

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
DEPARTAMENTO DE ENERGIA
ENGENHARIA ELÉTRICA - HABILITAÇÃO EM ROBÓTICA E AUTOMAÇÃO
INDUSTRIAL

Júlia Amin Francisco

**Predição da Energia Armazenada do Subsistema Sudeste/Centro-Oeste do
SIN Através de Redes Neurais Artificiais**

Juiz de Fora

2022

Júlia Amin Francisco

**Predição da Energia Armazenada do Subsistema Sudeste/Centro-Oeste do
SIN Através de Redes Neurais Artificiais**

Trabalho de Conclusão de Curso apresentado
à Faculdade de Engenharia da Universidade
Federal de Juiz de Fora, como requisito para
a obtenção de grau em Engenharia Elétrica.

Orientador: Leonardo Rocha Olivi

Juiz de Fora

2022

Júlia Amin Francisco

**Predição da Energia Armazenada do Subsistema Sudeste/Centro-Oeste do
SIN Através de Redes Neurais Artificiais**

Trabalho de Conclusão de Curso apresentado
à Faculdade de Engenharia da Universidade
Federal de Juiz de Fora, como requisito para
a obtenção de grau em Engenharia Elétrica.

Aprovada em 16 de Janeiro de 2023.

BANCA EXAMINADORA

Leonardo Rocha Olivi - Orientador
Universidade Federal de Juiz de Fora

Prof. Exuperry Barros Costa
Universidade Federal de Juiz de Fora



ATA DE APRESENTAÇÃO DE TRABALHO FINAL DE CURSO

DATA DA DEFESA: 16/01/2023 – 16H00

CANDIDATO: JÚLIA AMIN FRANCISCO

ORIENTADOR: PROF. LEONARDO ROCHA OLIVI

TÍTULO DO TRABALHO: PREDIÇÃO DA ENERGIA ARMAZENADA DO SUBSISTEMA SUDESTE/CENTRO-OESTE DO SIN ATRAVÉS DE REDES NEURAIAS ARTIFICIAIS

BANCA EXAMINADORA/INSTITUIÇÃO:

PRESIDENTE: PROF. LEONARDO ROCHA OLIVI / UFJF

AVALIADOR: PROF. EXUPERRY BARROS COSTA / UFJF


LOCAL: LABORATÓRIO LABRA – FACULDADE DE ENGENHARIA – UFJF

Nesta data, em sessão pública, após exposição oral de 19 minutos, o candidato foi arguido pelos membros da banca. Em decorrência desta arguição, a banca considerou o candidato:

APROVADO

REPROVADO

Na forma regulamentar foi lavrada a presente Ata que é abaixo assinada pelos membros da banca na ordem determinada e pelo candidato:

PRESIDENTE: 

AVALIADOR: 

CANDIDATO: 

AGRADECIMENTOS

Primeiramente a Deus por ter me colocado aqui e ter me dado saúde e força para superar os obstáculos.

Aos meus pais, pelo apoio incondicional em todas as decisões difíceis e em todas as etapas da minha vida.

Aos amigos que me incentivaram, me deram suporte e motivação para acreditar em mim e na minha capacidade.

Ao meu professor e orientador Leonardo Rocha Olivi, por sua dedicação ao ensino, sempre passando com excelência o conhecimento adiante e despertando nos alunos a vontade de crescer.

Obrigada.

“Uma educação de qualidade tem o poder de transformar sociedades em uma única geração; fornecer às crianças a proteção de que precisam contra os perigos da pobreza, exploração do trabalho e doenças; e dar-lhes o conhecimento, habilidades e confiança para atingir todo o seu potencial.”

Audrey Hepburn.

RESUMO

A Energia Armazenada (EA) é uma característica de sistemas elétricos que vem ganhando maior relevância conforme a mudança da matriz energética mundial para meios de geração sustentáveis, pois é uma forma de garantir que o sistema permaneça em funcionamento em períodos de maior demanda. Como no Brasil o armazenamento de energia é majoritariamente realizado por bombeamento hídrico para reservatórios, a energia armazenada torna-se um fator dependente não só da demanda, mas também da disponibilidade dos recursos hídricos. Logo, a predição a médio e longo prazo da Energia Armazenada pode garantir a melhor gestão dos recursos, apontar na direção da necessidade de ampliação do sistema e, até mesmo, mostrar em quais períodos é preciso o aumento da participação de usinas termelétricas ou nucleares no sistema. Este trabalho adota o uso de Redes Neurais Artificiais. O modelo proposto para o treinamento e a previsão da Energia Armazenada foi o de *Backpropagation + Momentum Based Learning*. Os resultados iniciais foram promissores, porém deixavam a desejar no quesito de uma previsão com maior horizonte temporal. Logo, optou-se por testar um algoritmo mais robusto, sendo então adotado o otimizador Adam, para o qual, a rede demonstrou desempenho satisfatório, mas também não sendo capaz de realizar previsões de até doze meses com menor erro. Já para uma predição a curto e médio prazo, ambos métodos se revelaram capazes. Portanto, ambos passam a ser uma ferramenta alternativa para o estudo da segurança no fornecimento de energia diante da transição energética.

Palavras-chave: Energia Armazenada, Redes Neurais Artificiais, Transição Energética, Segurança Energética.

ABSTRACT

The stored energy is one of the characteristics of an electrical system whose relevance has been growing with the worldwide shift in the energetic matrix for more sustainable ways of generating electricity since it's one strategy to ensure the system will be able to keep operating during times of higher demand. Already in Brazil, power storage is mainly achieved through water pumping into reservoirs, so it becomes a dependent variable not only of the demand but also of the availability of water resources. Therefore, a medium and long-term prediction of total stored energy may ensure better resource management, point out the necessity of system expansion, and also showcase in which time period it's needed to increase thermoelectric and nuclear power plants' contribution. This paper embraces the application of Artificial Neural Networks. The initial proposed model used to train the net and solve the stored energy predictions was Backpropagation plus Momentum Based Learning. The first results were promising. However, they lacked the consistency to succeed with long-term predictions. For this reason, a stronger method of training was applied, the Adam algorithm, for which the net did not show much improvement or lesser errors for twelve months' predictions. When it comes to a short and medium period of time forecast, both methods revealed themselves as capable of achieving. Thus, both could be alternative tools for studying the safety of energy delivery in the face of the global energy transition.

Keywords: Stored Energy, Artificial Neural Networks, Energy Transition, Energy security.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diferentes Arquiteturas de Redes Neurais	15
Figura 2 – Projeção da Capacidade Instalada no SIN	17
Figura 3 – Subsistemas Interligados do SIN	18
Figura 4 – Usina Hidrelétrica Reversível	20
Figura 5 – Neurônio	22
Figura 6 – Modelo básico de um neurônio artificial	23
Figura 7 – Funções: Tangente hiperbólica e Sigmoid	24
Figura 8 – Taxa de aprendizado	26
Figura 9 – Rede Neural Artificial Multicamadas	27
Figura 10 – Dados Colhidos	32
Figura 11 – Autocorrelação da Série Formada Pelos Dados de Energia Armazenada	32
Figura 12 – Processo de Normalização dos Dados	33
Figura 13 – Treinamento e Predição para 3 meses com 12 Entradas e 25 Neurônios na Camada Oculta	40
Figura 14 – Treinamento e Predição para 6 meses com 24 Entradas e 49 Neurônios na Camada Oculta	41
Figura 15 – Treinamento e Predição para 12 meses com 24 Entradas e 49 Neurônios na Camada Oculta	41
Figura 16 – Erros x Meses Preditos para 12 Entradas na RNA com <i>Backpropagation</i> + <i>Momentum</i>	42
Figura 17 – Erros x Meses Preditos para 24 Entradas na RNA com <i>Backpropagation</i> + <i>Momentum</i>	42
Figura 18 – Erros x Meses Preditos para 36 Entradas na RNA com <i>Backpropagation</i> + <i>Momentum</i>	43
Figura 19 – Erros Para Variação da Camada Oculta para Predição de 3 Meses . . .	44
Figura 20 – Treinamento e Predição para 3 meses com 36 Entradas e 73 Neurônios na Camada Oculta	44
Figura 21 – Treinamento e Predição para 6 meses com 12 Entradas e 25 Neurônios na Camada Oculta	45
Figura 22 – Treinamento e Predição para 6 meses com 36 Entradas e 73 Neurônios na Camada Oculta	45
Figura 23 – Treinamento e Predição para 12 meses com 12 Entradas e 25 Neurônios na Camada Oculta	46
Figura 24 – Erros x Meses Preditos para 12 Entradas na RNA com ADAM	47
Figura 25 – Erros x Meses Preditos para 24 Entradas na RNA com ADAM	47
Figura 26 – Erros x Meses Preditos para 36 Entradas na RNA com ADAM	48

Figura 27 – Erros x Número de Neurônios na Camada Oculta para ADAM	48
Figura 28 – Erros: ADAM x <i>Backpropagation</i> para 12 entradas na RNA	49
Figura 29 – Erros: ADAM x <i>Backpropagation</i> para 24 entradas na RNA	50
Figura 30 – Erros: ADAM x <i>Backpropagation</i> para 36 entradas na RNA	50
Figura 31 – Comparativo do Erro de Treinamento por N ^o de Neurônios na Camada Oculta para predição de 3 meses.	51
Figura 32 – Comparativo do Erro de Predição e de Validação por N ^o de Neurônios na Camada Oculta para predição de 3 meses.	52
Figura 33 – Ajuste e Predição: <i>Backpropagation</i> x Adam	53
Figura 34 – Predição de 3 meses: <i>Backpropagation</i> x Adam	54
Figura 35 – Predição de 3 meses Com Valores Desnormalizados: <i>Backpropagation</i> x Adam	55

LISTA DE TABELAS

Tabela 1 – Funcionamento de uma TDNN de duas entradas e uma saída	31
Tabela 2 – Topologia das TDNNs para Seleção do Melhor Horizonte Temporal. . .	36
Tabela 3 – Topologia das TDNNs Com Variação da Camada Oculta para Predição de 3 meses.	37
Tabela 4 – Resultados da Variação das Entradas e do Horizonte de Temporal do <i>Backpropagation</i>	38
Tabela 5 – Resultados da Variação das Entradas e do Horizonte de Temporal para o Adam.	38
Tabela 6 – Resultados para variação do N ^o de Neurônios na Camada Oculta na Previsão de 3 Meses para o <i>Backpropagation</i>	39
Tabela 7 – Resultados para variação do N ^o de Neurônios na Camada Oculta na Previsão de 3 Meses para o Adam	39
Tabela 8 – Tabela Comparativa da Predição x Dados Reais em GWh	55

LISTA DE ABREVIATURAS E SIGLAS

RNA	Rede Neural Artificial
TDNN	Time Delay Neural Network)
FAC	Função de Autocorrelação
OMS	Operador Nacional do Sistema Elétrico
ANEEL	Agência Nacional de Energia Elétrica
NNAR	Neural Network Autoregression
MLP	Multi-Layer Perceptron
MSE	Erro Quadrático Médio
ONS	Operador Nacional do Sistema Elétrico
SIN	Sistema Interligado Nacional
TES	Thermal Energy Storage
MLP	Multi-Layer Perceptron

LISTA DE SÍMBOLOS

net	Ativação
δ	<i>Bias</i>
∂	Derivada Parcial
s_y	Desvio Padrão de y
\bar{X}	Vetor de Entradas da RNA
\bar{X}	Vetor de Pesos da RNA
φ	Função de Ativação
S_t	Série Energia Armazenada Normalizada
N_{ce}	Neurônios da Camada Escondida
α	Parâmetro "massa" do Momento - <i>Backpropagation</i>
Q	Momento Newtoniano
μ	Primeiro Momento - Adam
s	Segundo Momento - Adam
β_1	Parâmetro do Primeiro Momento - Adam
β_2	Parâmetro do Segundo Momento - Adam
η	Taxa de Aprendizagem
φ	Função de Ativação
j	Número da Camada
m	Número de Entradas da Rede Neural
J	Função de Custo
y	Saída do Neurônio
u	Entrada do Neurônio
T	Iteração no Tempo
d	Valores de Treinamento e Validação

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Tema e Problema	14
1.2	Trabalhos Relacionados	16
1.3	Objetivo	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	O Sistema Elétrico Brasileiro	17
2.2	Sobre a Operação do Sistema e A Demanda de Energia	18
2.3	Energia Armazenada	19
2.4	Redes Neurais Artificiais	22
2.4.1	Neurônio Artificial	22
2.4.2	Arquitetura Básica - O Perceptron	23
2.4.3	Treinamento	25
2.4.4	Otimizadores para <i>Backpropagation</i>	28
2.4.4.1	<i>Momentum based learning</i>	28
2.4.4.2	ADAM - <i>Adaptive Moments</i>	29
3	METODOLOGIA	31
3.1	Seleção e Tratamento dos Dados	31
3.2	Treinamento e Predição	34
3.2.1	<i>Backpropagation com Momentum Based Learning</i>	35
3.2.2	ADAM	35
3.3	Ferramentas e Equipamentos	35
3.4	Modelo de Comparação e Avaliação dos Resultados	36
4	RESULTADOS E ANÁLISES	38
4.1	Visão Geral dos Resultados	38
4.2	Resultados - <i>Backpropagation + Momentum</i>	40
4.3	Resultados - Adam	44
4.3.1	Resultados Comparativo: Adam × <i>Backpropagation + Momentum</i>	49
5	CONCLUSÕES	57
5.1	Conclusões	57
5.2	Propostas de Trabalhos Futuros	59
	REFERÊNCIAS	60

1 INTRODUÇÃO

Relevante para suprir a demanda energética durante a oscilação na geração de energia renovável em sistemas interligados, o armazenamento de energia é uma tendência mundial na migração para um futuro sustentável e de baixa emissão de poluentes. Para garantir um suprimento constante e seguro de energia durante os períodos de menor rendimento de usinas fotovoltaicas e eólicas, ou ainda durante períodos de estiagem, é possível armazenar a energia gerada nos momentos de alto desempenho de diversas maneiras e então tirar proveito dela quando necessário (ONS, 2022b).

No Brasil, a forma costumeira de realizar o armazenamento de energia em larga escala é através do bombeamento hidrelétrico, o qual é capaz de, em um curto tempo de resposta, produzir grande quantidade de potência, trazendo estabilidade para o sistema e de maneira rentável. Contudo, assim como a vazão de água flutua durante o passar do tempo, o mesmo ocorre com a energia armazenada. Afinal, é necessário utilizar dessa para a manutenção do fornecimento, contudo nem sempre é possível recuperar o nível anterior devido a fatores como a estiagem, no caso do bombeamento hidrelétrico, ou de gasto dos aparelhos, como no uso de baterias (ONS, 2022a).

Sendo o armazenamento de energia fundamental para o fornecimento ininterrupto e ferramenta essencial para respaldar fontes de energia intermitentes, é proveitoso o estudo de métodos para a previsão do estoque de energia armazenada, facilitando a administração dos recursos disponíveis.

Ainda que não totalmente precisos, os métodos de previsão apontam a tendência de séries temporais, o que pode se tornar algo valioso para o sistema de produção e de distribuição de energia.

Para isso, modelos de previsão baseados em *deep learning*, tal qual o de redes neurais artificiais (RNA), o qual é baseado em modelos matemáticos do sistema nervoso central, fornecem uma boa opção, visto que permitem relações não lineares e complexas entre os dados de entrada e a variável de resposta. É possível alcançar resultados promissores implementando regressão linear, *feed-forward* e variando o número de neurônios nas camadas escondidas (HAYKIN, 2000).

Na abordagem de séries temporais, o uso de redes neurais auto regressivas (NNAR - *Neural Network Autoregression*) ou redes neurais de atraso de tempo (TDNN - *Time Delay Neural Network*) é bastante comum e ambas utilizam de dados temporais passados para a alimentação da rede.

A metodologia de (BOX; JENKINS, 1990), afirma que é possível realizar a previsão de uma série temporal apenas com os valores passados e atuais através da correlação de dados. Dado o comportamento da função de autocorrelação (FAC), pode-se dizer se a

previsão da série será possível por esses meios, ou não. Como o comportamento de séries temporais podem apresentar padrões complexos e de difícil identificação, por conta disso, a utilização da FAC nem sempre é a mais adequada, o que dá margem para a função de autocorrelação parcial. (BOX; JENKINS; REINSEL, 1994).

Este trabalho traz o estudo e a análise da implementação de redes neurais artificiais para a previsão da energia armazenada no subsistema Sudeste/Centro-Oeste. Assim como a relação e a análise do impacto da variação do número de entradas, de camadas ocultas e da taxa de aprendizagem nos resultados apresentados.

1.1 Tema e Problema

A viabilidade de realizar previsões fascina a humanidade há séculos. Na Mesopotâmia, há cerca de mil anos A.C eram utilizados mapas astrais, que relacionavam o arranjo celeste com acontecimentos cíclicos de maneira a prever o clima e o comportamento dos céus, como eclipses e luas sangrentas. Séculos depois, durante as cruzadas, os cristãos entraram em contato com os avanços realizados pelos árabes a partir da astrologia, dentre eles, a matemática e a alquimia. Pode-se dizer que a maneira de realizar previsões se dividiu durante os milênios entre práticas observacionais de repetição e reconhecimento de padrões e práticas voltadas para o misticismo e para o espiritual. De qualquer maneira, ainda na atualidade, as pessoas buscam respostas através da predição, seja ela realizada pelo lançar de búzios ou pela abstração da matemática e da estatística.

Para a área de engenharia, uma boa previsão pode ser extremamente útil no planejamento de projetos e na gestão de recursos, como no caso da construção de uma nova usina, caso a tendência indique o aumento da demanda energética, ou ainda, para a alocação de turbinas eólicas de acordo com o comportamento dos ventos de determinada região.

No âmbito das ciências exatas, para realizar uma previsão satisfatória, é necessário possuir um conjunto considerável de dados, assim como construir um modelo matemático que represente o sistema estudado, ou que seja capaz de aprender o comportamento deste e reproduzir suas mudanças no tempo. Para o caso específico da geração e aproveitamento de energia hidrelétrica, dentre os diversos fatores que podem ser estudados estão os dados hidrológicos, os quais compreendem: nível, vazão e volume. Além disso, é possível analisar também a demanda, a energia armazenada, o custo de operação dentre outros (HYNDMAN; ATHANASOPOULOS, 2018).

Os métodos de previsão podem variar de acordo com o número de amostras, o tempo de amostragem, o padrão de comportamento da curva, fatores correlacionados e outros. Uma forma simples de previsão é o *Naive Method*, ou, em tradução literal, método ingênuo, o qual usa como resultado da predição simplesmente a última e mais recente

observação. Já um método mais complexo se dá através da implementação de Redes Neurais ([HAYKIN, 2000](#); [HYNDMAN; ATHANASOPOULOS, 2018](#)).

Uma Rede Neural Artificial comporta-se de maneira parecida ao funcionamento do sistema nervoso central de seres vivos graças ao aprendizado de máquina. Os Neurônios Artificiais são os componentes da rede, podendo possuir várias entradas e com diferentes ponderações para cada uma delas. Após o processamento, a saída do neurônio pode servir de resposta para o problema ou de entrada para outro neurônio. Dependendo do arranjo dos neurônios, uma rede pode possuir diversas camadas de processamento, chamadas de *hidden layers*, ou camadas ocultas, antes de se encontrar uma conclusão para o dilema ([HAYKIN, 2000](#)).

O emprego de redes neurais para a previsão de séries se dá por meio de algumas arquiteturas, cujas principais são:

- *Time Delay Neural Networks* (TDNN);
- *Recurrent Neural Networks* (RNN);
- *Non-Linear Autoregressive* (NAR).

Essas, formam diferentes tipos de topologia, como alguns exemplos da Figura 1.

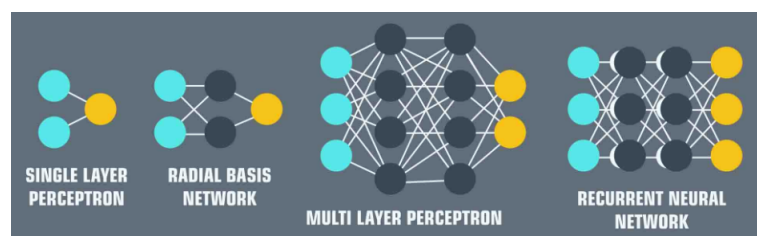


Figura 1 – Diferentes arquiteturas de RNAs

Fonte: ([ALLERIN, 2022](#)).

Para uma TDNN, sendo que, de acordo com ([CAUDILL, 1990](#)), o número de camadas escondidas deve ser escolhido com cautela, contudo, desde que haja número suficiente de neurônios e de dados, para realizar uma previsão satisfatória, basta uma camada oculta, ainda que isso tenha um custo maior de interações.

Concomitantemente, os autores de ([RUMELHART; WEIGEND., 1990](#)) defenderam um método de aprendizagem que se utiliza de um desdobramento da técnica do gradiente descendente, derivado da premissa de que quanto maior a rede, maior seu custo, tornando possível encontrar um balanço de custo benefício para a rede.

Assim, é viável implementar um algoritmo capaz de fazer isso, variar o número de entradas, e de camadas, e observar para qual cenário os resultados obtidos tiveram o melhor desempenho.

1.2 Trabalhos Relacionados

Ao levar em consideração a crescente busca por soluções para a iminente crise energética, é possível encontrar diversas publicações sobre a implementação de redes neurais para a previsão de fatores cruciais para a manutenção do abastecimento da rede.

Para realizar a previsão do Preço de Liquidação das Diferenças (PLD), os autores de (ROCHA, 2018) propuseram a utilização de Redes Neurais com Atraso de Tempo com a associação dos métodos de treinamento de Levenberg-Marquardt com *Momentum* e *Resilient-Backpropagation*.

Os autores de (RUAS et al.,) propuseram um sistema de previsão com base em redes neurais recorrentes para a previsão de demanda de energia a curto prazo. Para tal, foi escolhido o modelo de rede de Elman (ELMAN, 1990), através do qual é possível realizar a previsão de um passo à frente e então utilizar esse valor como retroalimentação para a rede.

Outra publicação com temática semelhante foi feita em (HIRUMA, 2016), o qual propôs modelos de previsão da velocidade do vento e da geração eólica de curto prazo através de redes neurais artificiais. Nesse trabalho, foi implementada uma rede *Multi-Layer Perceptron* (MLP) com etapas progressivas e regressivas.

1.3 Objetivo

Neste trabalho, busca-se, através da aplicação de diferentes meios de treinamento de redes neurais, realizar uma análise da melhor estrutura de rede e de otimização para a previsão da energia armazenada nos subsistema brasileiro do Sudeste/Centro-Oeste. Ou seja, pretende-se empregar diferentes algoritmos, variando fatores como taxa de aprendizagem, número de entradas e número de neurônios na camada oculta. Assim como fazer a alimentação dos dados, treinar a rede e verificar seus resultados de maneira comparativa.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 O Sistema Elétrico Brasileiro

O setor elétrico brasileiro é dividido entre geração, transmissão e distribuição. É considerado um dos setores mais confiáveis e que atende ao maior número de usuários, cerca de 99,7% do brasileiros segundo o IBGE (IBGE, 2022). Para garantir seu funcionamento é necessária uma enorme infraestrutura assim como uma boa organização, gerida através das instituições:

- Agência Nacional de Energia Elétrica (ANEEL, 2022): É o órgão regulador e fiscalizador, caracterizado por ser uma autarquia vinculada ao Ministério de Minas e Energia.
- Operador Nacional do Sistema Elétrico (ONS): É o órgão responsável pela gestão das instalações de geração e de transmissão de energia elétrica no Sistema Interligado Nacional (SIN), assim como dos sistemas isolados do país.
- Mercado atacadista de Energia Elétrica: Ambiente para comercialização da energia.

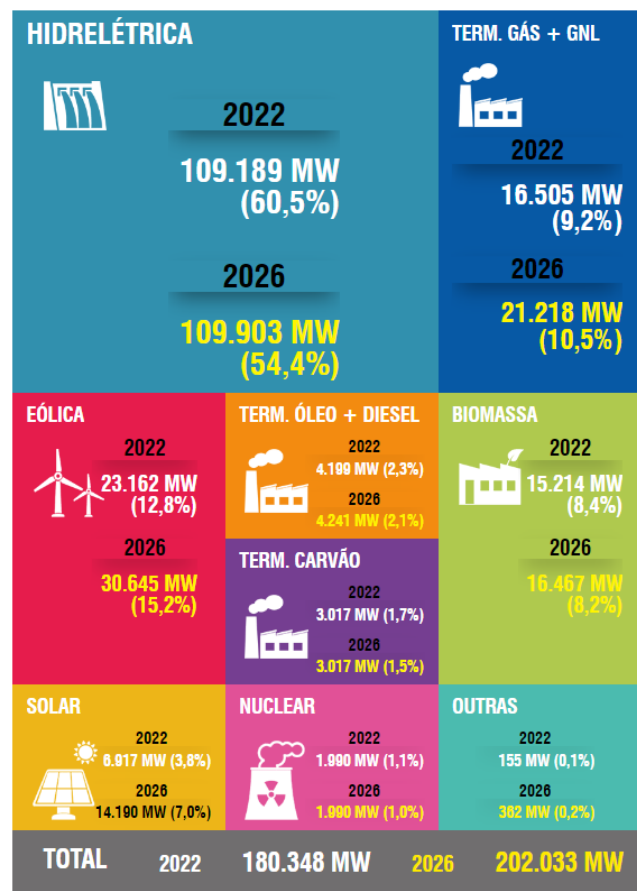


Figura 2 – Projeção da Capacidade Instalada no SIN

Fonte: (ONS, 2022c).

Uma das características predominantes do SIN é sua formação hidro-termo-eólica, Figura 2, com forte prevalência de hidrelétricas, tendo em vista as dezesseis bacias hidrográficas distribuídas em diferentes lugares do país. Em Dezembro de 2022, a capacidade de geração do SIN era de 180,348 GW com previsão de crescimento de 21,683 GW para o mesmo mês de 2026, sendo que dentre as fontes renováveis, a solar e a eólica, são as que possuem maior previsão de aumento percentual da capacidade instalada. Ou seja, cada vez mais se torna imprescindível adaptar métodos de aproveitar a energia gerada por fontes intermitentes.

O intercâmbio de energia entre as regiões é possível através do SIN, Figura 3, o qual é formado por quatro subsistemas: Sul, Sudeste/Centro-Oeste, Nordeste e Norte (ENERGIA, 2022).



Figura 3 – Subsistemas Interligados do SIN

Fonte: (ENERGIA, 2022).

2.2 Sobre a Operação do Sistema e A Demanda de Energia

Existem diversos fatores que geram uma alteração da demanda energética, dentre eles: a economia, o clima e as condições sociais do país. As mudanças na demanda podem ter características de variação horária, diária, mensal e até mesmo aleatórias. Ou seja, o sistema está a todo momento sujeito a distúrbios e, para que consiga se manter em

funcionamento, é necessário equilibrar a demanda com a quantidade de energia gerada. Para que isso seja possível, é preciso ter os recursos disponíveis para conseguir injetar no sistema a potência requerida.

É preciso ter consciência de que os equipamentos e instalações elétricas são projetadas para trabalhar dentro de faixas específicas de frequência e de tensão. Caso ocorra uma falha no fornecimento de energia, pode-se esperar um pior desempenho e até mesmo a perda de alguns equipamentos. Diante disso, o operador precisa atender certas exigências para garantir a segurança do sistema.

Geralmente, a frequência da rede é automaticamente controlada nos próprios geradores através da regulação da velocidade dos rotores. Quanto ao controle dos níveis de tensão, esse pode ser realizado em diferentes pontos da transmissão, da distribuição e até nas próprias usinas.

Para um sistema interligado, o despacho de geração entre diferentes usinas durante oscilações na demanda em diferentes regiões têm um papel fundamental para a manutenção do fornecimento seguro e de qualidade.

Dependendo do país ou da região abordada, esta troca de cargas entre as usinas pode estar relacionada ao tipo de geração implementada e sua capacidade de fornecimento ininterrupto. Na Europa, mesmo com o crescimento da participação de fontes renováveis, como a solar e a eólica, é indispensável o acionamento de termelétricas quando não há formas de garantir que a geração sustentável, e a energia armazenada advinda dela, responda à demanda do sistema. Logo, há um redirecionamento da carga para diferentes usinas conforme as variações da geração e da demanda.

2.3 Energia Armazenada

Diante da necessidade de uma transição energética para a preservação ambiental e para a independência do sistema em relação à recursos para a geração que não sejam renováveis, cresce significativamente a demanda por meios de armazenamento de energia, já que fontes renováveis obedecem determinados ciclos naturais e esses, por sua vez, podem ser intermitentes e variáveis no tempo. Ou seja, para conseguir ter uma injeção constante de potência em um sistema, é essencial que parte da energia seja armazenada para suprir a demanda durante a redução da disponibilidade dos recursos naturais.

No momento, contudo, não existe uma forma simples, prática e de baixo custo para realizar o armazenamento de energia elétrica, pois esse, intrinsecamente, ainda não é viável com os meios atuais. Sendo assim, é necessário converter a energia elétrica em outra forma de energia, tal qual química, mecânica ou potencial. Porém, as técnicas disponíveis para tal apresentam problemas quando o tema é armazenamento em larga escala, como o custo, o tempo de vida e a velocidade de resposta. Nesse sentido, o método de armazenamento

que melhor satisfaz as exigências de uma geração em grande escala é o de bombeamento, também chamado de hidrelétrica reversível, Figura 4.

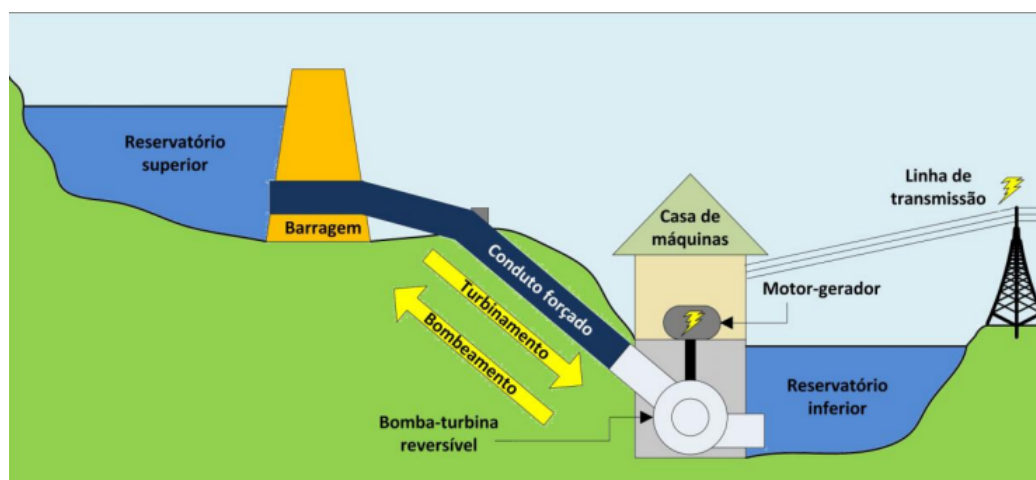


Figura 4 – Usina Hidrelétrica Reversível

Fonte: Centro Brasileiro de Infraestrutura

Devido aos períodos de estiagem, uma das maiores adversidades para a geração hidrelétrica é a existência de um suprimento de água estável. Nesse caso, o armazenamento ocorre da seguinte forma, em momentos de baixa demanda energética, é realizado o bombeamento da água de um reservatório inferior para um superior. Assim, quando a demanda for máxima, é possível utilizar do recurso estocado.

Diferentes propostas para esse tipo de reserva trazem a associação de mais um modo de geração: “Um novo conceito é a associação da energia eólica com usinas reversíveis, onde as variações da potência eólica podem ser niveladas usando o vento para bombear água ao reservatório superior, e gerando eletricidade através das turbinas hidrelétricas.” (PASQUALI, 2006).

Todavia, nem todas as usinas podem ser construídas dessa maneira, dada as características geográficas de cada região do país. Nesses casos, o nível do reservatório é ainda mais dependente do volume de água que chega pelos rios, e isso tende a oscilar durante o ano e por conta de fenômenos climáticos. Portanto, a energia armazenada passa a ser uma variável a ser analisada, pois, ao compreender seu comportamento se torna possível prever a quantidade de potência disponível e assim administrar melhor os recursos do sistema.

Outra forma de armazenamento, geralmente empregada em sistemas menores, é através de baterias de íons de lítio, pois são eficazes para guardar a energia por períodos de duas a oito horas. Atualmente, a Tesla é a maior fornecedora desse tipo de bateria para armazenamento em rede e apenas em 2020 foi possível implementar plantas com capacidade para armazenar mais de 3GWh. (CASARIN, 2021).

As baterias, contudo, não são a alternativa mais alinhada com a preservação do meio ambiente, visto que elas ainda apresentam um curto tempo de vida quando comparadas a um reservatório, além de dificuldades para o descarte e para a reciclagem, pois são inflamáveis. Há de se levar em consideração que, fora as questões relacionadas ao fim da vida útil da bateria, é preciso uma linha de produção que vai desde a mineração dos compostos não renováveis, majoritariamente em países subdesenvolvidos, ao transporte até o consumidor.

Formas alternativas de armazenamento que não envolvem descarte químico e que ainda não são utilizadas em larga escala estão em constante desenvolvimento, como através de energia térmica (TES - *Thermal Energy Storage*) ou de baterias de hidrogênio.

Existem diversas maneiras de implementar o TES, sendo que a energia pode ser transformada em calor sensível, calor latente ou até mesmo química. O armazenamento de calor sensível é alcançado através do aquecimento do meio armazenamento, que pode ser, por exemplo, sal fundido ou água pressurizada. Já o de calor latente, envolve a mudança de fase do meio, dessa forma entrega uma capacidade térmica maior. Em ambos, de maneira simplificada, a energia é liberada na troca de calor com um fluido que passa por tanque quente e por um tanque frio, ativando turbinas no percurso.

Já as baterias de hidrogênio envolvem a produção desse, geralmente através do processo de eletrólise da água, e o estoque em cilindros ou tanques pressurizados, podendo o elemento estar em estado gasoso ou líquido. A energia que o hidrogênio libera chega a ser o dobro da do gás natural por unidade de massa. Além disso, o hidrogênio pode ser usado tanto para a combustão em motores, quanto para a geração de eletricidade através de um catalisador de hidrogênio e oxigênio, tal que o subproduto do processo é a água (BACH, 2013).

Independente da forma escolhida para o armazenamento, a utilização do recurso está atrelada a demanda do sistema, a qual varia durante períodos do dia e do ano. Sendo assim, a quantidade de energia armazenada disponível para injetar a potência necessária para que o sistema funcione passa a flutuar durante esses ciclos. Portanto, a energia armazenada passa a ser uma variável a ser analisada, pois, ao compreender seu comportamento, se torna possível prever a quantidade de potência que o sistema consegue fornecer de maneira constante e assim administrar melhor os recursos disponíveis.

Esse aspecto se torna ainda mais relevante diante da necessidade de um planejamento energético sólido durante a transição energética. E até mesmo a análise para saber se essa transição sequer é viável dependendo da região do país ou do mundo. A crise energética que a Europa Ocidental enfrenta atualmente devido a conflitos internacionais é um exemplo dessa situação.

A proposta de auxiliar uma gestão e um planejamento energético adequado, pode

ser feita através de previsões geradas por dados colhidos. Um dos métodos utilizados é por meio de Redes Neurais Artificiais, através das quais pode-se estimar valores futuros para determinadas variáveis de interesse, como a Energia Armazenada.

2.4 Redes Neurais Artificiais

Inspiradas pelas formas de aprendizado dos seres vivos, redes neurais artificiais são como um modelo do sistema nervoso humano, sendo os neurônios as células, Figura 5, que compõem a rede de processamento de informações. Cada neurônio pode ser subdividido em partes responsáveis pelo recebimento e pelo envio de informações, o dendrito e o axônio.

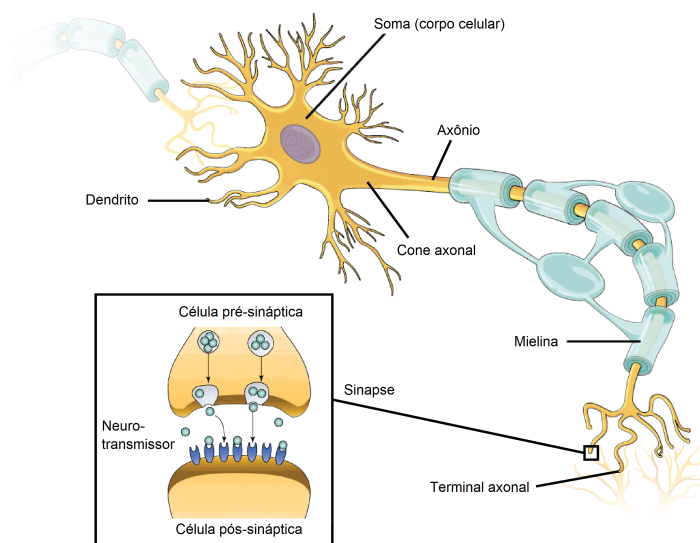


Figura 5 – Neurônio biológico

Fonte: (KHAN, 2022)

A região de conexão entre os neurônios são chamadas de sinapses e essas podem variar de acordo com os estímulos sofridos pelas células. Graças a essa capacidade de variação de peso nas sinapses, é possível o aprendizado (AGGARWAL, 2018).

2.4.1 Neurônio Artificial

Dado o neurônio biológico, é possível simular sua estrutura através de um número de entradas com diferentes pesos, análogas ao dendritos, e uma saída referente ao axônio, Figura 6. Já o corpo celular é um somatório seguido por uma função de ativação.

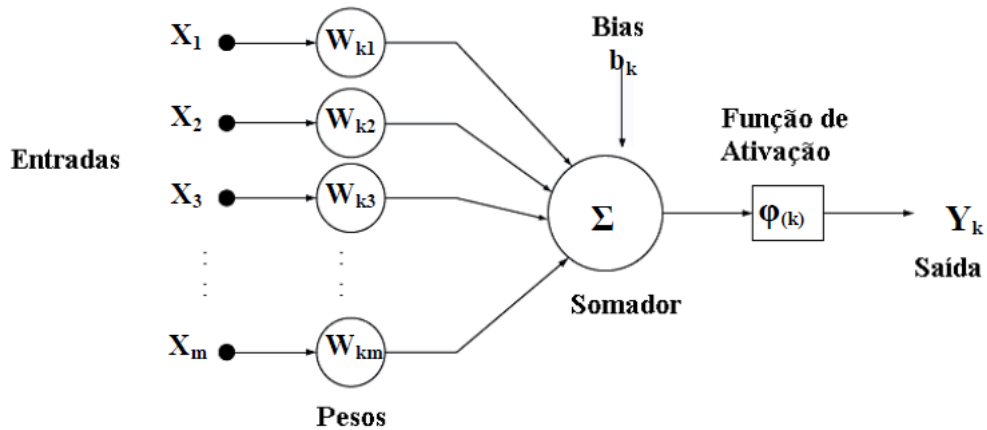


Figura 6 – Modelo básico de um neurônio artificial

Fonte: ResearchGate.

O organismo biológico se ajusta e passa a prever determinados resultados a partir de ações tomadas graças às experiências adquiridas ao longo da vida, sejam elas positivas ou negativas. Um exemplo básico disso é o sorriso, os bebês, por volta dos três meses de idade, passam a usar do sorriso como habilidade de interação social. Isso é um indicativo muito importante para o desenvolvimento da criança, pois aqui é possível avaliar o grau de evolução do sistema nervoso do bebê, sua capacidade de reconhecer o rosto da mãe e suas expressões faciais. Para que o desenvolvimento ocorra de maneira ideal, é essencial a socialização e a imersão em um ambiente repleto de informações, visto que uma parte importante do aprendizado ocorre dessa forma. É como se o cérebro fosse aos poucos atribuindo valores às cargas emocionais de cada experiência.

Tal característica do aprendizado pode ser emulada para o modelo matemático do neurônio através do ajuste minucioso dos pesos sinápticos da função do neurônio. A alimentação do modelo com vários dados de entradas e suas respectivas saídas é capaz de calibrar o sistema para que esse consiga prever a saída de uma entrada não utilizada no treinamento. Essa capacidade é conhecida como *model generalization*, pois o aprendizado de máquina é primariamente uma capacidade de generalização a partir dos dados de treinamento (AGGARWAL, 2018).

A agregação das informações ponderadas, então, deve passar pela função de ativação, a qual é responsável pelo juízo de valor do resultado e conseqüentemente pela ativação ou não do neurônio. Dependendo do modelo e da função escolhida é, ainda, possível determinar o grau de ativação desse.

2.4.2 Arquitetura Básica - O Perceptron

O modelo mais simplista de uma rede neural possui apenas uma camada de entrada e um único nó de saída, o Perceptron. Como pode ser visto na Figura 6. Nessa estrutura,

uma única camada é responsável por processar e transmitir para o nó de saída os recursos usados na alimentação. Às entradas, \bar{X} , são adicionadas valorações, \bar{W} , com o objetivo de calibrar o neurônio para uma função específica, dependendo dos dados que o Perceptron precisa classificar esses pesos são diferentes. Logo:

$$\bar{X} = [x_1 \quad x_2 \quad x_3 \quad \dots \quad x_m] \quad (2.1)$$

$$\bar{W}_j = [w_{j1} \quad w_{j2} \quad w_{j3} \quad \dots \quad w_{jm}] \quad (2.2)$$

O chamado *bias*, b_j , funciona como um coeficiente linear do neurônio e pode ser adicionado ao vetor \bar{W} como um w_0 , servindo para deslocar o ponto de ativação da função, mudar sua amplitude e angulação tornando a RNA mais abrangente e adaptável. Assim, a função linear dada por $\bar{W} \cdot \bar{X}$ é encaminhada pelo nó de saída e passa pela função de ativação φ_j .

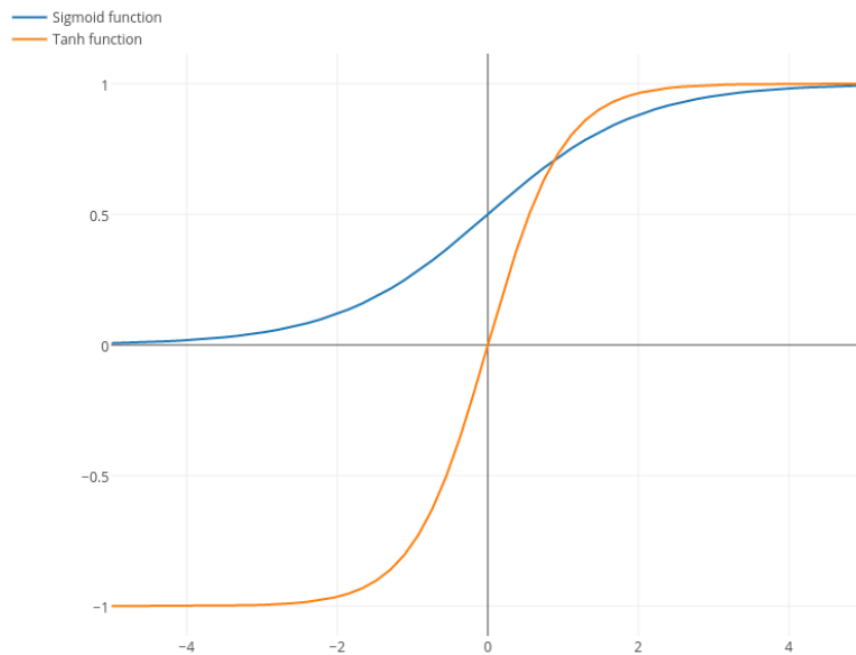


Figura 7 – Possíveis funções de ativação: Tangente hiperbólica e Logística.

Fonte: Stackexchange

Existem diferentes tipos de função de ativação, como as da Figura 7, e a seleção da mais adequada está ligada ao modelo de aprendizagem de máquina que se pretende aplicar. Para o caso do Perceptron as funções sinal, Degrau, ou ainda, uma Sigmoid, como a tangente hiperbólica, podem ser aplicadas para a situação de classificação binária de elementos. Tal que, a predição da saída, \hat{y} , é dada por:

$$\hat{y} = \varphi(\bar{W} \cdot \bar{X}) = \varphi \left(\sum_{i=1}^m w_i \cdot x_i \right) \quad (2.3)$$

É importante ressaltar que, conforme se aumenta o número de camadas de uma rede, a aplicação de funções de ativação não lineares, como a tangente hiperbólica, ganha destaque, principalmente quando se aborda a técnica de Gradiente Descendente para o treinamento da RNA, pois essa exige que a função de ativação seja derivável ao menos uma vez ao longo de seu domínio (RUMELHART; HINTON; WILLIAMS, 1986).

Todo esse processo configura a fase chamada de *forward propagation* e para que ela seja bem sucedida na predição é imprescindível que os pesos de \bar{W} estejam otimamente ajustados. Para tornar isso viável, é realizado o treinamento da Rede Neural até que ela fique calibrada.

2.4.3 Treinamento

Até os anos 80, não havia uma solução clara e objetiva para treinar redes neurais com múltiplas camadas. Sendo que, encontrar a maneira de calibrar os pesos adequados é parte intrínseca do processo de aprendizado de máquina e, conseqüentemente, das RNAs. Logo, durante décadas, o estudo e a aplicação do tema foram deixados de lado até que Rumelhart propôs a aplicação do algoritmo de *backpropagation*. Associada aos recursos atuais de *hardware* e grandes bancos de dados, a implementação dessa técnica se tornou algo viável e um dos métodos comumente utilizados envolve o Gradiente Descendente de uma função objetivo, que nesse caso toma a forma de uma função de custo, ou do erro quadrático médio, MSE, dividido por dois. Sendo \bar{d} o vetor de dados utilizados para o treinamento:

$$J = \frac{1}{2n} \sum_{i=1}^n (d_i - y_i)^2 \quad (2.4)$$

Como a característica do gradiente é apontar na direção de máximo crescimento de uma função e sendo o objetivo minimizar o custo, utiliza-se o sentido oposto ao do gradiente.

$$-\nabla f(x) = - \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, \dots, \frac{\partial f}{\partial x_m} \right] \quad (2.5)$$

Dado que a equação do neurônio é tal que:

$$u_j = \omega_j \cdot x_j + \omega_{0j} \cdot x_{0j} \quad (2.6)$$

E saída esperada:

$$y_{j+1} = \tanh(u_{j+1}) \quad (2.7)$$

A aplicação do gradiente descendente através da regra da cadeia toma a seguinte forma:

$$\frac{\partial J}{\partial \omega} = \frac{\partial J}{\partial y} \frac{\partial y}{\partial u} \frac{\partial u}{\partial \omega} \quad (2.8)$$

Contudo, ao se aplicar o gradiente, esse nem sempre irá ser bem sucedido em seu apontamento, pois ao lidar com números finitos, conforme o percurso se aproxima do menor custo, alguns imprevistos podem ocorrer, como oscilações em zig-zag, Figura 8, próximas da base da curva de custo.

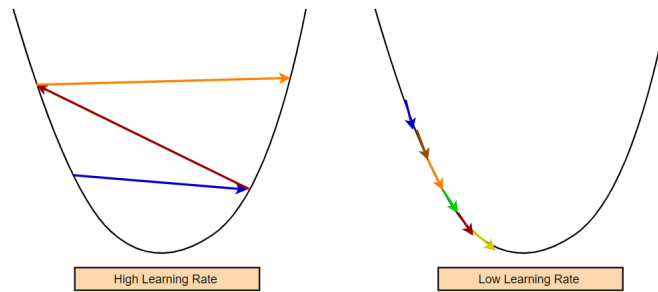


Figura 8 – Taxa de aprendizado e sua relação com oscilações no Gradiente Descendente.

Fonte: (SHUKLA, 2022)

Uma das maneiras de contornar esse problema é a aplicação de uma taxa de aprendizado, η . Tal que, a atualização dos pesos sinápticos a cada iteração, T , ocorre da seguinte maneira:

$$\omega_{jm_{T+1}} = \omega_{jm_T} + \eta \Delta \omega_{jm} \quad (2.9)$$

Esse procedimento pode ser ampliado e aplicado para uma rede neural de múltiplas camadas, inicialmente encontrando o gradiente para cada nó de saída e então progredindo para suas respectivas derivadas pela regra da cadeia. Dado uma rede genérica com j camadas e m entradas:

$$\nabla J = \frac{\partial J}{\partial y_j} = (y_j - d_j) \quad (2.10)$$

Logo:

$$\nabla J_j = \frac{\partial J}{\partial \omega_{jm}} = \frac{\partial J}{\partial y_j} \frac{\partial y_j}{\partial u_j} \frac{\partial u_j}{\partial \omega_{jm}} \quad (2.11)$$

E:

$$\nabla J_{j-1} = \frac{\partial J}{\partial \omega_{jm}} = \frac{\partial J}{\partial y_j} \frac{\partial y_j}{\partial u_j} \frac{\partial u_j}{\partial y_{j-1}} \frac{\partial y_{j-1}}{\partial u_{j-1}} \frac{\partial u_{j-1}}{\partial w_{(j-1)m}} \quad (2.12)$$

Os cálculos são realizados para todas as camadas da rede e também para o *bias*. Assim, em observância aos termos obtidos nas equações dos gradientes para cada camada, a atualização dos pesos sinápticos para cada camada irá se repetir até que o custo seja aceitável.

$$\bar{W}_{jT+1} = \bar{W}_{jT} - \eta \nabla J_j \quad (2.13)$$

Através desse processo, os pesos de cada entrada podem ser atualizados, mesmo aqueles pertencentes às camadas escondidas, graças à agregação dos gradientes dos nós. Ou seja, o método permanece similar, é apenas expandido, desde que os nós estejam conectados, para levar em consideração o custo adicional de cada nó e fazer a correção de seus respectivos pesos. Após a calibração, a fase *forward* é aplicada para fazer a previsão.

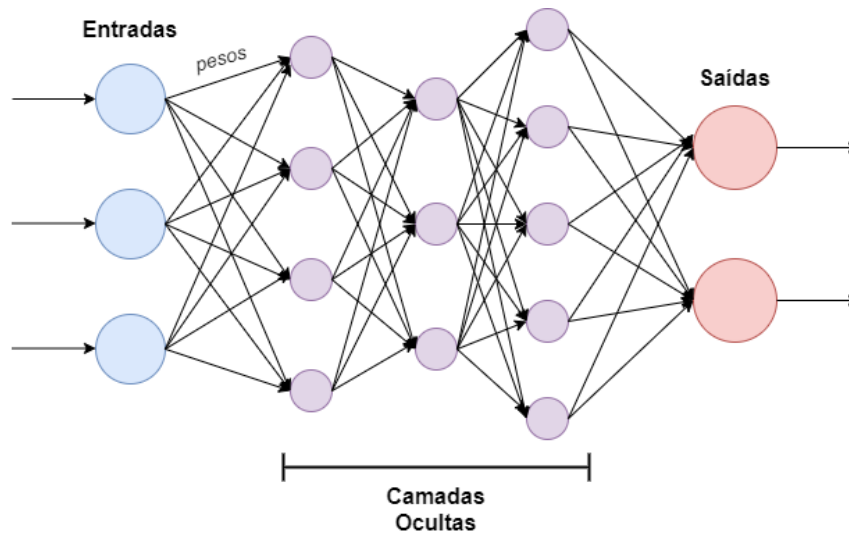


Figura 9 – Rede Neural Artificial Multicamadas.

Fonte: ([NASCIMENTO; OLIVI; LIMA, 2018](#))

Um aspecto das redes neurais de múltiplas camadas é que elas são capazes de reconhecer padrões não lineares e funções complexas. Contudo, redes diretas como a *Multi-Layer Perceptron* (MLP), Figura 9, não incorporam o tempo em seu processamento, o que as tornam inadequadas para a previsão de processos mutáveis, como séries temporais.

Para solucionar essa questão, são utilizadas janelas de tempo, ou seja, as entradas da rede neural passam a utilizar trechos da série temporal que se deseja prever. Para que a visão da rede não se torne limitada à sequência de dados usada na alimentação e, com isso, incapaz de identificar padrões distantes do trecho, é necessário introduzir ao menos uma nova camada escondida, que funciona como uma linha de atraso no tempo, visto que a entrada dessas camadas são as saídas das camadas anteriores. Ou seja, são inseridos níveis de memória na rede e com isso a rede neural passa a aprender não apenas com uma entrada, mas com uma sequência de entradas, como se fosse uma fila formada

pelas camadas escondidas que representam as posições dos dados do presente e do passado. Esse modelo adaptado é conhecido como *Time Delay Neural Network* (TDNN). (HAYKIN, 2000).

2.4.4 Otimizadores para *Backpropagation*

O algoritmo de *backpropagation* através do Gradiente Descendente foi uma das primeiras alternativas encontradas para o treinamento de redes neurais. A partir dela, visando aprimorar o método e contornar problemas típicos da retropropagação, alguns otimizadores foram desenvolvidos, como: *Learning rate decay*, *Momentum based learning*, Adam, RMSProp e outros. Essas diferentes ramificações trazem soluções para alguns dos obstáculos encontrados no algoritmo base, como o de *vanishing gradient*, o de *zigzagging* e o de ruídos computacionais.

2.4.4.1 *Momentum based learning*

A técnica baseada em momento é capaz de auxiliar o *backpropagation* para que durante a varredura da superfície da função de custo não ocorra minimização dos passos por conta de *zigzagging*, o que leva a redução da velocidade de aprendizado. Além disso, a técnica é eficaz para evitar mínimos locais, visto que utiliza de artifícios matemáticos para atribuir inércia na direção de minimização do custo a partir de um histórico de momentos.

De acordo com (ORR GENEVIEVE; MÜLLER, 1998), o emprego do momento acelera a velocidade de aprendizagem visto que, pontos de alta curvatura da superfície, provoca um efeito que pode ser comparado ao atrito e, com isso, atenua o tamanho dos passos. Contudo, em regiões mais planas, ainda é capaz de manter uma taxa de aprendizado eficaz.

Na física o momento é definido como produto da massa pela velocidade, para traduzir isso no algoritmo de *backpropagation* com *Momentum based learning* é realizada a escolha de uma unidade de massa e a velocidade, dada como o vetor v .

$$Q = \alpha v \tag{2.14}$$

Como para uma tarefa composta pela varredura da superfície por uma partícula simbólica, pode-se assumir uma dinâmica newtoniana, para qual a variação da posição no tempo nos leva à velocidade. Sendo a posição, nesse caso, os pesos sinápticos.

$$v = \frac{\partial}{\partial t} \bar{W}(t) \tag{2.15}$$

Numericamente, pode-se resolver como:

$$v = \bar{W}_t - \bar{W}_{t+1} \quad (2.16)$$

Logo, atualização dos pesos sinápticos irá contar com uma parcela extra, a do momento, tal que:

$$\bar{W}_{j_{t+1}} = \bar{W}_{j_t} - \eta \nabla J_j + \alpha (\bar{W}_t - \bar{W}_{t+1}) \quad (2.17)$$

Onde, α , que corresponde a massa da partícula, é um hiperparâmetro contido no intervalo $[0,1)$ e que contribui exponencialmente para a minimização do custo. Valores comumente atribuídos para α são: 0,5; 0,9 e 0,99. (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.4.4.2 ADAM - *Adaptive Moments*

Proposto por (KINGMA; BA, 2015), é um algoritmo baseado em métodos estocásticos que pode ser interpretado como uma associação entre o *Momentum based learning* e o RMSProp, à essa é adicionado um processo de normalização dos gradientes encontrados. (GOODFELLOW; BENGIO; COURVILLE, 2016).

Nesse modelo são utilizados o primeiro e o segundo momento do gradiente como uma parcela para encontrar taxas de aprendizagem adaptativas para os cálculos dos pesos sinápticos da rede neural. Assim, primeiro momento é a média e o segundo, a variação não centralizada dos gradientes. À eles, respectivamente, são associados os fatores β_1 e β_2 , que são os aspectos de decaimento. Devido a agregação de momentos adaptativos, o Adam responde bem para a previsão de dados com ruídos. Sendo assim, suas atualizações são realizadas através de:

$$\mu_{j_t} = \beta_1 \cdot \mu_{j_{t-1}} + (1 - \beta_1) \cdot \nabla J_{j_t} \quad (2.18)$$

Tal que μ é o primeiro momento, a média. E, para o segundo momento, s:

$$s_{j_t} = \beta_2 \cdot s_{j_{t-1}} + (1 - \beta_2) \cdot (\nabla J_{j_t})^2 \quad (2.19)$$

Assim, a correção pode ser aplicada da seguinte forma:

$$\hat{\mu}_j = \frac{\mu_j}{1 - \beta_1^t} \quad (2.20)$$

$$\hat{s}_j = \frac{s}{1 - \beta_2^t} \quad (2.21)$$

Somente então é possível fazer a atualização dos pesos sinápticos, tal que:

$$\bar{W}_{j_{t+1}} = \bar{W}_{j_t} - \frac{\eta}{\sqrt{\hat{s}_t + \varepsilon}} \hat{\mu}_t \quad (2.22)$$

Onde ε representa um valor pequeno com intuito de evitar a ocorrência de divisões por zero.

O algoritmo Adam costuma fornecer um desempenho superior a outros métodos, com uma rápida minimização do erro e capacidade de superar obstáculos mencionados anteriormente.

3 METODOLOGIA

3.1 Seleção e Tratamento dos Dados

Esse Capítulo trata da apresentação das técnicas implementadas no processo de predição da Energia Armazenada do subsistema Sudeste/Centro-Oeste do SIN com horizonte de longo prazo. Os objetos de análise passaram pelo número de meses preditos, o número de entradas da rede neural, a taxa de aprendizagem e o número de neurônios na camada escondida. Dado que os dados colhidos são variantes no tempo, o modelo de rede neural selecionado foi o de TDNN, visto que esse comporta a previsão de séries temporais, por conta de usar como alimentação uma janela temporal dinâmica com atraso de tempo. Tal que, para um rede genérica com duas entradas $X1$ e $X2$, uma série $S = [S_1, S_2, S_3]$, um vetor d para o validação, e a saída y , ocorre o processo descrito na Figura ?? para 3 atrasos de tempo:

TDNN com duas entradas (X1 e X2) e uma saída (y)			
Tempo (t)	t=1	t=2	t=3
X1	S1	S2	S3
X2	S2	S3	y4
d	S3	-	-
y	y3	y4	y5

Tabela 1 – Funcionamento de uma TDNN de duas entradas e uma saída

Tal comportamento é expandido para o número de entradas escolhidas para o treinamento, porém a dinâmica do processo permanece inalterada, visto que a arquitetura da rede não é modificada.

A base de dados utilizados foi obtida na página oficial do Operador Nacional do Sistema Elétrico e abrange o período de Janeiro de 2002 à Outubro de 2022, totalizando 250 meses referentes a Energia Armazenada em GWh. A característica da curva pode ser observada na Figura 10.

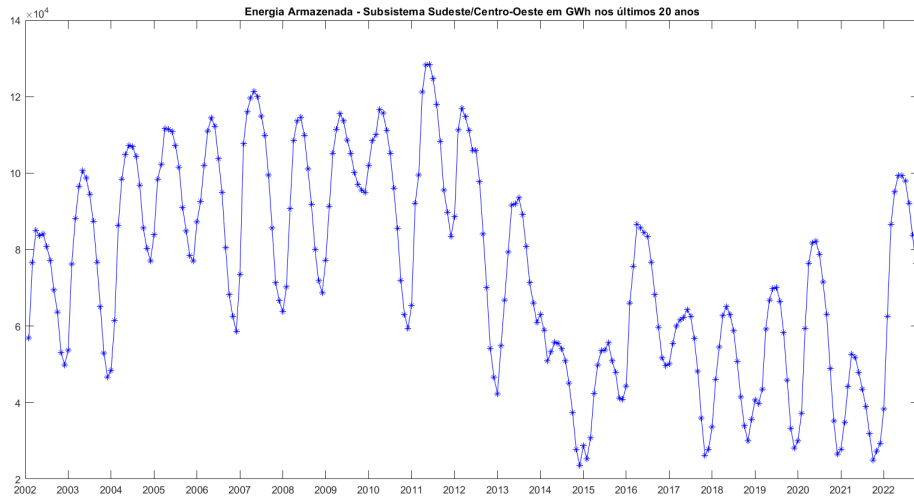


Figura 10 – Dados Colhidos

Para proceder com esse conjunto de informações é primeiramente investigada a autocorrelação da série temporal. A autocorrelação é uma medida estatística que considera o quanto os dados da série estão relacionados entre si, sendo capaz de apontar a existência de periodicidade mesmo em sinais contaminados por ruídos. Essa análise é necessária, pois indica se a predição será possível e qual o grau de dificuldade de realizá-la.

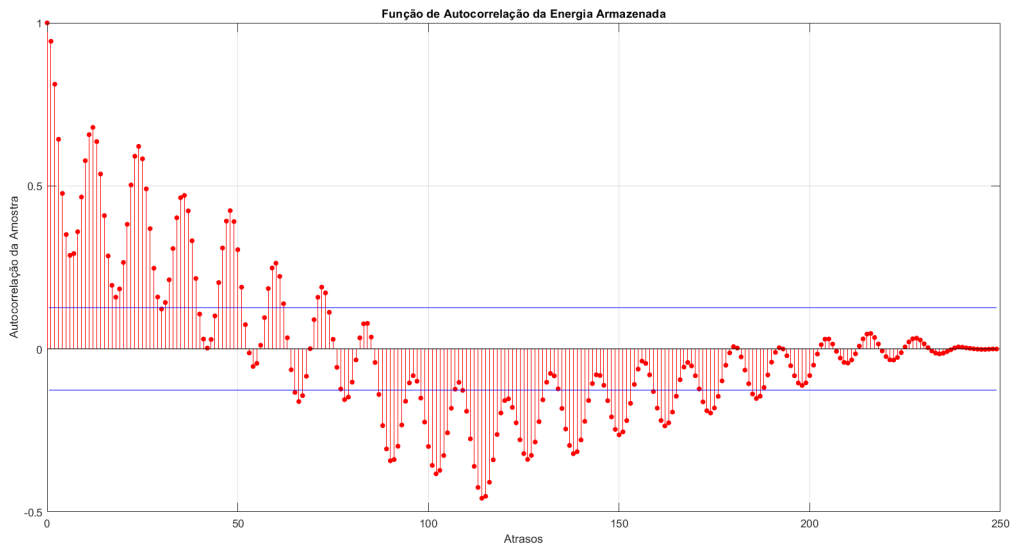


Figura 11 – Autocorrelação da Série Formada Pelos Dados de Energia Armazenada

Observa-se na Figura 11 que, além de a série adotada apresentar bom nível de autocorrelação para diversas unidades de atraso, o gráfico decai e em seguida apresenta

um comportamento oscilatório em torno de zero. Esse comportamento aponta para uma série temporal estacionária. (ROMÃO et al., 2020)

Para a função de ativação, optou-se pela tangente hiperbólica, visto que essa apresenta simetria em torno de zero assim como os dados escolhidos quando normalizados, o que indica maior velocidade de convergência. Além disso, o comportamento do gradiente da tangente hiperbólica em torno de zero é quatro vezes maior que o da função logística, ou seja, para obter maiores passos no treinamento, a tangente hiperbólica é mais indicada. Ademais, ela apresenta o benefício de evitar os problemas relacionados ao *exploding gradient* (quando o gradiente se torna grande demais e pode levar ao *zigzagging*), visto que apresenta continuidade, formato de ‘S’ na origem, quando comparada com a função degrau ou a sinal (KOLMOGOROV, 1957).

Tendo em conta a escala dos valores, para realizar uma predição com eficiência e evitar a saturação das saídas dos neurônios, foi feita a normalização de dados para que esses ficassem contidos nos limites da função de ativação selecionada, que no caso pertence ao intervalo $[-1,1]$. Sendo que, o processo de normalização se dá através de:

$$S_t = \varphi_{max} - \frac{(\varphi_{max} - \varphi_{min})(EA_{max} - EA_t)}{(EA_{max} - EA_{min})} \quad (3.1)$$

Onde, em 3.1, S_t é referente ao vetor de valores utilizados para a análise da TDNN, EA_{max} e EA_{min} são referentes limites superiores e inferiores dos valores dos dados que se deseja utilizar nesta etapa. Quanto a φ , continua a representar a função de ativação, seu valor máximo e valor mínimo, no caso 1 e -1. Ou seja, a série original passa pelo processamento da Figura 12 para só então ser utilizada como dado de entrada dos neurônios que possuem uma Sigmoide como função de ativação.

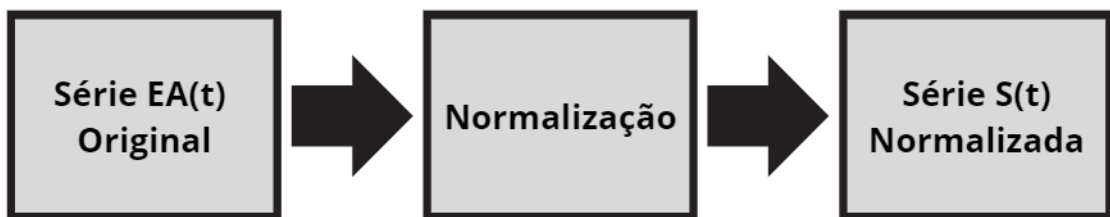


Figura 12 – Processo de Normalização dos Dados

A série dada por $S(t)$ foi então subdividida em duas parcelas, os dados de treinamento e os de predição. Os dados de treinamento foram separados entre aprendizagem e validação na proporção de 85%.

Quanto aos dados preditos, esses foram estudados para três meses, seis meses e doze meses, sendo que cada amostra colhida corresponde ao valor de Energia Armazenada no respectivo mês.

O número de entradas da Rede Neural também foi variado entre doze, vinte e quatro e trinta e seis, o que corresponde a um ano, dois anos e três anos de dados colhidos respectivamente.

Para a previsão de três meses, foram obtidos os resultados mais satisfatórios. Esse tempo de predição foi utilizado para outra análise visando estudar melhor a influência da camada oculta (N_{ce}). Sendo então realizada a variação do número de neurônios dela, tanto acima quando abaixo da regra de Komolgorov.

Dada a topologia de uma rede *feedforward* com múltiplas camadas, figura 9, e sabendo que, de acordo com o Teorema da Aproximação Universal, para a previsão de séries temporais com funções de ativação sigmóides basta uma camada escondida, o treinamento da rede é realizado através das etapas:

- Cálculo da saída y para pesos sinápticos inicializados de maneira aleatória através da fase *forward*;
- As saídas são comparadas com os valores de validação, d , e a diferença entre essas gera o erro quadrático médio (MSE);
- O MSE é avaliado e, para um valor acima do desejado, torna-se necessária a aplicação da fase *backwards*, na qual se realiza o ajuste dos pesos sinápticos. Sendo que, para isso, o algoritmo mais utilizado é o de *backpropagation* e suas variações conforme o otimizador aplicado;
- Repetição do processo até que o MSE esteja dentro do aceitável.

3.2 Treinamento e Predição

Os pesos sinápticos aplicados na fase *forward* são inicializados com valores aleatórios bem pequenos, pois isso faz com que os neurônios partam com uma inclinação de aproximadamente 45 graus e isso gera um ajuste mais otimizado para a rede.

Quando ao treinamento realizado, para ambos métodos explorados, ele foi submetido a determinados critérios de parada, sendo eles:

- Erro de Validação = 0.001;
- Erro de Treinamento = 0.001;
- Época Máxima = 10000;

Após o processamento dos dados, treinamento e predição, para visualizar os valores reais encontrados, é preciso fazer o processo de desnormalização. O qual é descrito pela equação:

$$EA_{predito} = EA_{max} - \frac{(EA_{max} - EA_{min})(S_{max} - S_{predito})}{S_{max} - S_{min}} \quad (3.2)$$

3.2.1 *Backpropagation com Momentum Based Learning*

Como mencionado nos tópicos 2.4.3 e 2.4.4, esse é um método de treinamento tradicional que funciona bem para diversas aplicações, além de ser mais simples para compreensão e implementação. A adição do momento auxilia a contornar questões que poderiam ser um impedimento para a aplicação de um *Backpropagation* puro, sendo que a série estudada pode apresentar ruídos e isso poderia prender o algoritmo em algum mínimo local devido à redução do gradiente.

Logo, como passo inicial, optou-se por utilizar este modelo, pois de acordo com a função de autocorrelação, Figura 11, os dados colhidos da Energia Armazenada seriam fáceis de prever.

Os resultados foram obtidos a partir da variação do número de meses previstos e do número de entradas. Como a melhor previsão foi para três meses, deu-se prosseguimento na variação do número de neurônios da camada oculta para esse tempo de predição.

Além disso, alguns parâmetros específicos foram ajustados para esse algoritmo, tal qual a taxa de momentum do treinamento, que seria semelhante a massa do objeto pontual que percorre a superfície da curva. Essa foi escolhida como sendo $\alpha = 0,95$.

3.2.2 ADAM

Como esse é um algoritmo que se utiliza do chamado gradiente descendente estocástico, como abordado na seção 2.4.4.2, sua implementação promete possibilitar um treinamento mais otimizado para a rede. Sendo assim, espera-se que esse método traga uma convergência mais rápida, levando a um menor erro em número de épocas reduzido quando em comparação com o *Backpropagation com Momentum*.

Seus parâmetros ligados ao primeiro e segundo momento foram selecionados como sendo: $\beta_1 = 0,5$ e $\beta_2 = 0,5$. Já o valor responsável por evitar a divisão por zero, foi estabelecido como $\varepsilon = 10^{-5}$.

3.3 Ferramentas e Equipamentos

Para a execução desse trabalho foi utilizado o *software* interativo MATLAB R2020a, através do qual foi implementado códigos com a arquitetura de TDNN com os otimizadores abordados acima.

Os arquivos de dados obtidos, normalizados e preditos foram armazenados na plataforma Google Sheets, a qual foi utilizada para construir gráficos qualitativos e

comparativos dos resultados de ambos métodos de treinamento. É uma boa opção, pois possibilita o acesso aos dados e resultados em mais de uma CPU.

O *hardware* principal utilizado para o processamento e previsão possui as seguintes especificações:

- Processador: Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz - 2.00 GHz;
- RAM: 4,00 GB;
- Windows 10 Home Single Language: Sistema operacional de 64 bits, processador baseado em x64.

3.4 Modelo de Comparação e Avaliação dos Resultados

Para cada treinamento e previsão realizadas foram encontrados os seguintes pontos de referência: Erro de Treinamento, Erro de validação, Épocas Totais e Erro de Predição. Os erros são obtidos através da função do erro quadrático médio, Equação 2.4.

Esses valores, assim como as curvas de predição foram analisados em conjunto e colocados em gráficos para uma melhor visualização e comparação dos métodos, do tamanho da rede ideal e do melhor horizonte temporal.

Quanto aos indicadores utilizados, esses possuem as variações da Tabela 2 para as primeiras rodadas de treinamento.

Meses Preditos	Nº de Entradas	Nº de Neurônios na camada oculta $N = 2N + 1$
12	12	25
6	12	25
3	12	25
12	24	49
6	24	49
3	24	49
12	36	73
6	36	73
3	36	73

Tabela 2 – Topologia das TDNNs para Seleção do Melhor Horizonte Temporal.

Em seguida, apenas para a previsão de três meses, foi realizada uma análise do impacto da variação do número de neurônios da camada escondida, através da modificação da regra de Kolmogorov tanto para o aumento de neurônios, quanto para a redução. Tal que a variação segue o padrão:

- $N1 = N + N/2$ (Variação 1 - Redução da camada);
- $N1 = 2N + 1$ (Regra de Kolmogorov);
- $N1 = 3N - N/2$ (Variação 2 - Aumento da camada).

Logo, o padrão será dado pela Tabela 3:

Nº de Entradas	Nº de Neurônios na camada oculta
12	18
12	25
12	30
24	36
24	49
24	60
36	54
36	73
36	90

Tabela 3 – Topologia das TDNNs Com Variação da Camada Oculta para Predição de 3 meses.

4 RESULTADOS E ANÁLISES

Este Capítulo trata dos resultados alcançados, por meio da metodologia abordada no Capítulo 3, para os dados de Energia Armazenada obtidos através da página oficial da ONS.

O Capítulo irá apresentar primeiramente a visão geral do estudo para ambos algoritmos utilizados (*Backpropagation + Momentum* e ADAM). Em seguida, as métricas abordadas serão correlacionadas com os gráficos para cada método de treinamento individual. Somente então o desempenho de cada algoritmo será comparado lado a lado.

4.1 Visão Geral dos Resultados

Através da variação do horizonte temporal de previsão e do número de entradas da RNA, o qual acarreta na variação do número de neurônios na camada oculta dada a regra de Kolmogorov. O número de épocas para todas as variações de topologia foi o máximo permitido, 10 mil. Os resultados das Tabelas 4 e 5 foram obtidos:

Variação de Meses Preditos e Variação do Número de Entradas X Erros dos Algoritmos					
Meses Preditos	Nº de Entradas	N_{Ce} $N1 = 2N + 1$	Backpropagation + <i>Momentum</i>		
			Erro de Treinamento	Erro de Validação	MSE da Predição
12	12	25	0,8939	0,1609	0,6464
6	12	25	1,0159	0,0993	0,1029
3	12	25	1,0367	0,0911	0,0104
12	24	49	0,8180	0,1222	0,5392
6	24	49	0,9111	0,0705	0,0359
3	24	49	0,8671	0,1193	0,0136
12	36	73	0,7164	0,0924	0,5045
6	36	73	0,6736	0,1825	0,0401
3	36	73	0,7326	0,1298	0,0234

Tabela 4 – Resultados da Variação das Entradas e do Horizonte de Temporal do *Backpropagation*.

Variação de Meses Preditos e Variação do Número de Entradas X Erros dos Algoritmos					
Meses Preditos	Nº de Entradas	N_{Ce} $N1 = 2N + 1$	Adam		
			Erro de Treinamento	Erro de Validação	MSE da predição
12	12	25	0,2186	0,0515	0,367
6	12	25	0,3456	0,0582	0,0257
3	12	25	0,3460	0,01083	0,025
12	24	49	0,0439	0,0118	0,3821
6	24	49	0,0630	0,0084	0,3002
3	24	49	0,0390	0,0138	0,0174
12	36	73	0,0105	0,0026	0,2375
6	36	73	0,0311	0,0065	0,1009
3	36	73	0,0447	0,0118	0,0123

Tabela 5 – Resultados da Variação das Entradas e do Horizonte de Temporal para o Adam.

Por meio dos dados encontrados, é possível compreender que ambos algoritmos tiveram melhor desempenho na previsão de três meses, ou seja, a médio prazo. Isso

ocorre, pois, para fazer as previsões a rede precisa utilizar dos valores calculados por ela e alimentá-los em si, pois não há conhecimento disponível do futuro para isso. Como consequência disso, ocorre uma contaminação da rede por seus próprios erros, gerando a explosão deles.

Ainda assim, algumas configurações específicas conseguiram também alcançar uma previsão razoável para seis meses. Contudo, visando aprofundar a análise em meios que promovem a especialização da rede e a minimização dos erros, especialmente o de previsão, optou-se por promover alterações no número de neurônios da camada oculta para a previsão de três meses apenas. Os resultados podem ser vistos nas Tabelas 6 e 7.

Erros x N° de Neurônios na Camada Oculta na Previsão de 3 Meses				
N° de Entradas	N_{Ce}	<i>Backpropagation + Momentum</i>		
		Erro de Treinamento	Erro de Validação	MSE da previsão
12	18	0,9689	0,1565	0,011
12	25	1,0367	0,0911	0,0104
12	30	0,9597	0,1686	0,0104
24	36	0,869	0,118	0,0147
24	49	0,8671	0,1193	0,0136
24	60	0,8745	0,1133	0,0125
36	54	0,7228	0,1387	0,0252
36	73	0,7326	0,1298	0,0234
36	90	0,6628	0,1999	0,0233

Tabela 6 – Resultados para variação do N° de Neurônios na Camada Oculta na Previsão de 3 Meses para o *Backpropagation*

Erros x N° de Neurônios na Camada Oculta na Previsão de 3 Meses				
N° de Entradas	N_{Ce}	ADAM		
		Erro de Treinamento	Erro de Validação	MSE da previsão
12	18	0,3503	0,0418	0,0622
12	25	0,346	0,0108	0,025
12	30	0,2697	0,0515	0,0171
24	36	0,0386	0,0085	0,0487
24	49	0,039	0,0138	0,0174
24	60	0,0655	0,0117	0,0173
36	54	0,0107	0,0017	0,0197
36	73	0,0447	0,0118	0,0123
36	90	0,0443	0,0066	0,0823

Tabela 7 – Resultados para variação do N° de Neurônios na Camada Oculta na Previsão de 3 Meses para o Adam

Fazendo a observação dos resultados das Tabelas 6 e 7 é possível afirmar que o padrão adotado pelas RNAs não é linear. Isso é, não necessariamente ao aumentar a rede

melhores resultados serão obtidos. De todas as variações implementadas, o menor erro de predição foi para o *Backpropagation + Momentum* com doze entradas e vinte e cinco neurônios na camada oculta. Em seguida, com o mesmo valor de erro de predição vem o mesmo algoritmo, mas dessa vez com a configuração de doze entradas e trinta neurônios na N_{Ce} . Por fim, em terceiro lugar ficou o ADAM na configuração de trinta e seis entradas e setenta e três neurônios na camada escondida. Ambos casos seguem o padrão da regra de Kolmogorov.

O lado positivo de se obter uma rede menor que alcança resultados tão bons quanto, ou ainda melhores que uma rede aumentada, está no fato de que isso leva a um gasto menor de tempo de processamento.

4.2 Resultados - *Backpropagation + Momentum*

De fato, a TDNN foi capaz de acompanhar com baixíssimo erro a tendência da Energia Armazenada para três meses, como na Figura 13. Isso revela que esse modelo de predição seria uma boa opção a ser implementada para uma melhor gestão dos recursos disponíveis.

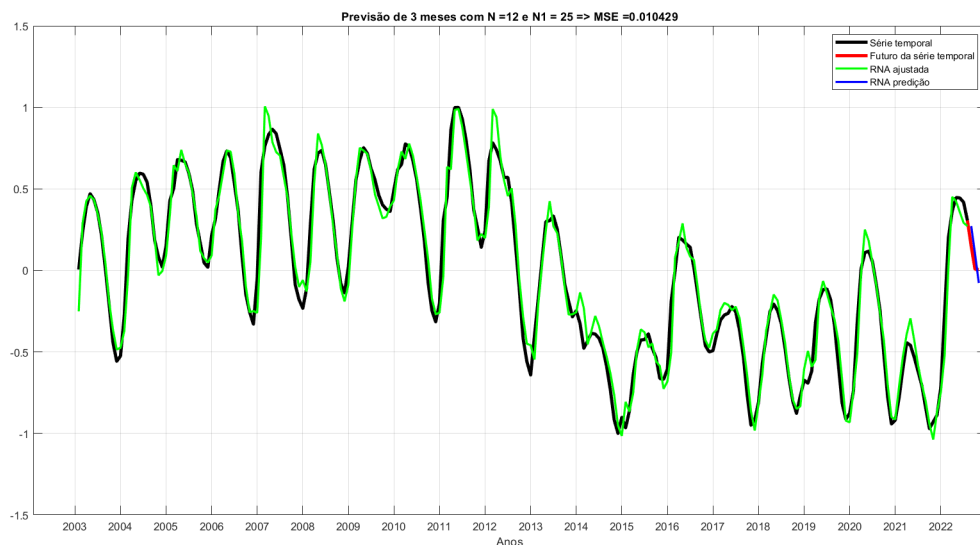


Figura 13 – Treinamento e Predição para 3 meses com 12 Entradas e 25 Neurônios na Camada Oculta

Já para seis meses, Figura 14, conforme a rede precisa realizar a predição nos meses finais, a tendência é de que a predição da RNA passe a se afastar do futuro da série de Energia Armazenada.

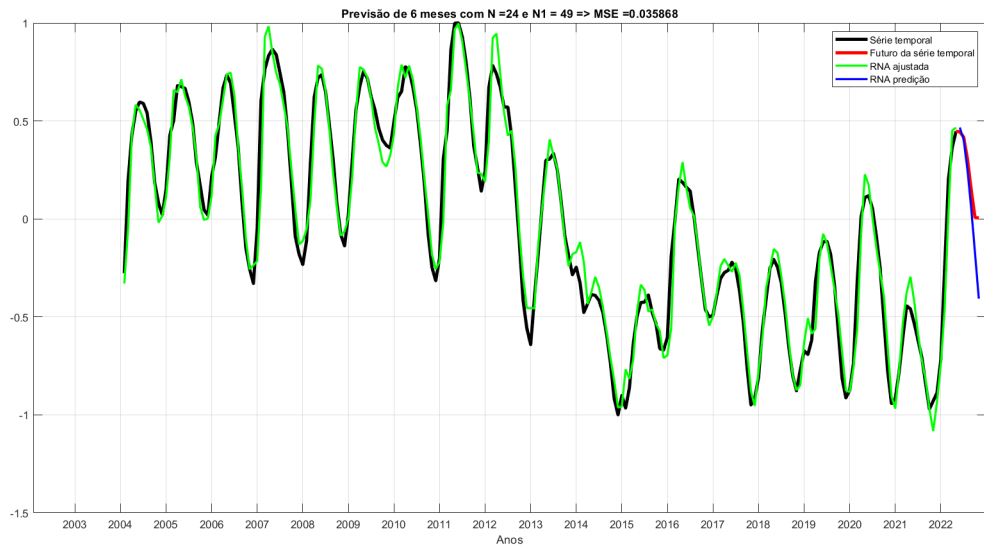


Figura 14 – Treinamento e Predição para 6 meses com 24 Entradas e 49 Neurônios na Camada Oculta

Quanto a relação para doze meses, vista na Figura 15, os erros explodem e, ainda que a tendência das curvas se assemelhem, o erro é muito alto.

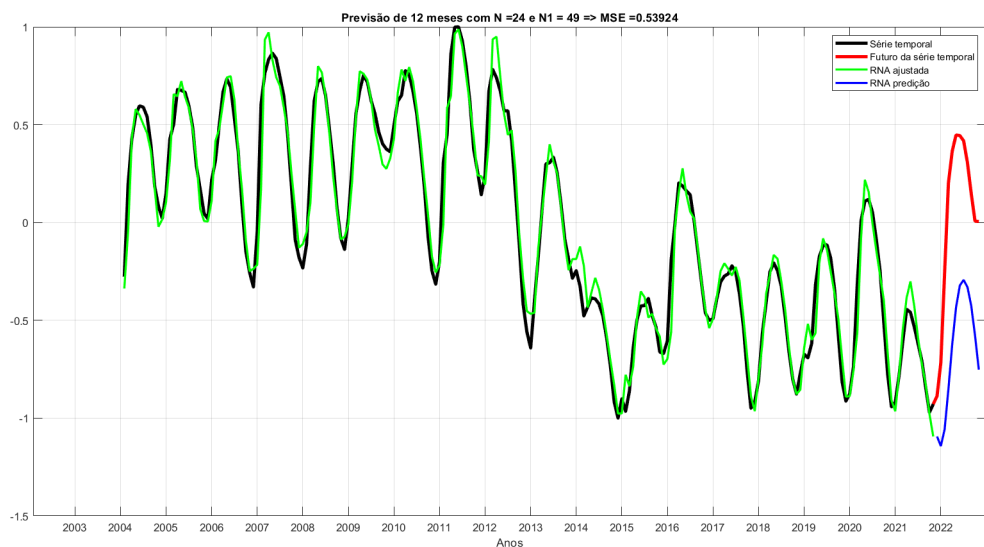


Figura 15 – Treinamento e Predição para 12 meses com 24 Entradas e 49 Neurônios na Camada Oculta

A relação entre número de entradas, horizonte temporal e erros pode ser melhor compreendida pelos gráficos presentes nas Figuras 16, 17 e 18.

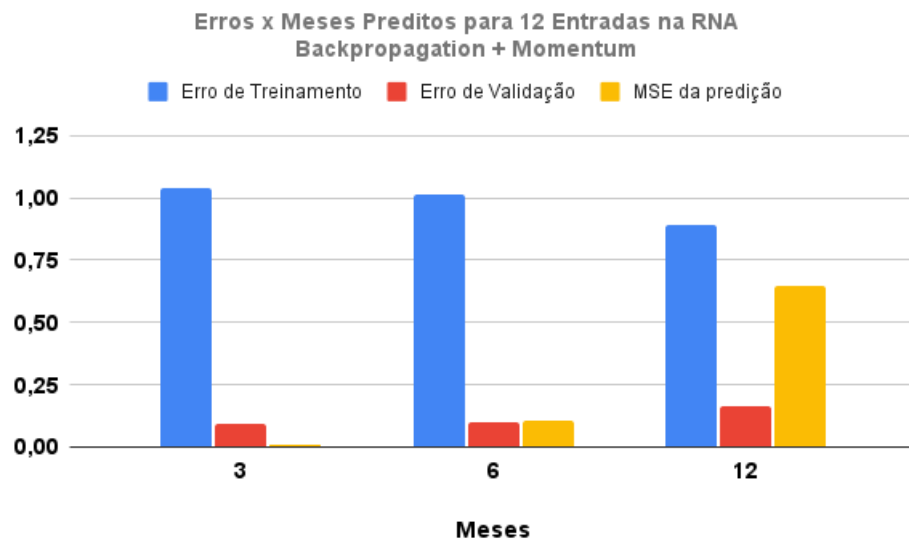


Figura 16 – Erros x Meses Preditos para 12 Entradas na RNA com *Backpropagation + Momentum*.

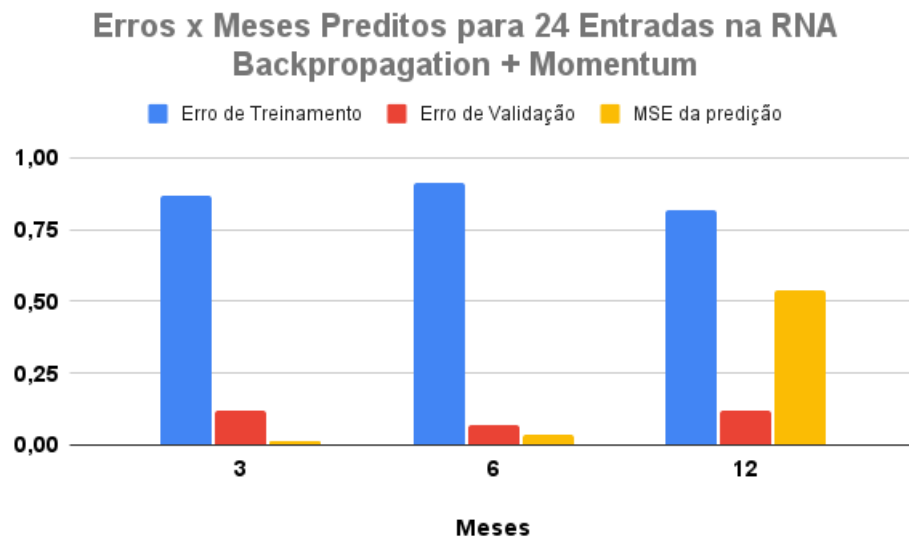


Figura 17 – Erros x Meses Preditos para 24 Entradas na RNA com *Backpropagation + Momentum*.

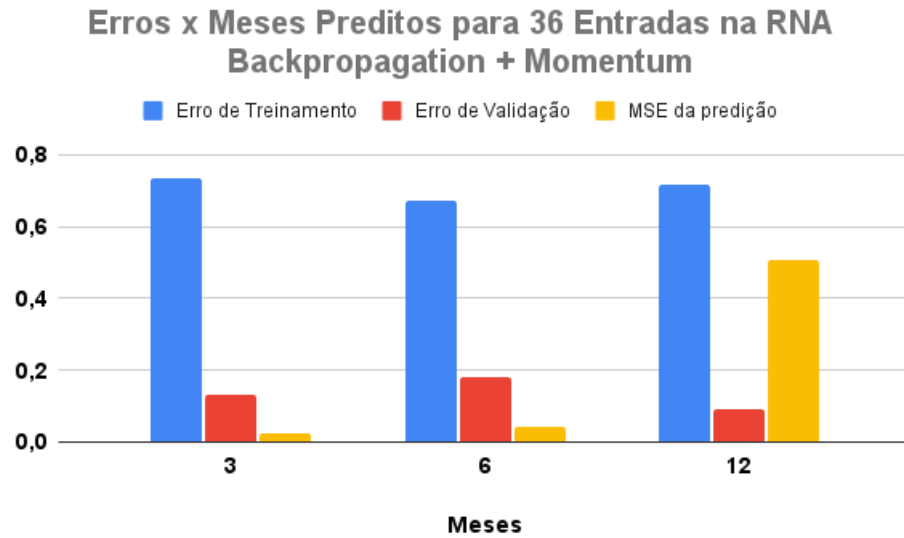


Figura 18 – Erros x Meses Preditos para 36 Entradas na RNA com *Backpropagation + Momentum*.

Aumentar o número de entradas não teve um impacto claramente positivo sobre os resultados, lembrando que o número de neurônios da camada escondida está relacionado a isso através da regra de Kolmogorov. Ou seja, sua variação para mais também não aponta significativa melhora dos resultados. Contudo, ao aumentar o período de predição, fica nítida a deterioração da qualidade da previsão da rede.

A partir da Figura 19 nota-se que, para a série estudada, não é vantajoso aumentar o número de neurônios na camada oculta. Ainda que o erro de treinamento passe a impressão de reduzir com o crescimento da rede, o erro de validação não revela a mesma tendência. Já em relação ao erro de predição, esse tende a piorar para um número de neurônios acima de 25 na camada oculta, o que aponta para um sobredimensionamento da rede.

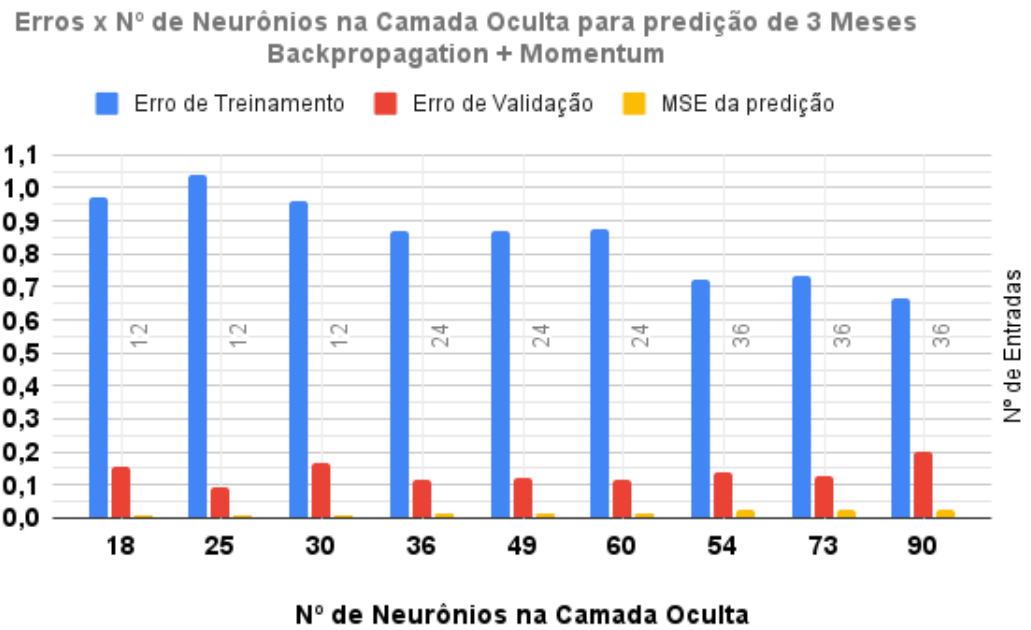


Figura 19 – Erros por Variação da Camada Oculta para Predição de 3 Meses.

4.3 Resultados - Adam

No gráfico da Figura 20, pode-se ver o padrão do melhor resultado obtido com Adam, é possível notar que o erro de treinamento e validação parecem muito baixos e a curva da RNA ajustada consegue acompanhar os dados de Energia Armazenada usados para o treinamento com pouco erro.

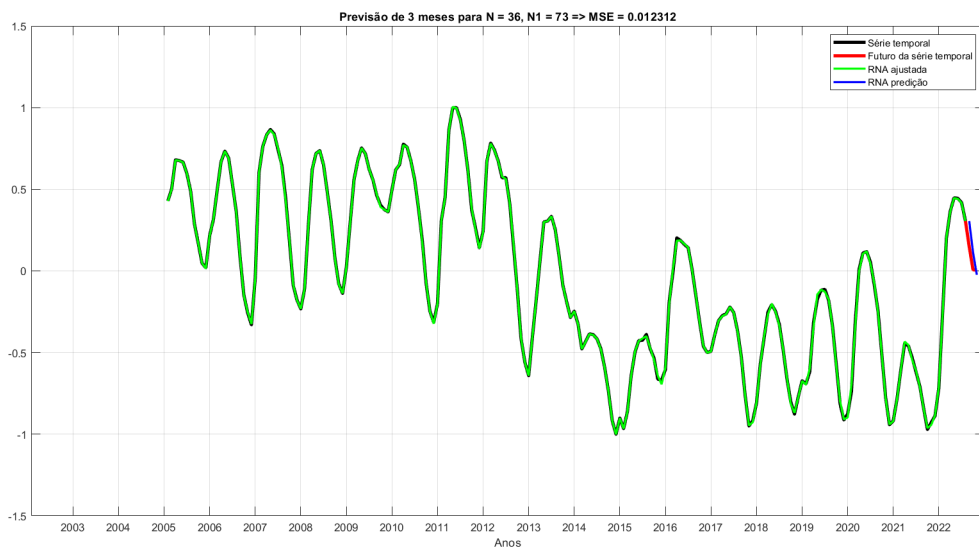


Figura 20 – Treinamento e Predição para 3 meses com 36 Entradas e 73 Neurônios na Camada Oculta

Na predição de seis meses, Figuras 21 e 22, fica evidente a piora qualitativa do erro de predição para um número maior de variáveis de entrada e também no aumento do número de neurônios na camada oculta. A previsão para seis meses precisa de um maior cuidado no dimensionamento da rede e as variações da topologia apresentaram uma maior influência sobre a qualidade da resposta.

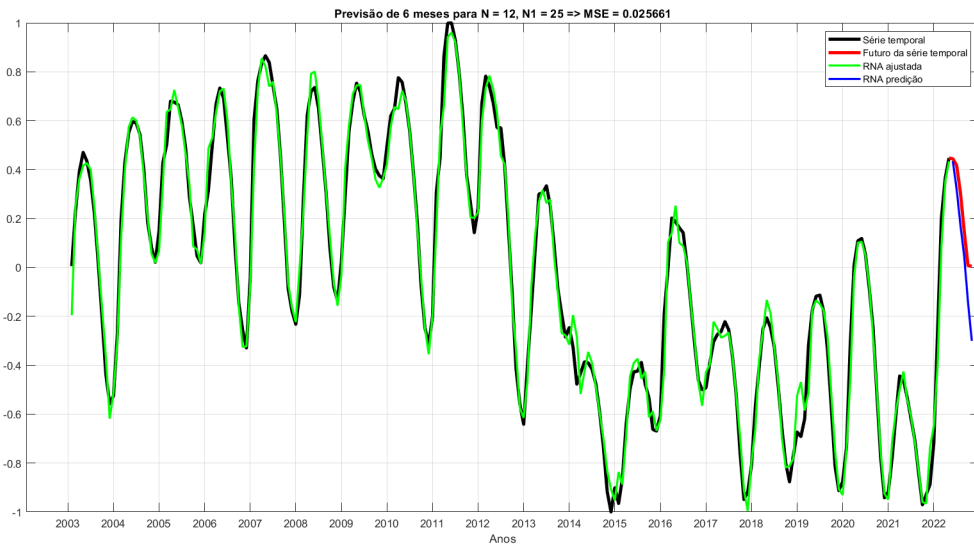


Figura 21 – Treinamento e Predição para 6 meses com 12 Entradas e 25 Neurônios na Camada Oculta

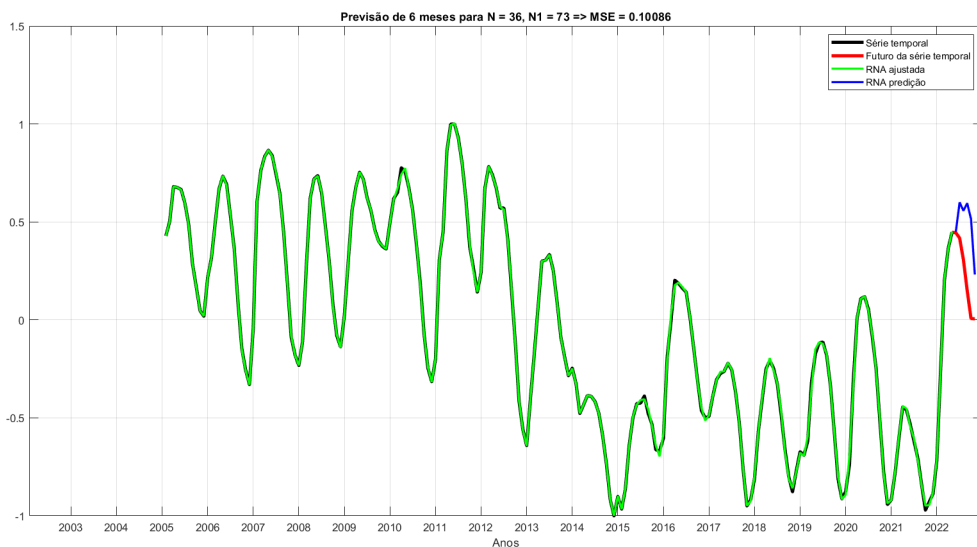


Figura 22 – Treinamento e Predição para 6 meses com 36 Entradas e 73 Neurônios na Camada Oculta

Já a predição de doze meses, assim como abordado anteriormente em 4.2, não foi muito bem sucedida para nenhuma das topologias adaptadas, pode-se ver, como exemplo disso, a Figura 23.

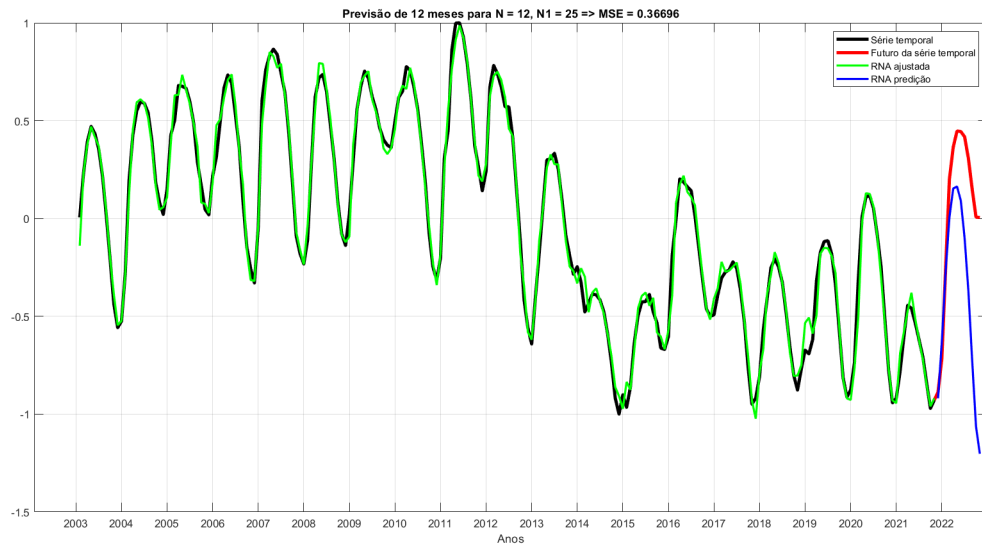


Figura 23 – Treinamento e Predição para 12 meses com 12 Entradas e 25 Neurônios na Camada Oculta

A procura de algum padrão, ou a observação da inexistência de um, se torna mais fácil através dos gráficos que relacionam os erros, como os das Figuras 24, 25 e 26. Além disso, nota-se que a topologia de 12 entradas apresentou bons resultados para horizontes temporais de 3 e seis meses. Já para 24 e 36 entradas, os resultados satisfatórios ficam restritos à 3 meses. O que fortalece a conclusão de que o aumento do tamanho da rede não necessariamente implica em melhores resultados. Isto se deve às particularidades de cada tipo de dado e também ao efeito do supertreinamento (*overtraining*), que leva a rede a se adaptar excessivamente, fornecendo boa conformação aos dados de treinamento mas uma adaptação ruim à predição (dados que não se encontram no grupo de treinamento).

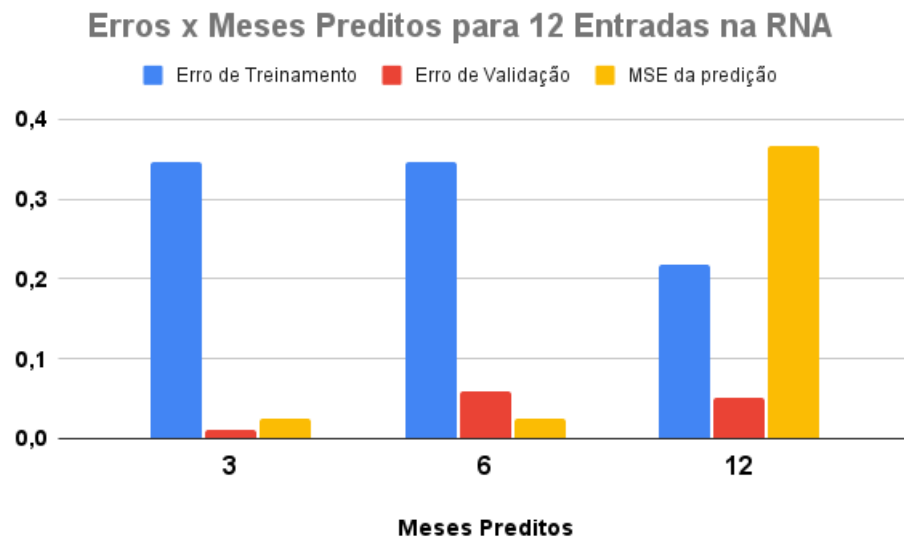


Figura 24 – Erros x Meses Preditos para 12 Entradas na RNA com ADAM

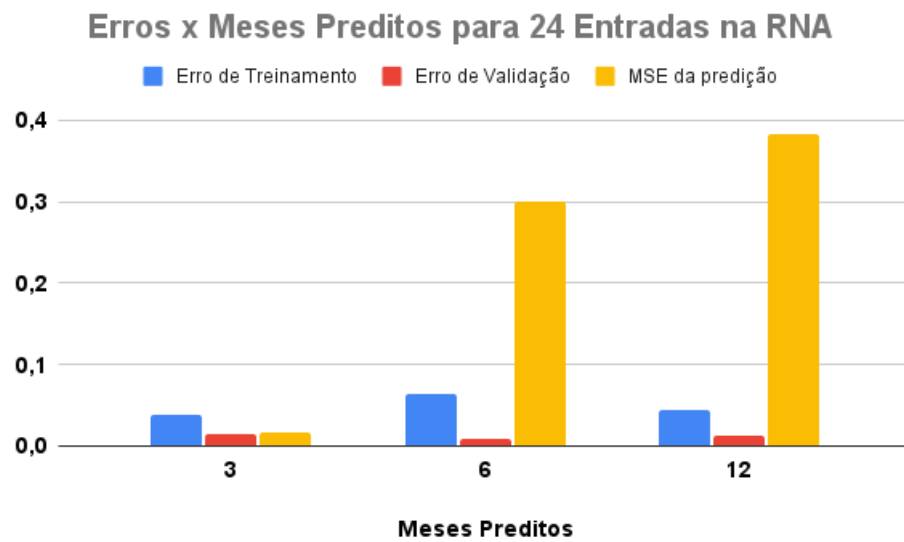


Figura 25 – Erros x Meses Preditos para 24 Entradas na RNA com ADAM

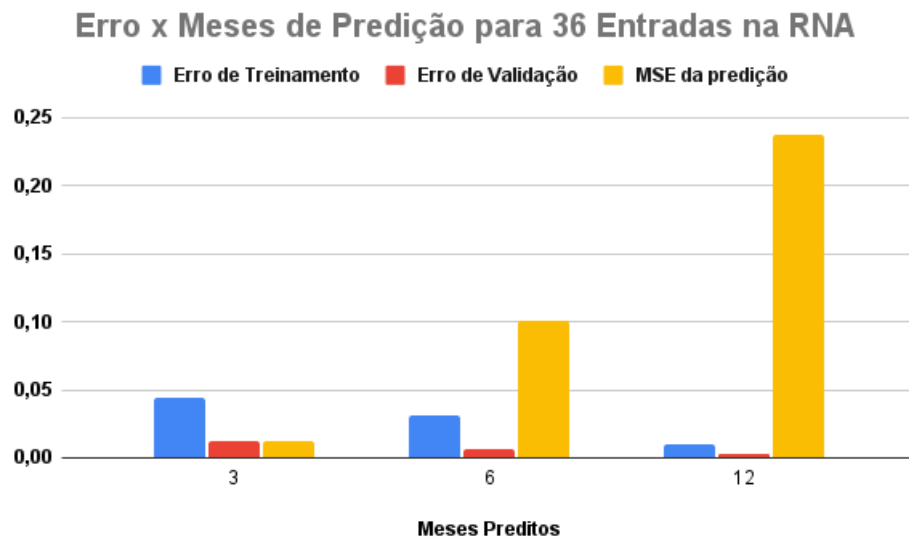


Figura 26 – Erros x Meses Preditos para 36 Entradas na RNA com ADAM

Em relação a variação da camada oculta para 3 meses de horizonte temporal, os resultados da Figura 27 foram obtidos. Para esse método de treinamento, a proporção entre o número de entradas e número de neurônios da camada oculta parece impactar positivamente o resultado até 73 neurônios na camada escondida. A partir desse ponto, o erro de predição sobe e indica um dimensionamento incorreto para a rede.

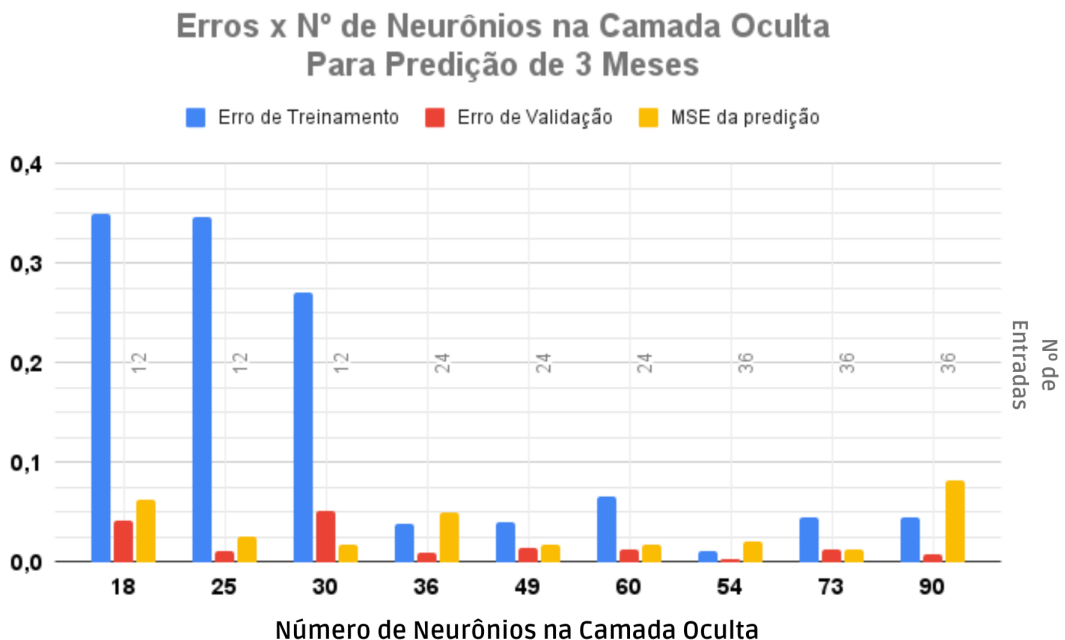


Figura 27 – Erros x Número de Neurônios na Camada Oculta para ADAM

Observe na Figura 27 que até 30 neurônios na camada escondida o erro de treinamento é significativamente maior, contudo isso não gera um impacto diretamente

relacionado ao erro de predição, o qual permanece dentro de uma faixa entre 0 e 0,1. Pode-se notar também que não necessariamente a redução de um dos erros implica na redução dos demais erros. Ou seja, o erro de validação pode ser baixo, assim como o de treinamento, mas o erro de predição ser maior do que para outras configurações topológicas.

4.3.1 Resultados Comparativo: Adam \times *Backpropagation* + *Momentum*

Diante a adoção de dois modelos de treinamento, um tradicional e um mais recentemente desenvolvido, para a mesma arquitetura de TDNN, torna-se interessante fazer uma comparação lado a lado da qualidade dos resultados obtidos por eles.

Para a etapa do estudo na qual se fez a análise dos resultados para a predição de diferentes horizontes temporais, um dos pontos que salta aos olhos é o erro de treinamento do *Backpropagation* que gira em torno de três vezes o erro de treinamento do Adam. Esse fato pode observado nas Figuras 28, 29, 30 e também no próprio traçado da RNA ajustada por cada método, o qual tem uma qualidade melhor para o Adam.

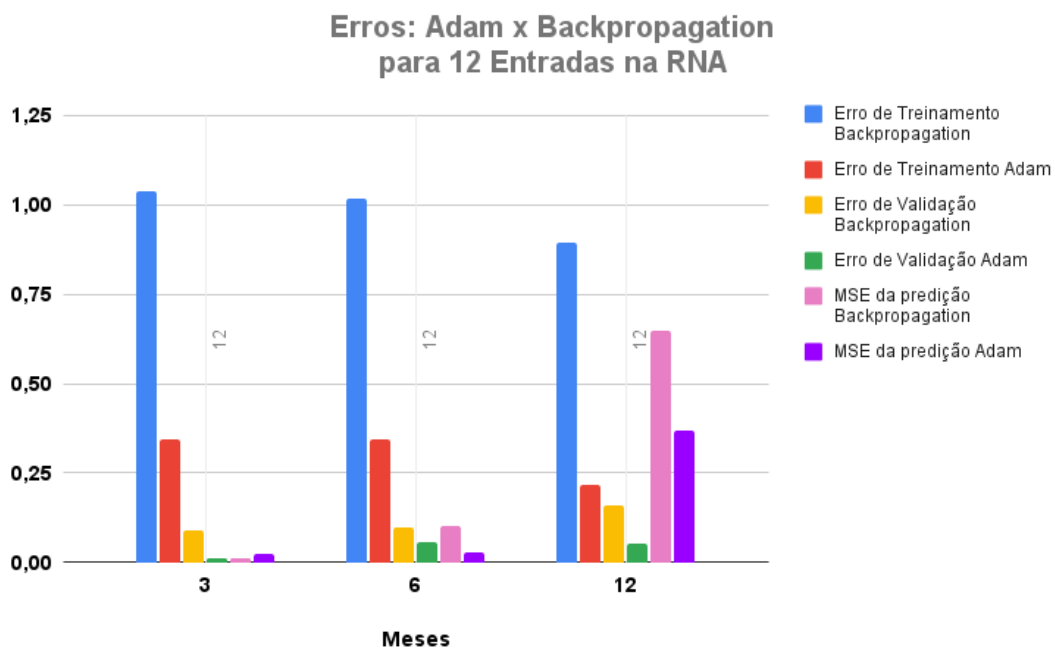


Figura 28 – Erros: ADAM x *Backpropagation* para 12 entradas na RNA

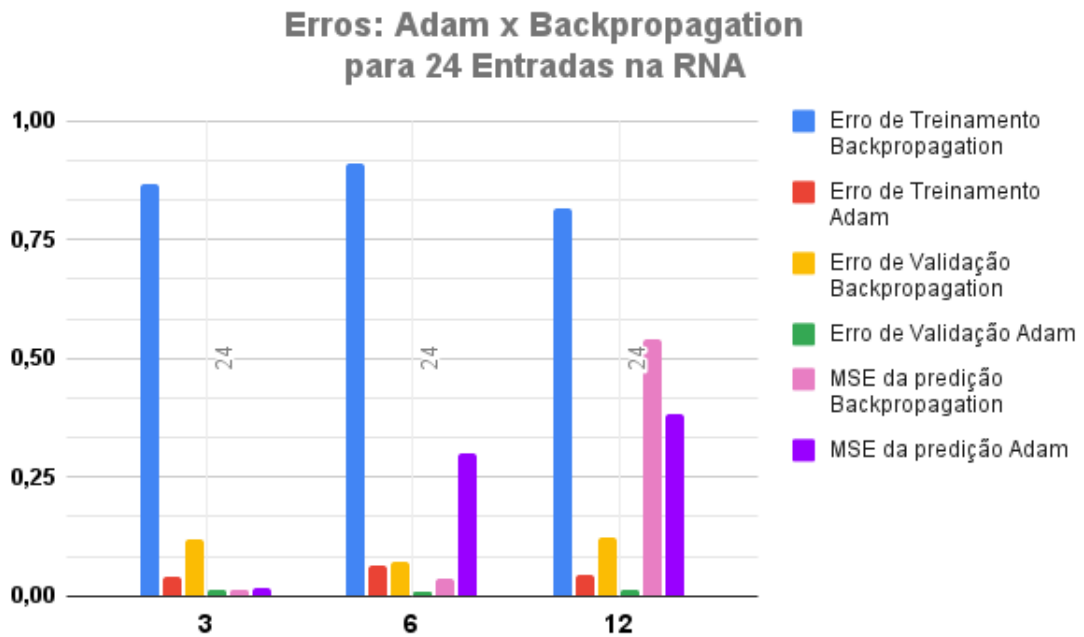


Figura 29 – Erros: ADAM x *Backpropagation* para 24 entradas na RNA

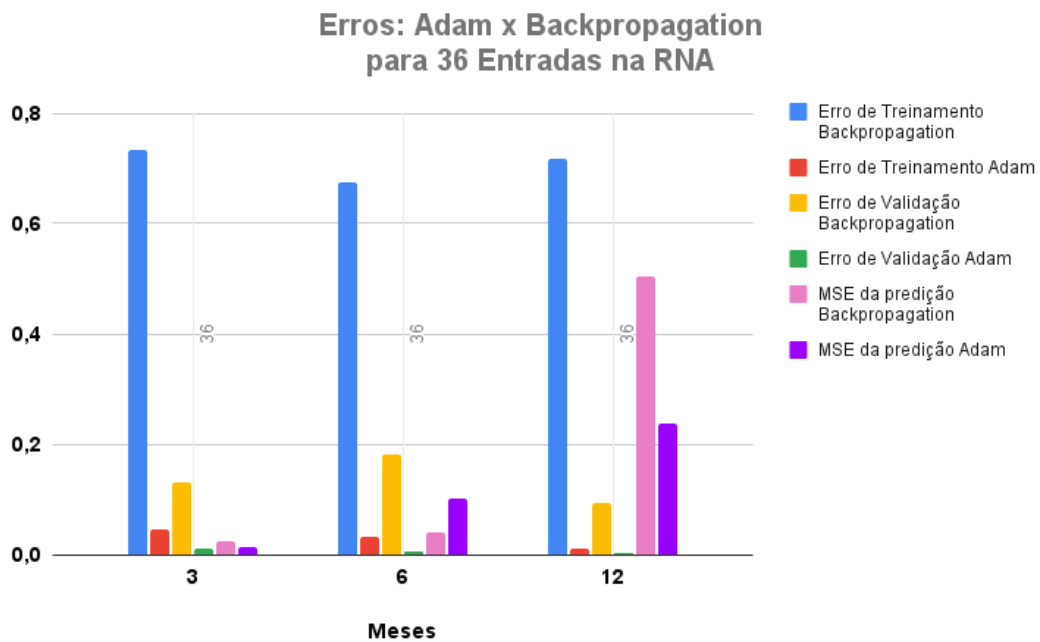


Figura 30 – Erros: ADAM x *Backpropagation* para 36 entradas na RNA

Outro aspecto interessante é que os erros de treinamento e de validação do algoritmo Adam são sempre menores em relação ao *Backpropagation*, porém em relação ao erro de predição o mesmo não ocorre. Ou seja, para a predição do futuro inacessível, ambos tiveram resultados muito semelhantes, especialmente para a previsão de três meses.

Já para seis meses, a topologia teve um papel fundamental, pois variando as entradas, e lembrando que nesta etapa a regra de Kolmogorov foi utilizada para estimar o número de neurônios na camada escondida, os resultados foram variados. Para doze entradas o erro da previsão do *Backpropagation* foi maior, mas já para 24 e 36 entradas o do Adam foi maior.

Quanto à previsão de doze meses, foi a que melhor indicou uma possível superioridade do Adam, visto que, para essa, todos os erros do *Backpropagation* foram maiores. Ou seja, esse algoritmo se destacou para a previsão de um horizonte temporal maior, especialmente para uma RNA mais encorpada, com 36 entradas e 73 neurônios na camada escondida.

Porém, como mencionado no Capítulo 3, e algo que também foi possível confirmar na comparação dos erros, as melhores previsões foram nitidamente para o intervalo de três meses. Logo, nas Figuras 31 e 32, se observa como a variação do número de neurônios na camada oculta impactou cada um dos algoritmos utilizados.

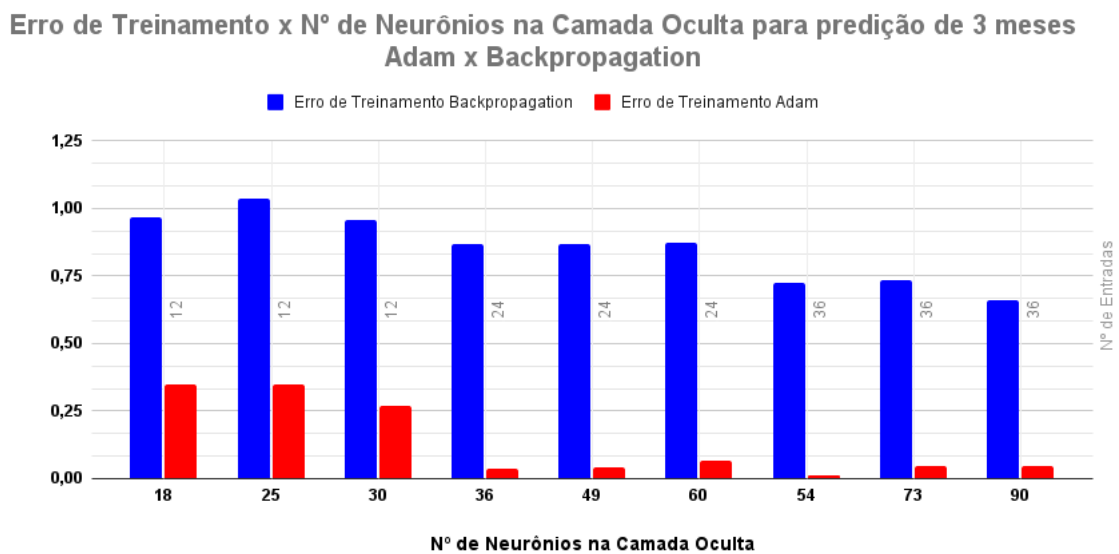


Figura 31 – Comparativo do Erro de Treinamento por N° de Neurônios na Camada Oculta para predição de 3 meses.

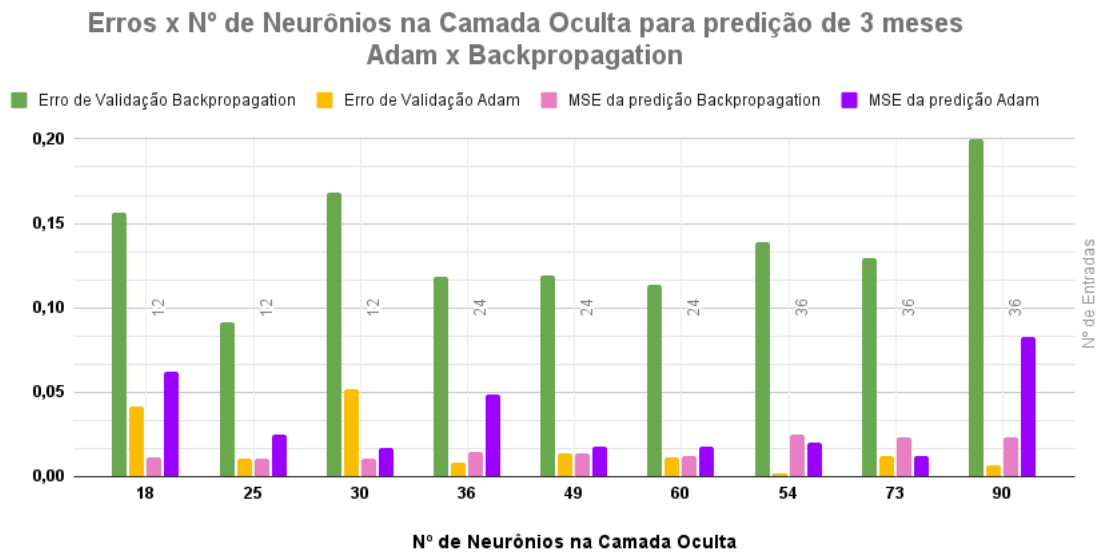


Figura 32 – Comparativo do Erro de Predição e de Validação por N° de Neurônios na Camada Oculta para predição de 3 meses.

Um dos fatos observados novamente é que não há um padrão claro que indique qual número de neurônios é o melhor para realizar a predição. E, muitas vezes, uma rede menor pode alcançar resultados mais satisfatórios.

Algo muito interessante, porém, se dá na performance do *Backpropagation*, o método obteve ótimas previsões para redes menores. O que indica que um algoritmo mais potente como o Adam não é necessário, na verdade é redundante para essas predições. Inclusive, seu maior valor nos erros da predição pode até mesmo estar relacionado com isso, a rede pode ter ficado especializada demais, ou seja, incapaz de fazer a generalização da maneira ideal, como indicam os erros baixos de validação e treinamento e isso levou ao prejuízo da capacidade de generalização na predição.

Dado que o melhor resultado de predição foi alcançado para o algoritmo de *Backpropagation* na topologia de 12 entradas com 25 neurônios na camada oculta. Observe nas Figuras 33 e 34 o ajuste da RNA e a predição de ambos os métodos para essa configuração específica.

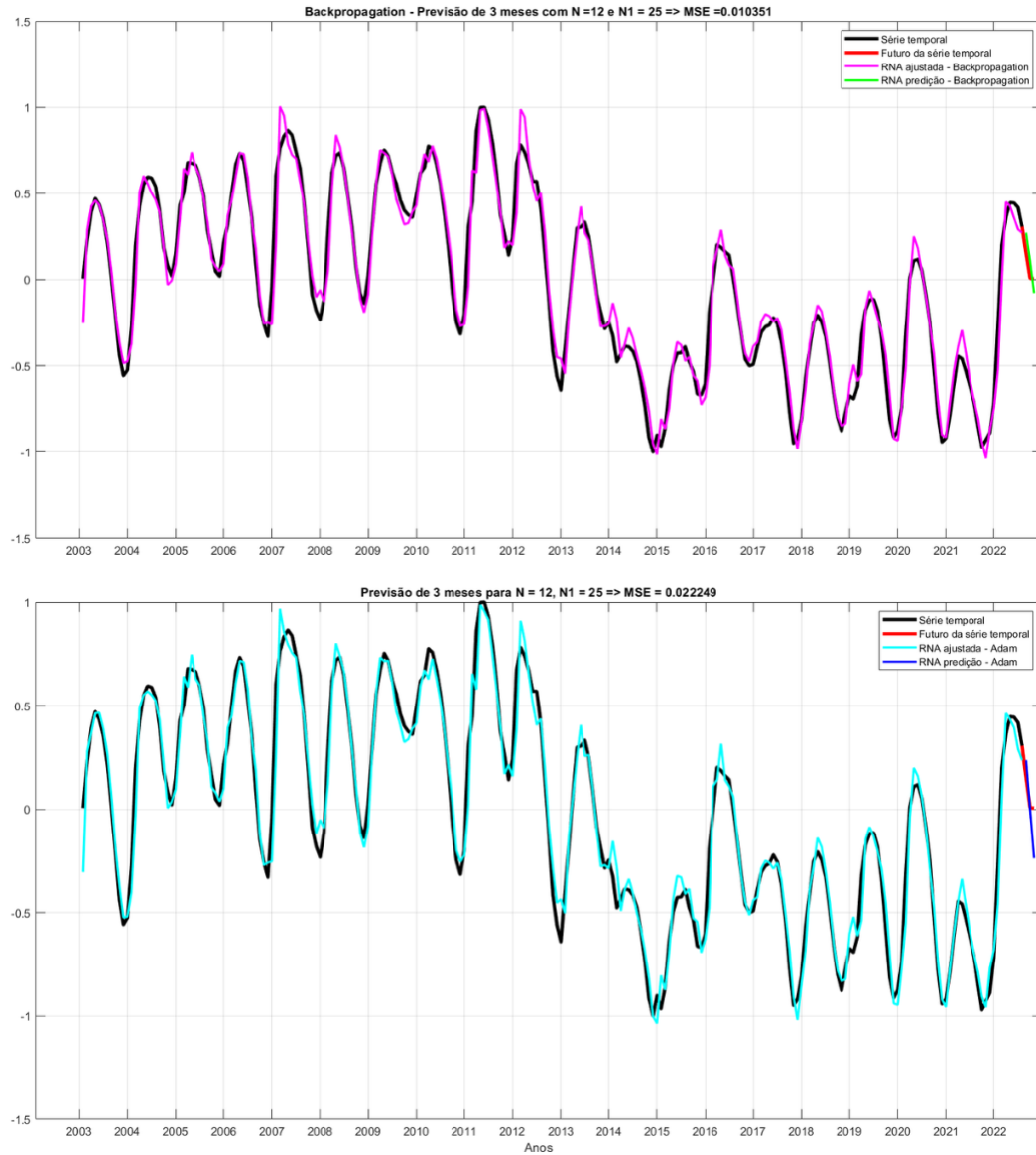


Figura 33 – Ajuste e Predição: *Backpropagation* x Adam.

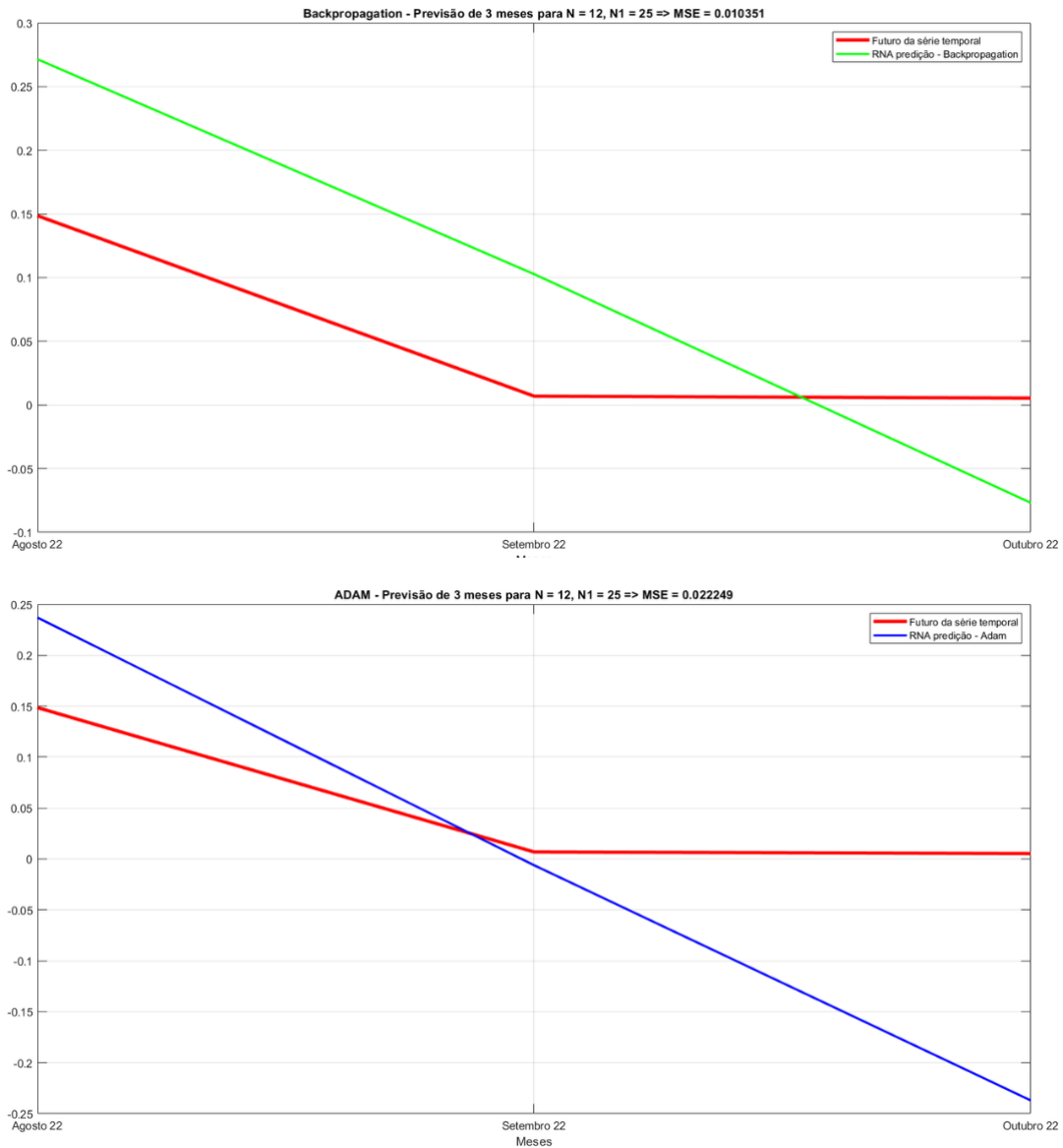


Figura 34 – Predição de 3 meses: *Backpropagation* x Adam

Apesar da diferença não ser facilmente captada ao observar as linhas dos gráficos completos, ao ampliar apenas os dados preditos pode-se ver que o *Backpropagation* conseguiu acompanhar melhor a tendência da curva de Energia Armazenada normalizada.

De maneira geral, ambos conseguiram uma boa adaptação aos dados, porém de fato, que para essa topologia, o erro de predição advindo do treinamento Adam foi mais que o dobro em relação ao do *Backpropagation*. Sendo que, para observar uma redução de erro para o Adam, foi necessário alterar a topologia da rede, aumentando o número de entradas e de neurônios na camada escondida. Procedimento, esse, que não seria vantajoso, pois o erro de predição ainda assim não alcançou o valor gerado no *Backpropagation* e, além disso, o crescimento da rede gera um gasto maior com processamento.

Ou seja, para algumas predições de séries temporais estacionárias mais vale um

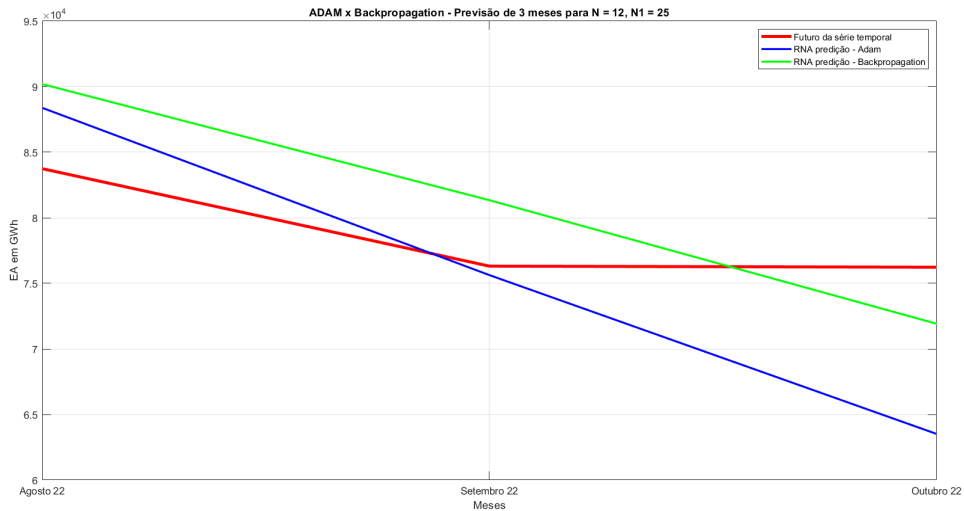
Tabela Comparativa da Predição x Dados Reais em GWh						
Mês	agosto 2022	Erro da Predição	setembro 2022	Erro da Predição	outubro 2022	Erro da Predição
Dados Reais de EA	83.745,00	-	76.312,00	-	76.231,00	-
Backpropagation	90.193,10	7,70%	81.339,40	6,59%	71.919,40	5,66%
Adam	88.384,60	5,54%	75.628,40	0,90%	63.516,60	16,68%

Tabela 8 – Tabela Comparativa da Predição x Dados Reais em GWh

método tradicional e simplista do que estratégias atuais prolíficas e altamente empregadas que prometem uma convergência mais rápida.

Levando em consideração a escala dos dados obtidos de Energia Armazenada, foi realizada a desnormalização dos dados preditos de ambas os modelos para a topologia de melhor resultado, tal que foi obtida a Tabela 8.

O gráfico com valores corrigidos para ambas as previsões pode ser visto na Figura 35 .

Figura 35 – Predição de 3 meses Com Valores Desnormalizados: *Backpropagation* x Adam

Ao desnormalizar os dados obtidos e compará-los aos valores que de fato ocorreram, pode-se compreender melhor como a imprecisão da previsão impacta os valores para números tão elevados. O erro absoluto da pior predição do *Backpropagation* chega a 6.448,1 GWh, já para o Adam, o erro da pior predição, alcança 12.714,4 GWh.

Algo que, também, ocorre é que a previsão por meio do algoritmo Adam obteve melhores resultados para Agosto e Setembro. Já para o mês de Outubro o *Backpropagation* foi melhor.

Em relação a comparação geral dos dois métodos de treinamento, ambos tiveram um desempenho aceitável diante do modelo de rede proposto e dos dados da série original. Além disso, ambos algoritmos funcionam melhor para a previsão a médio prazo de três

meses e conforme o aumento do horizonte temporal a resposta de ambos se deteriorou devido as redes ficarem contaminadas com os próprios erros.

5 CONCLUSÕES

5.1 Conclusões

Este trabalho abordou e comparou diferentes algoritmos de treinamento de redes neurais artificiais para a proposta de realizar a predição da energia armazenada da região Sudeste/Centro-Oeste do SIN, utilizando como dados de treinamento, validação e comparação os dados do histórico de operação do sistema fornecidos na página oficial da ONS.

Conceitos sobre o sistema elétrico brasileiro foram introduzidos, assim como diferentes métodos de armazenamento de energia e importância desta para garantir a segurança energética.

O significado de fazer uma predição foi discutido historicamente e como seus resultados finais possuem a capacidade de influenciar decisões e comportamentos sociais.

Seguidamente, noções sobre redes neurais artificiais e otimizadores, assim como seus meios de funcionamento, foram introduzidos. A modelagem matemática genérica de uma rede MLP, a qual foi implementada no trabalho, foi exposta, assim como as diferenças específicas dos para cada camada da rede.

Ainda sobre as RNA, foi explicado como ocorre o treinamento através das fases *feedforward* e *Backpropagation* e como o fazer o modelo de alimentação de dados de entrada para tornar uma rede MLP em TDNN.

Foi abordado, também, a necessidade de utilizar ao menos uma camada oculta para que a previsão de séries temporais seja bem sucedida, e como o dimensionamento do número de neurônios dessa camada pode ser realizado.

Foram, então, aplicados dois métodos de treinamento, um tradicional e outro moderno, *Backpropagation + Momentum based learning* e Adam respectivamente. Visando encontrar o melhor horizonte de tempo para a previsão, as características da rede foram variadas na expectativa de descobrir uma melhor configuração topológica, diferentes números de entradas e de neurônios na camada oculta foram testados. Sendo que, para ambos métodos de treinamento o melhor horizonte temporal foi o de três meses, para os demais, a rede não obteve uma capacidade satisfatória de generalização dos dados e os erros se mostraram maiores.

Na avaliação dos resultados obtidos através do treinamento por *Backpropagation + Momentum based learning*, a topologia que obteve o menor erro de predição foi a de 12 entradas, o que está alinhado com a função de autocorrelação da Energia Armazenada, e 25 neurônios na camada oculta, o que obedece a regra de Kolmogorov. Isso foi observado para a predição de três meses. Foi possível reparar que uma redução significativa dos erros de

treinamento e de validação não necessariamente resultam em um erro de predição menor, na verdade, isso pode levar a especialização da rede e a incapacidade de generalização dos dados como deve ocorrer. Além disso, foi possível a não linearidade da rede, visto que uma maior rede não gera obrigatoriamente um resultado com menor erro. A rede com 36 entradas, ficou sobredimensionada e isso chegou a elevar os erros de predição e de validação.

Quanto à observação das respostas alcançadas com o Adam, a melhor topologia foi para a previsão de três meses com 36 entradas e 73 neurônios na camada oculta. Além disso, seus erros de treinamento para até 30 neurônios na camada oculta foram maiores, o que possivelmente poderia indicar um subdimensionamento. Ou seja, para esse algoritmo, uma rede maior foi mais eficaz.

Ao comparar os dois métodos de treinamento o *Backpropagation + Momentum based learning*, para predição de 3 meses, obteve melhores resultados com uma rede reduzida. Algo que não se imaginaria visto que Adam é um algoritmo mais potente e que promete uma convergência mais rápida até mesmo para séries com ruídos. Adam, porém, para a configuração de 12 entradas e 25 neurônios na camada oculta conseguiu um menor erro de predição para seis meses. Além disso, para predições de 12 meses, conseguiu manter um erro de predição um pouco menor, porém não tão bom para ser utilizado para realizar previsões nas configurações impostas.

Ou seja, para doze entradas, ambos poderiam ser utilizados para realizar a previsão de até seis meses com sucesso dependendo da topologia adotada, afinal os dois conseguiram acompanhar a tendência da curva de Energia Armazenada, sendo que o *Backpropagation* se revelou melhor para horizontes menores.

Já para 24 entradas e 49 neurônios na camada oculta o *Backpropagation* se saiu melhor tanto para três quanto para seis meses de predição. Em 36 entradas e 73 neurônios na camada escondida, o Adam foi melhor para a predição de três meses e o *Backpropagation* para seis meses.

Um ponto em comum entre ambos os métodos foi que os menores erros individuais de predição foram para três meses de horizonte temporal. Por conta disso, conclui-se que esse seria o tempo ideal para se prever dentro o que foi proposto.

Quanto a comparação dos erros de treinamento e validação, foi observado que Adam de fato gera menores números para ambos em praticamente todas as topologias, contudo isso não se traduziu como um menor erro de predição. Na verdade, gerou a impressão de causar o oposto, no caso, de que Adam estaria especializado demais nos dados e isso prejudicou a generalização.

Concluí-se, portanto, que uma TDNN com treinamento tradicional através de *Backpropagation + Momentum based learning* é o melhor modelo para se utilizar na

predição da Energia Armazenada, dado que essa forma uma série temporal estacionária, de fácil processamento e predição.

5.2 Propostas de Trabalhos Futuros

Considerando os resultados obtidos e as limitações definidas pelo horizonte temporal, seguindo a linha de pesquisa voltada para a solução de problemas de bola de cristal, seria interessante averiguar diferentes técnicas para alcançar uma predição a longo prazo eficiente, dentre elas:

- Alteração da topologia da RNA para um modelo de rede recorrente como as de Elman e Jordan.
- Aplicar de maneira auxiliar diferentes metodologias de de inteligência artificial, como sistemas fuzzy e algoritmos genéticos.
- Utilizar dados relacionados à Energia Armazenada, como a demanda e dados hidrológicos para auxiliar a predição.

REFERÊNCIAS

- AGGARWAL, C. C. *Neural Networks and Deep Learning*. 1. ed. [S.l.]: Springer International Publishing AG, 2018.
- ALLERIN. *Recurrent Networks*. 2022. Disponível em <<https://www.allerin.com/>>. Acessado em 17 Dec. 2022.
- ANEEL. *Agência Nacional de Energia Elétrica*. 2022. Disponível em <<https://www.gov.br/aneel/pt-br>>. Acessado em 31 Nov. 2022.
- BACH, V. R. *DIMENSIONAMENTO DA CAPACIDADE DE PRODUÇÃO DE HIDROGÊNIO A PARTIR DA ENERGIA VERTIDA TURBINÁVEL DE ITAIPU E SUA UTILIZAÇÃO EM VEÍCULOS AUTOMOTIVOS*. 2013. Disponível em <<https://acervodigital.ufpr.br/bitstream/handle/1884/35300/TCC%20Vanessa%20Rossato%20Bach%20PDF.pdf?sequence=1&isAllowed=y>>. Acessado em 16 dec. 2022.
- BOX, G. E. P.; JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. [S.l.: s.n.], 1990.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. 4. ed. [S.l.: s.n.], 1994.
- CASARIN, R. *Tesla tem resultado recorde de instalações de sistemas de armazenamento*. 2021. Disponível em <<https://www.portalsolar.com.br/noticias/tecnologia/armazenamento/tesla-tem-resultado-recorde-de-instalacoes-de-sistemas-de-armazenamento>>. Acessado em 05 jan. 2023.
- CAUDILL. *Neural Network Primer*. [S.l.]: Miller Freeman Publications, 1990.
- ELMAN, J. L. *Finding Structure in Time*. 1990.
- ENERGIA, M. L. de. *Consumidores livres e especiais*. 2022. Disponível em <<https://www.mercadolivredeenergia.com.br/consumidores-livres-e-especiais/conceito/>>. Acessado em 17 Dec. 2022.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. 1. ed. [S.l.: s.n.], 2016.
- HAYKIN, S. *Redes Neurais: Princípios e Prática*. 2. ed. [S.l.: s.n.], 2000.
- HAYKIN, S. *Redes Neurais: Princípios e Prática*. 2. ed. [S.l.: s.n.], 2000.
- HIRUMA, F. S. *PREVISÃO DA GERAÇÃO DE ENERGIA EÓLICA COM REDES NEURAIIS ARTIFICIAIS*. 2016. Disponível em <<https://acervodigital.ufpr.br/bitstream/handle/1884/43520/R%20-%20D%20-%20FELIPE%20SHOGO%20HIRUMA.pdf?sequence=1&isAllowed=y>>. Acessado em 31 out. 2020.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. 2. ed. [S.l.: s.n.], 2018.
- IBGE. *Instituto Brasileiro de Geografia e Estatística*. 2022. Disponível em <<https://www.ibge.gov.br/>>. Acessado em 31 Nov. 2022.

- KHAN, S. *Khan Academy*. 2022. Disponível em <<https://khanacademy.org/>>. Acessado em 17 Dec. 2022.
- KINGMA, D.; BA, J. *Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations*, 2015.
- KOLMOGOROV, A. N. *On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of one Variable and Addition*. [S.l.]: Doklady Akademii, 1957.
- NASCIMENTO, F.; OLIVI, L.; LIMA, L. *Multi-step forecasting of the Brazilian energy market price using autoregressive neural networks. Simposio Brasileiro de Sistemas Elétricos (SBSE), Niteroi*, 2018.
- ONS. *Balanco de Energia*. 2022. Disponível em <<https://www.ons.org.br/paginas/energia-agora/balanco-de-energia>>. Acessado em 17 Dec. 2022.
- ONS. *Energia Armazenada*. 2022. Disponível em <https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/energia_armazenada.aspx>. Acessado em 17 Dec. 2022.
- ONS. *O sistema em números*. 2022. Disponível em <<https://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>>. Acessado em 17 Dec. 2022.
- ORR GENEVIEVE; MÜLLER, K.-R. *Neural networks : tricks of the trade*. 1. ed. [S.l.: s.n.], 1998.
- PASQUALI, L. M. *ESTUDO SOBRE A INFLUÊNCIA DA RESERVAÇÃO HIDRÁULICA EM SISTEMAS DE GERAÇÃO EÓLICA ISOLADOS*. 2006. Disponível em <<https://www.lume.ufrgs.br/bitstream/handle/10183/12888/000626342.pdf?sequence=1>>. Acessado em 05 nov. 2022.
- ROCHA, N. M. S. *Predição das diferenças dos preços de liquidação das diferenças entre as subregiões do SIN utilizando RNA*. 2018.
- ROMÃO, E. L. et al. *Estudo Comparativo entre Modelos Autoregressivos Integrados de Médias Móveis e Redes Neurais Artificiais na Modelagem e Previsão de Séries Econométricas. XL ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO*, 2020.
- RUAS, G. I. S. et al. *Previsão de Demanda de Energia Elétrica Utilizando Redes Neurais Artificiais e Support Vector Regression*. Disponível em <https://www.cos.ufrj.br/~ines/enia07_html/pdf/27927.pdf>. Acessado em 04 Nov. 2022.
- RUMELHART; WEIGEND. *Predicting the Future: A Connectionist Approach*. [S.l.]: Stanford, 1990.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning Representations by Back-propagating Errors. *Nature*, v. 323, n. 6088, p. 533–536, 1986. Disponível em: <<http://www.nature.com/articles/323533a0>>.
- SHUKLA, P. *The Gradient Descent Algorithm*. 2022. Disponível em <<https://towardsai.net/p/tutorials/the-gradient-descent-algorithm>>. Acessado em 05 jan. 2023.