentes instâncias de tempo:

$$\frac{\partial J}{\partial W_{xh}} = \sum_{t=1}^T \frac{\partial J}{\partial W_{xh}^t} \quad (19)$$

$$\frac{\partial J}{\partial W_{hh}} = \sum_{t=1}^T \frac{\partial J}{\partial W_{hh}^t} \quad (20)$$

$$\frac{\partial J}{\partial W_{hy}} = \sum_{t=1}^T \frac{\partial J}{\partial W_{hy}^t} \quad (21)$$

O algoritmo BPTT foi desenvolvido antes da popularização das redes recorrentes e tem sido aplicado em sistemas dinâmicos, sistemas com atrasos temporais, sistemas com *feedback* instantâneo entre variáveis, como equações ordinárias artificiais e modelos simultâneos. Contudo, para aplicação desse método também são exigidos recursos computacionais elevados, pois é grande o consumo de memória para o processamento do algoritmo.

3.6 Métodos de Otimização

O treinamento das redes de inteligência artificial necessita de diversos cálculos numéricos para construção de um modelo preditivo. Normalmente, os algoritmos envolvidos nessa etapa exigem um processo iterativo matemático por meio da inicialização aleatória de parâmetros, estimativa e atualização de valores. Com o objetivo de aprimorar esse processo, algoritmos de otimização foram desenvolvidos para a minimizar a função de custo definida. Assim, serão apresentados os algoritmos comumente utilizados para o treinamento das redes convolutivas e recorrentes, dentre os quase destacam-se o SGD (*Stochastic Gradient Descent*), SGDM (*Stochastic Gradient Descent With Momentum*) e o ADAM (*Adaptive Moment Estimation*).

3.6.1 *Stochastic Gradient Descent (SGD)*

A técnica do gradiente descendente estocástico (SGD) consiste em um dos métodos de otimização mais utilizados para o treinamento das redes neurais artificiais. É um método muito simples, porém poderoso para encontrar os pesos que minimizam a função de custo, e a aprimorar a predição do modelo.

A ideia principal do SGD é encontrar o mínimo global por meio do cálculo das derivadas parciais da função de custo. Em cada iteração, um passo é dado em direção ao valor mínimo. O tamanho do passo é definido por meio do hiperparâmetro definido como taxa de aprendizagem (η). Dessa forma, os pesos (w) da rede são somados a variação (Δw) provenientes do SGD, conforme as equações:

$$w := w + \Delta w \quad (22)$$

$$\Delta w = -\eta \frac{\partial J}{\partial w} \quad (23)$$

A derivada da função de custo em relação a cada peso é descrita na equação a seguir, onde L é a função de custo:

$$\nabla_{\theta} J(\theta) = \frac{1}{m} \sum_{i=1}^m \nabla_{\theta} L(x_i, y_i, \theta) \quad (24)$$

Assim, a função do gradiente descendente estocástico que realiza a amostragem no mini-lote para o cálculo do gradiente é apresentada na equação a seguir, em que m' é a quantidade de amostras e θ o parâmetro interno da rede:

$$g = \frac{1}{m'} \nabla_{\theta} \sum_{i=1}^{m'} \nabla_{\theta} L(x_i, y_i, \theta) \quad (25)$$

No entanto, Goodfellow e colaboradores (2016) apontam que uma desvantagem do SGD é o tempo de convergência para otimização dos valores. Assim, foi desenvolvido o método SGDM que possui variáveis de controle dos parâmetros que otimizam os pesos da rede neural artificial.

3.6.2 Stochastic Gradient Descent With Momentum (SGDM)

O método SGDM foi projetado para acelerar o aprendizado, mesmo que o gradiente seja ruidoso, com muitas variações ou com valores muito pequenos. O algoritmo acumula uma média móvel do gradiente passado de forma exponencial decrescente. Na Figura 18 é apresentado o comportamento do SGD, em um gráfico bidimensional, em comparação com o SGDM.

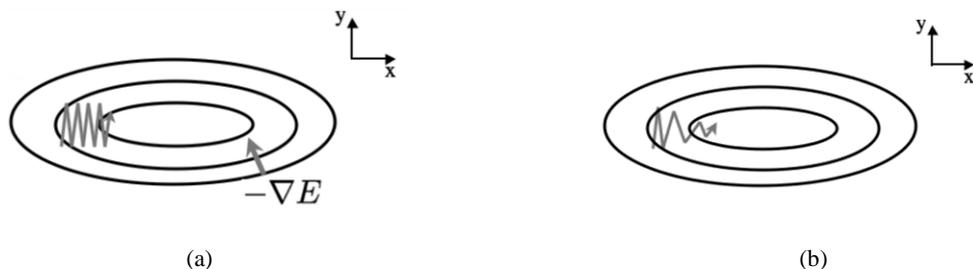


Figura 18 - Comparação do método SGD convencional (a) e SGDM (b). Fonte: (MARSLAND, 2015).

É possível observar que a técnica torna a dinâmica de otimização mais estável, melhorando a convergência, enquanto a técnica SGD apresenta linhas com amplitudes constantes e mais iterações para chegar ao valor mínimo (MARSLAND, 2015).

As equações 26 e 27 a seguir apresentam a atualização do vetor de pesos \bar{W} por meio do SGDM e o vetor \bar{V} que controla a velocidade do gradiente, para uma função de custo $J(W)$. A variável α é a taxa de aprendizado e β é o fator de suavização, com valores entre 0 e 1.

$$\bar{W} \leftarrow \bar{W} + \bar{V} \quad (26)$$

$$\bar{V} \leftarrow \beta \bar{V} - \alpha \frac{\partial J}{\partial \bar{W}} \quad (27)$$

Assim, valores elevados de β contribuem para uma velocidade constante de \bar{V} na direção correta. Quando o valor de β é zero, há o retorno para o algoritmo do SGD convencional. O parâmetro β também é referido como momento, analogicamente a física clássica em que um gradiente negativo é uma força que move uma partícula através do espaço de parâmetros (GOODFELLOW; BENGIO; COURVILLE, 2016).

Pode ser mostrado que, além de reduzir as oscilações laterais, conforme mostrado na figura anterior, o gradiente descendente baseado em momento acelera a taxa de convergência em regiões de gradiente constante.

3.6.3 Adaptive Moment Estimation (ADAM)

O ADAM é um algoritmo de otimização que combina os métodos *Root Mean Square Propagation* (RMSProp) e SGDM, com algumas modificações. No RMSProp é acumulado uma média de decaimento exponencial do quadrado do gradiente. O decaimento é dado pela taxa ρ . A variável r acumula os valores das médias, com valor inicial igual a zero, conforme a equação a seguir:

$$r \leftarrow \rho r + (1 - \rho)g \odot g \quad (28)$$

No otimizador ADAM, são incorporados dois estimadores de momento, ρ_1 e ρ_2 , com valores entre 0 e 1. Esses estimadores são aplicados no cálculo gradiente descendente e acumulados nas variáveis s e r , como mostram as equações a seguir.

$$s \leftarrow \rho_1 s + (1 - \rho_1)g \quad (29)$$

$$r \leftarrow \rho_2 r + (1 - \rho_2)g \odot g \quad (30)$$

Em seguida, são realizados ajustes nos estimadores calculados, no tempo t , por meio das equações:

$$\hat{s} \leftarrow \frac{s}{1 - \rho_1^t} \quad (31)$$

$$\hat{r} \leftarrow \frac{r}{1 - \rho_2^t} \quad (32)$$

A superioridade do otimizador ADAM é demonstrada no trabalho pioneiro dos autores Kingma e Ba (2014), onde o desempenho do algoritmo é comparado com otimizadores mais populares no campo de aprendizagem de máquina, dentre eles o AdaGrad (*Adaptive Gradient Algorithm*) e RMSProp, aplicados em uma rede neurais multicamadas e uma rede convolutiva. Assim, esse método tem se mostrado adequado para cenários em que há um grande conjunto de dados ou parâmetros de alta dimensão.

4 MATERIAIS E MÉTODOS

Neste capítulo serão apresentados os materiais e métodos utilizados para o desenvolvimento deste trabalho. A Seção 4.1 apresenta uma visão geral dos procedimentos e das técnicas aplicadas. A Seção 4.2 apresenta a modelagem da base de dados, incluindo a análise das variáveis e o pré-processamento aplicado. A Seção 4.3 apresenta as arquiteturas desenvolvidas para treino, teste e validação. A Seção 4.4 expõe o ambiente de simulação definido para a realização dos experimentos. A Seção 4.5 descreve a configuração do ambiente de simulação para a realização dos experimentos. Por fim, a Seção 4.6 apresenta as métricas adotadas para análise de desempenho.

4.1 Introdução

Prodanov (2013) define o método científico com um conjunto de processos e operações mentais que devem ser empregadas em uma investigação científica, sendo uma linha de raciocínio adotada no processo de pesquisa.

Neste trabalho, adota-se o método hipotético-dedutivo, que se inicia com uma lacuna que precisa ser preenchida no conhecimento científico, seguida pela formulação de hipóteses e por inferência dedutiva. Além disso, adota-se a abordagem quantitativa, definida por Severino (2013) como uma relação de causa e efeito que pode ser medida por meio de uma função matemática. Quanto aos procedimentos técnicos, é aplicado o modelo experimental que, segundo Severino (2013), consiste em submeter os objetos de estudo à influência de variáveis, em condições controladas e conhecidas pelo investigador, observando os resultados obtidos.

A Figura 19 apresenta, em diagrama de blocos, as etapas utilizadas nesse trabalho para o desenvolvimento do sistema proposto com o objetivo de prever a demanda de energia para o próximo dia, no cenário da COVID-19.

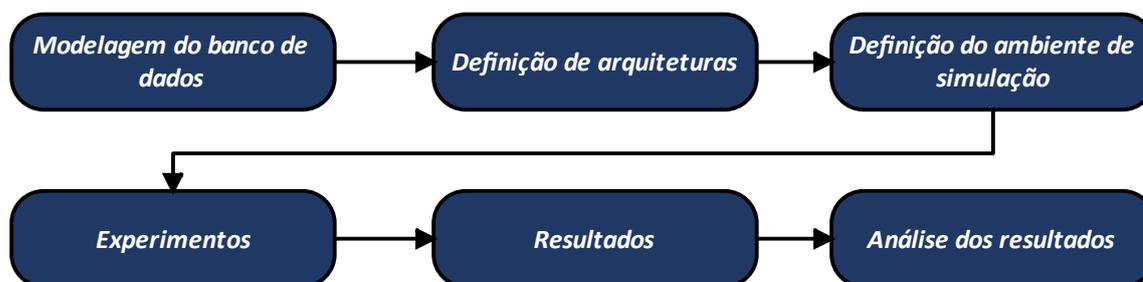


Figura 19 - Etapas de desenvolvimento do sistema proposto.

4.2 Modelagem do banco de dados

O banco de dados utilizado nesse trabalho foi o disponibilizado pela competição “*IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*”, realizado no período de dezembro de 2020 a maio de 2021. Essa competição teve o objetivo de promover uma análise detalhada dos impactos causados pelas medidas relacionadas a COVID-19 na demanda de energia elétrica. Para isso, foi disponibilizado um banco de dados, contendo informações de demanda de energia e variáveis climáticas de uma concessionária de energia (FARROKHABADI et al., 2022).

Os valores de demanda e das variáveis meteorológicas foram disponibilizados com intervalos de amostragem de 1 hora, no período de 18 de março de 2017 a 16 de fevereiro de 2021. Foram disponibilizadas também as previsões meteorológicas do dia seguinte, com 16 a 40 horas de antecedência. A Tabela 1 detalha as variáveis ofertadas, por nome, unidade e característica temporal.

Tabela 1 - Variáveis disponibilizadas pela competição.

Variável	Unidade	Característica temporal
Demanda	kW	Valor real
Pressão do ar	kPa	Valor real e previsão meteorológica
Cobertura de Nuvens	%	Valor real e previsão meteorológica
Umidade	%	Valor real
Temperatura	°C	Valor real e previsão meteorológica
Direção do vento	graus	Valor real e previsão meteorológica
Velocidade do vento	km/h	Valor real e previsão meteorológica
Tempo	Ano-Mês-Dia / HH-MM-SS	Valor real

Fonte: (FARROKHABADI et al., 2022).

Os dados disponibilizados de cada variável foram analisados separadamente, através da observação do comportamento gráfico das mesmas ao longo do tempo. Tal procedimento foi realizado com o objetivo de identificar valores discrepantes, valores ausentes ou ruídos presentes nos dados. Como exemplo, na Figura 20 a variável demanda é apresentada, em Gigawatts. É possível identificar que, no trecho hachurado em vermelho claro, ocorreu uma queda no consumo de energia, como consequência do período de *lockdown*. Esse período está compreendido entre fevereiro a junho de 2020, aproximadamente. A região vermelha escura

corresponde ao maior impacto na demanda do período de *lockdown*, ocorrendo no mês de março de 2020.

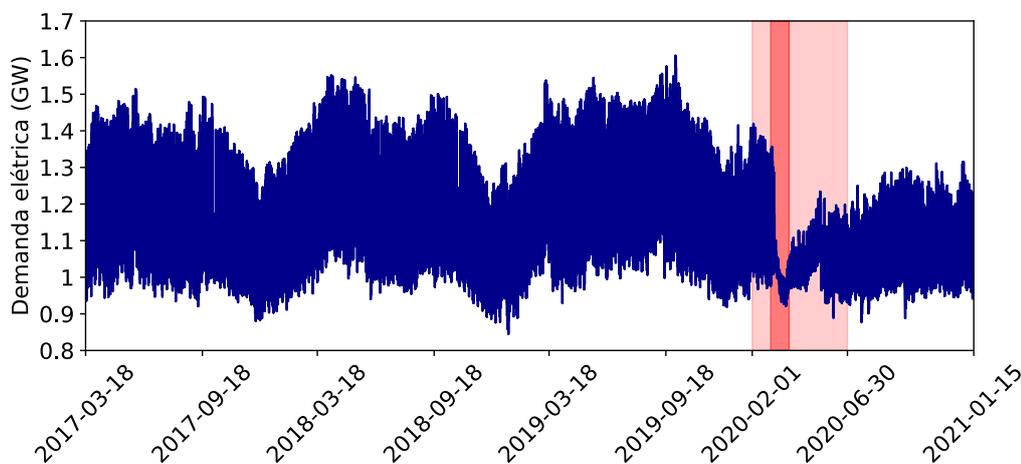
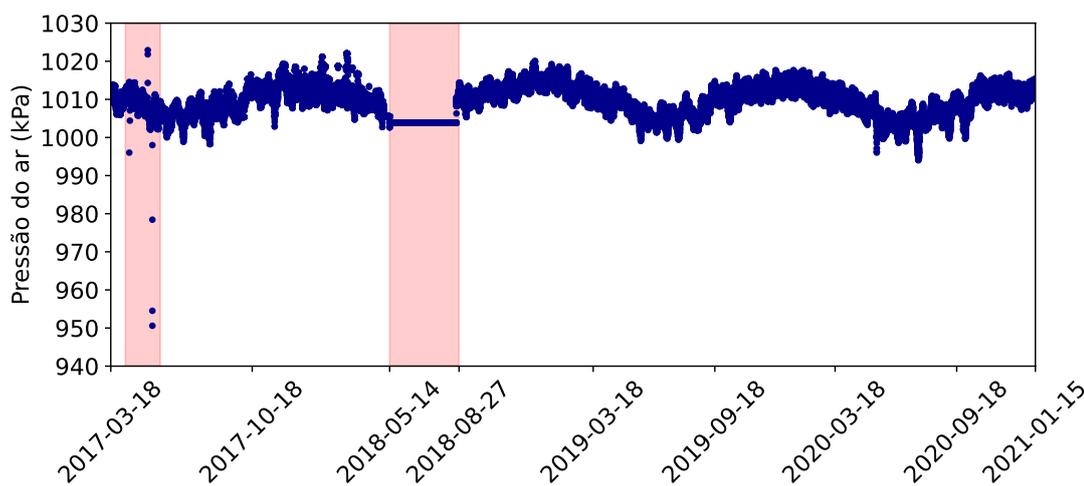
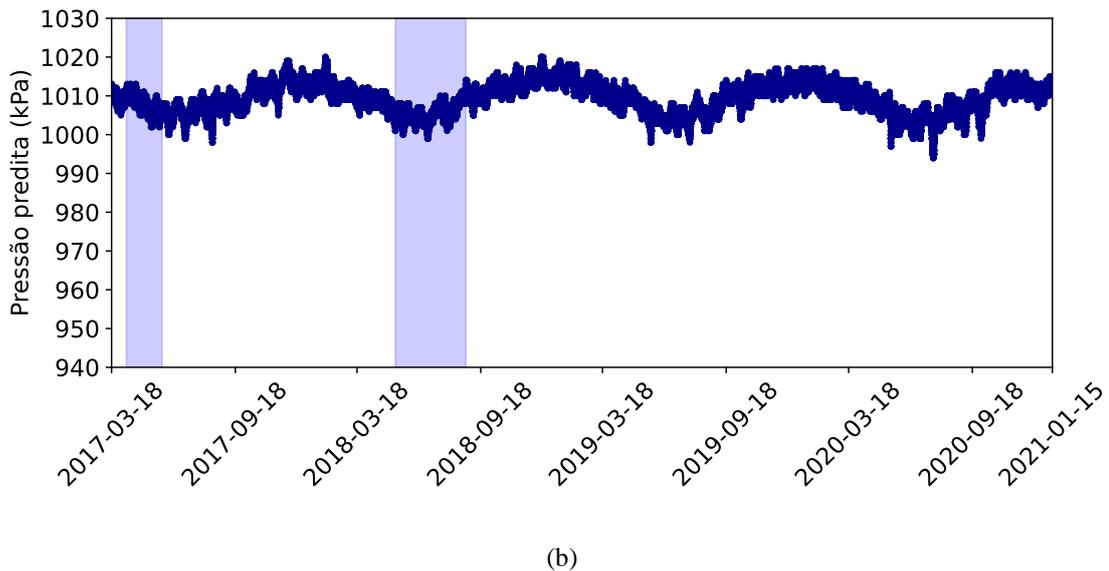


Figura 20 - Curva de demanda de eletricidade disponibilizada. Fonte: (FARROKHABADI et al., 2022) .

É observado que a variável pressão do ar possui valores medidos atípicos, como mostram as regiões hachuradas em vermelho claro na Figura 21(a). A região hachurada que se encontra mais à esquerda mostra valores de pressão do ar discrepantes, enquanto a região hachurada mais à direita mostra que, durante um período de 3 meses, a pressão medida permaneceu constante. A Figura 21(b) apresenta o valor de pressão do ar previsto pela previsão meteorológica, destacando por meio da região hachurada azul clara, o mesmo período de análise.



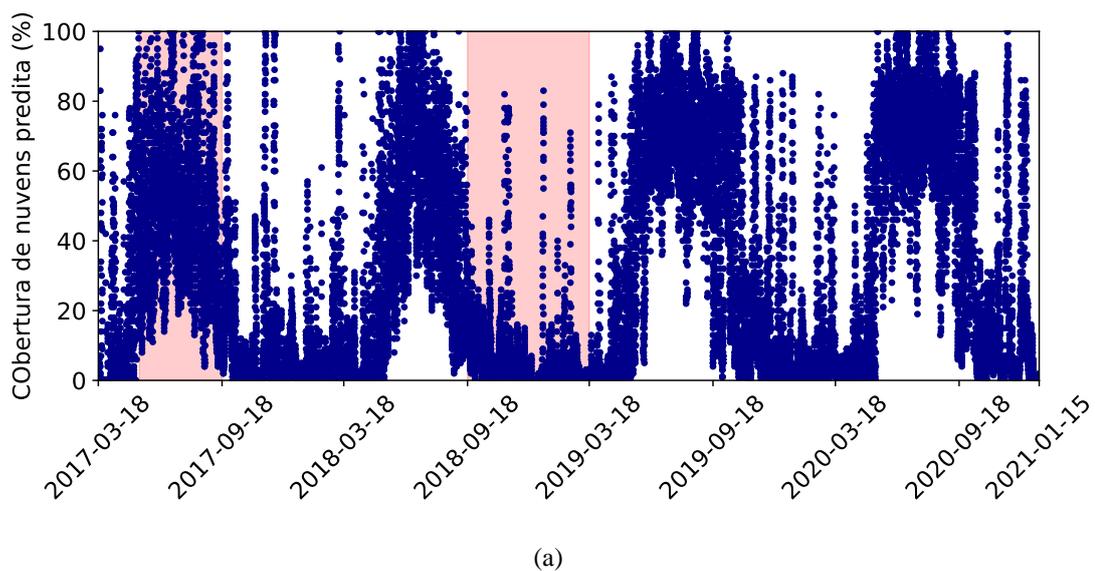
(a)



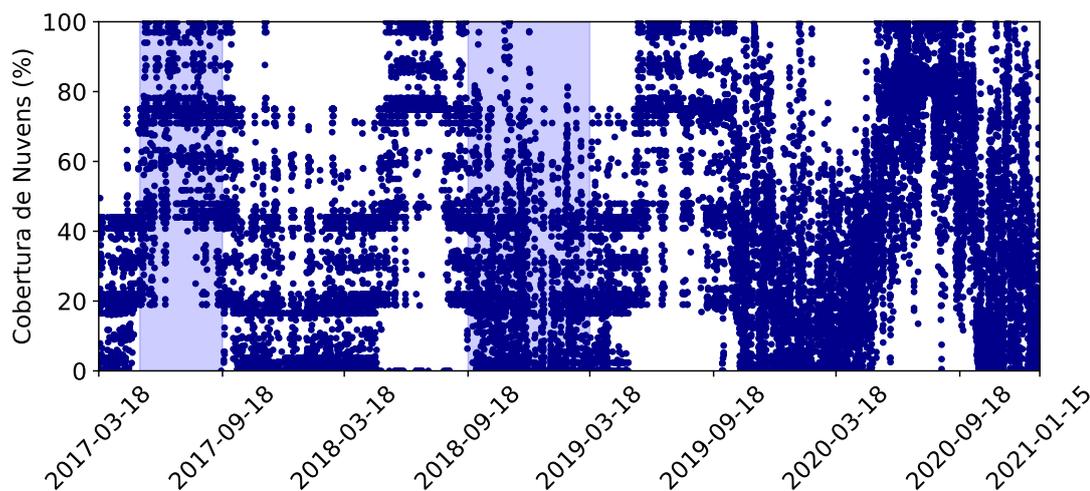
(b)

Figura 21 - Curvas de pressão disponibilizadas. (a) Pressão do ar real. (b) Pressão do ar pela previsão meteorológica. Fonte: (FARROKHABADI et al., 2022).

As variáveis meteorológicas previstas também possuem valores discrepantes no banco de dados. A cobertura de nuvens prevista, por exemplo, apresenta uma densidade de valores diferente se comparado ao valor medido, conforme mostrado através das regiões hachuradas em vermelho claro, na Figura 22(a). A região hachurada mais à esquerda apresenta uma cobertura de nuvens com densidade de valores concentrados entre 20% e 80%, no período de maio a setembro de 2017. A região hachurada mais à direita apresenta a densidade de cobertura de nuvens concentrada entre 0% e 20%, com alguns pontos fora da curva chegando a 80%, no período de setembro de 2018 a março de 2019.



(a)

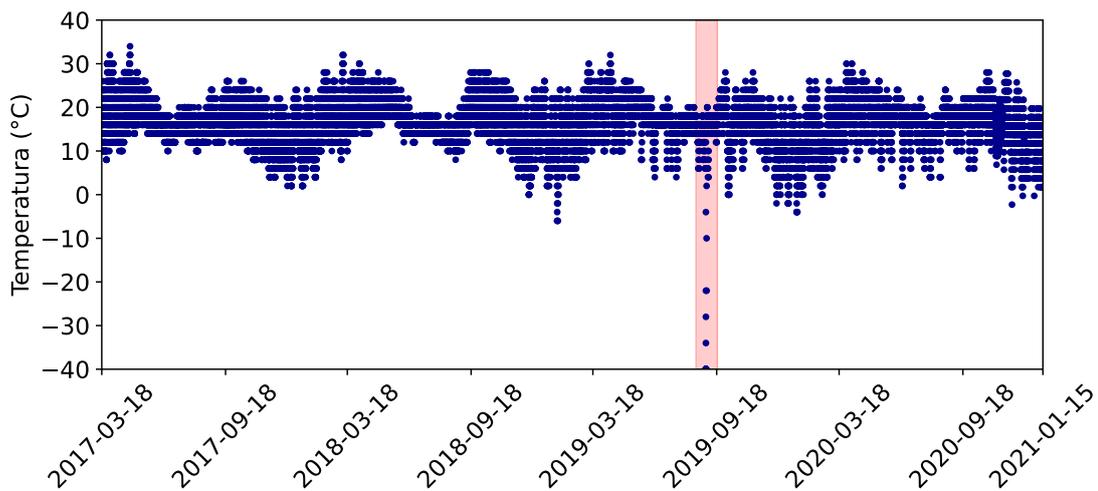


(b)

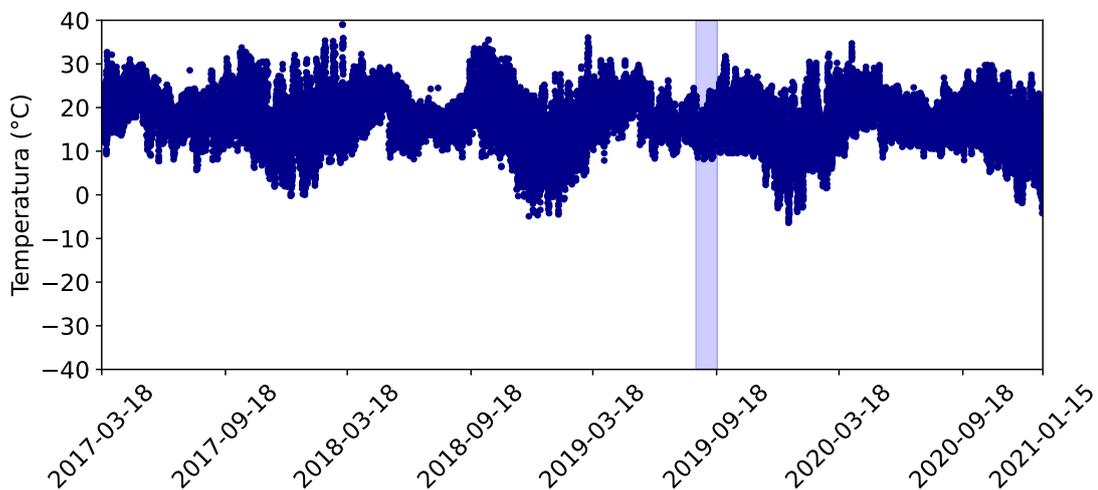
Figura 22 - Curvas de cobertura de nuvens do banco de dados. (a) Cobertura de nuvens predita. (b) Cobertura de nuvens real. Fonte: (FARROKHABADI et al., 2022).

Comparativamente, observa-se uma divergência de valores previstos e medidos. Na Figura 22(b) é identificado, por meio da região hachurada azul clara, o mesmo período de análise da Figura 22(a). Os valores de cobertura de nuvens medidos estão mais concentrados entre 80% e 70%, na região hachurada à esquerda. Na região mais à direita, os valores possuem um comportamento esparsos, se comparados ao mesmo período da Figura 22(a).

A curva da variável temperatura predita é apresentada na Figura 23(a). Nota-se que há presença de valores discrepantes, no período de início de setembro de 2019, no qual a temperatura alcança o valor de $-40\text{ }^{\circ}\text{C}$, assinalado na região vermelha. Além disso, observa-se que há valores não preenchidos ao longo da curva. Na Figura 23(b) é apresentado a temperatura medida, com as regiões hachuradas no mesmo período em azul claro. Não se observa a ocorrência de temperaturas próximas ao valor de $-40\text{ }^{\circ}\text{C}$, nem a ausência de valores ao longo da curva.



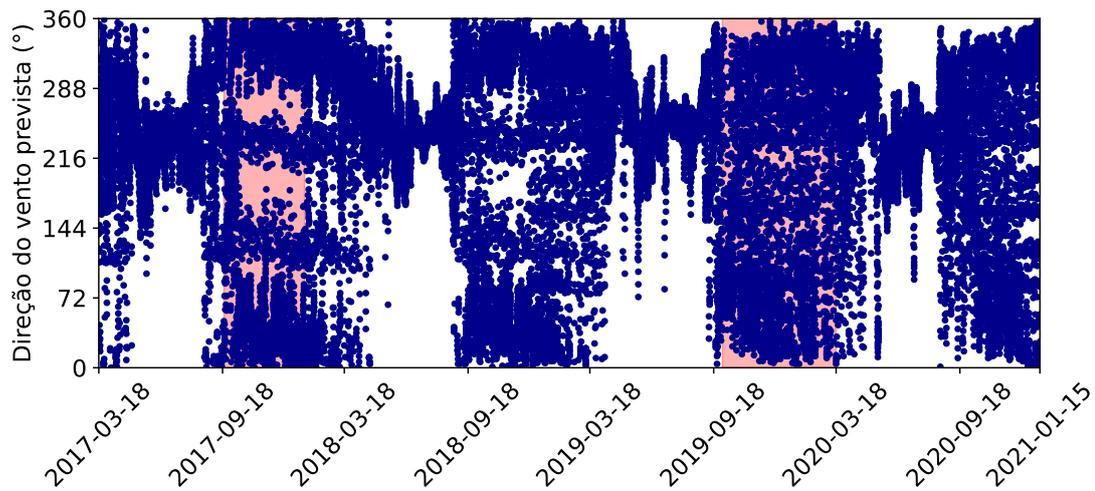
(a)



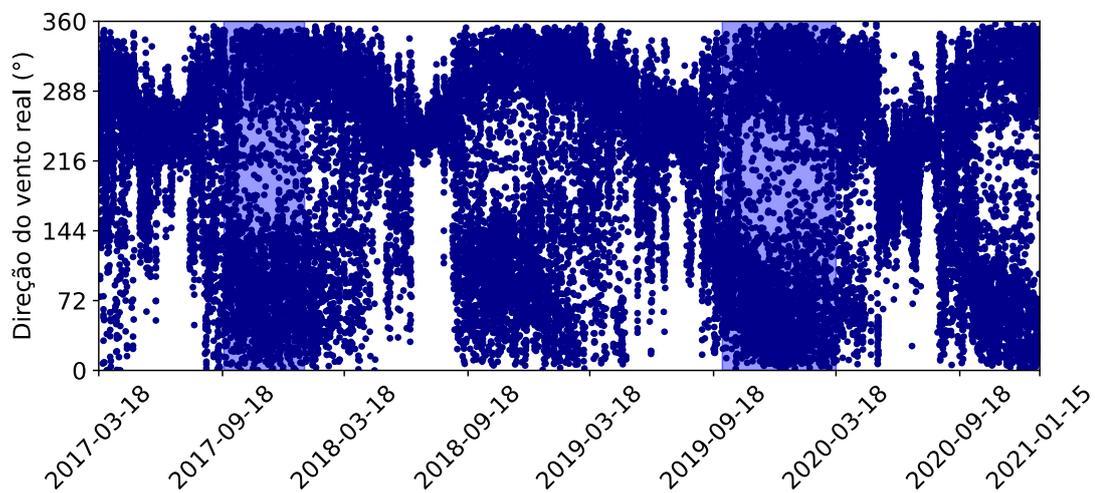
(b)

Figura 23 - Curvas de temperatura do banco de dados. (a) Temperatura predita. (b) Temperatura medida. Fonte: (FARROKHABADI et al., 2022).

A previsão e o valor medido da direção do vento são apresentados na Figura 24(a)-(b), respectivamente. É possível observar um comportamento divergente em ambas as curvas, em determinados períodos. Na Figura 24(a), a região hachurada mais à esquerda possui uma presença menor de valores compreendidos nas regiões vizinhas de 72° e no intervalo de 216° a 288° , no período de setembro de 2017 a janeiro de 2018. Na região hachurada mais à direita, os valores entre 216° e 288° são bem preenchidos. Em contrapartida, na Figura 24(b), a região hachurada à esquerda possui maior densidade de valores para o mesmo período da Figura 24(a). Também se observa na Figura 24(b) que na região hachurada à direita a densidade dos valores é menor para o mesmo período da Figura 24(a).



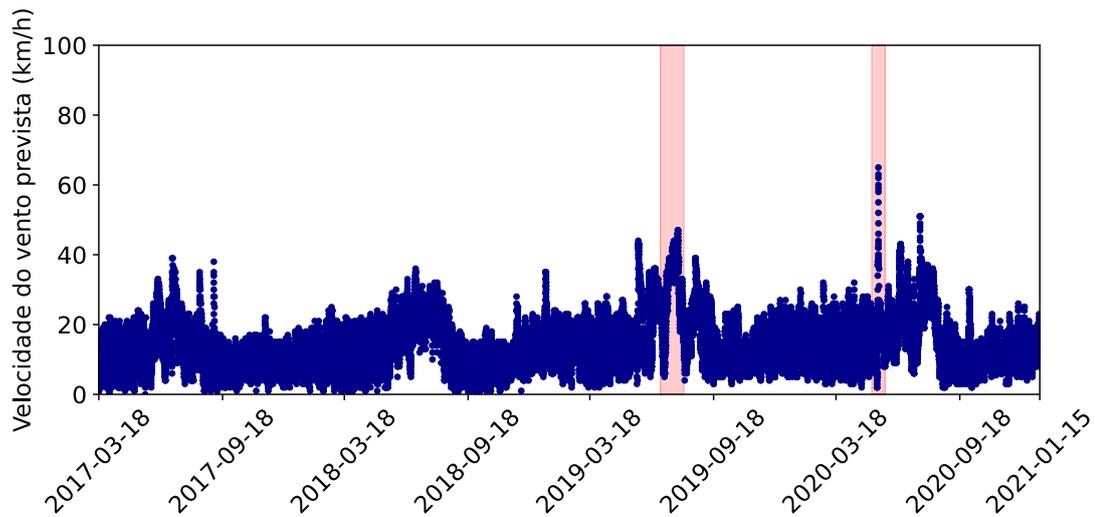
(a)



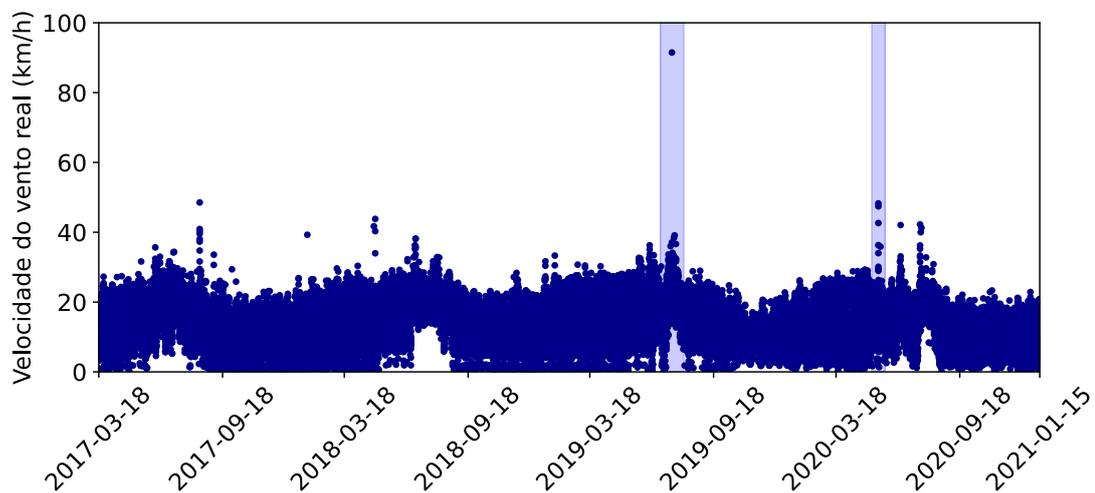
(b)

Figura 24 - Curvas de direção do vento do banco de dados. (a) Direção do vento prevista. (b) Direção do vento medida. Fonte: (FARROKHABADI et al., 2022).

Na Figura 25 são apresentadas as curvas de velocidade do vento (a) prevista e (b) medida respectivamente. Nota-se um comportamento divergente em alguns intervalos de cada curva. Na Figura 25(a), a região hachurada mais à esquerda apresenta uma quantidade baixa de valores menores que 20 km/h, no intervalo de julho a agosto de 2019. No início do mês de maio, atinge o valor próximo de 70 km/h, na região hachurada mais à direita. Esse mesmo intervalo foi hachurado de azul claro na Figura 25(b), no qual observa-se uma maior quantidade de valores próximos a 20 km/h e um ponto fora da curva próximo a 90 km/h, na região mais à esquerda. Na região à direita, observa-se que o valor máximo atingido é em torno de 50 km/h.



(a)



(b)

Figura 25 - Curvas de velocidade do vento do banco de dados. (a) Velocidade do vento prevista. (b) Velocidade do vento real. Fonte: (FARROKHABADI et al., 2022).

Assim, devido à falta de preenchimento de alguns dados e a discrepância de valores preditos pela meteorologia, bem como às anomalias observadas nos dados medidos, a variável pressão não foi utilizada na modelagem feita. Foram selecionadas apenas as variáveis demanda (kW), cobertura de nuvens (%), umidade (%), temperatura (°C), direção do vento (graus) e velocidade do vento (km/h).

Além disso, para o treinamento das arquiteturas de redes profundas propostas nesse estudo, foi observado também a necessidade de identificar o período de *lockdown*. Dessa forma, foi adicionada uma variável no banco de dados, chamada “*lockdown*”, que assume o valor 0 nos períodos normais e o valor 1 nos períodos de *lockdown*, considerando a data e hora de ocorrência do evento.

Ao observar-se os valores das variáveis, registra-se a ocorrência de diferentes faixas dinâmicas. Por exemplo, a carga varia entre 845,471 MW e 1605,192 MW. A temperatura varia entre -6.42 °C e 39.07 °C. Uma convergência mais rápida dos algoritmos de otimização de redes profundas é obtida através da normalização das variáveis utilizadas na entrada das mesmas (MOHAMMAD; AHMED; KIM, 2021).

Assim, um procedimento de normalização foi aplicado a todas as variáveis, conforme mostrado na equação 33. Após o procedimento de normalização, os valores de demanda foram multiplicados por 10. Portanto, a faixa dinâmica da demanda foi [0,10], enquanto a faixa dinâmica das demais variáveis foi [0,1]. Como resultado, a demanda de eletricidade tornou-se a característica mais significativa na entrada da rede.

$$x_{novo} = \left(\frac{x_{antigo} - x_{min}}{x_{max} - x_{min}} \right) \quad (33)$$

Em seguida, um método de janela deslizante foi empregado com o objetivo de estabelecer o período de antecedência, em horas, utilizado para prever a demanda da próxima hora. O tamanho de janela Δ_t utilizada foi de 24 horas. A Figura 26 ilustra o método aplicado. Os blocos verdes contêm os dados para predição correspondentes a Δ_t horas. O bloco laranja representa a hora em que a demanda de energia é predita, correspondente a hora $t_p = t_i + \Delta_t$, em que t_i é a hora inicial da janela. Observa-se ainda, que entre uma janela e a janela seguinte, ocorre a superposição de dados de previsão de $\Delta_t - 1$ horas. Considerando a hora inicial de uma janela como t_i , a previsão para essa janela é feita na hora $t_p = t_i + \Delta_t$.

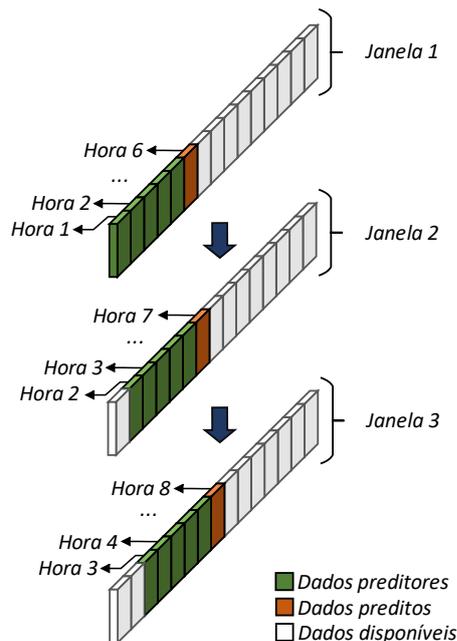


Figura 26 - Método de janela deslizante aplicado nos dados de entrada.

4.3 Definição das arquiteturas

Três arquiteturas de redes neurais profundas são propostas nesse estudo para predição da demanda no contexto da pandemia. Essas arquiteturas combinam redes neurais convolucionais e redes neurais recorrentes.

4.3.1 Arquitetura 1 (CNN)

Os dados de entrada da rede possuem dimensão $33577 \times 24 \times 6$. Cada linha é constituída pela quantidade de horas amostradas pela janela deslizante, totalizando 33577 linhas com 24 horas cada. Esse procedimento foi aplicado em todas as variáveis de entrada e resultou em 6 matrizes contendo os valores de demanda, temperatura, velocidade do vento, direção do vento, cobertura de nuvens e *lockdown*. Essas matrizes foram empilhadas e constituem a profundidade dos dados de entrada.

Na primeira arquitetura, uma rede neural convolutiva, 4 camadas convolucionais e função de ativação ReLU foram utilizadas para extração de características. As duas primeiras camadas convolucionais possuem dimensão 24×100 e são conectadas sequencialmente. Em seguida, há uma camada *max pooling* com *kernel* de tamanho 2×2 e passo de 2. Posteriormente, duas camadas convolucionais de dimensão 12×100 são conectadas em sequência. A última camada é constituída por um neurônio com função de ativação linear. A Figura 27 apresenta o diagrama de blocos dessa arquitetura.

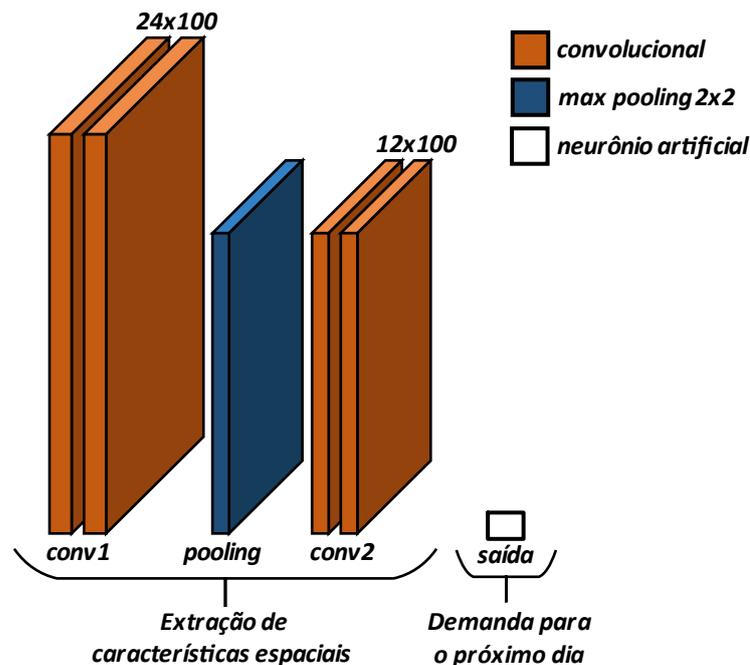


Figura 27 - Primeira arquitetura de rede neural profunda proposta.

Com essa arquitetura, pretende-se avaliar o desempenho da CNN na predição da demanda de eletricidade. Nessa rede, as características são extraídas através das camadas convolucionais, enquanto a predição é feita pela camada inteiramente conectada formada por apenas pelo neurônio de saída.

4.3.2 Arquitetura 2 (LSTM)

Os dados de entrada para treinamento da arquitetura 2 possui dimensão $33577 \times 24 \times 6$, assim como na arquitetura 1. A segunda arquitetura consiste em uma rede neural recorrente, composta por duas camadas LSTMs de extração de características. A primeira camada contém 64 células, seguida da segunda camada com 32 células, sequencialmente. Em seguida, uma camada de linearização (*reshape*) é aplicada e a dimensão dos dados é convertida para 1×32 . A última camada, também é constituída por 1 neurônio. A Figura 28 apresenta essa arquitetura em diagrama de blocos.

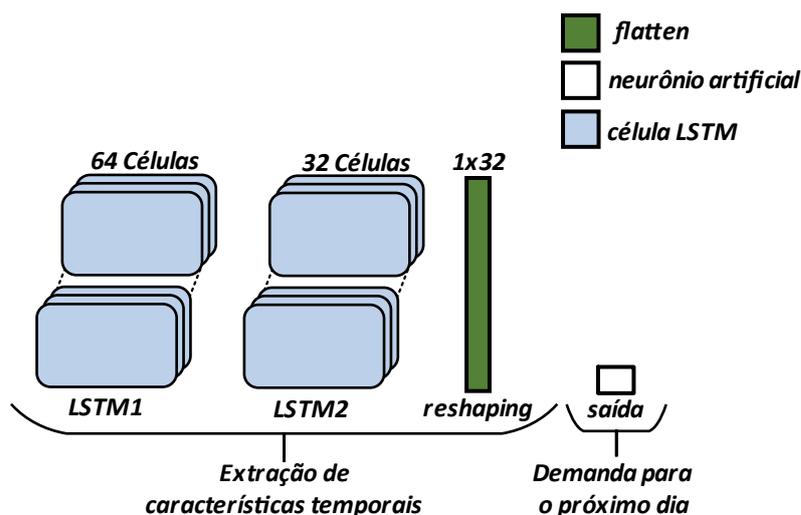


Figura 28 - Segunda arquitetura de rede neural profunda proposta.

O objetivo dessa arquitetura é avaliar o desempenho de um modelo na extração de características temporais dos dados de entrada. Essa estratégia justifica-se devido a variação da demanda de eletricidade ao longo do tempo. É possível observar essa relação durante um dia de consumo, por exemplo, em que a demanda tende a ser maior durante o horário comercial.

4.3.3 Arquitetura 3 (CNN + LSTM)

Assim como na arquitetura 1 e 2, os dados de entrada possuem dimensão $33577 \times 24 \times 6$. Essa terceira arquitetura é mista, convolucional-recorrente. Trata-se de uma combinação das arquiteturas 1 e 2, com as 4 camadas convolucionais, seguidas pelas 2 camadas recorrentes. As

dimensões das camadas são mantidas iguais às anteriores, sendo as convolucionais iguais 24×100 e 12×100 e as recorrentes iguais a 64 células e 32 células, respectivamente. Utiliza-se também uma camada *max pooling* com *kernel* de tamanho 2×2 e passo de 2, entre as camadas convolucionais. Por fim, uma camada de linearização (*reshape*) e um neurônio de saída, com ativação linear, são acrescentados. A Figura 29 apresenta a estrutura dessa arquitetura.

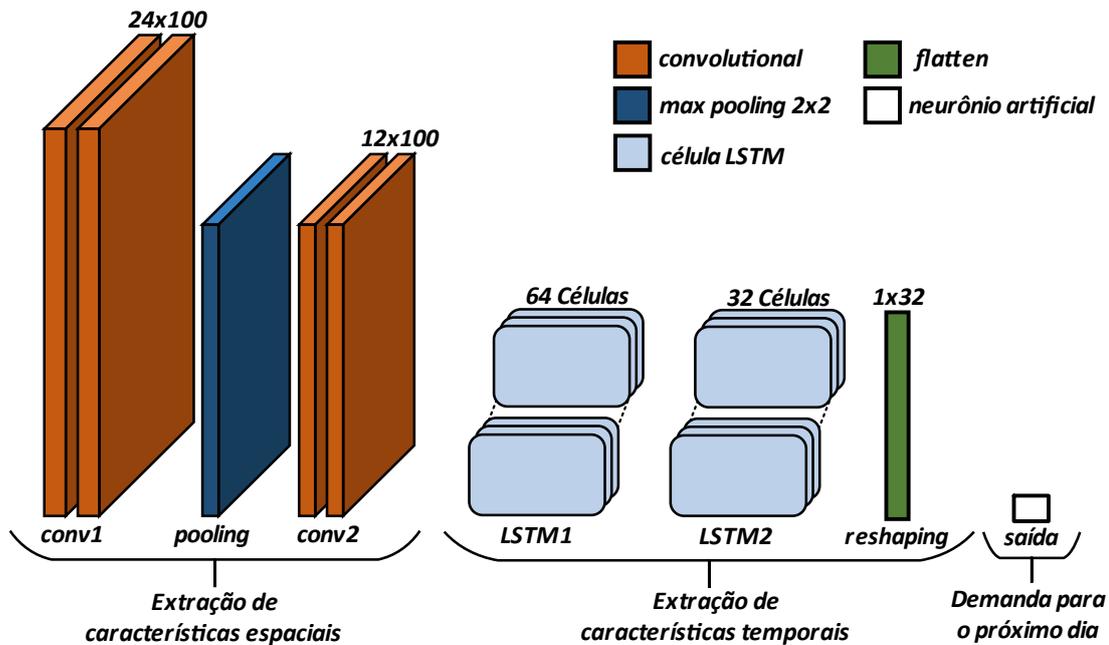


Figura 29 - Terceira arquitetura de rede neural profunda proposta.

Com essa arquitetura, busca-se analisar o desempenho de um modelo em que se realiza tanto a extração de características espaciais como a extração de características temporais dos dados de entrada. A combinação das duas técnicas constitui um modelo híbrido e tem se mostrado promissora em trabalhos relacionados a predição de demanda de energia (GHALEHKHONDABI et al., 2017; NTI et al., 2020; RAZA; KHOSRAVI, 2015).

4.4 Definição do ambiente de simulação

O sistema computacional empregado para as simulações consistiu em um processador do modelo Intel Core i5-11400H, com *clock* de 2.70GHz, 16 GB de memória RAM, GPU GeForce GTX 1650 de 4 GB e 500 GB de SSD.

O ambiente colaborativo Google Colab foi utilizado para a realização da modelagem do banco de dados, realização dos experimentos e análise de desempenho. Adotou-se a linguagem de programação *python*. Foram importados os pacotes Tensorflow v2.9.1, Keras v2.9.0, Pandas v1.3.5, Numpy v1.21.6 e Matplotlib v3.2.2.

4.5 Experimentos

Com a modelagem da base de dados, onde foram realizadas a análise gráfica dos dados de entrada, seleção de características e adição de rótulo do período de *lockdown*, os experimentos de cada arquitetura foram iniciados.

Nesta etapa de treinamento, realiza-se um estudo de ablação, através da variação dos parâmetros utilizados no treinamento das redes profundas. Tal estudo tem como objetivo obter a rede com um melhor desempenho. Os parâmetros utilizados nas três arquiteturas foram os mesmos, visando a realização do *benchmark* e seleção da melhor arquitetura. Na Figura 30, são apresentados os parâmetros que foram testados e avaliados durante os primeiros treinamentos dos modelos, dentre eles o método de otimização, o tipo de regularização, o tamanho do minilote e a taxa de aprendizado. Os blocos na cor verde destacam os parâmetros que produziram os melhores resultados.

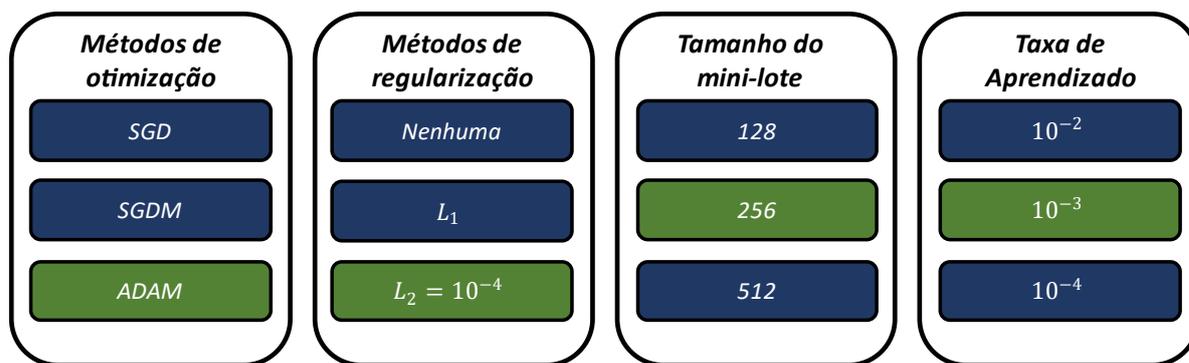


Figura 30 - Parâmetros variados durante o treinamento das redes de aprendizado profundo. Os blocos verdes foram os métodos e valores adotados.

O critério de parada aplicado durante a aprendizagem foi o número de épocas do treinamento, que foi igual a 300. Foi adotada uma função auxiliar que supervisiona o desempenho da arquitetura e retorna a rede com o melhor desempenho, ao término de todas as iterações. A Figura 31 apresenta a sistemática dessa função. No final de cada treinamento, os dados de validação são aplicados e o desempenho da rede atual é comparado com o desempenho da rede anterior. Caso o desempenho seja superior, a função auxiliar substitui o modelo anterior pelo modelo atual. Dessa forma, os pesos e polarizações do modelo são salvos e os próximos treinamentos não interferem no modelo salvo, até que um desempenho superior seja alcançado.

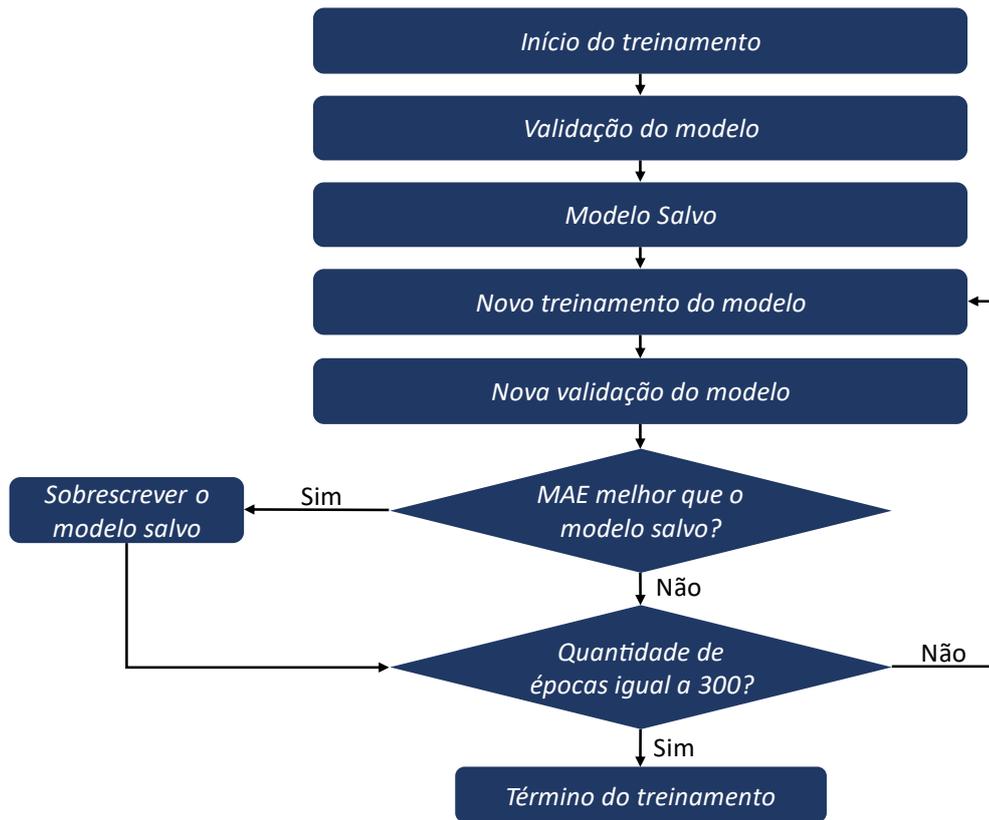


Figura 31 - Sistemática de funcionamento da função auxiliar implementada, que supervisiona o desempenho da arquitetura e retorna a rede com o melhor desempenho.

4.6 Métricas de desempenho

Com o objetivo de viabilizar o *benchmark* com os modelos da competição, a métrica MAE foi aplicada para avaliação de desempenho das arquiteturas propostas anteriormente. A definição dessa métrica é feita através da equação a seguir, em que x_{real} é o valor esperado, $x_{predito}$ é o valor predito pelo modelo e n é a quantidade de dados observados.

$$MAE = \frac{\sum |x_{real} - x_{predito}|}{n} \quad (34)$$

O MAE é monitorado durante o treinamento e validação das redes profundas através da função auxiliar. Na etapa de teste, o valor MAE obtido é comparado com os resultados dos modelos desenvolvidos pelos participantes da competição do IEEE.

5 RESULTADOS E DISCUSSÕES

Esta seção apresenta o desempenho das três arquiteturas propostas. São mostrados os resultados obtidos da etapa de treinamento dos modelos, incluindo as validações e testes, utilizando o banco de dados disponível. Os resultados também são discutidos e uma análise comparativa é apresentada.

5.1 Resultados da arquitetura 1

A Figura 32 apresenta a curva de aprendizado da arquitetura 1. O eixo x representa a quantidade de épocas e o eixo y o valor MAE da demanda normalizada e com faixa dinâmica de 0 a 10. As simulações das 3 arquiteturas foram realizadas com os mesmos parâmetros de treinamento, com taxa de aprendizado de 10^{-3} , minilote de 256, regularização L_2 de 10^{-4} e otimizador ADAM. A cor verde tracejada indica a perda no treinamento do modelo e a curva azul indica a perda na validação.

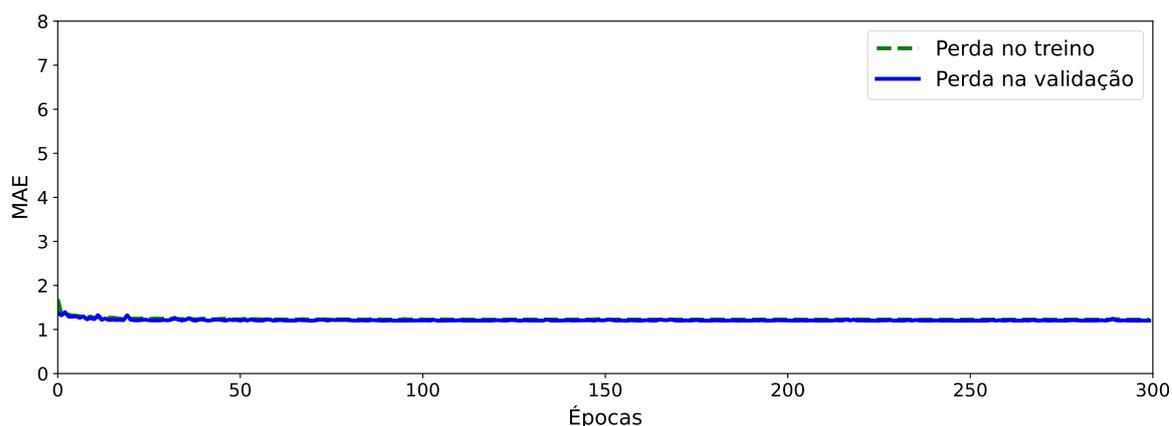


Figura 32 - Gráfico de convergência da arquitetura 1 proposta.

Observa-se que a arquitetura 1 apresentou uma curva de aprendizado estável, com os valores da métrica MAE bem próximos, nos conjuntos de validação e treinamento. Contudo, nesse treinamento, o valor final da métrica MAE ainda ficou com valor acima de 1. Ao converter esse resultado para a escala original, obtêm-se o valor de 93968,08 kW. Esse erro é elevado se comparado aos melhores resultados obtidos pelos competidores, como foi apresentado no Quadro 3.

5.2 Resultados da arquitetura 2

A Figura 33 apresenta o desempenho da arquitetura 2 proposta. Observa-se uma pequena instabilidade nas primeiras épocas, porém o valor final da métrica MAE situou-se

próxima a 0. Porém, observa-se um aumento da métrica MAE no conjunto de validação, após 75 épocas, caracterizando o *overfitting* do modelo.

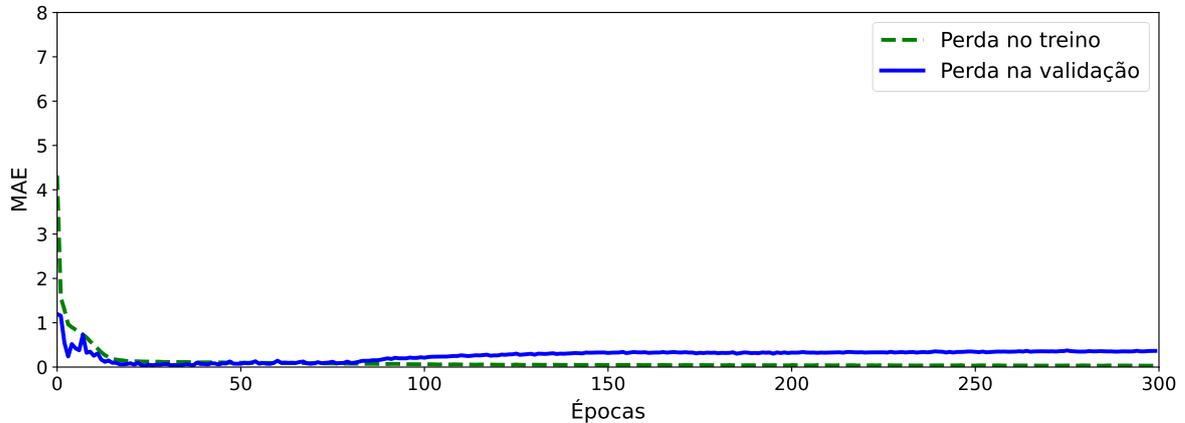


Figura 33 - Gráfico de convergência da arquitetura 2 proposta.

Ao converter o resultado MAE para a escala original, têm-se o valor de 2530,15 kW. Esse resultado supera os valores obtidos pelos primeiros colocados na competição. Contudo, o desempenho do modelo durante o aprendizado mostra que é necessário a implementação de melhorias para evitar o sobreajuste do modelo.

5.3 Resultados da arquitetura 3

A Figura 34 apresenta o desempenho de treinamento e validação da arquitetura 3. É observado uma estabilidade da curva nas primeiras épocas de treinamento e o valor da métrica MAE converge para próximo de 0. Esse comportamento é semelhante tanto na curva de treinamento, quanto na curva de validação.

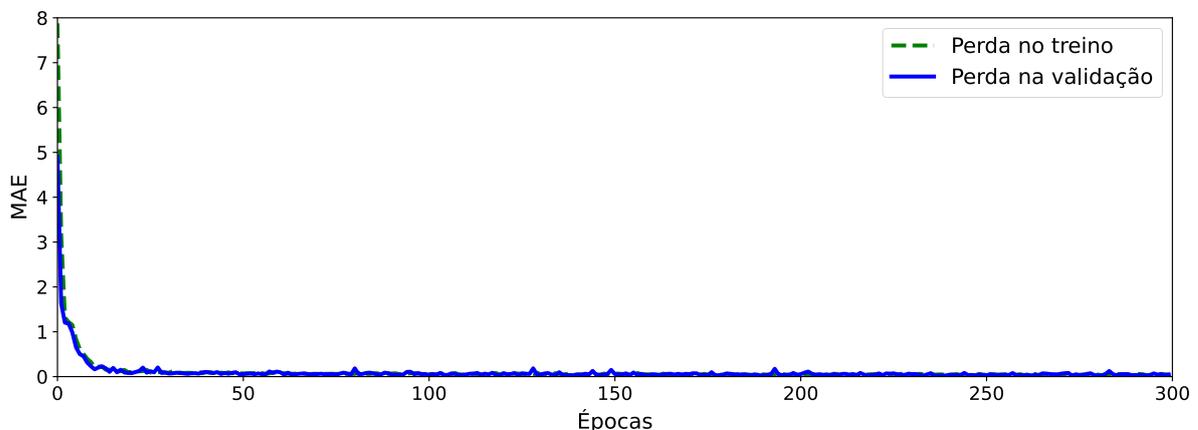


Figura 34 - Gráfico de convergência da arquitetura 3 proposta.

O valor da métrica MAE convertido para a escala original é igual a 2361,84 kW. Esse resultado também é superior aos valores obtidos pelos primeiros colocados na competição.

Além disso, o desempenho do modelo é superior ao da Arquitetura 2, proposta neste trabalho, resolvendo o problema de aumento de erro no conjunto de validação durante o treinamento, o erro de *overfitting*.

A Tabela 2 apresenta uma comparação entre os valores obtidos para a métrica MAE pelos 3 primeiros colocados na competição com os valores alcançados pelas arquiteturas simuladas. Também são apresentados os métodos aplicados por cada autor, comparando com os métodos propostos em cada arquitetura.

Tabela 2 - Desempenho dos modelos da competição e do modelo proposto.

Rank	Método	MAE (kW)
1	Conjunto multi-modelos de aprendizado de máquina (72 modelos)	10844
2	Combinação de dias semelhantes por meio do tipo do dia e temperatura de pico	11849
3	Conjunto multi-modelos de aprendizado de máquina (674 modelos)	11890
Arquitetura 1	CNN	93968,08
Arquitetura 2	LSTM	2530,15
Arquitetura 3	CNN+LSTM	2361,84

A Tabela 3 apresenta um resumo dos valores de demanda preditos pela arquitetura 3 em comparação ao valor real da demanda, em Gigawatts. Observa-se a proximidade dos valores preditos, demonstrando a alta capacidade de predição da rede.

Tabela 3 - Resumo dos valores reais do banco de dados e valores preditos pela arquitetura 3.

Valor real (GW)	Valor predito (GW)
1,009	1,010
0,989	0,990
0,976	0,975
0,967	0,966
0,959	0,955
⋮	⋮
1,187	1,186
1,147	1,147
1,121	1,123
1,101	1,100
1,078	1,077

A Figura 35 apresenta a curva de previsão de demanda diária pela arquitetura 3 e a demanda real do banco de dados no mesmo período. Nota-se a alta similaridade das curvas, indicando uma previsão de carga próxima ao valor esperado. Assim, a arquitetura 3 foi a que obteve o melhor desempenho.

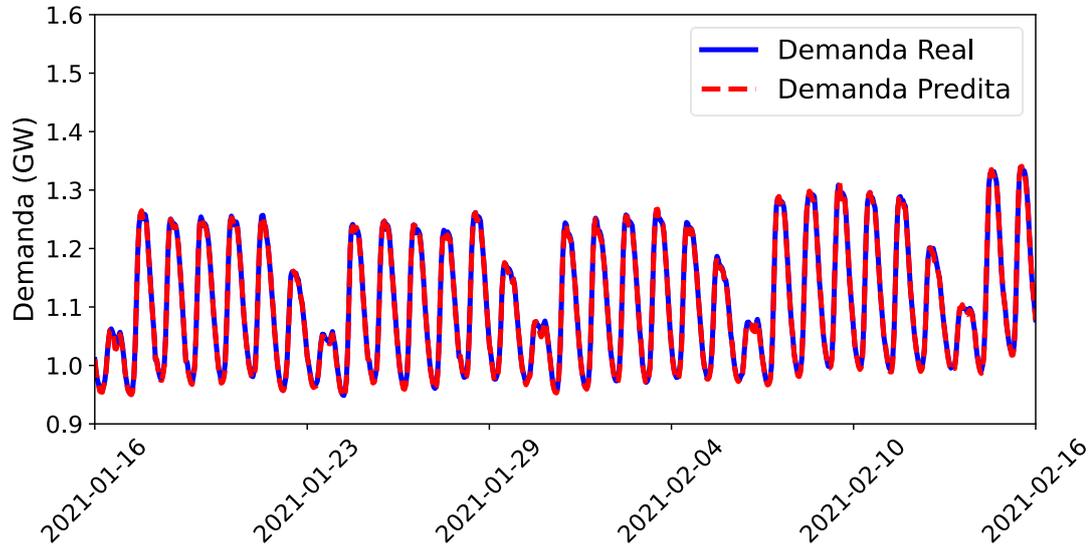


Figura 35 - Curva de previsão de demanda da arquitetura 3. A linha azul é o valor da demanda de energia real e a linha vermelha tracejada é o valor da demanda predita.

5.4 Discussão dos resultados

Os resultados apresentados buscam identificar qual arquitetura possui melhor desempenho na previsão da demanda para o dia seguinte. Na etapa de treinamento, ocorre um ajuste nos pesos da rede, de modo que o valor de perda diminuiu a cada minilote processado, convergindo o MAE para um ponto de mínimo. O método de otimização utilizado foi o método ADAM. Como os pesos são iniciados de forma aleatória, um novo treinamento de uma mesma arquitetura pode resultar em pesos diferentes ao final do experimento. A inicialização aleatória dos pesos no espaço de estados pode resultar em um ponto próximo a um mínimo global ou a mínimos locais. O treinamento em minilote, assim como o treinamento online, confere ao método de otimização a possibilidade do mesmo não ficar preso nos mínimos locais, e, portanto, permite ao método de otimização a busca de um mínimo global.

O processo iterativo da função auxiliar permitiu a identificação e armazenamento do modelo com o melhor resultado. Esse processo realizado dentro do minilote de cada época de treinamento, colaborou para o monitoramento da métrica MAE em tempo real, viabilizando a análise de convergência do modelo antes do término do experimento. Além disso, a taxa de aprendizado, combinada com a regularização das camadas da rede profunda, mostrou ser uma

estratégia adequada ao se observar a suavização da convergência do modelo, durante o treinamento.

Na etapa de modelagem, a análise preliminar dos dados resultou na exclusão de características que apresentavam anomalias, como as variáveis de pressão medida e as demais variáveis preditas pela meteorologia. A aplicação dessa técnica demonstrou ser fundamental para a mitigação de problemas decorrentes da presença de dados espúrios no banco de dados, o que poderia acarretar em um erro maior da demanda predita.

A normalização dos dados de entrada, com a adição de um peso maior para os dados de demanda, foi fundamental para o bom desempenho do modelo. Essa técnica proporcionou a valorização, na entrada das redes, da característica de maior importância e manteve os demais valores de menor importância na mesma escala. Essa estratégia não foi utilizada em outros trabalhos publicados na literatura que se destacaram na competição.

O método de janelamento estabeleceu a quantidade de dados antecedentes necessários para prever a demanda do próximo dia. O período de 24 horas de dados anteriores se mostrou suficiente para a interpretação da dependência temporal pelo modelo, pois a demanda e as variáveis climáticas variam, conforme as horas do dia e ao longo do ano, com alta relação de causalidade.

A arquitetura 1 apresentou uma curva estável de aprendizagem e convergência dos valores logo nas primeiras épocas, não havendo uma evolução no aprendizado da rede nas épocas seguintes. Esse comportamento, resultou em um valor elevado da métrica MAE, ao se comparar com os demais trabalhos da competição. Assim, o desempenho da rede demonstra que a aplicação de redes neurais convolucionais não são suficientes para prever a demanda para o dia seguinte. Esse fato está relacionado a forte relação temporal dos dados de entrada.

A arquitetura 2 apresentou instabilidade nas primeiras épocas do treinamento. Em seguida, nota-se a estabilidade nos valores, entre as épocas 20 e 75, aproximadamente. No entanto, nas épocas seguintes o valor da perda aumenta no conjunto de validação, enquanto o valor da perda no conjunto de treinamento mantém-se próxima de zero. Esse comportamento caracteriza o *overfitting* do modelo e inviabiliza a aplicação dessa arquitetura na previsão da demanda. Apesar do valor MAE obtido ter um valor menor se comparado aos melhores trabalhos da competição, o desempenho do modelo, como um todo, demonstra que essa arquitetura precisa de melhorias.

A arquitetura 3 apresentou um treinamento bem comportado ao longo das 300 épocas utilizadas, sem que tenha ocorrido erro de *overfitting*. O excelente desempenho da métrica MAE no conjunto de validação, durante o treinamento, foi refletido no valores obtidos para a mesma

no conjunto de teste, conforme mostrado na Tabela 2 e 3 e no gráfico da Figura 35. Os resultados obtidos, tanto no conjunto de validação, como no conjunto de teste, comprovam o acerto na escolha de uma arquitetura mista de redes profundas, convolutiva + recorrente, bem como dos hiperparâmetros da mesma, para a predição da demanda diária.

Por fim, ressalta-se que a principal contribuição deste trabalho foi realizar a predição de demanda de energia elétrica para o dia seguinte e comprovar a efetividade da rede mesmo em um cenário pandêmico, como o causado pelas medidas restritivas da COVID-19. A contribuição para o preenchimento de uma lacuna existente na literatura e a avaliação do modelo desenvolvido, de forma comparativa, também justificam a importância desse estudo. A capacidade de predição de demanda quando se utiliza uma rede neural convolutiva e recorrente também é confirmada por meio do resultado da terceira arquitetura proposta.

6 CONCLUSÕES

Neste trabalho, foi proposto um método de predição de demanda para o próximo dia aplicando redes neurais convolutivas e redes recorrentes. Os resultados alcançados, inclusive no cenário pandêmico da COVID-19, fundamentam a contribuição deste trabalho, frente a poucos trabalhos encontrados na literatura que envolvam este cenário e a preocupação global na emissão de CO₂.

A utilização da base de dados da competição “*IEEE DataPort Competition Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*” permitiu a análise *benchmark* dos modelos desenvolvidos neste trabalho. Na análise comparativa, a arquitetura 3 foi superior aos resultados obtidos pelos melhores colocados na competição e sobre as demais arquiteturas propostas neste trabalho. A arquitetura 3 combina a extração de características espaciais e temporais dos dados de entrada. O desempenho superior desse modelo mostra que uma estrutura mista de redes profundas é a melhor técnica para esse cenário.

A aplicação de uma análise prévia das variáveis de entrada permitiu a caracterização do banco de dados e contribuiu significativamente para o resultado alcançado, por meio da exclusão de valores discrepantes e ruídos presentes nos dados. A aplicação da normalização, seguido da ponderação da variável demanda por um fator de 10, possibilitou que o modelo interpretasse essa variável como a mais importante, durante a etapa de treinamento, o que minimizou o erro médio absoluto.

A definição de um intervalo de 24 horas, por meio do método de janelamento, mostrou ser a quantidade de dados satisfatória para predizer a demanda da próxima hora. Além disso, a aplicação de uma função auxiliar foi fundamental para armazenar a matriz de pesos do modelo com o melhor desempenho, possibilitando a realização da etapa de testes e experimentos futuros, não se limitando ao emprego da plataforma Google Colab[®], utilizada nesse trabalho.

Em aplicações futuras, busca-se investigar a aplicação do modelo desenvolvido em outros cenários, por meio da transferência de conhecimento da rede. Nesse contexto, é possível aprimorar a gestão de energia em sistemas vinculados a uma demanda contratada, permitindo a predição da demanda requerida do sistema. Além disso, previsões de pico de demanda podem ser aplicadas, viabilizando o uso de energias renováveis, como energia solar fotovoltaica, para compensar esses períodos críticos. Também pode ser investigada a influência de outras variáveis para a predição da demanda, por meio de uma análise de correlação, o que permite a otimização de recursos computacionais necessários para treinamento do modelo.

Portanto, a aplicação de redes neurais híbridas (convolutivas-recorrente) constitui-se em uma abordagem adequada em aplicações de predição de demanda de energia para o dia

seguinte. Além disso, o modelo proposto obteve um bom desempenho mesmo com anomalias no comportamento na curva de demanda, causados pelas medidas restritivas do *lockdown*. Por fim, quando aplicados em sistemas operacionais de tempo real, o modelo de previsão de demanda desenvolvido pode proporcionar um melhor gerenciamento de recursos e contribuir para o avanço de sistemas energéticos autossustentáveis.

7 REFERÊNCIAS

- AGGARWAL, C. C. **Neural Networks and Deep Learning**. 1. ed. Suíça: Springer Nature, 2018a. v. 1
- AGGARWAL, C. C. **Instructor’s Solution Manual for “Neural Networks and Deep Learning”**. 1. ed. Suíça: Springer Nature, 2018b. v. 1
- AHMAD, M.; JOE, J.; HAN, D. CortexNet: Convolutional Neural Network with Visual Cortex in human brain. 2018.
- ALASALI, F. et al. Impact of the covid-19 pandemic on electricity demand and load forecasting. **Sustainability (Switzerland)**, v. 13, n. 3, p. 1–22, 1 fev. 2021.
- AL-SAUDI, K.; DEGELER, V.; MEDEMA, M. Energy consumption patterns and load forecasting with profiled cnn-lstm networks. **Processes**, v. 9, n. 11, 1 nov. 2021.
- ALZUBAIDI, L. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, n. 1, 1 dez. 2021.
- ANEEL. Resolução Normativa ANEEL nº 1.000, de 7 de dezembro de 2021. **Agência Nacional Reguladora de Energia Elétrica**, 7 dez. 2021.
- AYUB, N. et al. Big data analytics for short and medium-term electricity load forecasting using an AI techniques ensembler. **Energies**, v. 13, n. 19, 1 out. 2020.
- BARMAN, M.; DEV CHOUDHURY, N. B. A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. **Sustainable Cities and Society**, v. 61, 1 out. 2020.
- BEDI, J.; TOSHNIWAL, D. Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting. **IEEE Access**, v. 6, p. 49144–49156, 28 ago. 2018.
- CHRISTOPHER M. BISHOP. **Pattern Recognition and Machine Learning**. 1. ed. Nova York: Springer Science, 2006.
- Day-ahead electricity demand forecasting: Post-COVID paradigm**. Disponível em: <<https://ieeedataport.org/competitions/day-ahead-electricity-demand-forecastingpost-covid-paradigm>>. Acesso em: 1 dez. 2021.
- DE FIGUEIREDO ROSA, J. G. et al. **Previsão de Demanda de Energia Elétrica Utilizando Modelos Lineares**. Sociedade Brasileira de Automatica, 10 dez. 2020.
- DE VILMAREST, J.; GOUDE, Y. State-Space Models for Online Post-Covid Electricity Load Forecasting Competition. **IEEE Open Access Journal of Power and Energy**, 2022.
- EMPRESA DE PESQUISA ENERGÉTICA. **Balanco Energético Nacional 2021: Ano base 2020**. Disponível em: <<https://www.epe.gov.br/sites-pt/publicacoes-dados->

abertos/publicacoes/PublicacoesArquivos/publicacao-675/topico-638/BEN2022.pdf>. Acesso em: 25 nov. 2022.

ENEAPOLITAN XIA JIANG, R. **Artificial Intelligence With an Introduction to Machine Learning SECOND EDITION**. [s.l.: s.n.]. Disponível em: <<https://www.crcpress.com/Chapman--HallCRC-Artificial-Intelligence-and-Robotics-Series/book-series/ARTILRO>>.

ENERGY AGENCY, I. **Global Energy Review: CO2 Emissions in 2021 Global emissions rebound sharply to highest ever level**. [s.l.: s.n.]. Disponível em: <www.iea.org/t&c/>.

ERTEL, W. **Undergraduate Topics in Computer Science Introduction to Artificial Intelligence**. [s.l.: s.n.]. Disponível em: <<http://www.springer.com/series/7592>>.

FARROKHABADI, M. et al. Day-Ahead Electricity Demand Forecasting Competition: Post-COVID Paradigm. **IEEE Open Access Journal of Power and Energy**, 2022.

FERREIRA, R. V. **Previsão de Demanda: Um Estudo de Caso para o Sistema Interligado Nacional**. Dissertação de Mestrado—Belo Horizonte: Universidade Federal de Minas Gerais, 2016.

GHALEHKHONDABI, I. et al. An overview of energy demand forecasting methods published in 2005–2015. **Energy Systems**, v. 8, n. 2, p. 411–447, 1 maio 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1. ed. [s.l.] The MIT Press, 2016.

HAFEEZ, G. et al. A novel accurate and fast converging deep learning-based model for electrical energy consumption forecasting in a smart grid. **Energies**, v. 13, n. 9, 1 maio 2020.

HAYKIN, S. S. **Neural networks and learning machines**. [s.l.] Prentice Hall/Pearson, 2009.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, 1997.

HUANG, L. et al. Prediction-based analysis on power consumption gap under long-term emergency: A case in China under COVID-19. **Applied Energy**, v. 283, 1 fev. 2021.

INSTITUTO ACENDE BRASIL. Impactos da covid-19 sobre o setor elétrico e medidas para mitigar seus efeitos. **White Paper 23**, 2020.

JAWAD, M. et al. Machine Learning Based Cost Effective Electricity Load Forecasting Model Using Correlated Meteorological Parameters. **IEEE Access**, v. 8, p. 146847–146864, 2020.

JIANG, Q. et al. A Stacking Learning Model Based on Multiple Similar Days for Short-Term Load Forecasting. **Mathematics**, v. 10, n. 14, p. 2446, 13 jul. 2022.

KAGAN, N.; OLIVEIRA, C.; ROBBA, E. **Introdução aos Sistemas de Distribuição de Energia Elétrica**. 1. ed. São Paulo: Blusher, 2005. v. 1

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. 22 dez. 2014.

KONDAIAH, V. Y.; SARAVANAN, B. Short-Term Load Forecasting with a Novel Wavelet-Based Ensemble Method. **Energies**, v. 15, n. 14, p. 5299, 21 jul. 2022.

LE, T. et al. Improving electric energy consumption prediction using CNN and Bi-LSTM. **Applied Sciences (Switzerland)**, v. 9, n. 20, 1 out. 2019.

LECUN. LeCun 1989 - Backpropagation applied to. **Neural Computation**, 1989.

LECUN, Y. et al. **Gradient-Based Learning Applied to Document Recognition**. [s.l.: s.n.].

LOPES DE OLIVEIRA, E. **Previsão de demanda de energia elétrica com redes neurais artificiais e análise por série de Fourier**. Dissertação de Mestrado—Minas Gerais: Universidade Federal de Minas Gerais, 2016.

LU, H.; MA, X.; MA, M. A hybrid multi-objective optimizer-based model for daily electricity demand prediction considering COVID-19. **Energy**, v. 219, 15 mar. 2021.

MARSLAND, S. **Machine Learning an Algorithmic Perspective**. 2. ed. [s.l.] Chapman & Hall/CRC, 2015. v. 1

MD SALLEH, N. S.; SULIMAN, A.; JORGENSEN, B. N. **A Systematic Literature Review of Machine Learning Methods for Short-term Electricity Forecasting**. 2020 8th International Conference on Information Technology and Multimedia, ICIMU 2020. **Anais...**Institute of Electrical and Electronics Engineers Inc., 24 ago. 2020.

MOHAMMAD, F.; AHMED, M. A.; KIM, Y. C. Efficient energy management based on convolutional long short-term memory network for smart power distribution system. **Energies**, v. 14, n. 19, 1 out. 2021.

NTI, I. K. et al. Electricity load forecasting: a systematic review. **Journal of Electrical Systems and Information Technology**, v. 7, n. 1, dez. 2020.

OBST, D.; DE VILMAREST, J.; GOUDE, Y. Adaptive Methods for Short-Term Electricity Load Forecasting during COVID-19 Lockdown in France. **IEEE Transactions on Power Systems**, v. 36, n. 5, p. 4754–4763, 1 set. 2021.

Operador Nacional do Sistema Elétrico. Disponível em: <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/curva_carga_horaria.aspx>. Acesso em: 25 nov. 2022.

ÖZBAY, H.; DALCALI, A. Effects of COVID-19 on electric energy consumption in Turkey and ANN-based short-term forecasting. **Turkish Journal of Electrical Engineering and Computer Sciences**, v. 29, n. 1, p. 78–97, 1 jan. 2021.

PENG, J. et al. A flexible potential-flow model based high resolution spatiotemporal energy demand forecasting framework. **Applied Energy**, v. 299, 1 out. 2021.

PRODANOV, C. **Metodologia do trabalho científico [recurso eletrônico]: métodos e técnicas da pesquisa e do trabalho científico**. 2. ed. Novo Amburgo: Feevale, 2013.

RASCHKA, S.; MIRJALILI, V. **Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2**. [s.l.: s.n.].

RAZA, M. Q.; KHOSRAVI, A. **A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings**. *Renewable and Sustainable Energy Reviews* Elsevier Ltd, , 18 jun. 2015.

ROMÁN-PORTABALES, A.; LÓPEZ-NORES, M.; PAZOS-ARIAS, J. J. **Systematic review of electricity demand forecast using ann-based machine learning algorithms**. *Sensors* MDPI AG, , 1 jul. 2021.

RUMELHART DAVID; HINTON GEOFFREY; WILIAMS RONALD. Learning representations by back-propagating erros. *Nature*, v. 323, p. 533–536, 1986.

SCARABAGGIO, P. et al. **Analyzing the Effects of COVID-19 Pandemic on the Energy Demand: The Case of Northern Italy**. 12th AEIT International Annual Conference, AEIT 2020. *Anais...*Institute of Electrical and Electronics Engineers Inc., 23 set. 2020.

SENJYU, T. et al. One-hour-ahead load forecasting using neural network. *IEEE Transactions on Power Systems*, v. 17, n. 1, p. 113–118, fev. 2002.

SEVERINO, A. **Metodologia do trabalho científico**. 1. ed. São Paulo: Cortex Editora, 2013.

STEVENSON, W. **Elementos de Análise de Sistemas de Potencia**. 2. ed. São Paulo: McGraw Hill, 1986.

THEODORIDIS, S. **Machine Learning: A Bayesian and Optimization Perspective**. 2. ed. [s.l.] Elsevier Ltd, 2020.

TUDOSE, A. M. et al. Short-term load forecasting using convolutional neural networks in covid-19 context: The Romanian case study. *Energies*, v. 14, n. 13, 1 jul. 2021.

VEERAMSETTY, V. et al. Short Term Electric Power Load Forecasting Using Principal Component Analysis and Recurrent Neural Networks. *Forecasting*, v. 4, n. 1, p. 149–164, 24 jan. 2022.

VERWIEBE, P. A. et al. **Modeling energy demand—a systematic literature review**. *Energies* MDPI, , 1 dez. 2021.

WERBOS, P. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, v. 78, n. 10, p. 1550–1560, 1990.

XU, Y.; VAZIRI-PASHKAM, M. Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications*, v. 12, n. 1, 1 dez. 2021.

YU, C. N.; MIROWSKI, P.; HO, T. K. A Sparse Coding Approach to Household Electricity Demand Forecasting in Smart Grids. **IEEE Transactions on Smart Grid**, v. 8, n. 2, p. 738–748, 1 mar. 2017.

ZHU, K.; GENG, J.; WANG, K. A hybrid prediction model based on pattern sequence-based matching method and extreme gradient boosting for holiday load forecasting. **Electric Power Systems Research**, v. 190, 1 jan. 2021.

ZIEL, F. Smoothed Bernstein Online Aggregation for Short-Term Load Forecasting in IEEE DataPort Competition on Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm. **IEEE Open Access Journal of Power and Energy**, 2022.