

**UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ  
CAMPUS DE CASCAVEL  
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA**

**DESENVOLVIMENTO E VALIDAÇÃO DE MODELOS DE REDES NEURAIS PARA  
ESTIMATIVA DE EVAPOTRANSPIRAÇÃO DE REFERÊNCIA NO PARANÁ**

**ADRIANA POSTAL**

**CASCAVEL - PR  
2024**

ADRIANA POSTAL

**DESENVOLVIMENTO E VALIDAÇÃO DE MODELOS DE REDES NEURAIS PARA  
ESTIMATIVA DE EVAPOTRANSPIRAÇÃO DE REFERÊNCIA NO PARANÁ**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutor em Engenharia Agrícola, área de concentração Recursos Hídricos e Saneamento Ambiental.

**Orientador** - Prof. Dr. Silvio César Sampaio

**Coorientador** - Dr. Prof. Marcio Antonio Villas Boas

**CASCADEL - PR  
MARÇO - 2024**

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Postal, Adriana

Desenvolvimento e validação de modelos de redes neurais para estimativa de evapotranspiração de referência no Paraná / Adriana Postal; orientador Silvio César Sampaio; coorientador Marcio Antonio Vilas Boas. -- Cascavel, 2024. 102 p.

Tese (Doutorado Campus de Cascavel) -- Universidade Estadual do Oeste do Paraná, Centro de Ciências Exatas e Tecnológicas, Programa de Pós-Graduação em Engenharia Agrícola, 2024.

1. redes neurais na agricultura. 2. manejo da irrigação. 3. dados meteorológicos. I. Sampaio, Silvio César, orient. II. Vilas Boas, Marcio Antonio, coorient. III. Título.



**unioeste**

Universidade Estadual do Oeste do Paraná  
Reitoria  
CNPJ 78.680.337/0001-84  
Rua Universitária, 1619, Jardim Universitário  
Tel.: (45) 3220-3000 - Fax: (45) 3225-4590 - www.unioeste.br  
CEP: 85819-110 - Cx. P.: 701  
Cascavel - PARANÁ



## **ADRIANA POSTAL**

### **DESENVOLVIMENTO E VALIDAÇÃO DE MODELOS DE REDES NEURAIIS PARA ESTIMATIVA DE EVAPOTRANSPIRAÇÃO DE REFERÊNCIA NO PARANÁ**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutora em Engenharia Agrícola, área de concentração Recursos Hídricos e Saneamento Ambiental, linha de pesquisa Recursos Hídricos, APROVADO(A) pela seguinte banca examinadora:



Documento assinado digitalmente  
**SILVIO CESAR SAMPAIO**  
Data: 13/03/2024 10:29:27-0300  
Verifique em <https://validar.iti.gov.br>

**Orientador(a) - Silvio César Sampaio**

**Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)**



Documento assinado digitalmente  
**MARCIO ANTONIO VILAS BOAS**  
Data: 11/03/2024 09:10:12-0300  
Verifique em <https://validar.iti.gov.br>

**Coorientador (a) - Marcio Antonio Vilas Boas**

**Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)**



Documento assinado digitalmente  
**GUSTAVO HENRIQUE PAETZOLD**  
Data: 08/03/2024 11:31:16-0300  
Verifique em <https://validar.iti.gov.br>

**Gustavo Henrique Paetzold**

**Universidade Tecnológica Federal do Paraná (UTFPR)**

*Claudia Brandelero Rizzi*

**Claudia Brandelero Rizzi**

**Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)**



Documento assinado digitalmente  
**CARLOS JOSE MARIA OLGUIN**  
Data: 06/03/2024 15:21:06-0300  
Verifique em <https://validar.iti.gov.br>

**Carlos José Maria Olguin**

**Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)**



Documento assinado digitalmente  
**RALPHO RINALDO DOS REIS**  
Data: 13/03/2024 09:42:54-0300  
Verifique em <https://validar.iti.gov.br>

**Ralpho Rinaldo dos Reis**

**Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)**

**Cascavel, 06 de março de 2024.**

## **BIOGRAFIA**

Adriana Postal é brasileira, nascida em 12 de janeiro de 1978 na cidade de Cascavel, Paraná. cursou o segundo grau no Colégio Estadual João XXIII em Clevelândia-PR. Possui graduação em Informática pela Universidade Estadual do Oeste do Paraná (2002) e mestrado em Engenharia Elétrica pela Universidade Federal de Santa Catarina (2004). É professora assistente da Universidade Estadual do Oeste do Paraná desde 2003, atuando como docente no curso de Ciência da Computação. Tem experiência na área de Ciência da Computação, com ênfase em Inteligência Artificial, atuando principalmente nos seguintes temas: algoritmos genéticos, redes neurais, computação evolutiva e robótica inteligente, e desde 2020 realiza estudos de doutorado no Programa de Pós-Graduação em Engenharia Agrícola, área de concentração: Recursos Hídricos e Saneamento Ambiental, na Universidade Estadual do Oeste do Paraná.

*A Josué e Rafael, os amores da minha vida.*

## **AGRADECIMENTOS**

Agradeço ao meu esposo Josué, pelo carinho, amor e pela força dada em todos os momentos. E por percorrer comigo esse caminho, principalmente nos momentos mais difíceis, ajudando a encontrar soluções para este trabalho, sempre ao meu lado.

Agradeço ao meu filho Rafael, pelo carinho, apoio dados, e por compreender a minha ausência em muitos momentos nesses 4 anos.

À minha família, que mesmo longe, me deu forças e apoio.

Ao meu orientador, Prof. Silvio, e ao meu coorientador, Prof. Marcio, pelo apoio e confiança.

Por fim, agradeço a todos, que de uma forma ou outra, participaram da realização deste trabalho: amigos, colegas, professores.

Postal, Adriana **Desenvolvimento e validação de modelos de redes neurais para estimativa de evapotranspiração de referência no Paraná**. Orientador: Silvio César Sampaio. Coorientador: Marcio Antonio Villas Boas. 2024. 100 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel-Paraná, 2024.

## RESUMO

Esta tese abordou a estimativa da evapotranspiração de referência ( $ET_o$ ) no contexto agrícola, com foco no manejo eficiente da irrigação. O objetivo central foi desenvolver e validar modelos baseados em Redes Neurais Artificiais (RNAs) do tipo MLP (*Multilayer Perceptron*), para a previsão da  $ET_o$  no estado do Paraná, utilizando dados do Instituto Nacional de Meteorologia (INMET) e de um produtor na região oeste. A pesquisa iniciou-se com uma revisão sobre o manejo e o método de Penman-Monteith para estimativa de  $ET_o$  e o uso de RNAs, destacando desafios e abordagens relevantes. A metodologia envolveu a coleta e organização de dados meteorológicos do INMET, fundamentais para o treinamento dos modelos, o desenvolvimento e treinamento de RNAs, utilizando dois otimizadores (SGD e Adam), explorando variáveis meteorológicas como entrada para prever a  $ET_o$ , além da validação dos modelos via dados do produtor. A análise considerou métricas estatísticas, incluindo Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e Coeficiente de Determinação ( $R^2$ ). Os resultados revelaram que as configurações RNA<sub>9</sub> (com 8 variáveis de entrada) e RNA<sub>10</sub> (com 4 variáveis de entrada) apresentaram desempenho promissor na estimativa da  $ET_o$ . A validação com dados de um produtor destacou a relevância prática dos modelos. As métricas MAE, MSE, RMSE e  $R^2$  indicaram eficácia, enquanto o Teste de Kolmogorov-Smirnov contribuiu para avaliar a aderência das previsões ao comportamento estatístico dos dados reais. Os resultados das métricas mostraram que a melhor configuração foi a RNA<sub>9</sub>, com um erro médio absoluto de apenas 0,01 mm/dia com os dois otimizadores e a configuração com menos variáveis foi a RNA<sub>10</sub>, com um erro médio absoluto de apenas 0,03 mm/dia, também com os dois otimizadores. Com  $R^2$  variando entre 0,99 e 1, é possível afirmar que os modelos se adequaram aos dados do estado do Paraná, mesmo quando confrontados com os dados do produtor, que não foram utilizados no treinamento. Sugestões para pesquisas futuras incluem o desenvolvimento de um sistema integrado de coleta automatizada de dados meteorológicos do INMET e a implementação de um aplicativo para manejo da irrigação baseado nos modelos mais eficientes. Ao fim deste trabalho, foi possível desenvolver e validar os modelos de RNAs para previsão simplificada e precisa da  $ET_o$  na agricultura, promovendo uma abordagem acessível para otimizar a gestão da irrigação no estado do Paraná.

**Palavras-chave:** redes neurais na agricultura; manejo da irrigação; dados meteorológicos.

Postal, Adriana **Development and validation of neural network models for the estimation of reference evapotranspiration in Paraná.** Advisor: Silvio César Sampaio. Co-advisor: Marcio Antonio Villas Boas. 2024. 100 f. Thesis (Doctorate in Agricultural Engineering) - Western Paraná State University, Cascavel-Paraná, 2024.

## ABSTRACT

This thesis addressed the estimation of reference evapotranspiration ( $ET_o$ ) in the agricultural context, focusing on efficient irrigation management. The main objective was to develop and validate models based on Multilayer Perceptron Artificial Neural Networks (MLP ANNs) for  $ET_o$  prediction in the state of Paraná, using data from the National Institute of Meteorology (INMET) and a producer in the western region. The research began with a review of irrigation management and the Penman-Monteith method for  $ET_o$  estimation, as well as the use of ANNs, highlighting challenges and relevant approaches. The methodology involved the collection and organization of meteorological data from INMET, essential for model training, and the development and training of ANNs using two optimizers (SGD and Adam), exploring meteorological variables as input to predict  $ET_o$ , along with model validation using producer data. The analysis considered statistical metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination ( $R^2$ ). The results revealed that the configurations RNA<sub>9</sub> (with 8 input variables) and RNA<sub>10</sub> (with 4 input variables) showed promising performance in  $ET_o$  estimation. Validation with producer data highlighted the practical relevance of the models. The MAE, MSE, RMSE, and  $R^2$  metrics indicated efficacy, while the Kolmogorov-Smirnov test contributed to evaluating the predictions' adherence to the statistical behavior of real data. The metric results showed that the best configuration was RNA<sub>9</sub>, with a mean absolute error of only 0.01 mm/day with both optimizers, and the configuration with fewer variables was RNA<sub>10</sub>, with a mean absolute error of only 0.03 mm/day, also with both optimizers. With  $R^2$  ranging between 0.99 and 1, it is possible to affirm that the models suited the data from the state of Paraná, even when confronted with producer data not used in training. Suggestions for future research include developing an integrated system for automated collection of meteorological data from INMET and implementing an irrigation management application based on the most efficient models. At the end of this work, it was possible to develop and validate ANN models for simplified and accurate  $ET_o$  prediction in agriculture, promoting an accessible approach to optimize irrigation management in the state of Paraná.

**Keywords:** neural networks in agriculture; irrigation management; meteorological data.

## SUMÁRIO

<b>LISTA DE TABELAS</b>	<b>x</b>
<b>LISTA DE FIGURAS</b>	<b>xii</b>
<b>LISTA DE ABREVIATURAS E SIGLAS</b>	<b>xiii</b>
<b>LISTA DE SÍMBOLOS</b>	<b>xv</b>
<b>1 INTRODUÇÃO</b>	<b>17</b>
<b>2 OBJETIVOS</b>	<b>19</b>
2.1 Objetivo geral . . . . .	19
2.2 Objetivos específicos . . . . .	19
<b>3 REVISÃO BIBLIOGRÁFICA</b>	<b>20</b>
3.1 Manejo da irrigação . . . . .	20
3.1.1 Indicadores para manejo de água de irrigação . . . . .	22
3.2 Evapotranspiração de referência . . . . .	23
3.2.1 Métodos de estimativa da evapotranspiração de referência . . . . .	25
3.2.2 Previsão da evapotranspiração de referência . . . . .	27
3.3 Redes neurais artificiais . . . . .	27
3.3.1 Arquiteturas de RNAs . . . . .	30
3.3.2 Funções de ativação . . . . .	33
3.3.3 Processos de aprendizado de uma RNA . . . . .	34
3.3.4 Otimizadores . . . . .	36
3.4 Trabalhos de RNAs na Estimativa da $ET_o$ . . . . .	37
3.4.1 Trabalhos de referência para configurações de entrada . . . . .	39
<b>4 MATERIAL E MÉTODOS</b>	<b>43</b>
4.1 Dados meteorológicos utilizados . . . . .	43

4.1.1	Dados das estações do INMET . . . . .	43
4.1.2	Dados do produtor rural . . . . .	45
4.2	Pré-processamento dos dados . . . . .	47
4.2.1	Limpeza dos dados . . . . .	49
4.2.2	Ajuste de granularidade . . . . .	49
4.2.3	Inclusão de novas variáveis . . . . .	49
4.2.4	Normalização dos dados . . . . .	52
4.3	Cálculo da $ET_o$ . . . . .	53
4.4	Variáveis de entrada . . . . .	53
4.4.1	Configurações 1 a 3 . . . . .	56
4.4.2	Configurações 4 a 7 . . . . .	56
4.4.3	Configurações 8 e 9 . . . . .	57
4.4.4	Configurações 10 e 11 . . . . .	57
4.4.5	Configuração 12 . . . . .	57
4.4.6	Configurações 13 e 14 . . . . .	58
4.5	Projeto da arquitetura da RNA . . . . .	58
4.6	Avaliação da arquitetura . . . . .	61
4.6.1	Escolha da função de erro e métrica de avaliação . . . . .	61
4.6.2	Teste Kolmogorov-Smirnov . . . . .	63
4.6.3	Descrição dos hiperparâmetros . . . . .	64
4.7	Experimentos realizados . . . . .	65
4.7.1	Divisão de dados para treinamento e validação . . . . .	65
4.7.2	Execução dos experimentos . . . . .	66
4.7.3	Seleção das melhores configurações . . . . .	66
4.7.4	Teste de generalização . . . . .	66
4.7.5	Seleção final dos modelos . . . . .	66
4.8	Ferramentas computacionais utilizadas . . . . .	67
4.9	Síntese da metodologia . . . . .	67
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>70</b>
5.1	Avaliação das RNAs projetadas . . . . .	70
5.2	Análise das configurações testadas . . . . .	74
5.2.1	Configurações 1 a 3 . . . . .	74
5.2.2	Configurações 4 a 7 . . . . .	76

5.2.3	Configurações 8 e 9 . . . . .	77
5.2.4	Configurações 10 e 11 . . . . .	79
5.2.5	Configuração 12 . . . . .	80
5.2.6	Configurações 13 e 14 . . . . .	80
5.3	Análise das 5 melhores configurações . . . . .	81
5.4	Análise das configurações com os dados do produtor . . . . .	84
5.5	Análise das métricas e do teste Kolmogorov-Smirnov . . . . .	87
5.6	Divergências nas estimativas da $ET_o$ entre o método de Penman-Monteith e os dados observados . . . . .	89
<b>6</b>	<b>CONCLUSÕES</b>	<b>91</b>
<b>7</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>92</b>
7.1	Limitações do trabalho . . . . .	92
7.2	Trabalhos futuros . . . . .	93
	<b>REFERÊNCIAS</b>	<b>95</b>

## LISTA DE TABELAS

Tabela 1	Localização geográfica e número de amostras das estações meteorológicas do INMET no estado do Paraná . . . . .	44
Tabela 2	Variáveis meteorológicas disponíveis nas estações do INMET . . . . .	45
Tabela 3	Variáveis climatológicas coletadas na estação meteorológica particular .	48
Tabela 4	Funções utilizadas para o ajuste de granularidade dos dados das estações do INMET . . . . .	50
Tabela 5	Funções utilizadas para o ajuste de granularidade (horário e diário) dos dados da estação particular . . . . .	51
Tabela 6	Variáveis meteorológicas utilizadas no treinamento dos modelos de RNAs	54
Tabela 7	Configurações de entrada escolhidas para os testes das RNAs . . . . .	55
Tabela 8	Número de neurônios para a camada de entrada e limites para o número de neurônios na camada oculta, para cada configuração testada . . . . .	60
Tabela 9	Bibliotecas do Python utilizadas e suas versões . . . . .	68
Tabela 10	Arquitetura utilizada para o treinamento das configurações de entrada .	71
Tabela 11	Resumo das métricas das configurações para o otimizador SGD . . . . .	72
Tabela 12	Resumo das métricas das configurações para o otimizador Adam . . . . .	73
Tabela 13	Resultados da 1ª execução para as configurações 1 a 3 . . . . .	75
Tabela 14	Comparação das configurações 4 a 7 com o trabalho de Sowmya, Kumar e Ambat (2020) . . . . .	76
Tabela 15	Comparação das configurações 8 e 9 com o trabalho de Ferraz (2014) .	78
Tabela 16	Comparação das configurações 10 e 11 com os trabalhos de Santos (2020) ( $M_1$ e $M_6$ ) e Walls <i>et al.</i> (2020) (Mod. 2) . . . . .	80
Tabela 17	Comparação das configurações 13 e 14 com o trabalho de Arca, Benincasa e Vincenzi (2001) . . . . .	81
Tabela 18	Sumarização das 5 melhores configurações . . . . .	82
Tabela 19	Variáveis utilizadas pelas 5 melhores configurações . . . . .	83

Tabela 20	Sumarização das configurações com os dados do produtor . . . . .	84
Tabela 21	Menores e maiores diferenças entre os valores reais e preditos para a configuração RNA <sub>9</sub> . . . . .	86
Tabela 22	Menores e maiores diferenças entre os valores reais e preditos para a configuração RNA <sub>10</sub> . . . . .	87
Tabela 23	Comparação das 4 métricas utilizadas com o valor do KS e Valor-p . . .	88
Tabela 24	Métricas que mostram as divergências entre o valor da $ET_o$ da estação do produtor e os valores calculados e preditos . . . . .	90

## LISTA DE FIGURAS

Figura 1	Representação simplificada do neurônio biológico (Data Science Academy, 2019) . . . . .	28
Figura 2	Exemplo de neurônio artificial . . . . .	29
Figura 3	Rede <i>feedforward</i> de uma única camada . . . . .	31
Figura 4	Rede <i>feedforward</i> com uma camada oculta e uma camada de saída . .	32
Figura 5	Rede recorrente sem neurônios ocultos . . . . .	33
Figura 6	Exemplos de funções de ativação . . . . .	34
Figura 7	Localização da propriedade rural . . . . .	46
Figura 8	Estação meteorológica particular . . . . .	47
Figura 9	Modelo geral da RNA . . . . .	59
Figura 10	Síntese da metodologia utilizada . . . . .	68

## LISTA DE ABREVIATURAS E SIGLAS

Adam	<i>Adaptive Moment Estimation</i>
Alt	Altitude
ANN	<i>Artificial Neural Networks</i>
ECDF	<i>Empirical Cumulative Distribution Function</i>
ET	Evapotranspiração
$ET_c$	Evapotranspiração da cultura
$ET_{FAO}$	Evapotranspiração de referência padrão FAO
$ET_o$	Evapotranspiração de referência padrão
$ET_p$	Evapotranspiração potencial
FAO	<i>Food and Agriculture Organization of the United Nations</i>
GRU	<i>Gated Recurrent Unit</i>
IA	Inteligência Artificial
INMET	Instituto Nacional de Meteorologia
Lat	Latitude
Lon	Longitude
LSTM	<i>Long Short-Term Memory</i>
KS	Teste de Kolmogorov-Smirnov
MAE	<i>Mean Absolute Error</i>
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Squared Error</i>
PM-FAO	Método Penman-Monteith FAO

$R^2$	Coeficiente de determinação
ReLU	<i>Rectified Linear Unit</i>
RMSE	<i>Root Mean Squared Error</i>
RNA	Rede Neural Artificial
RNN	<i>Recurrent Neural Network</i>
SGD	<i>Stochastic Gradient Descent</i>
Ur	Umidade Relativa

## LISTA DE SÍMBOLOS

$K_c$	Coeficiente de cultura
$T_{max}$	Temperatura máxima
$T_{min}$	Temperatura mínima
$R_s$	Radiação solar incidente
$k_{R_s}$	Coeficiente empírico de proporcionalidade
$R_a$	Radiação solar no topo da atmosfera
$dr$	Distância inversa relativa entre a Terra e o Sol
$\omega_s$	Ângulo horário ao nascer do sol
$\varphi$	Latitude do local
$\delta$	Declinação solar
$z$	Valor padronizado do atributo
$x$	Valor original do atributo
$\mu$	Média dos valores do atributo
$\sigma$	Desvio padrão dos valores do atributo
$\Delta$	Declividade da curva de pressão de vapor
$R_n$	Saldo de radiação na superfície da cultura
$G$	Fluxo de calor no solo
$\gamma$	Constante psicrométrica
$u_2$	Velocidade do vento à 2 m de altura
$e_s$	Pressão de valor de saturação
$e_a$	Pressão de vapor
$T$	Temperatura média diária
$n$	Número de neurônios de entrada
$s$	Número de neurônios de saída
$X$	Variáveis de entrada para o treinamento da RNA
$Y$	Variável alvo ou saída desejada para o treinamento da RNA

$N$	Número de amostras
$x_i$	Média diária da $ET_o$ , calculada pelo método de Penman-Monteith
$y_i$	Média diária da $ET_o$ predita pelo modelo de RNA
$D$	Estatística de teste KS
$F_n(x)$	Função de distribuição acumulativa empírica das observações
$F(x)$	Função de distribuição acumulativa teórica
$Ur_{max}$	Umidade relativa máxima
$Ur_{min}$	Umidade relativa mínima
$P$	Ocorrência de precipitação
$Ro$	Radiação solar global incidente diária
$B_s$	Duração de brilho solar diário
$\Delta e$	Deficit de saturação do ar

## 1 INTRODUÇÃO

A gestão eficiente da irrigação desempenha um papel fundamental na otimização da produção agrícola, especialmente em cenários onde a água é um recurso cada vez mais precioso e escasso. A prática adequada da irrigação não apenas promove o aumento da produtividade das culturas, mas também desempenha um papel importante na conservação dos recursos hídricos, evitando seu uso excessivo e desperdício.

A motivação para este estudo surge da necessidade de aprimorar a eficiência da gestão da irrigação, enfrentando os desafios associados à complexidade e à demanda considerável de variáveis meteorológicas no método tradicional de Penman-Monteith, escolhido pela FAO (*Food and Agriculture Organization of the United Nations*) para estimativas de evapotranspiração de referência ( $ET_o$ ). Este método, embora preciso, muitas vezes encontra limitações práticas, uma vez que exige um conjunto extenso de variáveis meteorológicas que podem não estar facilmente disponíveis para agricultores, especialmente em ambientes agrícolas. A superação dessas limitações é importante para a implementação eficaz de práticas sustentáveis de irrigação.

Diante dessa complexidade, as Redes Neurais Artificiais (RNAs) surgem como uma abordagem interessante para previsões em diversos contextos, incluindo estimativas meteorológicas. A capacidade das RNAs de aprender padrões complexos a partir de dados torna-as especialmente adequadas para modelar relações não lineares inerentes aos processos climáticos.

Este trabalho focou na utilização de dados provenientes das estações meteorológicas do Instituto Nacional de Meteorologia (INMET) no estado do Paraná, fundamentais para o treinamento e validação de modelos de RNAs na estimativa da  $ET_o$ . Incorporando dados da região oeste do Paraná não utilizados no processo de refinamento dos modelos, o estudo desenvolveu modelos capazes de oferecer previsões de fácil compreensão e aplicáveis para os agricultores locais, contribuindo diretamente para a otimização da gestão da irrigação nessa região.

Em resumo, a pesquisa é motivada pela necessidade de aprimorar a eficiência da ges-

tão da irrigação, superando as limitações práticas do método tradicional de Penman-Monteith. A utilização de RNAs representa uma abordagem para alcançar previsões úteis com um conjunto mais acessível de variáveis meteorológicas, contribuindo para práticas sustentáveis de irrigação na agricultura.

Além disso, esta pesquisa ofereceu como contribuição à Engenharia Agrícola abordagens inovadoras baseadas em RNAs para a simplificação do cálculo da  $ET_o$ , utilizando um número reduzido de variáveis climáticas, superando desafios práticos enfrentados por agricultores, especialmente em ambientes onde a disponibilidade de dados meteorológicos é limitada. Ao utilizar a eficiência e adaptabilidade das RNAs, este estudo visou proporcionar ferramentas acessíveis e precisas para otimizar o manejo da irrigação, refletindo uma abordagem sustentável no uso dos recursos hídricos.

## 2 OBJETIVOS

### 2.1 Objetivo geral

Este trabalho desenvolveu e validou modelos de redes neurais artificiais (RNAs) para estimar a evapotranspiração de referência ( $ET_o$ ) no estado do Paraná, utilizando dados meteorológicos provenientes de estações do Instituto Nacional de Meteorologia (INMET) e de um produtor localizado na região oeste do estado.

### 2.2 Objetivos específicos

1. Desenvolver um modelo de previsão de  $ET_o$  baseado em RNAs, simplificando o cálculo da  $ET_o$  por meio de um conjunto reduzido de variáveis climáticas.
2. Organizar dados meteorológicos provenientes das estações do Instituto Nacional de Meteorologia (INMET), essenciais para o treinamento dos modelos de RNAs.
3. Desenvolver e treinar modelos de RNAs para a estimativa de  $ET_o$ , utilizando variáveis meteorológicas relevantes como entrada.
4. Avaliar as variáveis de entrada para a RNA, buscando identificar a configuração ótima que equilibre precisão e simplicidade na estimativa da  $ET_o$ , proporcionando uma abordagem acessível para agricultores no uso do modelo.
5. Avaliar o desempenho dos modelos por meio de métricas estatísticas como o Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e Coeficiente de Determinação ( $R^2$ ).
6. Validar os modelos de RNAs com dados de  $ET_o$  de um produtor na região oeste do Paraná, comparando as estimativas com os valores observados.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Manejo da irrigação

O manejo da irrigação consiste na determinação do momento, da quantidade e de como aplicar água na plantação, garantindo o uso eficiente da água, evitando desperdícios e diminuindo os gastos com energia elétrica sem afetar a produção (SENAR, 2019). A irrigação de uma cultura deve levar em consideração vários aspectos do sistema produtivo, como o controle fitossanitário, as condições meteorológicas, econômicas e as estratégias de condução da cultura (Gutierrez; Neves, 2021). Segundo Silveira e Stone (1994, p. 7-8), o manejo tem os seguintes objetivos:

**Maximização da produtividade da cultura:** O manejo da irrigação visa aprimorar a produtividade da cultura, assegurando a oferta adequada de água em todas as fases de seu desenvolvimento. Esse fornecimento ideal de água contribui para aumentar o rendimento da cultura.

**Minimização do uso de água e custos de energia:** Um dos objetivos-chave é evitar o uso excessivo de água e energia na irrigação. Isso não apenas economiza recursos hídricos, mas também reduz os custos relacionados à energia e ao bombeamento de água.

**Aumento da eficiência do uso de fertilizantes:** O manejo adequado da irrigação contribui para melhorar a eficiência dos fertilizantes aplicados. A presença adequada de água no solo viabiliza uma melhor absorção de nutrientes, como nitrogênio e fósforo, pelas plantas.

**Redução da incidência de doenças:** O excesso de água no solo durante o ciclo da cultura pode aumentar a incidência de doenças, especialmente as causadas por fungos do solo. Portanto, o manejo da irrigação visa minimizar o risco dessas doenças.

**Manutenção das condições do solo:** O manejo adequado da irrigação desempenha um papel importante na prevenção de problemas no solo, como a lixiviação de nutrientes, a erosão e a desagregação. Isso contribui para manter as condições químicas e físicas do solo, assegurando sua fertilidade e produtividade a longo prazo.

Segundo Sousa *et al.* (2011), o acesso do irrigante à tecnologia é frequentemente limitado por diversos fatores. Primeiramente, a falta de informação e o custo dos equipamentos e instrumentos destinados ao manejo da irrigação representam barreiras significativas. Além disso, em muitas regiões, o baixo nível de instrução dos irrigantes dificulta a implementação de estratégias mínimas de manejo.

Outros desafios também se apresentam para os irrigantes, incluindo:

**Custo da água:** historicamente, a água é frequentemente barata ou sem custo (Silveira; Stone, 1994), mas essa realidade está mudando gradualmente. Em 1997, o governo federal instituiu a Política Nacional de Recursos Hídricos pela lei nº 9.433/97<sup>1</sup>. Em 2016, em rios de domínio da União, a cobrança foi implementada na Bacia do Rio Paraíba do Sul, nas Bacias dos Rios Piracicaba, Capivari e Jundiaí, na Bacia do Rio São Francisco e na Bacia do Rio Doce<sup>2</sup>.

**Custo da energia elétrica:** O custo da energia elétrica para irrigação representa uma parcela significativa dos custos variáveis do sistema de irrigação, podendo chegar a até 70%, como indicado no trabalho de Silva (2020).

**Carência de dados edafoclimáticos:** A falta de análises físico-hídricas do solo e de dados climatológicos adequados dificulta o correto manejo da irrigação (Silveira; Stone, 1994).

**Prioridade das atividades:** O produtor enfrenta uma série de preocupações em seu dia-a-dia, como o calendário de aplicação de adubos, o controle de pragas e doenças e o desenvolvimento da cultura, o que pode diminuir a atenção dada ao manejo da irrigação (Silveira; Stone, 1994).

**Consultoria técnica e/ou especializada:** A assistência técnica disponível muitas vezes não oferece opções de manejo adequadas, seja por falta de conhecimento ou interesse (Sousa *et al.*, 2011).

**Metodologia:** Mesmo quando disponíveis, as informações e metodologias de manejo podem não ser apresentadas de forma acessível ou compreensível para os irrigantes, limitando sua aplicação prática (Sousa *et al.*, 2011).

Dessa forma, compreender os desafios enfrentados pelos irrigantes é importante para promover melhorias efetivas no manejo da irrigação. A falta de acesso à tecnologia, os custos

<sup>1</sup> [http://www.planalto.gov.br/ccivil\\_03/leis/19433.htm](http://www.planalto.gov.br/ccivil_03/leis/19433.htm)

<sup>2</sup> <https://agrosmart.com.br/blog/cobranca-pelo-uso-da-agua-agricultura/>

operacionais crescentes, a carência de dados e consultoria técnica especializada, juntamente com as múltiplas demandas enfrentadas pelos produtores, mostram a necessidade urgente de soluções acessíveis e eficazes. Ao enfrentar esses desafios de forma proativa, podemos avançar na busca por práticas sustentáveis de manejo da irrigação, beneficiando não apenas os produtores, mas também a produtividade agrícola e a preservação dos recursos naturais (Sousa *et al.*, 2011).

### 3.1.1 Indicadores para manejo de água de irrigação

A irrigação geralmente é realizada quando a deficiência de água na planta começa a afetar seu desenvolvimento e produtividade. Na prática, o manejo da irrigação é adaptado a cada situação específica, levando em conta indicadores relacionados à planta, ao solo e/ou à atmosfera (Sousa *et al.*, 2011).

A quantidade de água aplicada por irrigação pode ser determinada com base na avaliação da disponibilidade de água no solo (por meio de sensores de umidade, por exemplo) ou na quantidade de água evapotranspirada pela cultura entre duas irrigações consecutivas (Gutierrez; Neves, 2021).

Segundo Sousa *et al.* (2011), existem vários indicadores disponíveis para determinar o momento de irrigar e a quantidade de água a ser aplicada, agrupados como segue: indicadores de manejo com base na planta; indicadores de manejo com base no solo e indicadores de manejo com base na atmosfera (evapotranspiração).

Os indicadores baseados na planta ajudam a determinar o momento ideal para irrigar, enquanto os baseados na atmosfera auxiliam na estimativa da quantidade de água necessária para a irrigação. Por outro lado, os indicadores do solo geralmente são utilizados para decidir tanto o momento quanto a quantidade de água a ser aplicada. Na prática, é frequente o uso conjunto de múltiplos indicadores do sistema solo-planta-atmosfera para orientar o manejo da irrigação.

A escolha dos indicadores baseados na atmosfera, especificamente a evapotranspiração de referência ( $ET_o$ ), foi fundamentada em sua ampla aceitação como uma medida robusta e confiável da demanda evaporativa da atmosfera sobre uma superfície vegetada padrão. A  $ET_o$ , que será descrita na Seção 3.2, é calculada a partir de variáveis meteorológicas, como temperatura do ar, umidade relativa, velocidade do vento e radiação solar, capturando de forma abrangente as condições atmosféricas que influenciam o processo de evapotranspiração das plantas. Essa abordagem, juntamente com outros indicadores, permite uma determinação

mais precisa do momento e da quantidade de água a ser aplicada durante a irrigação. Além disso, a  $ET_o$  é amplamente empregada em estudos e práticas de irrigação em todo o mundo, facilitando a comparação e a aplicação dos resultados deste trabalho em diferentes contextos e regiões.

### 3.2 Evapotranspiração de referência

A combinação de dois processos separados em que a água é perdida, por um lado, da superfície do solo por evaporação e, por outro lado, da cultura por transpiração, é chamada de evapotranspiração (ET). Segundo Allen *et al.* (1998), esse processo é descrito em três conceitos:

**Evaporação:** é o processo pelo qual a água líquida é convertida em vapor de água (vaporização) e removida da superfície de evaporação (remoção do vapor). A água evapora de uma variedade de superfícies, como lagos, rios, calçadas, solos e vegetação úmida.

**Transpiração:** consiste na vaporização da água líquida contida nos tecidos vegetais e na remoção do vapor para a atmosfera. A água, junto com alguns nutrientes, é captada pelas raízes e transportada pela planta. Quase toda a água captada é perdida pela transpiração e apenas uma pequena fração é usada dentro da planta.

**Evapotranspiração:** é um processo combinado que engloba tanto a evaporação quanto a transpiração, ocorrendo simultaneamente. A taxa de evaporação do solo em uma área cultivada é influenciada principalmente pela quantidade de radiação solar incidente na superfície do solo. Enquanto a cultura cresce e forma um dossel mais denso, a quantidade de radiação solar que atinge o solo diminui, tornando a transpiração o processo dominante para a perda de água. Inicialmente, a evaporação do solo é mais proeminente, mas à medida que a cultura se desenvolve e cobre completamente o solo, a transpiração assume um papel principal nesse processo.

A evapotranspiração de referência ( $ET_o$ ) é um indicador que expressa a demanda hídrica de uma região e é influenciada principalmente pelas condições climáticas locais. Este parâmetro serve como uma referência para determinar a quantidade de água necessária para diferentes culturas, sendo quantificado em milímetros de água (mm) (SENAR, 2019).

A  $ET_o$  foi definida por Allen *et al.* (1998) como sendo a taxa de evapotranspiração que ocorre em uma superfície de referência cujas características são: extensa superfície com ve-

getação em crescimento ativo, cobrindo totalmente o solo, com altura da vegetação de 0,12 m, índice de área foliar de 2,88, albedo igual a 0,23 e com uma resistência superficial de  $70 \text{ s m}^{-1}$ . Logo, a  $ET_o$  é um elemento indicativo da demanda hídrica das culturas de um determinado local e período. A cultura de referência utilizada é uma cultura hipotética, cujas características se assemelham, bem de perto, à evapotranspiração da grama. Isto permite que tais características (valores numéricos) mantenham-se como parâmetros constantes adotados no cálculo da  $ET_o$  (Carvalho *et al.*, 2011).

Os métodos para determinar a  $ET_o$  podem ser de medição direta ou indireta. Enquadram-se em métodos diretos os lisímetros e os métodos de balanço de água no solo. Como indiretos, os modelos micrometeorológicos, teóricos e empíricos, que são baseados na utilização de dados climáticos (Lucas, 2019).

O método mais preciso para medir a  $ET_o$  é o lisímetro, um equipamento utilizado para medir a evapotranspiração de plantas ou a taxa de perda de água do solo. O lisímetro consiste em uma estrutura composta por um recipiente de solo ou uma parcela de solo delimitada, que é isolada do ambiente externo para permitir a medição precisa da quantidade de água que entra e sai do sistema solo-planta-atmosfera (Krishna, 2019). Além de ser um equipamento caro, a realização de medições é demorada e de difícil execução em condições de campo, justificando seu uso apenas para desenvolver e calibrar métodos de estimativas.

Existem vários modelos disponíveis para estimar a  $ET_o$  por métodos indiretos, desenvolvidos e adotados em diferentes condições agro-climáticas. Dentre eles, o método Penman-Monteith foi selecionado pela FAO (*Food and Agriculture Organization of the United Nations*)<sup>3</sup> por sua capacidade de proporcionar resultados similares aos que se obtém pelas medições diretas em diversos tipos de clima (Lucas, 2019). Assim, a evapotranspiração de referência padrão FAO ( $ET_{FAO}$ ), estimada pela equação de Penman-Monteith parametrizada de acordo com a superfície de referência (Equação 1), tornou-se referência em substituição às medidas lisimétricas (Allen *et al.*, 1998).

$$ET_{FAO} = \frac{0.408\Delta(R_n - G) + \gamma \left( \frac{900}{T+273} \right) u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

Em que:

$ET_{FAO}$  – evapotranspiração de referência padrão,  $\text{mm dia}^{-1}$

$\Delta$  – declividade da curva de pressão de vapor,  $\text{kPa } ^\circ\text{C}^{-1}$

<sup>3</sup> Organização das Nações Unidas para a Alimentação e a Agricultura.

$R_n$  – saldo de radiação na superfície da cultura, MJ m<sup>-2</sup> dia<sup>-1</sup>

$G$  – fluxo de calor no solo, MJ m<sup>-2</sup> dia<sup>-1</sup>

$\gamma$  – constante psicrométrica, kPa °C<sup>-1</sup>

$u_2$  – velocidade do vento à 2 m de altura, ms<sup>-1</sup>

$e_s$  – pressão de valor de saturação, kPa

$e_a$  – pressão de vapor, kPa

$T$  – temperatura média diária, °C

A Equação 1 utiliza parâmetros climáticos que podem ser medidos por estações meteorológicas automáticas: temperatura do ar, umidade do ar, radiação solar (luz solar) e velocidade do vento. Os demais parâmetros podem ser estimados por equações específicas. Um roteiro para o cálculo da  $ET_{FAO}$  e as equações para estimar os valores não disponíveis na estação meteorológica pode ser encontrado em Conceição (2006).

A  $ET_o$  é um parâmetro climático essencial na gestão da irrigação, representando a demanda hídrica potencial da atmosfera em uma determinada região. Ao calcular a  $ET_o$  com precisão, os agricultores podem obter estimativas confiáveis das necessidades de água de suas culturas, permitindo uma irrigação mais eficiente e sustentável. Embora a  $ET_o$  seja influenciada principalmente pelas condições climáticas, é importante lembrar que as práticas de irrigação devem ser adaptadas às características específicas da cultura para garantir um uso ótimo da água e maximizar a produtividade agrícola. Assim, a compreensão da  $ET_o$  e sua correta aplicação são elementos-chave para promover a sustentabilidade e o sucesso na agricultura irrigada (Lucas, 2019).

### 3.2.1 Métodos de estimativa da evapotranspiração de referência

O termo evapotranspiração (ET) foi introduzido por Thornthwaite e Wilm em 1944, mas foi no final da década de 1940 que dois trabalhos importantes surgiram: o de Thornthwaite (1948), que realizou uma pesquisa analítica, empírica e fundamentada na análise dos fatos, e o de Penman (1948), que fez uma abordagem sintética, científica e firmada no conhecimento físico que rege o fenômeno.

Como Sedyama (1996) descreve em seu trabalho sobre o histórico e a evolução da ET, Thornthwaite tinha a preocupação de explicar as variações sazonais do balanço de água

no solo e de tentar definir as diferenças regionais do clima, por isso é um método que utiliza apenas a temperatura média do ar e a duração do dia, em média mensal. Já Penman preocupava-se com os processos físicos envolvidos na evaporação e com o desenvolvimento de um método que, a partir de elementos meteorológicos, expressasse a estimativa da taxa de evaporação da água em contato livre com a atmosfera, da umidade da superfície do solo e da vegetação.

Na década de 1960, Monteith propôs um novo método para calcular diretamente a evapotranspiração da cultura ( $ET_c$ ), baseado no método de Penman, que ficou conhecido como método de Penman-Monteith, enquanto Doorenbos e Pruitt, em 1977, introduziram o conceito de evapotranspiração de referência ( $ET_o$ ), em substituição ao termo evapotranspiração potencial ( $ET_p$ ), usado desde a década de 1940 (Carvalho *et al.*, 2011).

Em 1990, a FAO promoveu um encontro de especialistas para, entre outros objetivos, analisar os conceitos e procedimentos de metodologias de cálculos para a evapotranspiração. Neste encontro, a  $ET_o$  tornou-se o conceito utilizado para o cálculo da evapotranspiração e o método recomendado para sua estimativa foi o desenvolvido por Monteith (1965) que, após parametrização, passou a denominar-se Penman-Monteith FAO (PM-FAO), o qual foi bastante aceito internacionalmente (Carvalho *et al.*, 2011).

Durante muitos anos, uma variedade de métodos foi desenvolvida para estimar a  $ET_o$ . Essa diversidade de métodos surge em resposta a diferentes necessidades, como adaptabilidade às condições climáticas locais, facilidade de uso e a disponibilidade limitada de dados meteorológicos necessários para alimentar esses métodos (Carvalho *et al.*, 2011). Muitos desses métodos possuem variantes específicas, ajustadas e calibradas para atender às peculiaridades de determinadas regiões, resultando em uma ampla gama de opções disponíveis. Carvalho *et al.* (2011) oferecem uma síntese dos métodos mais reconhecidos para estimativa da  $ET_o$ , destacando os principais elementos utilizados por cada um e discutindo o desempenho de alguns desses métodos.

Além dos métodos matemáticos tradicionais, há uma crescente investigação sobre a aplicação de outras técnicas, como as redes neurais artificiais (RNAs), para estimar a  $ET_o$ . As RNAs têm se destacado como ferramentas importantes em diversas áreas do conhecimento, especialmente na modelagem de processos complexos, incluindo aqueles pertinentes ao clima e à agricultura. No contexto deste trabalho, optou-se por utilizar RNAs (Seção 3.3) devido à sua capacidade de lidar com relações não lineares e complexas presentes nos dados meteorológicos, bem como pela sua flexibilidade em lidar com conjuntos de dados incompletos ou

ruidosos.

### 3.2.2 Previsão da evapotranspiração de referência

Segundo Lucas (2019), a estimativa precisa da  $ET_o$  é essencial para o planejamento de irrigação e gerenciamento dos recursos hídricos. Prever o valor dessa variável permite conhecer antecipadamente as necessidades de água das culturas, auxiliando agricultores na tomada de decisão de atividades como gerenciamento dos sistemas de abastecimento de água em áreas onde a produção agrícola é o principal usuário da água (Mariño; Tracy; Taghavi, 1993), redução do custo da energia elétrica, visto que é possível planejar a irrigação em horários com tarifas mais baratas (Perea *et al.*, 2015), identificação de possível estresse futuro das culturas (Zhao *et al.*, 2019) e operação de canais e reservatórios, sendo possível a previsão da quantidade de água que deve ser entregue (Bachour *et al.*, 2016).

### 3.3 Redes neurais artificiais

Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano. Elas consistem em unidades interconectadas chamadas neurônios, organizadas em camadas. Cada conexão entre neurônios tem um peso associado, que é ajustado durante o treinamento da rede (Fausett, 1994).

As RNAs consistem em um método de solucionar problemas de inteligência artificial (IA), construindo um sistema que tenha elementos (os neurônios) que simulam o cérebro humano, inclusive seu comportamento, ou seja, aprendendo, errando e fazendo descobertas (Braga; Carvalho; Ludemir, 2012). Elas foram desenvolvidas como generalização de modelos matemáticos da cognição humana ou da neurobiologia, baseadas nas seguintes suposições (Haykin, 2009): a) o processamento da informação ocorre em elementos simples, chamados neurônios; b) sinais são propagados entre neurônios sobre elos de conexão chamados conexões sinápticas; c) cada conexão tem um peso associado, o qual, em uma rede neural típica, multiplica o sinal transmitido; d) cada neurônio aplica uma função de ativação (geralmente não-linear) à soma ponderada de suas entradas para produzir o seu sinal de saída.

O neurônio biológico possui três tipos de componentes que são de particular interesse no entendimento de um neurônio artificial: os dendritos, soma e axônio. Os vários dendritos recebem sinais de outros neurônios. Os sinais são impulsos elétricos que são transmitidos através de lacunas (*gap*) sinápticas por processos químicos. O soma, ou corpo da célula, faz a adição dos sinais de entrada. Quando entradas suficientes são recebidas, as células

disparam, isto é, transmitem os sinais para os dendritos das outras células através do axônio. O axônio serve para conectar a célula nervosa a outras do sistema nervoso (Fausett, 1994). O modelo de um neurônio biológico é mostrado na Figura 1.

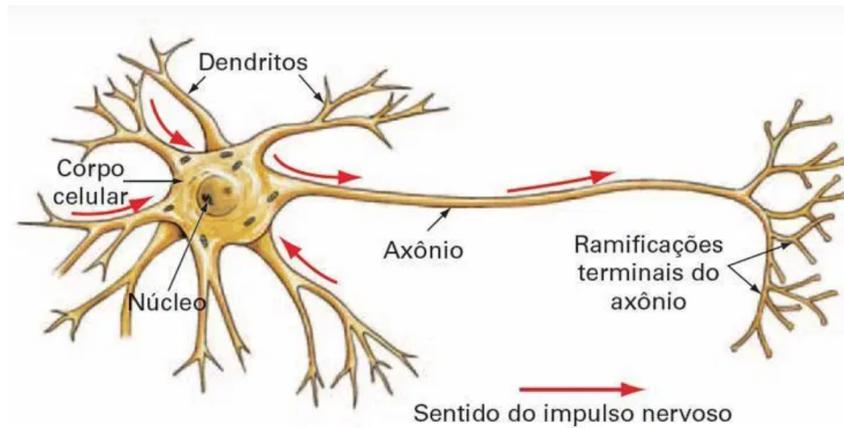


Figura 1 Representação simplificada do neurônio biológico (Data Science Academy, 2019)

Em 1943, McCulloch e Pitts (1943) criaram o primeiro neurônio artificial, mais simples do que poderia ser diante da informação que naquela época já estava disponível sobre o comportamento elétrico da célula nervosa. O primeiro neurônio era binário, produzindo saída pulso ou não pulso, e suas várias entradas poderiam ser do tipo excitatório, aumentando a atividade do neurônio, ou inibitório, diminuindo essa atividade. Para determinar a saída do neurônio, calculava-se a soma ponderada das entradas com os respectivos ganhos como fatores de ponderação. Se este resultado fosse maior ou igual a um certo limiar então a saída do neurônio era pulso, caso contrário, não pulso (McCulloch; Pitts, 1943).

O diagrama em blocos da Figura 2 mostra o modelo do neurônio artificial. É possível identificar os três elementos básicos do modelo neuronal (Haykin, 2009): um conjunto de sinapses ou conexões de entrada, sendo cada entrada ponderada por um peso sináptico; uma junção de soma, responsável pela combinação aditiva das entradas ponderadas pelos pesos sinápticos; e uma função de ativação (Seção 3.3.2), geralmente não-linear, responsável pela ativação da saída ou resposta do elemento processador. Também é possível observar o *bias*, que é aplicado externamente (representado por  $b_k$ ), e tem intuito de aumentar o grau de liberdade desta função e, conseqüentemente, a capacidade de aproximação da rede (Haykin, 2009).

Os pesos sinápticos desempenham um papel fundamental nas RNAs. Esses pesos são atribuídos às conexões entre neurônios em uma RNA e representam a força ou importância

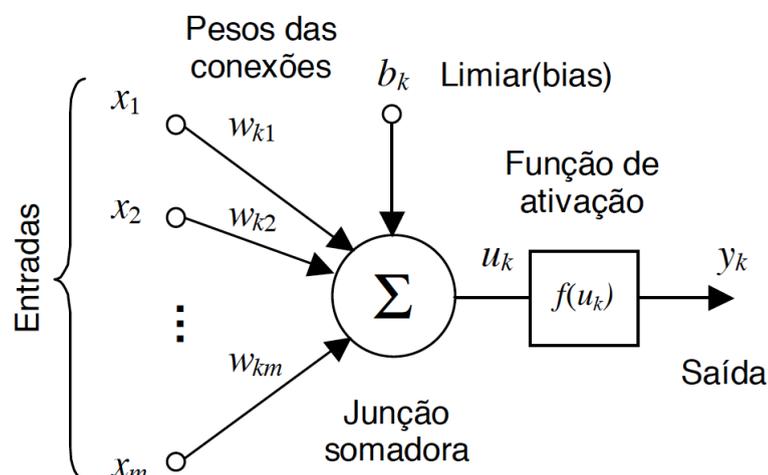


Figura 2 Exemplo de neurônio artificial

Fonte: Adaptado de Haykin (2009)

relativa de cada conexão. Em termos simples, os pesos sinápticos determinam o impacto que a saída de um neurônio tem sobre os neurônios subsequentes na rede. Durante o treinamento de uma RNA, os pesos sinápticos são ajustados iterativamente para otimizar o desempenho da rede na tarefa desejada. O processo de aprendizado em uma RNA é, em grande parte, o processo de ajustar esses pesos para que a RNA seja capaz de realizar tarefas complexas, como classificação, previsão ou reconhecimento de padrões, com alta precisão (Goodfellow; Bengio; Courville, 2016).

Os pesos sinápticos em uma rede neural podem ser inicializados de várias maneiras, e a escolha da estratégia de inicialização dos pesos desempenha um papel importante no treinamento bem-sucedido da rede (Haykin, 2009). Inicialmente, os pesos podem ser definidos de forma aleatória, com valores pequenos distribuídos uniformemente ou seguindo uma distribuição normal. No entanto, métodos de inicialização específicos foram desenvolvidos para melhorar o processo de treinamento, como o *he\_normal*, que inicializa os pesos com valores amostrados de uma distribuição normal com média zero e um desvio padrão calculado com base no número de entradas para a camada. Isso ajuda a evitar problemas de saturação nas camadas iniciais da rede e acelera o processo de convergência durante o treinamento (Goodfellow; Bengio; Courville, 2016).

Neurônios individuais possuem capacidade computacional limitada, mas quando em conjunto, em forma de uma rede (neural), são capazes de resolver problemas de complexidade elevada (Braga; Carvalho; Ludemir, 2012). Os neurônios são organizados em 3 camadas:

**Camada de Entrada:** onde os padrões são apresentados à rede.

**Camadas Intermediárias ou Ocultas:** onde é feita a maior parte do processamento, através das conexões ponderadas. Estas camadas podem ser consideradas como extratoras de características.

**Camada de Saída:** onde o resultado final é concluído e apresentado.

### 3.3.1 Arquiteturas de RNAs

Uma RNA é um sistema de neurônios ligados por conexões sinápticas e divididas em neurônios de entrada, que recebem estímulo do meio externo, neurônios internos ou ocultos (*hidden*) e neurônios de saída, que se comunicam com o exterior. A presença dos neurônios internos é muito importante para o funcionamento adequado da RNA, uma vez que sua ausência torna a resolução de problemas linearmente não separáveis praticamente impossível. Em termos simples, uma RNA é composta por múltiplas unidades de processamento conectadas por canais de comunicação, cada um associado a um peso específico (Goodfellow; Bengio; Courville, 2016).

A escolha da arquitetura de uma RNA está relacionada com o tipo de problema a ser resolvido e é definida por 4 parâmetros principais: número de camadas da rede, número de neurônios em cada camada, tipo de conexão entre neurônios e a topologia da rede (Fiorin *et al.*, 2011).

A maneira pela qual os neurônios de uma RNA estão estruturados está diretamente ligada ao algoritmo de treinamento da rede (Seção 3.3.3). Existem três principais arquiteturas: redes *Feedforward* com uma única camada, redes *Feedforward* de múltiplas camadas e redes Recorrentes.

#### **Rede *Feedforward* com uma única camada**

É o caso mais simples: uma camada de entrada e uma camada de saída. Geralmente os neurônios de entrada são lineares: eles propagam o sinal de entrada para a próxima camada. A rede chama-se *feedforward* porque a propagação do sinal ocorre apenas da entrada para a saída (sentido positivo). Esse tipo de arquitetura é capaz de resolver problemas multi-variáveis de múltiplas funções acopladas, mas com algumas restrições de complexidade, por serem de uma única camada (Braga; Carvalho; Ludemir, 2012). A Figura 3 mostra esta arquitetura.

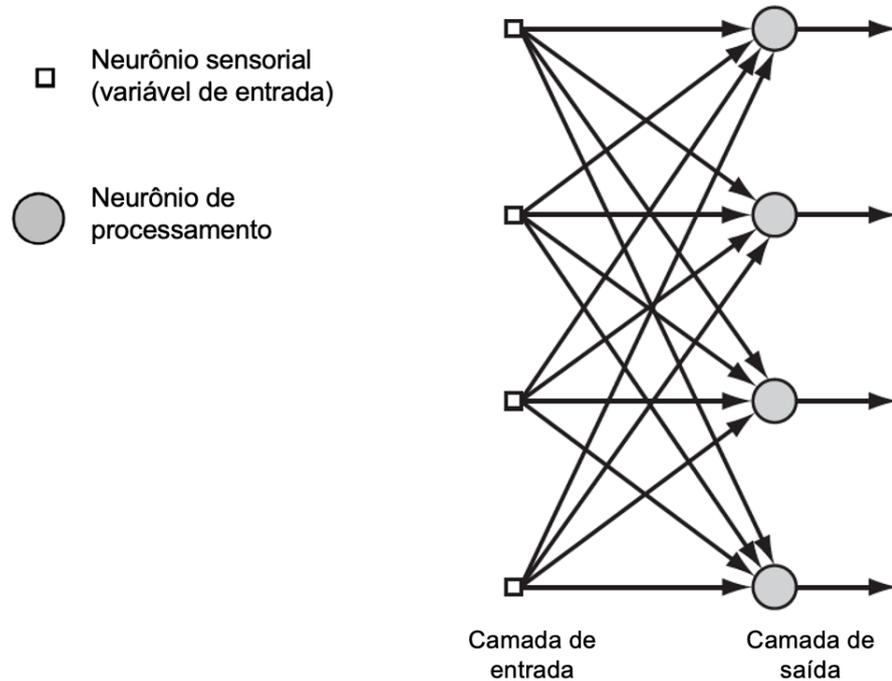


Figura 3 Rede *feedforward* de uma única camada

Fonte: Adaptado de Haykin (2009)

### Rede *Feedforward* de múltiplas camadas

Também chamadas de MLPs (*Multilayer Perceptron*), possuem uma ou mais camadas ocultas (intermediárias/escondidas), com a saída de cada camada oculta sendo utilizada como entrada para a próxima camada. A camada intermediária confere à RNA uma maior capacidade computacional e universalidade na aproximação de funções contínuas (Braga; Carvalho; Ludemir, 2012). A Figura 4 mostra esta arquitetura.

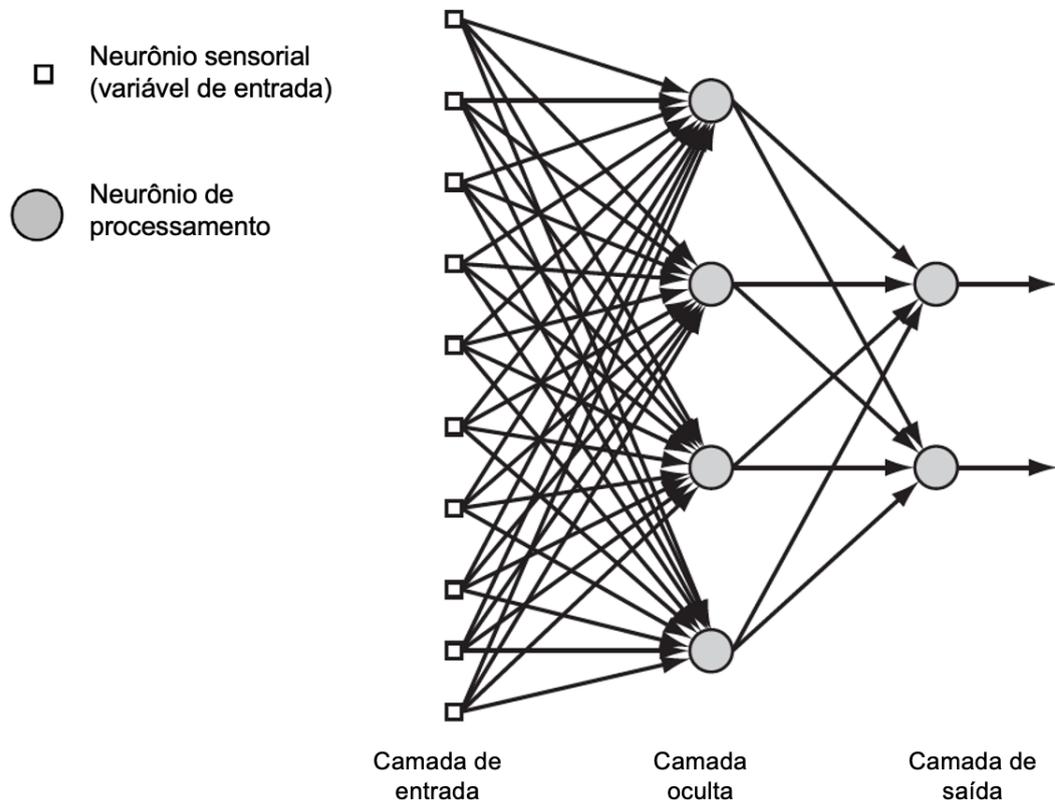


Figura 4 Rede *feedforward* com uma camada oculta e uma camada de saída

Fonte: Adaptado de Haykin (2009)

### Redes recorrentes

As redes neurais recorrentes (RRNs - *Recurrent Neural Network*) são um tipo de RNA com um ou mais laços de realimentação, e consistem em um conjunto de neurônios, um conjunto correspondente de atrasos unitários, além de um número de laços de realimentação igual ao número de neurônios (Haykin, 2009). Assim, a saída de cada neurônio é a entrada para os demais neurônios da rede, como é ilustrado na Figura 5. As RRNs são conhecidas por sua capacidade de processar sequências de dados, mantendo estados internos. Entre as versões mais comumente usadas das RRNs estão a *Long Short-Term Memory* (LSTM) e a *Gated Recurrent Unit* (GRU), que são especialmente eficazes em lidar com problemas de dependência de longo prazo e com o problema de desaparecimento do gradiente, comuns em RRNs tradicionais (Goodfellow; Bengio; Courville, 2016).

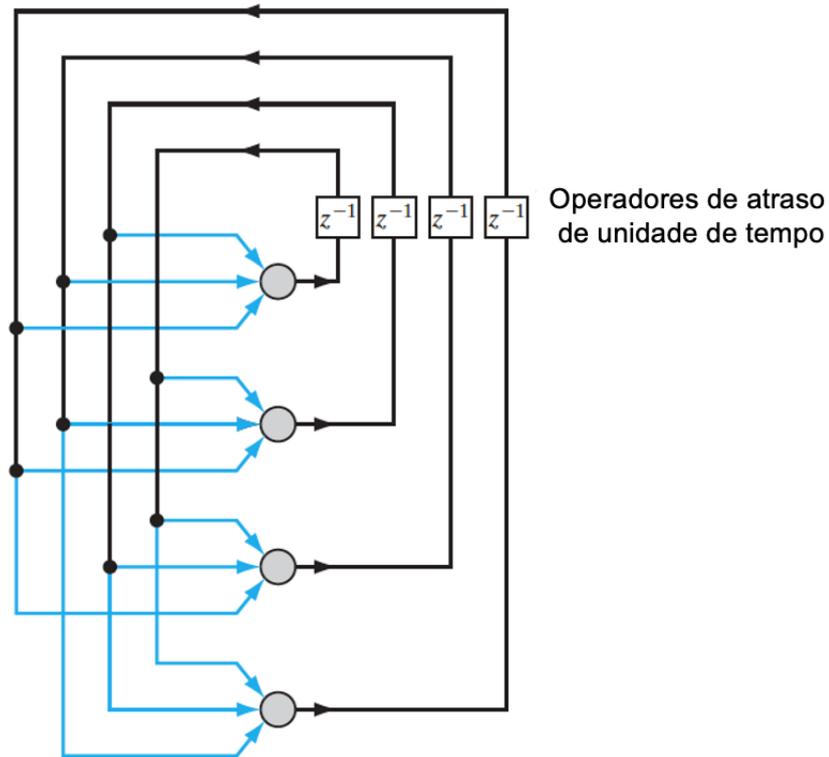


Figura 5 Rede recorrente sem neurônios ocultos

Fonte: Adaptado de Haykin (2009)

### 3.3.2 Funções de ativação

A função de ativação é responsável por gerar a saída  $y_k$  a partir dos vetores de peso ( $w$ ) e de entrada ( $x$ ), sendo responsável pelo sinal a ser transmitido por um determinado neurônio aos demais a ele conectados (Braga; Carvalho; Ludemir, 2012). As funções mais utilizadas são a linear (Figura 6a), sigmoide (Figura 6b), tangente hiperbólica (Figura 6c) e linear retificada (ReLU) (Figura 6d). A escolha das funções de ativação de uma rede neural é uma consideração importante, uma vez que define como devem ser seus dados de entrada (Reis, 2018).

Essas funções, inseridas em cada neurônio da rede, determinam como os sinais de entrada são transformados e propagados por meio da rede. Elas introduzem não linearidades essenciais, permitindo que as redes capturem relações não lineares nos dados, tornando-as capazes de modelar uma ampla variedade de fenômenos (Goodfellow; Bengio; Courville, 2016).

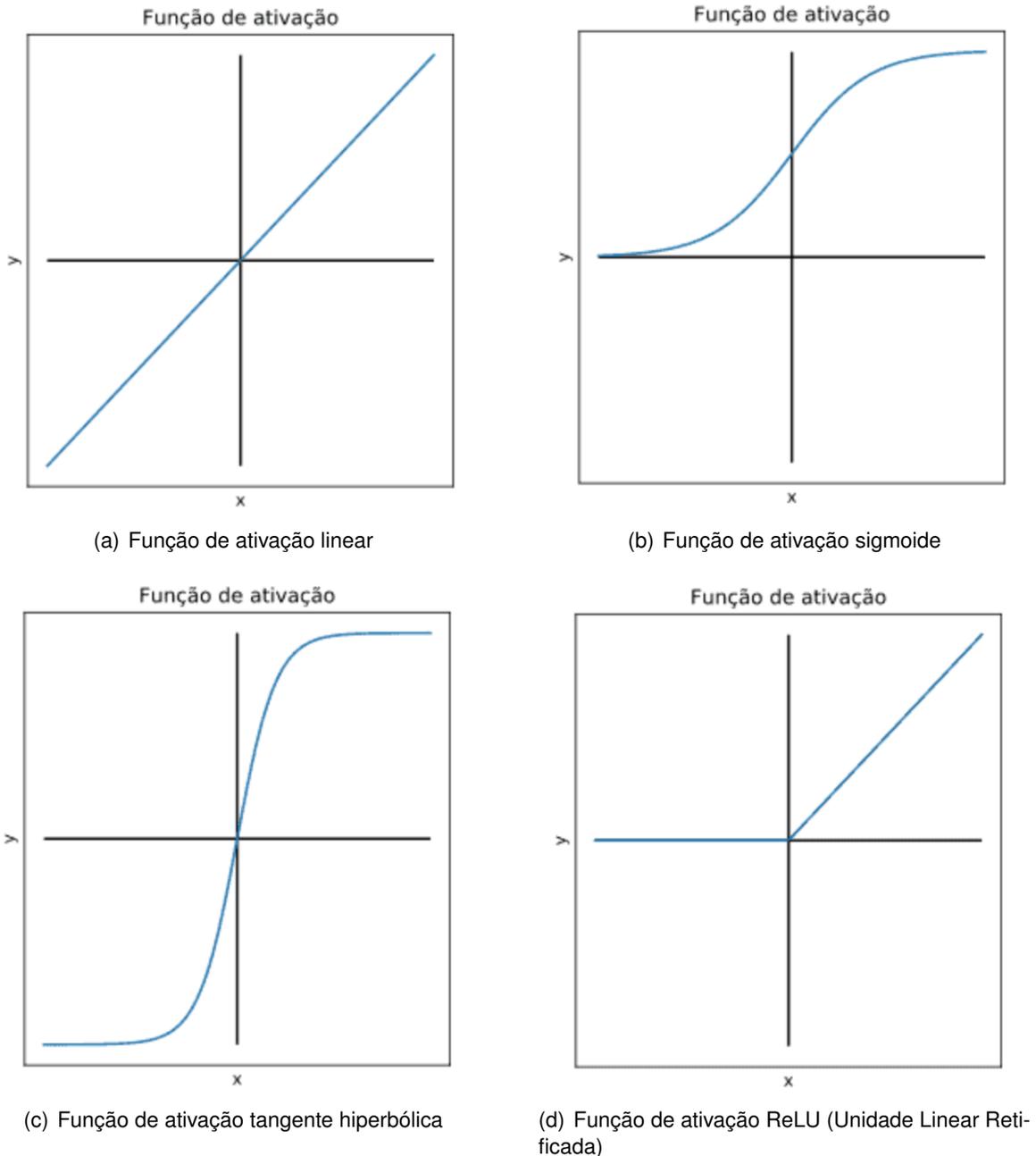


Figura 6 Exemplos de funções de ativação

Fonte: Ceccon (2020)

### 3.3.3 Processos de aprendizado de uma RNA

A propriedade mais importante das RNAs é a habilidade de aprender em seu ambiente e com isso melhorar seu desempenho. Para isso, a maioria dos modelos de RNAs possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem por meio de exemplos. A

RNA passa por um processo de treinamento a partir dos casos reais conhecidos adquirindo, a partir deste treinamento, a sistemática necessária para executar adequadamente o processo desejado dos dados fornecidos. No entanto, é importante ressaltar que a compatibilidade do método de aprendizado depende tanto da arquitetura da RNA quanto da natureza da tarefa a ser executada. Depois do treinamento, a RNA deve ser capaz de extrair regras básicas a partir de dados reais, diferindo da computação tradicional onde é necessário um conjunto de regras rígidas pré-fixadas e algoritmos (Haykin, 2009), (Braga; Carvalho; Ludemir, 2012).

A RNA se baseia nos dados para extrair um modelo geral; por isso, a fase de treinamento deve ser rigorosa, a fim de se evitar modelos espúrios. Todo o conhecimento de uma RNA está armazenado nas sinapses, ou seja, nos pesos atribuídos às conexões entre os neurônios. De 50 a 90% do total de dados deve ser separado para o treinamento da RNA, dados estes escolhidos aleatoriamente, a fim de que a rede “aprenda” as regras e não “decore” exemplos. O restante dos dados só é apresentado à RNA na fase de testes a fim de que ela possa “deduzir” corretamente o inter-relacionamento entre os dados (Braga; Carvalho; Ludemir, 2012). Os principais paradigmas de aprendizado são o supervisionado, o não-supervisionado e o aprendizado por reforço.

O aprendizado supervisionado desempenha um papel importante no contexto das RNAs. Nessa abordagem, uma RNA é treinada para aprender a mapear entradas para saídas com base em um conjunto de dados de treinamento rotulado. O resultado desejado para cada entrada é apresentado à RNA, sendo que muitas iterações devem ser realizadas até que a rede aprenda (Fausett, 1994). Durante o treinamento, a RNA otimiza seus parâmetros internos, como os pesos sinápticos, de forma a minimizar a diferença entre suas previsões e os rótulos reais (Haykin, 2009). Isso é geralmente alcançado por meio de algoritmos de otimização que buscam encontrar os melhores parâmetros da rede (Seção 3.3.4). O aprendizado supervisionado é amplamente utilizado em redes neurais para tarefas de classificação, regressão e diversas outras aplicações, tornando-se um pilar central da capacidade das redes neurais de realizar previsões precisas e aprender relações complexas nos dados.

Enquanto no aprendizado supervisionado os modelos são treinados com a orientação de rótulos de saída conhecidos para realizar tarefas específicas, o aprendizado não-supervisionado lida com dados não rotulados (Fausett, 1994). O objetivo principal do aprendizado não-supervisionado é descobrir estruturas, padrões ou relações intrínsecas nos dados por conta própria: durante o processo de aprendizado os padrões de entrada são apresentados continuamente à rede, e a existência de regularidades nesses dados faz com que o

aprendizado seja possível (Braga; Carvalho; Ludemir, 2012).

Embora o aprendizado supervisionado seja amplamente utilizado em muitas aplicações, o aprendizado não-supervisionado desempenha um papel fundamental em áreas como mineração de dados, processamento de linguagem natural e análise de dados exploratória, onde a estrutura subjacente dos dados muitas vezes não é conhecida a priori (Haykin, 2009).

No aprendizado por reforço, as RNAs atuam como agentes que aprendem a tomar decisões sequenciais em ambientes dinâmicos. Ao contrário do aprendizado supervisionado, onde os modelos são treinados com rótulos de saída conhecidos, e do aprendizado não-supervisionado, que busca estruturas intrínsecas em dados não rotulados, este aprendizado envolve um processo de tentativa e erro, onde as RNAs aprendem a tomar ações sequenciais para maximizar uma recompensa acumulada ao longo do tempo, explorando diferentes estratégias para encontrar as ações que levam às melhores recompensas (Goodfellow; Bengio; Courville, 2016).

Essa abordagem é aplicada em diversos domínios, desde jogos de videogames até robótica e veículos autônomos, onde as RNAs precisam aprender a tomar decisões ótimas em ambientes dinâmicos e incertos. O aprendizado por reforço com RNAs destaca a importância da exploração, aproveitamento e tomada de decisões sequenciais, e continua a ser uma área ativa de pesquisa na inteligência artificial (Haykin, 2009).

#### 3.3.4 Otimizadores

Os otimizadores desempenham um papel importante no treinamento de RNAs. Em essência, eles são algoritmos ou métodos que ajustam iterativamente os pesos sinápticos da rede durante o processo de aprendizado. O principal objetivo dos otimizadores é minimizar a função de erro (perda), que quantifica o quão bem a RNA está se ajustando aos dados de treinamento. Isso é alcançado por meio da atualização gradual dos pesos sinápticos com base na derivada da função de erro em relação aos pesos. Em termos simples, os otimizadores orientam a RNA na busca pelo mínimo global da função de erro, resultando em um modelo capaz de realizar previsões precisas (Goodfellow; Bengio; Courville, 2016).

Existe uma ampla gama de métodos otimizadores disponíveis na literatura, cada um com suas próprias características e aplicações. Um resumo dos otimizadores mais utilizados pode ser visto nos trabalhos de Ribeiro e Junior (2020) e de Santos, Moreira e Junior (2021). Para este estudo em particular, serão abordados dois otimizadores amplamente utilizados: o Adam (*Adaptive Moment Estimation*) e o SGD (*Stochastic Gradient Descent*).

Criado por Kingma e Ba (2015), o otimizador Adam (*Adaptive Moment Estimation*) é uma das escolhas populares quando se trata de treinamento de RNAs. Ele é amplamente utilizado devido à sua capacidade de se adaptar dinamicamente às taxas de aprendizado ao longo do processo de treinamento, combinando as taxas de aprendizado de forma adaptativa para cada parâmetro com a técnica de *momentum*, que ajuda a acelerar a convergência (Goodfellow; Bengio; Courville, 2016). O algoritmo mantém duas médias móveis exponenciais, uma para os gradientes das variáveis e outra para os quadrados dos gradientes. Essas médias móveis são usadas para calcular as correções adaptativas das taxas de aprendizado, resultando em uma convergência mais rápida e estável durante o treinamento (Kingma; Ba, 2015).

Proposto por Sutskever *et al.* (2013), o Gradiente Descendente Estocástico (SGD - *Stochastic Gradient Descent*) é um dos métodos mais fundamentais e amplamente utilizados para otimização em treinamento de RNAs. O SGD funciona atualizando os pesos da rede com base no gradiente da função de erro calculada em um subconjunto aleatório dos dados de treinamento, em vez de usar o conjunto de dados completo. Essa abordagem estocástica permite um treinamento eficiente em grandes conjuntos de dados (Goodfellow; Bengio; Courville, 2016).

O SGD apresenta alguns desafios, como a escolha apropriada da taxa de aprendizado: taxas muito altas podem levar à divergência, enquanto taxas muito baixas podem resultar em convergência lenta. Além disso, o SGD pode sofrer de oscilações em direção à convergência, o que pode ser aliviado com técnicas como *momentum* e taxa de aprendizado adaptativa (Sutskever *et al.*, 2013).

Apesar de suas limitações, o SGD continua sendo uma ferramenta valiosa para a otimização de RNAs, especialmente quando combinado com técnicas de regularização e inicialização de pesos adequadas. É frequentemente usado como uma referência para comparar o desempenho de otimizadores mais avançados, como o Adam, em uma variedade de tarefas de aprendizado (Ribeiro; Junior, 2020).

### 3.4 Trabalhos de RNAs na Estimativa da $ET_o$

A evapotranspiração é um fenômeno complexo e não linear porque depende de vários fatores climatológicos que interagem, como temperatura, umidade, velocidade do vento, radiação, tipo e estágio de crescimento da cultura, etc (Kumar *et al.*, 2002). Por possuir estas características, vários estudos têm utilizado redes neurais artificiais para prever com mais exatidão os valores da evapotranspiração em função de elementos climáticos. Alguns exem-

plos são os trabalhos de Arca, Beniscasa e Vincenzi (2001), Odhiambo *et al.* (2001), Kumar *et al.* (2002), Silva (2002), Sudheer, Gosain e Ramasastrri (2003), Trajkovic, Todorovic e Stankovic (2003) e Zanetti *et al.* (2007), os quais encontraram resultados satisfatórios, melhores do que os obtidos com o método convencional de Penman-Monteith (Kumar *et al.*, 2002). Uma observação importante feita por Zanetti *et al.* (2008) é que todos esses trabalhos, com exceção do trabalho de Arca, Beniscasa e Vincenzi (2001), não se preocuparam em simplificar as variáveis de entrada para as RNAs, o que pode tornar a utilização da técnica dependente de todos os dados do modelo que, como já mencionado, nem sempre estão disponíveis (Conceição, 2006).

No final da década de 2000, novas pesquisas surgiram na tentativa de simplificar os dados necessários para a estimação da evapotranspiração. Entre esses trabalhos, podemos citar Zanetti *et al.* (2008), cujo objetivo foi realizar essa estimativa para o Estado do Rio de Janeiro, utilizando latitude, longitude, altitude, temperatura média do ar, amplitude térmica diária e o dia sequencial do ano, de 17 estações meteorológicas. Os autores afirmam que a evapotranspiração pôde ser estimada com segurança nessas estações.

O trabalho de Irigoyen (2010) estimou a evapotranspiração de referência e a evapotranspiração da limeira ácida a partir de dados de Piracicaba-SP, com RNAs e técnicas de regressão, utilizando os dados de radiação solar global, saldo de radiação ou radiação no topo da atmosfera em combinação com a temperatura do ar, déficit de pressão de vapor no ar e velocidade do vento, obtendo bom desempenho com as RNAs, mesmo na falta de alguns desses dados, mostrando melhor resultado do que os métodos de regressão.

No trabalho de Ferraz (2014) o objetivo foi estimar a  $ET_o$  com base em conjuntos de dados climáticos com variáveis limitadas (latitude, longitude, altitude, dia do ano, temperatura máxima diária, temperatura mínima diária, umidade relativa do ar média diária, velocidade do vento, ocorrência de precipitação, radiação solar global incidente diária, duração de brilho solar diário e deficit de saturação do ar) e coordenadas geográficas, para o Estado do Rio Grande do Sul. Os resultados obtidos foram considerados ótimos, porém o próprio autor considera que ainda seria possível resultados semelhantes utilizando menor quantidade de variáveis de entrada.

O trabalho de Lucas (2019) apresenta uma previsão de séries temporais de  $ET_o$ , utilizando modelos de aprendizagem profunda, como as redes neurais convolucionais. O trabalho utilizou as variáveis de temperatura (máxima e mínima), velocidade do vento, umidade relativa média e insolação, além de arquiteturas diferentes de RNAs. Ao final, o resultado foi satisfatório.

Jimenez *et al.* (2021) realizou um estudo onde utilizou redes neurais de Memória de Longo Prazo (LSTM) para auxiliar nas prescrições de irrigação, com resultados excelentes para previsão com uma hora de antecedência, usando como entrada os valores de sensores de potencial matricial do solo em três profundidades, zona de manejo, valores de chuva e de irrigação. Além de resultados satisfatórios para seis horas de antecedência utilizando apenas três entradas, os valores de sensores de potencial matricial do solo em três profundidades. Esse estudo mostra que é possível diminuir as variáveis necessárias para uma boa predição da irrigação.

Enquanto alguns trabalhos citados utilizaram uma ampla gama de variáveis climáticas e métodos específicos de estimativa da  $ET_o$ , o presente estudo adotou uma abordagem diferente ao simplificar as variáveis utilizadas para a estimativa da  $ET_o$ , utilizando RNAs, fornecendo uma alternativa eficiente para otimizar o manejo da irrigação.

#### 3.4.1 Trabalhos de referência para configurações de entrada

Nesta seção, descrevem-se os trabalhos relacionados que desempenharam um papel fundamental na fundamentação das configurações de entrada adotadas para as RNAs na presente investigação. Como as estações meteorológicas possuem uma variação muito grande em relação às variáveis coletadas, algumas variáveis utilizadas nos trabalhos de referência não estão disponíveis nos conjuntos de dados utilizados, por isso, modificações foram realizadas nas configurações escolhidas, que serão descritas com detalhes, incluindo as modificações realizadas, na Seção 4.4. Os resultados encontrados por estes trabalhos serão comparados aos experimentos realizados e discutidos no Capítulo 5.

O trabalho de Sowmya, Kumar e Ambat (2020), que serviu de base para as configurações de entrada 4 a 7 (Seção 4.4.2), propõe um método de previsão de  $ET_o$ , empregando dados meteorológicos mínimos. Eles utilizam dois conjuntos de dados do Sistema de informações de gerenciamento de irrigação da Califórnia, totalizando 21.544 amostras: o primeiro conjunto são dados da cidade de Santa Mônica e o segundo, da estação de Oakville. Para o treinamento, os dados coletados são do período de 1989 a 2014, e para o teste do modelo, o período ficou entre 2014 e 2019. Eles selecionaram 4 configurações para as variáveis de entrada:

**DnnV1:** com apenas uma variável de entrada: radiação solar ( $R_s$ )

**DnnV2:** utiliza radiação solar ( $R_s$ ) e temperatura máxima ( $T_{max}$ )

**DnnV3:** utiliza radiação solar ( $R_s$ ), temperatura máxima ( $T_{max}$ ) e mínima ( $T_{min}$ )

**DnnV6:** utiliza seis variáveis: radiação solar ( $R_s$ ), temperaturas máxima ( $T_{max}$ ) e mínima ( $T_{min}$ ), velocidade do vento ( $U_2$ ) e umidade relativa máxima ( $UR_{max}$ ) e mínima ( $UR_{min}$ )

No trabalho de Ferraz (2014), foram testadas seis configurações de entrada, utilizando a combinação das variáveis latitude (Lat), longitude (Lon), altitude (Alt), dia do ano (DA), temperatura máxima diária ( $T_{max}$ ), temperatura mínima diária ( $T_{min}$ ), umidade relativa do ar média diária ( $Ur$ ), velocidade do vento ( $U_2$ ), ocorrência de precipitação (P), radiação solar global incidente diária (Ro), duração de brilho solar diário ( $B_s$ ) e deficit de saturação do ar ( $\Delta e$ ), conforme lista:

**ANN1:** Lat, Lon, Alt, DA,  $T_{min}$ ,  $T_{max}$ ,  $Ur$

**ANN2:** Lat, Lon, Alt, DA,  $T_{min}$ ,  $T_{max}$ ,  $Ur$ ,  $U_2$

**ANN3:** Lat, Lon, Alt, DA,  $T_{min}$ ,  $T_{max}$ ,  $Ur$ ,  $U_2$ , Ro

**ANN4:** Lat, Lon, Alt, DA,  $T_{min}$ ,  $T_{max}$ ,  $Ur$ ,  $U_2$ , Ro,  $B_s$

**ANN5:** Lat, Lon, Alt, DA,  $T_{min}$ ,  $T_{max}$ ,  $Ur$ ,  $U_2$ , Ro,  $B_s$ ,  $\Delta e$

**ANN6:** Lat, Lon, Alt, DA,  $T_{min}$ ,  $T_{max}$ ,  $Ur$ ,  $U_2$ , P, Ro,  $B_s$ ,  $\Delta e$

Ferraz (2014) utilizou os dados das estações meteorológicas do estado do Rio Grande do Sul, no período de 2008 a 2012, disponíveis na página do INMET. O autor também afirma que o uso de RNAs para a previsão de  $ET_o$  se mostra como uma boa alternativa para situações onde não há disponibilidade de dados meteorológicos. O trabalho de Ferraz (2014) foi utilizado como base para as configurações de entrada 8 e 9 (Seção 4.4.3).

As configurações 10 e 11 (Seção 4.4.4) foram baseadas em dois trabalhos: o de Santos (2020) e o de Walls *et al.* (2020). Santos (2020) realizou a estimativa de  $ET_o$  para os municípios de Mendes e Teresópolis, ambos no estado do Rio de Janeiro, utilizando dados do CPTEC/INPE<sup>4</sup>, no período de 2010 e 2015, utilizando os seguintes dados: radiação solar acumulada (RadSol), temperatura máxima (Tmax), temperatura média (Tmed), temperatura mínima (Tmin), umidade relativa do ar (UmiRel) e velocidade do vento medida a 10m de altura (VelVento). Com esses dados, o autor criou seis configurações de entrada, descritas como:

$M_1$ : RadSol + Tmax + Tmed + Tmin + UmiRel + VelVento

<sup>4</sup> Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais.

$M_2$ : RadSol + Tmax + Tmed + Tmin + UmiRel

$M_3$ : RadSol + Tmax + Tmin + UmiRel

$M_4$ : RadSol + Tmax + Tmed + Tmin

$M_5$ : Tméd + UmiRel + VelVento

$M_6$ : RadSol + Tmed + UmiRel + VelVento

Já o trabalho de Walls *et al.* (2020) previu a  $ET_o$  para a cidade de Nissouri Creek, no condado de Oxford, Canadá, com dados de abril a julho de 2018. Os modelos testados foram uma combinação das variáveis meteorológicas: radiação líquida (Rn), temperatura do ar (T), fluxo de calor do solo (G), velocidade do vento (u2) e umidade relativa (UR). Foram sete modelos, descritos abaixo:

**Modelo 1:** Rn, T, G, u2, UR

**Modelo 2:** Rn, T, G, u2

**Modelo 3:** Rn, T, G

**Modelo 4:** Rn, T

**Modelo 5:** Rn, G

**Modelo 6:** Rn, u2

**Modelo 7:** Rn, UR

A configuração 12 (Seção 4.4.5) foi baseada no trabalho de Silva (2002), que calculou a  $ET_o$  utilizando dados da estação meteorológica do INMET localizada em Viçosa-MG. Foram realizados diversos experimentos com seis variáveis de entrada: temperatura média diária (Tmed), temperatura mínima diária (Tmin), temperatura máxima diária (Tmax), umidade relativa diária (Ur), velocidade do vento diária medida a 10m de altura (Vv) e número de horas de brilho solar diária (Inso), coletados no período de 1961 a 2001.

Por fim, o trabalho de Arca, Beniscasa e Vincenzi (2001) dá sustentação às configurações 13 e 14 (Seção 4.4.6). O estudo foi realizado com base em dados meteorológicos coletados de 1992 a 1995 por uma estação meteorológica CNR-IATA<sup>5</sup>, localizada em Florença,

<sup>5</sup> Instituto de Agrometeorologia e Análise Ambiental Aplicada à Agricultura.

Itália, e utilizou as variáveis: temperatura do ar (T), umidade relativa do ar (Ur), radiação solar global (Rsg), radiação difusa (Rd), radiação líquida (Rl), velocidade (VV) e direção do vento (DV), intensidade da chuva (IC) e pressão atmosférica (Pressao), além de usar os valores de  $ET_o$  calculados pelo método de Penman-Monteith como entrada em seus modelos.

Este capítulo ofereceu uma revisão bibliográfica, além dos trabalhos correlatos, fornecendo a base teórica para as escolhas desta pesquisa, estabelecendo a fundamentação teórica necessária para o desenvolvimento prático deste estudo. No próximo capítulo, a estrutura experimental, a coleta de dados e a arquitetura das RNAs serão detalhadas, transitando da teoria para a implementação prática. Essa transição é fundamental para compreender como os princípios teóricos se manifestam na prática, contribuindo significativamente para o avanço do conhecimento nesta área.

## 4 MATERIAL E MÉTODOS

Neste capítulo é apresentada uma visão detalhada do aparato experimental, ferramentas computacionais e abordagens metodológicas empregadas na pesquisa. Também são descritos os recursos, dados e procedimentos que foram utilizados para coletar informações e realizar análises. Além disso, ele delinea as etapas do processo experimental, desde a aquisição de dados até a aplicação das técnicas analíticas. A compreensão completa dessa etapa é essencial para que os leitores possam reproduzir os experimentos e avaliar a validade dos resultados obtidos.

### 4.1 Dados meteorológicos utilizados

Para esse trabalho, foram utilizados dois conjuntos de dados: o primeiro, resultado da coleta das estações do INMET, ligado ao Ministério da Agricultura e Pecuária, e o segundo, resultado da estação meteorológica em uma pequena propriedade. Os dados e sua utilização neste trabalho serão descritos nas próximas subseções.

#### 4.1.1 Dados das estações do INMET

O Instituto Nacional de Meteorologia (INMET), ligado ao Ministério da Agricultura e Pecuária fornece, em sua página, dados meteorológicos de diversas estações espalhadas por todo o território nacional. No estado do Paraná, estão disponíveis dados de 24 estações automáticas distribuídas em todas as regiões. No entanto, quatro dessas estações foram descartadas devido ao número insuficiente de amostras. As 20 estações restantes, incluindo suas coordenadas geográficas e o número total de amostras (após o pré-processamento detalhado na Seção 4.2), estão descritas na Tabela 1.

Os dados aqui coletados foram utilizados no projeto, treinamento, avaliação e escolha das arquiteturas de Redes Neurais, processo que será detalhado na Seção 4.5. As variáveis meteorológicas disponíveis nas estações automáticas podem ser vistas na Tabela 2, sendo que os nomes das variáveis são os mesmos disponibilizados nas planilhas do Instituto. Os

dados utilizados foram coletados de hora em hora, no período de 01/07/2018 a 30/06/2022, totalizando 22.387 amostras (após o pré-processamento).

Tabela 1 Localização geográfica e número de amostras das estações meteorológicas do INMET no estado do Paraná

<b>Nome</b>	<b>Latitude</b>	<b>Longitude</b>	<b>Altitude</b>	<b>Total dados coletados</b>
Campina da Lagoa	-24,57	-52,80	598	825
Cidade Gaúcha	-23,36	-52,93	365,79	1311
Clevelândia	-26,42	-52,35	966,06	1361
Colombo	-25,32	-49,16	950	1461
Curitiba	-25,45	-49,23	922,91	1049
Dois Vizinhos	-25,70	-53,10	545,95	825
Foz do Iguaçu	-25,60	-54,48	235,09	1032
General Carneiro	-26,40	-51,35	1009,01	1331
Ivaí	-25,01	-50,85	803,58	1461
Japira	-23,77	-50,18	692,88	1045
Joaquim Távora	-23,51	-49,95	512,67	1442
Laranjeiras do Sul	-25,37	-52,39	835	838
Marechal Cândido Rondon	-24,53	-54,02	392,07	1426
Maringá	-23,41	-51,93	548,51	1405
Morretes	-25,51	-48,81	50,31	1149
Nova Fátima	-23,42	-50,58	664,31	851
Paranapoema	-22,66	-52,13	308,7	1165
Planalto	-25,72	-53,75	399,1	812
São Mateus do Sul	-25,84	-50,37	780,21	1065
Ventania	-24,28	-50,21	1093,41	533

Fonte: INMET (<https://tempo.inmet.gov.br>)

Tabela 2 Variáveis meteorológicas disponíveis nas estações do INMET

<b>Variável</b>	<b>Unidade de Medida</b>
Data	dd/mm/aa
Hora	UTC
Temperatura instantânea	°C
Temperatura máxima	°C
Temperatura mínima	°C
Umidade relativa instantânea	%
Umidade relativa máxima	%
Umidade relativa mínima	%
Ponto de orvalho instantâneo	°C
Ponto de orvalho máximo	°C
Ponto de orvalho mínimo	°C
Pressão atmosférica instantânea	hPa
Pressão atmosférica máxima	hPa
Pressão atmosférica mínima	hPa
Velocidade do vento	m/s
Direção do vento	m/s
Rajada de vento	m/s
Radiação	KJ/m <sup>2</sup>
Chuva	mm/dia

Fonte: A autora

#### 4.1.2 Dados do produtor rural

Os dados coletados foram obtidos em uma pequena propriedade localizada a 20 km ao sul da sede do município de Cascavel-PR (Figura 7), com as coordenadas geográficas Latitude 25°08'06.6" Sul, Longitude 53°27'20.5" Oeste, e altitude de 524 metros em relação ao nível do mar, em um solo classificado como Latossolo Vermelho Distroférico típico. A classificação climática do município de Cascavel, segundo Köppen, é Cfa, sendo a temperatura média no mês mais frio inferior a 18 °C (mesotérmico) e a temperatura média no mês mais quente acima de 22 °C (Wrege *et al.*, 2012).



Figura 7 Localização da propriedade rural

Fonte: A autora

Os dados aqui coletados foram utilizados para a validação dos modelos desenvolvidos para estimar a  $ET_o$ . A validação consiste em verificar se os modelos, treinados e avaliados com dados do INMET, são capazes de fazer previsões úteis quando aplicados a um conjunto de dados independente, não utilizado durante o treinamento. Essa abordagem ajuda a garantir a robustez e a capacidade de generalização dos modelos para diferentes conjuntos de dados, fornecendo uma avaliação mais abrangente de seu desempenho.

A estação meteorológica particular disponível na propriedade é da marca Davis, modelo Vantage Pro2 GroWeather Wireless Sensor Suite, mostrada na Figura 8, com sistema *datalogger* para armazenamentos dos dados, que coleta as variáveis climatológicas mostradas na Tabela 3. A estação está instalada a dois metros do solo. Os dados foram coletados entre 05/10/2018 e 31/12/2021, totalizando 1.144 amostras após o pré-processamento da Seção 4.2. O tempo de coleta variou durante esse período, conforme mostrado abaixo:

- De 5 em 5 minutos:
  - 05/10/2018 a 14/02/2019
  - 20/06/2020 a 24/07/2020
- De 15 em 15 minutos:
  - 23/02/2019 a 20/06/2020
  - 24/07/2020 a 31/12/2021

- De hora em hora: 16/02/2019 a 21/02/2019



Figura 8 Estação meteorológica particular

Fonte: Proprietário rural

Apesar da estação coletar o valor da evapotranspiração ( $ET_o$ ), não foi possível verificar qual o método que a mesma utiliza para realizar esse cálculo. No manual da estação é possível encontrar as medidas para a instalação dos sensores (DAVIS INSTRUMENTS, 2017) e quais as variáveis que ela utiliza para calcular a  $ET_o$  (DAVIS INSTRUMENTS, 2023). Por esse motivo, decidiu-se estimar a  $ET_o$  pelo método de Penman-Monteith, como descrito na Seção 4.3. Desta forma, ambos conjuntos de dados tiveram a  $ET_o$  estimada pelo mesmo método, o que torna o treinamento e testes com as Redes Neurais comparáveis.

## 4.2 Pré-processamento dos dados

Os dados coletados, em ambos conjuntos, precisaram passar por um pré-processamento para torná-los adequados para uso posterior. Essas etapas podem incluir a limpeza dos dados (preencher valores faltantes, reduzir ruídos nos dados, etc.), integração,

Tabela 3 Variáveis climatológicas coletadas na estação meteorológica particular

<b>Variável</b>	<b>Unidade de Medida</b>
Date & Time	dd/mm/aa hh:mm
Pressão atmosférica	hPa
Temperatura	°C
Temperatura máxima	°C
Temperatura mínima	°C
Umidade	%
Temperatura de ponto de orvalho	°C
Temperatura de bulbo úmido	°C
Velocidade do vento	km/h
Direção do vento	–
Distância percorrida pelo vento	km
Rajada de vento	km/h
Direção da rajada de vento	–
Sensação térmica	°C
Índice de calor	°C
Índice THW	°C
Precipitação	mm
Índice pluviométrico	mm/h
Irradiação solar	W/m <sup>2</sup>
Energia Solar	Ly
Irradiação solar máxima	W/m <sup>2</sup>
Evapotranspiração	mm
Graus Dias de Aquecimento	–
Graus Dias de Resfriamento	–

Fonte: A autora

transformação e/ou redução de dados, além da normalização (Castanedo, 2015). As etapas realizadas neste trabalho para os dados utilizados são descritas nas próximas subseções.

#### 4.2.1 Limpeza dos dados

Em ambos os conjuntos, foram retiradas as variáveis que não foram coletadas pela estação (muitas amostras sem valores naquela variável), que possuíam erro de leitura (coleta não realizada em determinados horários, por exemplo), os dias que não possuíam coleta para as 24 h e variáveis que não são utilizadas na maioria dos métodos de estimativa de  $ET_o$  (um resumo das variáveis mais utilizadas nos diversos métodos de estimativa pode ser conferido no trabalho de Carvalho *et al.* (2011)).

As variáveis não utilizadas em cada conjunto estão listadas a seguir. Para as estações do INMET, as variáveis excluídas foram: hora, ponto de orvalho máximo e mínimo, pressão atmosférica máxima e mínima, direção e rajada de vento, e chuva. Em relação à estação particular, as variáveis descartadas incluíram: direção do vento, distância percorrida pelo vento, direção da rajada de vento, graus dias de aquecimento, graus dias de resfriamento, e temperatura de bulbo úmido.

#### 4.2.2 Ajuste de granularidade

Ambos conjuntos de dados não possuem amostragem diária, conforme descrito nas Seções 4.1.1 e 4.1.2. Para a estimativa da  $ET_o$  pelo método de Penman-Monteith, os dados meteorológicos devem possuir amostragem contínua, por isso, os valores foram agrupados por dia (Ferraz, 2014), conforme mostrado na Tabela 4, para os dados das estações do INMET e na Tabela 5, para os dados da estação particular. No caso da estação particular foram necessários dois ajustes: para os valores horários e depois para os valores diários (Orssatto, 2021).

#### 4.2.3 Inclusão de novas variáveis

Além das variáveis disponíveis nas estações do INMET e da estação particular, foi necessário incluir outras variáveis que são importantes para o aprendizado da Rede Neural. Essas variáveis são descritas abaixo:

**Coordenadas geográficas das estações:** para cada estação do INMET e para a estação particular foram incluídas a latitude, a longitude e a altura.

Tabela 4 Funções utilizadas para o ajuste de granularidade dos dados das estações do INMET

Variável	Ajuste de granularidade
Pressão atmosférica	Média do dia
Temperatura máxima ( $T_{max}$ )	Valor máximo do dia
Temperatura mínima ( $T_{min}$ )	Valor mínimo do dia
Temperatura média	$(T_{max} + T_{min})/2$
Umidade Relativa Média	Média do dia
Umidade Relativa Máxima	Valor máximo do dia
Umidade Relativa Mínima	Valor mínimo do dia
Temperatura de ponto de orvalho	Média do dia
Velocidade do vento	Média do dia
Radiação solar	Média do dia

Fonte: A autora

**Dia Juliano:** conforme Conceição (2006), o Dia Juliano<sup>1</sup> é uma das variáveis necessárias para estimar o valor da radiação, usada para calcular a  $ET_o$  pelo método de Penman-Monteith.

**Estimativa da  $ET_o$ :** conforme o método de Penman-Monteith, descrito na Seção 3.2 pela Equação 1.

**Radiação estimada:** para a estimativa da  $ET_o$ , é necessário considerar diversos tipos de radiação, como a radiação solar incidente, a radiação refletida pela superfície da cultura (albedo), a radiação de onda longa emitida pela atmosfera e a radiação de onda curta refletida pela atmosfera (Leivas *et al.*, 2007). No entanto, devido à falta de informação sobre o tipo de radiação coletada pelas estações meteorológicas, optou-se por estimar a radiação total usando as Equações 2 e 3, conforme o método descrito por Conceição (2006).

<sup>1</sup> Forma de representar a data do calendário gregoriano como um número inteiro, contando o número de dias decorridos desde o início de um determinado ano. Por exemplo, 1º de janeiro corresponde ao dia juliano 1, enquanto 31 de dezembro corresponde ao dia juliano 365 (ou 366 em anos bissextos). Essa representação é frequentemente usada em astronomia e em cálculos relacionados a eventos cíclicos, como a estimativa da  $ET_o$ .

Tabela 5 Funções utilizadas para o ajuste de granularidade (horário e diário) dos dados da estação particular

Variável	Ajuste horário	Ajuste diário
Pressão atmosférica	Média da hora	Média do dia
Temperatura máxima ( $T_{max}$ )	Valor máximo da hora	Valor máximo do dia
Temperatura mínima ( $T_{min}$ )	Valor mínimo da hora	Valor mínimo do dia
Temperatura	Média da hora	$(T_{max} + T_{min})/2$
Umidade	Média da hora	Média do dia
Temperatura de ponto de orvalho	Média da hora	Média do dia
Velocidade do vento	Média da hora	Média do dia
Rajada de vento	Valor máximo da hora	Valor máximo do dia
Sensação térmica	Média da hora	Média do dia
Índice de calor	Média da hora	Média do dia
Índice THW	Média da hora	Média do dia
Precipitação	Somatório da hora	Somatório do dia
Índice pluviométrico	Somatório da hora	Somatório do dia
Irradiação solar	Média da hora	Média do dia
Energia Solar	Média da hora	Média do dia
Irradiação solar máxima	Valor máximo da hora	Valor máximo do dia
Evapotranspiração	Somatório da hora	Somatório do dia

Fonte: A autora

$$R_s = k_{Rs} R_a \sqrt{(T_{max} - T_{min})} \quad (2)$$

Em que:

$R_s$  – radiação solar incidente, MJ m<sup>-2</sup> dia<sup>-1</sup>

$k_{Rs}$  – coeficiente empírico de proporcionalidade, adimensional, cujo valor para a região Sul do Brasil é  $\approx 0,15$  (Xavier *et al.*, 2017)

$R_a$  – radiação solar incidente no topo da atmosfera, MJ m<sup>-2</sup> dia<sup>-1</sup>

$T_{max}$  – temperatura máxima, °C

$T_{min}$  – temperatura mínima, °C

$$Ra = \frac{118,08}{\pi} dr [\omega_s \text{sen}(\varphi) \text{sen}(\delta) + \cos(\varphi) \cos(\delta) \text{sen}(\omega_s)] \quad (3)$$

Em que:

$Ra$  – radiação solar no topo da atmosfera,  $\text{MJ m}^{-2} \text{dia}^{-1}$

$dr$  – distância inversa relativa entre a Terra e o Sol, em radianos

$\omega_s$  – ângulo horário ao nascer do sol, em radianos

$\varphi$  – latitude do local, em radianos

$\delta$  – declinação solar, em radianos

#### 4.2.4 Normalização dos dados

A normalização dos dados é um passo importante no treinamento de Redes Neurais quando os dados de entrada estão em escalas diferentes, que é o caso de dados meteorológicos (Hastie; Tibshirani; Friedman, 2009). A normalização torna as magnitudes dos atributos comparáveis, já que diferenças extremas nas escalas podem fazer com que um atributo com valores maiores domine as etapas do treinamento, levando a um desempenho desequilibrado (Louzada, 2013).

O método usado para normalizar os dados foi o `StandardScaler`, da biblioteca `scikit-learn`<sup>2</sup> do Python, que segue a distribuição normal (Gaussiana) dos dados, transformando os atributos de forma que a média seja 0 e o desvio padrão seja 1 e é dada pela Equação 4.

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

Em que:

$z$  – valor padronizado do atributo

$x$  – valor original do atributo

$\mu$  – média dos valores do atributo

$\sigma$  – desvio padrão dos valores do atributo

<sup>2</sup> O método pode ser acessado no link: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

### 4.3 Cálculo da $ET_o$

Como visto na Seção 3.2, o método Penman-Monteith foi selecionado pela FAO por sua capacidade de proporcionar resultados similares aos que se obtém pelas medições diretas em diversos tipos de clima (Lucas, 2019).

Para o cálculo da  $ET_o$  implementou-se um programa em linguagem C, o qual seguiu o roteiro dado por Conceição (2006) para o cálculo da Equação 1.

### 4.4 Variáveis de entrada

Como vimos na Seção 3.2, as variáveis meteorológicas são muito importantes para a estimativa da  $ET_o$ . Considerando as variáveis disponíveis nas estações do INMET e do produtor rural, este trabalho explorou 19 configurações diferentes destas variáveis. Essas configurações foram derivadas da literatura especializada, considerando estudos anteriores que identificaram a relevância dessas variáveis na estimativa da  $ET_o$ .

Dentre todas as variáveis disponíveis nos dados coletados, foram selecionadas 14 para a criação das 19 combinações. Essas variáveis são mostradas na Tabela 6, juntamente com o nome utilizado na codificação e suas unidades de medida.

É importante ressaltar, que dentro da literatura utilizada para dar suporte à seleção das 19 configurações, os dados disponíveis em cada estudo são, em grande parte, diferentes das disponíveis neste trabalho. Por esse motivo, as variáveis escolhidas para cada configuração foram adaptadas ao contexto da presente pesquisa, mas mantendo a coerência com os trabalhos escolhidos e levando em consideração as particularidades dos conjuntos de dados disponíveis.

A Tabela 7 mostra as variáveis escolhidas em cada uma das 19 configurações. As variáveis levam o nome utilizadas na codificação, conforme mostrado na Tabela 6.

Nas próximas subseções serão descritas cada configuração escolhida, as adaptações realizadas, as variáveis utilizadas e as fontes bibliográficas que dão suporte às escolhas. A Seção 3.4.1 descreve de forma mais detalhada os trabalhos utilizados como base para as configurações.

Tabela 6 Variáveis meteorológicas utilizadas no treinamento dos modelos de RNAs

<b>Variável</b>	<b>Nome Variável no Código</b>	<b>Unidade de Medida</b>
Dia Juliano (calculado)	DiaJ	dd
Temperatura Média	T	°C
Temperatura Máxima	Tmax	°C
Temperatura Mínima	Tmin	°C
Umidade Relativa Média	Ur	%
Umidade Relativa Máxima	Urmax	%
Umidade Relativa Mínima	Urmin	%
Pressão Atmosférica	Pressao	kPa
Velocidade do Vento	Vento	m/s
Radiação (da estação)	Radiacao	Kj/m <sup>2</sup>
Latitude	Lat	Graus decimais
Longitude	Long	Graus decimais
Altitude	Alt	Metros
Radiação (estimada)	Rs	Mj/m <sup>2</sup>
$ET_o$ (calculada)	ET0	mm/dia

Fonte: A autora

Tabela 7 Configurações de entrada escolhidas para os testes das RNAs

<b>Configurações de entrada</b>	<b>DiaJ</b>	<b>T</b>	<b>Tmax</b>	<b>Tmin</b>	<b>Ur</b>	<b>Urmax</b>	<b>Urmin</b>	<b>Pressao</b>	<b>Vento</b>	<b>Radiacao</b>	<b>Lat</b>	<b>Long</b>	<b>Alt</b>	<b>Rs</b>
RNA <sub>1</sub>	X	X	X	X							X			
RNA <sub>1a</sub>		X	X	X							X			
RNA <sub>2</sub>	X	X	X	X							X		X	X
RNA <sub>2a</sub>		X	X	X							X			X
RNA <sub>3</sub>	X	X	X	X					X		X			
RNA <sub>3a</sub>		X	X	X					X		X			
RNA <sub>4</sub>														X
RNA <sub>5</sub>		X	X											X
RNA <sub>6</sub>		X	X	X										X
RNA <sub>7</sub>		X	X	X		X			X					X
RNA <sub>8</sub>	X	X	X	X	X						X	X	X	
RNA <sub>8a</sub>		X	X	X	X						X	X	X	
RNA <sub>9</sub>	X	X	X	X	X				X		X	X	X	
RNA <sub>9a</sub>		X	X	X	X				X		X	X	X	
RNA <sub>10</sub>		X			X				X					X
RNA <sub>11</sub>	X	X			X				X	X				
RNA <sub>12</sub>	X	X	X	X	X				X					
RNA <sub>13</sub>	X	X			X			X	X	X				
RNA <sub>14</sub>	X	X	X	X	X			X	X					X

Fonte: A autora

#### 4.4.1 Configurações 1 a 3

Segundo Conceição (2006), para estimar o valor da radiação, utilizada para o cálculo da  $ET_o$ , são necessárias as seguintes variáveis meteorológicas:

- Dia Juliano (DiaJ)
- Temperatura mínima ( $T_{min}$ )
- Temperatura máxima ( $T_{max}$ )
- Latitude da estação (Lat)

O trabalho de Lima *et al.* (2019) mostra como a radiação global e a temperatura do ar são as variáveis mais importantes para o método de Penman-Monteith. Por esse motivo, essas configurações utilizam as 4 variáveis citadas acima e a própria radiação (a disponível nas estações e a estimada), conforme lista abaixo. Como a variável Dia Juliano é um valor adimensional, que indica apenas o dia do ano, decidiu-se testar a mesma configuração com e sem esse valor<sup>3</sup>, para avaliar sua importância no treinamento.

**RNA<sub>1</sub> e RNA<sub>1a</sub>:** menor número de variáveis necessárias para estimar a radiação: DiaJ,  $T_{min}$ ,  $T_{max}$ , Lat.

**RNA<sub>2</sub> e RNA<sub>2a</sub>:** são as mesmas variáveis da configuração 1, mas inclui a radiação estimada: DiaJ,  $T_{min}$ ,  $T_{max}$ , Lat, Rs.

**RNA<sub>3</sub> e RNA<sub>3a</sub>:** mesmas variáveis da configuração 2, mas agora utilizando a radiação disponível nas estações no lugar da estimada: DiaJ,  $T_{min}$ ,  $T_{max}$ , Lat, Radiacao.

#### 4.4.2 Configurações 4 a 7

Essas configurações foram criadas baseadas no trabalho de Sowmya, Kumar e Ambat (2020), que utilizou 4 combinações de variáveis: a primeira só utilizou a radiação solar, a segunda utilizou radiação solar e temperatura máxima, a terceira, radiação solar e temperaturas máxima e mínima e a última, que usou as mesmas variáveis da terceira combinação e incluiu velocidade do vento e umidade relativa mínima e máxima.

Para este trabalho, as configurações de entrada são as listadas abaixo:

---

<sup>3</sup> As configurações que possuem a letra a em sua nomeação não utilizam o Dia Juliano.

**RNA<sub>4</sub>**: esta configuração utiliza apenas a radiação estimada:  $R_s$ .

**RNA<sub>5</sub>**: além de  $R_s$  utiliza também  $T_{max}$ .

**RNA<sub>6</sub>**:  $R_s$ ,  $T_{max}$ ,  $T_{min}$ .

**RNA<sub>7</sub>**: esta configuração utiliza as mesmas da anterior, mas inclui  $V_{ento}$ ,  $U_{rmax}$ ,  $U_{rmin}$ .

#### 4.4.3 Configurações 8 e 9

Essas configurações foram baseadas no trabalho de Ferraz (2014), e como elas também usam o Dia Juliano, utilizou-se a mesma ideia da Seção 4.4.1. As configurações estão descritas abaixo:

**RNA<sub>8</sub> e RNA<sub>8a</sub>**:  $Lat$ ,  $Long$ ,  $Alt$ ,  $DiaJ$ ,  $T_{min}$ ,  $T_{max}$ ,  $U_r$ .

**RNA<sub>9</sub> e RNA<sub>9a</sub>**: são as mesmas variáveis da configuração 8, mas inclui a velocidade do vento,  $V_{ento}$ .

#### 4.4.4 Configurações 10 e 11

A ideia para essas duas configurações foi baseada nos trabalhos de Santos (2020) e Walls *et al.* (2020), com algumas modificações: no lugar de temperatura e umidade relativa máxima e mínima, utilizou-se a temperatura e umidade médias. Walls *et al.* (2020) também utilizaram fluxo de calor do solo, variável que não está disponível nos dados deste trabalho. As duas pesquisas usam algum tipo de radiação, por esse motivo, utilizou-se a radiação da estação e a estimada, conforme listado abaixo:

**RNA<sub>10</sub>**: temperatura média,  $T$ , umidade relativa,  $U_r$ ,  $V_{ento}$  e a radiação solar estimada,  $R_s$ .

**RNA<sub>11</sub>**: mesmas variáveis da configuração anterior mas no lugar da radiação estimada, utilizou-se a radiação disponível na estação,  $R_{diacao}$ .

#### 4.4.5 Configuração 12

Esta configuração foi baseada no trabalho de Silva (2002), que utilizou as temperaturas média, máxima e mínima ( $T$ ,  $T_{max}$ ,  $T_{min}$ ), a umidade relativa média ( $U_r$ ) e a direção do vento ( $V_{ento}$ ). O estudo também aplicou o número de horas de brilho solar diária (ou insolação), mas como os dados disponíveis não possuem essa variável, optou-se por não estimá-la e usar apenas as 5 variáveis citadas.

#### 4.4.6 Configurações 13 e 14

As configurações 13 e 14 são baseadas no trabalho de Arca, Beniscasa e Vincenzi (2001), que utiliza, além das variáveis que foram escolhidas para estas configurações, 3 radiações diferentes: radiação solar global, radiação difusa, radiação líquida. Por não ser possível saber qual radiação as estações fazem uso, utilizou-se uma configuração com a radiação da estação e uma com a estimada. Outra diferença entre estas configurações é o trabalho original, que utiliza direção do vento e intensidade da chuva, variáveis não disponíveis. Essas são as únicas configurações que utilizam a variável de pressão atmosférica. Considerando essas diferenças, as duas configurações são como segue:

**RNA<sub>13</sub>**: o trabalho original utiliza a temperatura do ar: nesta configuração utilizou-se a temperatura média ( $T$ ). As outras variáveis são  $U_r$ , Vento, Pressao, Radiacao.

**RNA<sub>14</sub>**: mesmas variáveis da configuração anterior mas no lugar da radiação da estação, utilizou-se a radiação estimada,  $R_s$ .

Outra informação que foi incluída nos testes, é o valor da  $ET_o$  calculada por outros métodos: Hargreaves, Hargreaves-Samani e Blaney-Criddle (Arca; Beniscasa; Vincenzi, 2001).

### 4.5 Projeto da arquitetura da RNA

Na Seção 4.4, as variáveis de entrada foram definidas, estabelecendo assim a base para o desenvolvimento dos modelos de estimativa da  $ET_o$ . Nesta seção, os aspectos arquitetônicos examinados e refinados durante as investigações são explorados em detalhes. A Figura 9 ilustra o modelo geral que foi utilizado como base para o desenvolvimento dos modelos de estimativa da  $ET_o$ . Em seguida, são discutidos o tipo de rede neural, o número de camadas, entre outros aspectos relevantes para a arquitetura da RNA.

A arquitetura escolhida foi a Perceptron Multi-Camadas (Seção 3.3.1), com uma camada de oculta, escolha comum em problemas de aprendizado de máquina devido à sua eficácia comprovada em uma variedade de aplicações (Heaton, 2008), (Goodfellow; Bengio; Courville, 2016). Embora mais comumente associada aos problemas de classificação, como reconhecimento de padrões e classificação de imagens, a escolha pela MLP foi feita devido à sua capacidade comprovada de lidar com problemas de regressão (Carvalho; Delgado, 2022). Essa escolha mostrou-se adequada para o problema em questão, permitindo a modelagem eficaz da relação entre as variáveis de entrada e saída em um contexto de previsão.

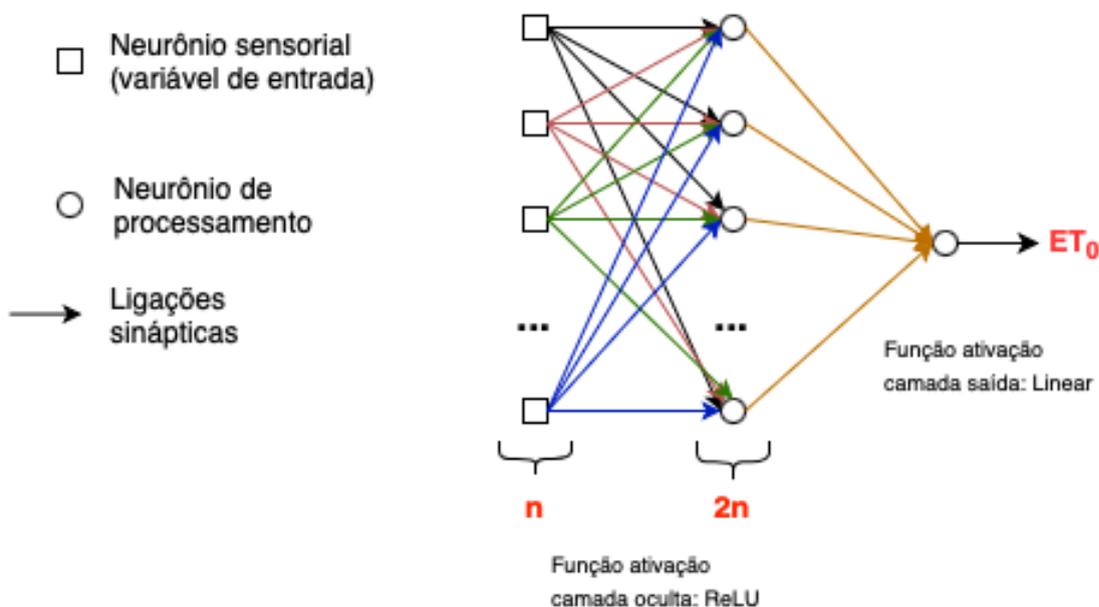


Figura 9 Modelo geral da RNA

Fonte: A autora

O número de neurônios da camada de entrada ( $n$ ) irá variar entre as configurações testadas, visto que o número de neurônios de entrada é igual ao número de variáveis da configuração. A escolha do número de neurônios na camada oculta, de acordo com Heaton (2008), também foi realizada de forma empírica, mas deve seguir algumas regras, levando em conta o número de neurônios de entrada ( $n$ ) e o número de neurônios de saída ( $s$ ), que neste trabalho é igual a 1 (um), visto que a saída é apenas um único valor, a estimativa da  $ET_o$ . Os valores mínimo, médio e máximo de neurônios da camada oculta são como segue:

**Número mínimo:**  $(n + s)/2$

**Número médio:**  $(n * (2/3)) + s$

**Número máximo:**  $2 * n$

Considerando as configurações escolhidas para teste e as regras acima, é possível mostrar o número de neurônios para cada configuração na Tabela 8. Os valores de neurônios foram arredondados. O valor escolhido para os testes foi o número máximo de neurônios.

Sobre as funções de ativação (Seção 3.3.2), para a camada oculta foi utilizada a função ReLU, devido à sua eficácia na captura de relações não-lineares complexas nos dados. Essa função é a mais indicada, já que permite que a RNA aprenda e modele de forma eficiente características não-lineares, o que é fundamental na estimativa da  $ET_o$ , já que as relações

Tabela 8 Número de neurônios para a camada de entrada e limites para o número de neurônios na camada oculta, para cada configuração testada

<b>Configuração entrada</b>	<b>Num. Entrada</b>	<b>Num. Mínimo</b>	<b>Num. Médio</b>	<b>Num. Máximo</b>
RNA <sub>1</sub>	4	3	4	8
RNA <sub>1a</sub>	3	2	3	6
RNA <sub>2</sub>	5	3	4	10
RNA <sub>2a</sub>	4	3	4	8
RNA <sub>3</sub>	5	3	4	10
RNA <sub>3a</sub>	4	3	4	8
RNA <sub>4</sub>	1	1	2	2
RNA <sub>5</sub>	2	2	2	4
RNA <sub>6</sub>	3	2	3	6
RNA <sub>7</sub>	6	4	5	12
RNA <sub>8</sub>	7	4	6	14
RNA <sub>8a</sub>	6	4	5	12
RNA <sub>9</sub>	8	5	6	16
RNA <sub>9a</sub>	7	4	6	14
RNA <sub>10</sub>	4	3	4	8
RNA <sub>11</sub>	4	3	4	8
RNA <sub>12</sub>	5	3	4	10
RNA <sub>13</sub>	5	3	4	10
RNA <sub>14</sub>	5	3	4	10

Fonte: A autora

climáticas e biológicas são frequentemente não-lineares (Lucas, 2019), além de ser computacionalmente eficiente no treinamento de RNAs.

Para a camada de saída foi utilizada a função linear, muito utilizada em problemas de regressão, já que permite que a RNA calcule diretamente a estimativa da  $ET_o$  com base nas características extraídas pelas camadas anteriores, tornando a interpretação e a utilidade da saída mais diretas, já que produz resultados em todo o domínio dos números reais (Heaton, 2008).

A inicialização dos pesos, que são ajustados durante o treinamento, pode ser feita de forma randômica ou utilizando algum método mais específico (Heaton, 2008). Neste trabalho,

optou-se por inicializar os pesos utilizando a função `he_normal` da biblioteca `tensorflow`<sup>4</sup> do Python.

## 4.6 Avaliação da arquitetura

A avaliação da arquitetura da RNA é uma etapa importante no desenvolvimento de qualquer sistema que utiliza essa técnica. Nesta seção, são exploradas as escolhas fundamentais feitas na configuração da RNA, incluindo a função de erro, métrica de avaliação e a exploração de hiperparâmetros, com destaque para a importância desse processo para o desempenho geral do modelo.

### 4.6.1 Escolha da função de erro e métrica de avaliação

Para avaliar o desempenho dos modelos de RNAs na estimativa da  $ET_o$ , utilizaram-se diversas métricas de avaliação, que abrangem diferentes aspectos do erro na estimativa da  $ET_o$ . Abaixo, a descrição de cada uma das métricas utilizadas (Santos, 2020).

#### Erro Médio Absoluto

O Erro Médio Absoluto (MAE - *Mean Absolute Error*), dado pela Equação 5, calcula a média das diferenças absolutas entre as estimativas da RNA e os valores reais da  $ET_o$ , fornecendo uma medida de erro que considera a magnitude das discrepâncias. Sua medida é dada pela mesma unidade da  $ET_o$ , mm/dia, quanto menor seu valor, menor o erro entre a saída da RNA e o valor da  $ET_o$ .

$$\frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad (5)$$

Em que:

$N$  – número de amostras

$x_i$  – média diária da  $ET_o$ , calculada pelo método de Penman-Monteith

$y_i$  – média diária da  $ET_o$  predita pelo modelo de RNA

---

<sup>4</sup> [https://www.tensorflow.org/api\\_docs/python/tf/keras/initializers/HeNormal](https://www.tensorflow.org/api_docs/python/tf/keras/initializers/HeNormal)

### Erro Quadrático Médio

O Erro Quadrático Médio (MSE - *Mean Squared Error*), dado pela Equação 6, computa a média dos quadrados dos erros, atribuindo pesos maiores às discrepâncias maiores, tornando-se sensível aos valores discrepantes. Quanto mais próximo de 0, melhor o modelo neural se adaptou à previsão.

$$\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (6)$$

Em que:

$N$  – número de amostras

$x_i$  – média diária da  $ET_o$ , calculada pelo método de Penman-Monteith

$y_i$  – média diária da  $ET_o$  predita pelo modelo de RNA

### Raiz do Erro Quadrático Médio

A Raiz do Erro Quadrático Médio (RMSE - *Root Mean Squared Error*), dada pela Equação 7, é derivada do MSE e expressa o erro na mesma unidade das previsões, mm/dia, tornando-se uma métrica útil para compreender a dispersão dos erros. Quanto menor seu valor, melhor.

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (7)$$

Em que:

$N$  – número de amostras

$x_i$  – média diária da  $ET_o$ , calculada pelo método de Penman-Monteith

$y_i$  – média diária da  $ET_o$  predita pelo modelo de RNA

## Coeficiente de Determinação

O coeficiente de determinação ( $R^2$ ) avalia o quão bem as previsões se ajustam aos dados reais, indicando a proporção da variabilidade na  $ET_o$  explicada pela RNA. Quanto mais próximo de 1, mais próximo dos valores reais está a previsão da RNA.

A combinação dessas métricas proporcionou uma visão abrangente do desempenho dos modelos, considerando diferentes aspectos da precisão das estimativas da  $ET_o$  (Carmo; Silva, 2023).

A escolha da função de erro e métrica de avaliação durante o treinamento desempenham papéis fundamentais na determinação de como a RNA ajusta seus pesos durante o treinamento e como a precisão das previsões resultantes é avaliada. Essas escolhas desempenham um papel importante para que a RNA aprenda adequadamente a mapear as relações entre as variáveis climáticas de entrada e a  $ET_o$  de saída. A função de erro escolhida foi o MSE e como métrica de avaliação foi escolhido o MAE (Goodfellow; Bengio; Courville, 2016).

### 4.6.2 Teste Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov (KS) foi empregado como uma ferramenta estatística auxiliar às métricas utilizadas, para avaliar quão bem as distribuições dos dados observados de  $ET_o$  estimada pela RNA se ajustam aos dados reais, considerando as condições de teste (LUIZ *et al.*, 2021). A estatística de teste  $D$ , dada pela Equação 8, será utilizada para calcular a maior discrepância entre a Função de Distribuição Acumulativa Empírica (ECDF - *Empirical Cumulative Distribution Function*) das estimativas de  $ET_o$  geradas pela RNA e a ECDF dos dados reais (Teixeira; Gomes; Santos, 2020).

$$D = \max |F_n(x) - F(x)| \quad (8)$$

Em que:

$D$  – estatística de teste

$F_n(x)$  – função de distribuição acumulativa empírica das observações.

$F(x)$  – função de distribuição acumulativa teórica

O teste foi aplicado para verificar se as amostras de  $ET_o$  estimadas pela RNA e as amostras reais de  $ET_o$  provêm da mesma distribuição de probabilidade. Além disso, foi calculado o valor-p associado à estatística de teste  $D$ , fornecendo uma medida adicional para avaliar a consistência entre as distribuições. Adotou-se um nível de significância de 0,05 para a comparação com o valor-p. Um valor-p menor que o nível de significância sugere evidências estatísticas de que as amostras não provêm da mesma distribuição de probabilidade, contribuindo para uma avaliação mais completa da adequação do modelo de RNA em reproduzir a distribuição dos dados observados (LUIZ *et al.*, 2021).

#### 4.6.3 Descrição dos hiperparâmetros

É importante ressaltar que as escolhas dos hiperparâmetros adotadas neste estudo, como a taxa de aprendizado, otimizadores, número máximo de épocas e tamanho do lote, foram fundamentadas em uma extensa revisão da literatura especializada em RNAs (Heaton, 2008), (Goodfellow; Bengio; Courville, 2016). Essas escolhas desempenham um papel crítico no desempenho e na estabilidade de modelos de RNA. Cada uma dessas configurações foi cuidadosamente selecionada com base em resultados empíricos de estudos anteriores que abordaram problemas semelhantes, como os trabalhos de Reis (2018), de Honnef (2020), além de diversos tutoriais disponíveis na Internet<sup>5 6 7 8</sup>. As decisões foram orientadas por práticas estabelecidas na comunidade de pesquisa em RNA. Ao adotar essas configurações, busca-se não apenas otimizar o desempenho do modelo, mas também garantir a replicabilidade e a confiabilidade dos resultados em conformidade com as melhores práticas reconhecidas na área de RNAs.

Considerando a literatura e as definições da Seção 3.3.3, o paradigma de aprendizado escolhido foi o supervisionado e a taxa de aprendizagem foi fixada em 0,001, devido a sua eficácia em facilitar a convergência do modelo. Foram explorados os dois otimizadores descritos na Seção 3.3.4, o Adam e o SGD, permitindo uma comparação entre essas estratégias de otimização. O valor de *momentum* foi fixado em 0,9, um valor que mostrou bom desempenho em manter o treinamento do modelo estável.

O número máximo de épocas de treinamento foi definido como 600, embora tenha sido implementado o mecanismo de parada antecipada, por meio da função `EarlyStopping` da

---

<sup>5</sup> <https://www.deeplearningbook.com.br/capitulo-26-como-escolher-os-hiperparametros-de-uma-rede-neural/>

<sup>6</sup> <https://aws.amazon.com/pt/what-is/hyperparameter-tuning/>

<sup>7</sup> <https://pt.linkedin.com/pulse/hiperparâmetros-jose-r-f-junior>

<sup>8</sup> <https://learn.microsoft.com/pt-br/azure/machine-learning/component-reference/neural-network-regression?view=azureml-api-2>; entre outros

biblioteca Keras<sup>9</sup>, que encerra o treinamento quando não há mais melhorias na métrica de validação, aumentando a eficiência do processo de treinamento. O tamanho do lote (*batch\_size*) foi estabelecido em 16, permitindo uma eficiente utilização dos recursos computacionais.

Cada combinação de otimizador foi avaliada cinco vezes, resultando em múltiplas execuções independentes. Essa abordagem de avaliação repetida permite obter estatísticas robustas do desempenho do modelo para cada configuração de hiperparâmetros, o que garante que os resultados sejam confiáveis e representativos.

## 4.7 Experimentos realizados

Nesta seção, apresentamos uma descrição detalhada dos experimentos realizados com o intuito de atingir os objetivos da pesquisa. Iniciamos abordando a estratégia de divisão dos dados, delineando os procedimentos experimentais adotados, desde a configuração das RNAs até a seleção das melhores configurações de entrada. Em seguida, exploramos o teste de generalização, empregando dados do produtor rural para avaliar a capacidade das RNAs em prever a  $ET_0$  em cenários do mundo real. Essa seção fornece as bases metodológicas para sustentar a análise dos resultados no Capítulo 5, aumentando a solidez e confiabilidade das conclusões.

### 4.7.1 Divisão de dados para treinamento e validação

O primeiro conjunto de dados (descritos na Seção 4.1.1), composto por 22.387 amostras, foi dividido em três conjuntos distintos: treinamento, validação e teste, considerando as variáveis de entrada (X) e o alvo (Y). O conjunto de treinamento, com 14.327 amostras (64% do total), foi utilizado para o treinamento da RNA com diferentes configurações de entrada, conforme descrito na Seção 4.4. O conjunto de validação, composto por 3.582 amostras (16%), desempenhou um papel importante no ajuste fino das configurações de hiperparâmetros, fornecendo uma avaliação objetiva do desempenho da RNA na previsão do alvo Y (valor de  $ET_0$ ). O conjunto de teste, composto por 4.478 amostras (20%), foi reservado para a avaliação final do desempenho do modelo e também incluiu as variáveis de entrada X correspondentes ao alvo Y.

---

<sup>9</sup> [https://keras.io/api/callbacks/early\\_stopping/](https://keras.io/api/callbacks/early_stopping/)

#### 4.7.2 Execução dos experimentos

Cada configuração de hiperparâmetros previamente detalhada na Seção 4.6.3 foi submetida a um rigoroso processo de experimentação. Para cada configuração, foram realizadas cinco execuções completas de treinamento. Em cada execução, as métricas de erro (MAE, MSE, RMSE e  $R^2$ ) foram registradas e armazenadas em uma planilha. A média desses erros foi calculada para cada configuração de entrada, fornecendo uma visão consolidada do desempenho do modelo em diferentes cenários.

#### 4.7.3 Seleção das melhores configurações

Com o objetivo de identificar as configurações de entrada mais promissoras, uma análise das configurações foi conduzida com base nas métricas de erro médias. Foram selecionadas as cinco melhores configurações de entrada para cada configuração de hiperparâmetros, resultando em um conjunto de 10 modelos distintos. Essas configurações destacaram-se devido ao desempenho superior observado nos conjuntos de treinamento e validação.

#### 4.7.4 Teste de generalização

Para verificar a capacidade de generalização dos modelos treinados, os 10 modelos selecionados foram posteriormente testados com um conjunto independente de dados (descritos na Seção 4.1.2), com 1.144 amostras. Essa etapa permitiu avaliar o desempenho dos modelos em dados não treinados, identificando as configurações mais robustas e adequadas à aplicação real.

#### 4.7.5 Seleção final dos modelos

Com base no desempenho no estudo de caso, dois modelos foram selecionados para comparações finais. O primeiro critério de escolha baseou-se nas métricas utilizadas, incluindo MAE, MSE, RMSE e  $R^2$ , com o modelo que apresentou os melhores resultados em todas essas métricas emergindo como a primeira opção. Além disso, a eficiência na utilização de variáveis de entrada foi considerada, levando à seleção do segundo modelo, que manteve um desempenho comparável, mas utilizou um número reduzido de variáveis. Essas escolhas refletem um equilíbrio entre alta precisão e simplicidade, atendendo às necessidades do produtor rural e contribuindo para uma gestão mais eficaz da irrigação.

## 4.8 Ferramentas computacionais utilizadas

Nesta seção, descrevem-se as principais ferramentas computacionais que desempenharam um papel central ao longo deste projeto de pesquisa. O foco está nas linguagens de programação e *softwares* específicos que foram essenciais, desde a preparação e processamento dos dados até a implementação e avaliação das RNAs.

Para a etapa inicial de pré-processamento de dados (Seção 4.2) foi utilizado o *software Microsoft Excel*. Essa ferramenta foi empregada para realizar a limpeza dos dados, bem como ajustar a granularidade das informações, preparando os dados brutos para as próximas etapas.

A inclusão de novas variáveis nos dados (Seção 4.2.3) e o cálculo da  $ET_o$  como variável alvo (Seção 4.3), foram aplicados utilizando a linguagem de programação C. Essas novas variáveis foram calculadas e adicionadas aos dados originais, proporcionando informações adicionais para a modelagem posterior.

A normalização dos dados (Seção 4.2.4) foi realizada com a biblioteca *scikit-learn* do Python. Essa etapa foi fundamental para promover a compatibilidade das escalas das variáveis de entrada, o que contribuiu significativamente para o desempenho eficaz do modelo de aprendizado de máquina.

A implementação e o treinamento das RNAs foram conduzidos no ambiente *Colab* do Google<sup>10</sup>, que se baseia na linguagem Python<sup>11</sup>. Para viabilizar funcionalidades específicas, foram utilizadas diversas bibliotecas. A versão utilizada do Python é a 3.10.12 e a Tabela 9 lista essas bibliotecas, acompanhadas de suas respectivas versões<sup>12</sup>.

## 4.9 Síntese da metodologia

Esta seção apresenta uma síntese da metodologia adotada neste estudo, fornecendo uma visão geral das etapas-chave do processo. A Figura 10 mostra as diferentes fases da metodologia, destacando as etapas essenciais envolvidas na coleta, processamento e análise dos dados, e pode ser resumida em 7 grandes etapas:

**Coleta de dados:** inicialmente, foram coletados dados de 20 estações meteorológicas localizadas no estado do Paraná.

---

<sup>10</sup> <https://colab.google>

<sup>11</sup> <https://python.org.br>

<sup>12</sup> Versões válidas em agosto/2023.

Tabela 9 Bibliotecas do Python utilizadas e suas versões

Biblioteca	Versão
pandas	1.5.3
numpy	1.23.5
tensorflow	2.12.0
sklearn	1.2.2
keras	2.12.0
sys	3.10.12
matplotlib	3.7.1

Fonte: A autora

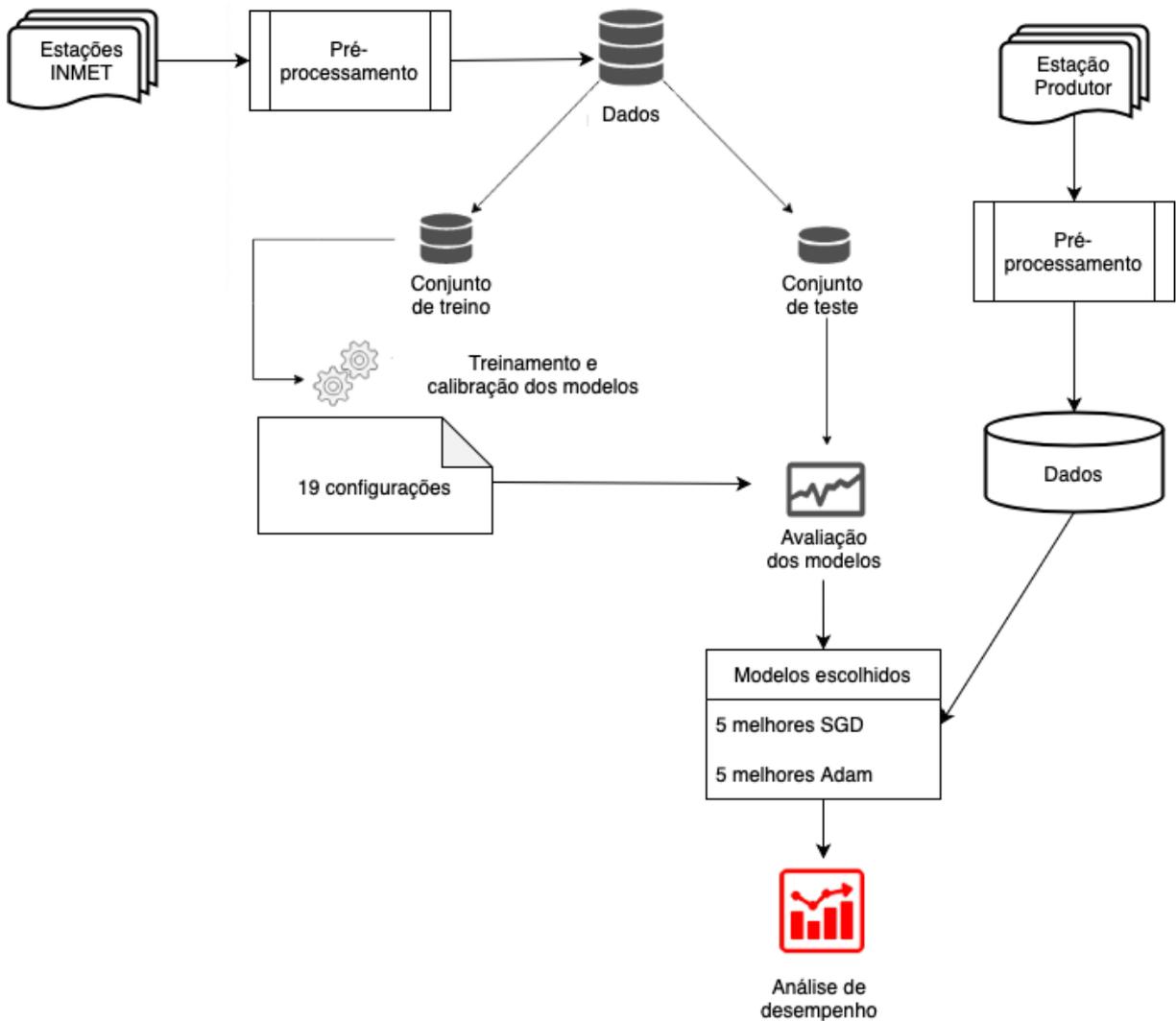


Figura 10 Síntese da metodologia utilizada

Fonte: A autora

**Pré-processamento dos dados:** os dados coletados foram pré-processados, limpos, sendo incluídas variáveis adicionais consideradas relevantes para o treinamento dos modelos de RNA.

**Divisão dos dados:** em seguida, os dados foram divididos em conjuntos de treinamento (64%), validação (15%) e teste (20%).

**Treinamento e calibração dos modelos:** utilizando os dados de treinamento, foram treinadas e calibradas 19 configurações diferentes de modelos de RNA, empregando dois otimizadores distintos e realizando cinco execuções para cada configuração. As métricas de erro foram armazenadas para posterior análise.

**Avaliação dos modelos:** com base nas métricas de erro obtidas, foram selecionadas as cinco melhores configurações para cada otimizador, totalizando 10 modelos.

**Teste com os dados do produtor:** os dados do produtor foram pré-processados e testados com os 10 modelos selecionados.

**Avaliação do desempenho dos modelos:** utilizando as métricas de erro e o teste de Kolmogorov-Smirnov, o desempenho dos modelos foi avaliado, resultando na escolha de duas configurações: a melhor entre todas e aquela que utilizou um número reduzido de variáveis.

Este capítulo apresentou a estrutura utilizada para a realização dos experimentos e a subsequente análise de resultados. No próximo capítulo, os resultados obtidos são discutidos e oferecerão uma visão abrangente das conclusões reveladas pelas investigações metodológicas conduzidas.

## 5 RESULTADOS E DISCUSSÃO

Neste capítulo, apresentam-se e interpretam-se os resultados derivados da aplicação das RNAs para a estimativa da  $ET_o$ . Os resultados foram analisados em detalhes, considerando diversas configurações de variáveis de entrada, arquiteturas de RNAs e métricas de avaliação, fornecendo uma visão abrangente do desempenho dos modelos, relacionando os achados com as principais práticas discutidas na revisão bibliográfica. Também é incluída uma avaliação da eficácia do modelo e sua capacidade de generalização para dados do produtor rural.

### 5.1 Avaliação das RNAs projetadas

Cada uma das configurações descritas na Seção 4.4 foi treinada com os dados provenientes das estações do INMET (descritos com detalhes na Seção 4.1.1) utilizando o projeto de RNA detalhado na Tabela 10.

Foram utilizados os dois otimizadores descritos na Seção 3.3.4, o SGD (Configuração 1 -  $C_1$ ) e o Adam (Configuração 2 -  $C_2$ ). O número de neurônios de entrada e o número de neurônios na camada oculta dependem do número de variáveis de entrada, conforme mostrado na Tabela 8 (utilizou-se o número máximo de neurônios). Para uma avaliação robusta, os resultados apresentados são a média de 5 execuções, realizadas em uma partição dos dados em treinamento (64%), validação (16%), e teste (20%). As métricas utilizadas, MAE, MSE, RMSE e  $R^2$ , descritas na Seção 4.6.1, fornecem uma visão abrangente do desempenho das configurações em ambas as estratégias de otimização e estão resumidas nas Tabelas 11 (para  $C_1$ ) e 12 (para  $C_2$ ).

Tabela 10 Arquitetura utilizada para o treinamento das configurações de entrada

<b>Arquitetura RNA</b>	<b>C<sub>1</sub></b>	<b>C<sub>2</sub></b>
Num. Camada Ocultas	1	1
Num. Neurônios de Saída	1	1
Otimizador	SGD	Adam
Função de Ativação Camada Oculta	Relu	Relu
Taxa Aprendizagem	0,001	0,001
<i>Momentum</i>	0,9	NA
Função de Erro	MSE	MSE
Métrica	MAE	MAE
<i>Cross-Validation (CV)</i>	5	5
Max. Num. Épocas (uso do <i>EarlyStopping</i> )	600	600
Tamanho do batch ( <i>batch_size</i> )	16	16

NA: não se aplica ao otimizador

Fonte: A autora

Para cada otimizador, foram escolhidas as 5 melhores configurações para testar com os dados do produtor. Para o otimizador SGD, as configurações escolhidas estão em destaque na Tabela 11 e para o otimizador Adam estão em destaque na Tabela 12 e em ambos as melhores configurações foram: RNA<sub>7</sub>, RNA<sub>8</sub>, RNA<sub>9</sub>, RNA<sub>10</sub> e RNA<sub>14</sub>.

Na próxima seção, são analisadas cada uma das configurações testadas, comparando-as com a literatura descrita na Seção 3.4.1. Após, serão discutidos os resultados entre as 5 melhores configurações com cada otimizador.

Tabela 11 Resumo das métricas das configurações para o otimizador SGD

Configuração	Variáveis		Num. Neur. Camada Oculta	Num. Épocas (média)	MAE (média)	MSE (média)	RMSE (média)	R <sup>2</sup> (média)
	Entrada	4						
RNA <sub>1</sub>	4	8	53,2	0,35	0,44	0,66	0,84	
RNA <sub>1a</sub>	3	6	53,2	0,56	0,58	0,76	0,74	
RNA <sub>2</sub>	5	10	68	0,31	0,41	0,64	0,86	
RNA <sub>2a</sub>	4	4	76,8	0,33	0,42	0,65	0,84	
RNA <sub>3</sub>	5	10	69,4	0,28	0,38	0,62	0,87	
RNA <sub>3a</sub>	4	4	66	0,42	0,50	0,71	0,80	
RNA <sub>4</sub>	1	2	23	1,87	1,10	1,04	-1,16*10 <sup>10</sup>	
RNA <sub>5</sub>	2	4	40,8	0,44	0,50	0,70	0,79	
RNA <sub>6</sub>	3	6	59,2	0,38	0,45	0,67	0,82	
<u>RNA<sub>7</sub></u>	<u>6</u>	<u>12</u>	<u>124,8</u>	<u>0,04</u>	<u>0,13</u>	<u>0,36</u>	<u>0,98</u>	
<u>RNA<sub>8</sub></u>	<u>7</u>	<u>14</u>	<u>92,4</u>	<u>0,16</u>	<u>0,27</u>	<u>0,52</u>	<u>0,92</u>	
RNA <sub>8a</sub>	6	12	66,8	0,40	0,49	0,70	0,81	
<u>RNA<sub>9</sub></u>	<u>8</u>	<u>16</u>	<u>296,4</u>	<u>0,01</u>	<u>0,08</u>	<u>0,28</u>	<u>0,99</u>	
RNA <sub>9a</sub>	7	14	49,4	0,26	0,40	0,63	0,88	
<u>RNA<sub>10</sub></u>	<u>4</u>	<u>8</u>	<u>139</u>	<u>0,03</u>	<u>0,12</u>	<u>0,34</u>	<u>0,99</u>	
RNA <sub>11</sub>	4	8	67	0,29	0,42	0,65	0,86	
RNA <sub>12</sub>	5	10	43,6	0,31	0,45	0,67	0,85	
RNA <sub>13</sub>	5	10	60,8	0,25	0,38	0,62	0,88	
<u>RNA<sub>14</sub></u>	<u>5</u>	<u>10</u>	<u>124,2</u>	<u>0,02</u>	<u>0,11</u>	<u>0,33</u>	<u>0,99</u>	

Fonte: A autora

Tabela 12 Resumo das métricas das configurações para o otimizador Adam

Configuração	Variáveis		Num. Neur. Camada Oculta	Num. Épocas (média)	MAE (média)	MSE (média)	RMSE (média)	R <sup>2</sup> (média)
	Entrada	4						
RNA <sub>1</sub>	4	8		110,8	0,34	0,43	0,66	0,84
RNA <sub>1,a</sub>	3	6		92,4	0,56	0,58	0,76	0,74
RNA <sub>2</sub>	5	10		99,4	0,31	0,40	0,64	0,85
RNA <sub>2,a</sub>	4	4		63,4	0,34	0,43	0,65	0,84
RNA <sub>3</sub>	5	10		82,2	0,27	0,38	0,61	0,87
RNA <sub>3,a</sub>	4	4		96,4	0,42	0,50	0,71	0,80
RNA <sub>4</sub>	1	2		29	1,01	0,75	0,86	0,53
RNA <sub>5</sub>	2	4		43,6	0,44	0,50	0,71	0,79
RNA <sub>6</sub>	3	6		82	0,38	0,46	0,68	0,82
<u>RNA<sub>7</sub></u>	<u>6</u>	<u>12</u>		<u>82,6</u>	<u>0,03</u>	<u>0,12</u>	<u>0,35</u>	<u>0,98</u>
<u>RNA<sub>8</sub></u>	<u>7</u>	<u>14</u>		<u>105,8</u>	<u>0,15</u>	<u>0,26</u>	<u>0,51</u>	<u>0,93</u>
RNA <sub>8,a</sub>	6	12		82,8	0,40	0,49	0,70	0,81
<u>RNA<sub>9</sub></u>	<u>8</u>	<u>16</u>		<u>102,6</u>	<u>0,01</u>	<u>0,08</u>	<u>0,27</u>	<u>0,99</u>
RNA <sub>9,a</sub>	7	14		59,2	0,25	0,39	0,63	0,88
<u>RNA<sub>10</sub></u>	<u>4</u>	<u>8</u>		<u>70,4</u>	<u>0,02</u>	<u>0,11</u>	<u>0,34</u>	<u>0,99</u>
RNA <sub>11</sub>	4	8		97,4	0,28	0,42	0,64	0,87
RNA <sub>12</sub>	5	10		50,4	0,31	0,44	0,66	0,85
RNA <sub>13</sub>	5	10		94,2	0,24	0,38	0,62	0,89
<u>RNA<sub>14</sub></u>	<u>5</u>	<u>10</u>		<u>80,4</u>	<u>0,02</u>	<u>0,11</u>	<u>0,33</u>	<u>0,99</u>

Fonte: A autora

## 5.2 Análise das configurações testadas

### 5.2.1 Configurações 1 a 3

Essas configurações foram escolhidas por considerarem apenas as variáveis meteorológicas necessárias para estimar o valor da radiação, sendo esta uma variável importante na estimativa da  $ET_o$  pelo método de Penman-Monteith, conforme destacado por Lima *et al.* (2019). Neste caso, a análise compreendeu as seis configurações em relação aos dois otimizadores, Adam e SGD. A Tabela 13 ilustra os doze testes, ordenados do menor para o maior erro, considerando as métricas definidas na Seção 4.6.1.

Na avaliação comparativa dessas seis configurações treinadas com os otimizadores Adam e SGD (totalizando doze testes), observou-se um desempenho superior nas configurações que incorporaram o Dia Juliano (DiaJ) como variável de entrada, em comparação com aquelas que não o fizeram (configurações identificadas com 'a' após o número). O Dia Juliano reflete a posição do dia no calendário, indicando a variação sazonal, que está relacionada à posição do Sol no céu, à quantidade de radiação solar recebida e a fatores climáticos sazonais, sendo uma variável muito útil para o cálculo da  $ET_o$ .

Nestas configurações não houve uma distinção significativa entre os otimizadores utilizados. O coeficiente de determinação ( $R^2$ ), indicador da adequação das previsões aos dados reais, variou entre 0,87 (RNA<sub>3</sub>, a melhor) e 0,74 (RNA<sub>1a</sub>, a pior). Da mesma forma, o Erro Médio Absoluto (MAE) oscilou de 0,27 mm/dia (melhor) a 0,56 mm/dia (pior). Importante notar que nenhuma das configurações se posicionou entre as cinco melhores de cada otimizador, e a configuração RNA<sub>1a</sub> figurou entre as piores, tanto com o otimizador SGD quanto com o Adam.

Tabela 13 Resultados da 1ª execução para as configurações 1 a 3

Configuração	Otimizador	Variáveis		Num. Neur. Camada Oculta	Num. Épocas (média)	MAE (média)	MSE (média)	RMSE (média)	R <sup>2</sup> (média)
		Entrada	10						
RNA <sub>3</sub>	Adam	5	10	10	82,2	0,27	0,38	0,61	0,87
RNA <sub>3</sub>	SGD	5	10	10	69,4	0,28	0,38	0,62	0,87
RNA <sub>2</sub>	SGD	5	10	10	68	0,31	0,41	0,64	0,86
RNA <sub>2</sub>	Adam	5	10	10	99,4	0,31	0,40	0,64	0,85
RNA <sub>2a</sub>	SGD	4	4	4	76,8	0,33	0,42	0,65	0,84
RNA <sub>2a</sub>	Adam	4	4	4	63,4	0,34	0,43	0,65	0,84
RNA <sub>1</sub>	SGD	4	8	8	53,2	0,35	0,44	0,66	0,84
RNA <sub>1</sub>	Adam	4	8	8	110,8	0,34	0,43	0,66	0,84
RNA <sub>3a</sub>	Adam	4	4	4	96,4	0,42	0,50	0,71	0,80
RNA <sub>3a</sub>	SGD	4	4	4	66	0,42	0,50	0,71	0,80
RNA <sub>1a</sub>	SGD	3	6	6	53,2	0,56	0,58	0,76	0,74
RNA <sub>1a</sub>	Adam	3	6	6	92,4	0,56	0,58	0,76	0,74

Fonte: A autora

## 5.2.2 Configurações 4 a 7

As configurações RNA<sub>4</sub> a RNA<sub>7</sub> foram inspiradas no estudo de Sowmya, Kumar e Ambat (2020), que examinou os dados meteorológicos de duas estações. Utilizando as métricas RMSE e R<sup>2</sup>, foram comparados os resultados das quatro variantes do modelo base (DnnV1, DnnV2, DnnV3 e DnnV6, referenciados como T1a e T1b para as estações Santa Monika e Oakville, respectivamente) com as quatro configurações da RNA desenvolvida (referenciadas como RNA<sub>4</sub> a RNA<sub>7</sub>, com as variantes C<sub>1</sub> e C<sub>2</sub> para os otimizadores SGD e Adam). Vale notar uma divergência nas variáveis de entrada, já que o trabalho base incorporou a radiação solar, enquanto o presente estudo utilizou a radiação estimada.

A Tabela 14 apresenta um resumo dos resultados, destacando as diferenças e as semelhanças nas performances dessas configurações. Importante ressaltar que uma investigação futura poderá explorar o impacto da utilização da radiação da estação nos resultados.

Tabela 14 Comparação das configurações 4 a 7 com o trabalho de Sowmya, Kumar e Ambat (2020)

<b>Métrica</b>	<b>DnnV1 (T1a)</b>	<b>DnnV1 (T1b)</b>	<b>RNA<sub>4</sub> (C<sub>1</sub>)</b>	<b>RNA<sub>4</sub> (C<sub>2</sub>)</b>
RMSE	0,46 mm/dia	0,61 mm/dia	1,04 mm/dia	0,86 mm/dia
R <sup>2</sup>	0,9	0,91	-1,16*10 <sup>10</sup>	0,53
<b>Métrica</b>	<b>DnnV2 (T1a)</b>	<b>DnnV2 (T1b)</b>	<b>RNA<sub>5</sub> (C<sub>1</sub>)</b>	<b>RNA<sub>5</sub> (C<sub>2</sub>)</b>
RMSE	0,36 mm/dia	0,52 mm/dia	0,70 mm/dia	0,71 mm/dia
R <sup>2</sup>	0,94	0,94	0,79	0,79
<b>Métrica</b>	<b>DnnV3 (T1a)</b>	<b>DnnV3 (T1b)</b>	<b>RNA<sub>6</sub> (C<sub>1</sub>)</b>	<b>RNA<sub>6</sub> (C<sub>2</sub>)</b>
RMSE	0,36 mm/dia	0,51 mm/dia	0,67 mm/dia	0,68 mm/dia
R <sup>2</sup>	0,94	0,94	0,82	0,82
<b>Métrica</b>	<b>DnnV4 (T1a)</b>	<b>DnnV4 (T1b)</b>	<b>RNA<sub>7</sub> (C<sub>1</sub>)</b>	<b>RNA<sub>7</sub> (C<sub>2</sub>)</b>
RMSE	0,3 mm/dia	0,43 mm/dia	0,36 mm/dia	0,35 mm/dia
R <sup>2</sup>	0,96	0,96	0,98	0,98

Fonte: A autora

Comparando as arquiteturas utilizadas, destaca-se que o trabalho de Sowmya, Kumar e Ambat (2020) adotou uma abordagem com quatro camadas ocultas, cada uma caracterizada

por uma disposição específica de neurônios ocultos. Para DnnV1, DnnV2, DnnV3 e DnnV6, os números de neurônios nas camadas ocultas variaram, e o número de épocas diferiu para cada variante, além de utilizar a ReLU como função de ativação. Em contrapartida, nas configurações 4 a 7 do presente estudo, uma abordagem mais simplificada foi adotada, com apenas uma camada oculta contendo o dobro do número de neurônios de entrada. A função de ativação ReLU também foi empregada. O número de épocas variou conforme a configuração e o otimizador, situando-se entre 23 e 125, em média, evidenciando a adaptação da rede às características específicas de cada configuração. Essa simplificação visou equilibrar eficiência e desempenho, explorando uma estrutura mais enxuta em comparação com o trabalho de Sowmya, Kumar e Ambat (2020).

Comparando os valores de cada configuração mostrados na Tabela 14, é possível verificar que a RNA<sub>7</sub>, com os dois otimizadores, obteve um RMSE de 0,35 e 0,36 mm/dia, enquanto a DnnV6 obteve um RMSE de 0,3 e 0,43 mm/dia, sugerindo uma maior precisão na capacidade das configurações da RNA em prever a  $ET_o$ . Essa configuração (RNA<sub>7</sub>) foi a 4ª melhor, com ambos otimizadores, entre as 10 melhores escolhidas para os testes com os dados do produtor.

Ao analisar o coeficiente de determinação ( $R^2$ ) vemos que, embora o modelo base apresente resultados expressivos, as configurações da RNA conseguem manter níveis competitivos de ajuste aos dados. Essa análise aponta para a eficácia das configurações mais simplificadas da RNA em alcançar desempenho comparável ao modelo base, evidenciando uma potencial eficiência na modelagem da  $ET_o$  com arquiteturas mais enxutas.

### 5.2.3 Configurações 8 e 9

Como descrito na Seção 4.4.3, essas configurações foram baseadas no trabalho de Ferraz (2014), que testou 6 configurações para a entrada das RNAs, como descrito na Seção 3.4.1: ANN1 a ANN6. O autor usou uma arquitetura MLP, com uma camada oculta e uma de saída. O número de neurônios na camada oculta variou entre 45 e 50 neurônios. A melhor configuração obtida por ele foi a ANN4, com 10 neurônios de entrada, 50 na camada oculta e com a função Sigmoide Logística como função de ativação.

As configurações RNA<sub>8</sub>, RNA<sub>8a</sub>, RNA<sub>9</sub> e RNA<sub>9a</sub> do presente estudo compartilham as mesmas variáveis de entrada que as ANN1 e ANN2 do trabalho de Ferraz (2014). No trabalho base, a ANN4, que utiliza 10 variáveis de entrada (Lat, Long, Alt, DiaJ, Tmin, Tmax, Ur, VV,

Ro<sup>1</sup>, N<sup>2</sup>), mostrou ser a configuração mais eficaz.

A Tabela 15 mostra a ANN1 em comparação com as configurações RNA<sub>8</sub> e RNA<sub>8a</sub>, a ANN2 com a RNA<sub>9</sub> e RNA<sub>9a</sub>, e destaca a ANN4 como a melhor configuração do trabalho base. Vale notar que, entre as configurações do presente estudo, a configuração RNA<sub>8</sub> alcançou a quinta melhor performance, enquanto a configuração RNA<sub>9</sub> se destacou como a melhor. Novamente fica evidente que o Dia Juliano melhora o desempenho da configuração que utiliza essa variável.

Tabela 15 Comparação das configurações 8 e 9 com o trabalho de Ferraz (2014)

<b>Métrica</b>	<b>ANN1</b>	<b>RNA<sub>8</sub> (C<sub>1</sub>)</b>	<b>RNA<sub>8</sub> (C<sub>2</sub>)</b>	<b>ANN2</b>	<b>RNA<sub>9</sub> (C<sub>1</sub>)</b>	<b>RNA<sub>9</sub> (C<sub>2</sub>)</b>	<b>ANN4</b>
RMSE (mm/dia)	0,59	0,52	0,51	0,53	0,28	0,27	0,52
R <sup>2</sup>	0,86	0,92	0,93	0,88	0,99	0,99	0,89

Fonte: A autora

Os resultados nas métricas RMSE e R<sup>2</sup> mostram um desempenho superior para as configurações propostas em comparação com as configurações do trabalho base. Ao analisar a métrica RMSE, a configuração RNA<sub>8</sub> mostra valores notavelmente inferiores (0,52 e 0,51, respectivamente para cada otimizador) em comparação com a configuração ANN1 do trabalho base (0,59). Da mesma forma, na métrica R<sup>2</sup>, ambas as configurações RNA<sub>8</sub> exibem uma melhoria substancial (0,92 e 0,93, respectivamente) em relação à configuração ANN1 (0,86).

Avaliando a configuração RNA<sub>9</sub> (em ambos otimizadores), os resultados permanecem favoráveis, mostrando um desempenho muito maior, com valores de RMSE de 0,28 nas duas variações, contrastando com o valor de 0,53 da configuração ANN2 do trabalho base. Na métrica R<sup>2</sup>, as configurações RNA<sub>9</sub> alcançam 0,99, indicando a adaptação dos modelos propostos aos dados reais.

Comparando com a melhor configuração do trabalho base, ANN4, que utiliza 10 variáveis de entrada, as configurações RNA<sub>8</sub> (com 7 variáveis) e RNA<sub>9</sub> (com 8 variáveis) mostraram um desempenho superior, ressaltando a eficácia e robustez das configurações propostas para aprimorar a precisão das estimativas da *ET<sub>o</sub>*.

<sup>1</sup> Radiação solar global incidente diária.

<sup>2</sup> Duração de brilho solar diário.

#### 5.2.4 Configurações 10 e 11

Como descrito na Seção 4.4.4, as configurações  $RNA_{10}$  e  $RNA_{11}$  foram comparadas com os estudos conduzidos por Santos (2020) e Walls *et al.* (2020) no âmbito da estimação da  $ET_o$ . Santos (2020) empregou uma rede neural do tipo MLP, composta por duas camadas ocultas (9 neurônios na primeira camada e 3 na segunda), utilizando a função de ativação logística. Já Walls *et al.* (2020) explorou diversas arquiteturas, destacando que a mais eficaz foi uma MLP com uma camada oculta contendo 4 neurônios e a função de ativação ReLU.

No estudo de Santos (2020), a melhor configuração para a cidade de Mendes-RJ, M6, incorporou as variáveis de temperatura média, umidade relativa, velocidade do vento e radiação solar como entradas para o modelo; para a cidade de Teresópolis, o melhor modelo foi o M1, que utilizou as variáveis radiação solar, temperatura mínima, máxima e média, umidade relativa e velocidade do vento. Enquanto isso, Walls *et al.* (2020) identificou que seu modelo mais bem-sucedido, o Modelo 2, utilizou como entradas a radiação líquida, temperatura do ar, fluxo de calor do solo, velocidade do vento e umidade relativa. Para as configurações  $RNA_{10}$  e  $RNA_{11}$ , propostas neste estudo, as variáveis de entrada escolhidas foram temperatura média, umidade relativa, velocidade do vento e radiação solar. Vale destacar que, devido às discrepâncias nos tipos de radiação entre os estudos de referência e os dados disponíveis, a configuração  $RNA_{10}$  empregou a radiação estimada, enquanto a configuração  $RNA_{11}$  utilizou a radiação da estação meteorológica. Essa escolha estratégica visou adaptar o modelo às condições específicas do presente estudo, considerando as limitações e peculiaridades dos dados disponíveis.

A Tabela 16 mostra as métricas dos modelos propostos e dos modelos de base. Os resultados da análise indicam que o modelo proposto,  $RNA_{10}$ , apresentou desempenho superior nas duas configurações ( $C_1$  e  $C_2$ ), classificando-se como o terceiro melhor entre os modelos testados. Esta conclusão é respaldada pelos valores comparativos de MAE e  $R^2$ , os quais revelam que este modelo proposto apresentou erros significativamente menores e uma capacidade de explicação da variabilidade dos dados comparável aos modelos  $M_1$  (6 variáveis de entrada) e  $M_6$  (4 variáveis de entrada) de Santos (2020) e do Modelo 2 de Walls *et al.* (2020) (5 variáveis de entrada), utilizando apenas 4 variáveis de entrada e uma arquitetura mais simplificada.

Em contraste, a performance do modelo proposto  $RNA_{11}$  nas duas configurações ( $C_1$  e  $C_2$ ) não alcançou os mesmos patamares de excelência. Em ambos os casos, os resultados de MAE e  $R^2$  indicam um desempenho inferior em comparação com os modelos de Santos (2020)

e de Walls *et al.* (2020), sugerindo que a  $RNA_{11}$  pode não ser tão adequada para a tarefa de estimativa de  $ET_o$ , ao menos nas condições e configurações avaliadas.

Tabela 16 Comparação das configurações 10 e 11 com os trabalhos de Santos (2020) ( $M_1$  e  $M_6$ ) e Walls *et al.* (2020) (Mod. 2)

<b>Métrica</b>	<b>M6</b>	<b>M1</b>	<b>Mod. 2</b>	<b><math>RNA_{10}</math> (<math>C_1</math>)</b>	<b><math>RNA_{10}</math> (<math>C_2</math>)</b>	<b><math>RNA_{11}</math> (<math>C_1</math>)</b>	<b><math>RNA_{11}</math> (<math>C_2</math>)</b>
MAE (mm/dia)	0,12	0,08	0,18	0,03	0,02	0,29	0,28
RMSE (mm/dia)	0,16	0,1	0,39	0,34	0,34	0,65	0,64
$R^2$	0,99	0,99	0,99	0,99	0,99	0,86	0,87

Fonte: A autora

### 5.2.5 Configuração 12

A configuração  $RNA_{12}$ , descrita na Seção 4.4.5, inspirada no trabalho de Silva (2002), apresentou desafios substanciais durante a análise comparativa. Uma compreensão mais aprofundada revelou que as métricas de avaliação utilizadas por Silva (2002) divergem significativamente das empregadas neste estudo, tornando a comparação direta impraticável. Além disso, é importante observar que a abordagem temporal para previsão da  $ET_o$  também difere substancialmente, com Silva (2002) fornecendo previsões para múltiplos dias no futuro, enquanto o presente estudo concentrou-se na estimativa para o dia seguinte. Devido a essas disparidades fundamentais, a configuração  $RNA_{12}$  foi excluída da análise detalhada, uma vez que as condições de avaliação e as métricas adotadas não permitiram uma comparação justa e significativa com as demais configurações avaliadas.

### 5.2.6 Configurações 13 e 14

Essas configurações foram descritas na Seção 4.4.6 e foram avaliadas em relação ao trabalho de Arca, Beniscasa e Vincenzi (2001), que emprega um conjunto abrangente de variáveis meteorológicas, incluindo temperatura do ar, umidade relativa, radiação solar global, radiação difusa, radiação líquida, velocidade e direção do vento, intensidade da chuva e pressão atmosférica. As configurações propostas utilizaram temperatura média, umidade relativa, velocidade do vento, pressão atmosférica e radiação (da estação na  $RNA_{13}$ , e a estimada em  $RNA_{14}$ ) como variáveis de entrada.

Os autores conduziram uma série de testes, explorando combinações dessas variáveis, além de incorporar a  $ET_o$  calculada por outros métodos como uma entrada adicional. A arquitetura de RNAs no trabalho base foi caracterizada por um número de neurônios de entrada equivalente ao número de variáveis de entrada, uma camada oculta com neurônios variando de 5 a 8, e um neurônio de saída. Contudo, a função de ativação adotada não foi explicitamente mencionada no estudo de referência.

A melhor combinação identificada em Arca, Beniscasa e Vincenzi (2001), denominada ANN 1, incorpora a  $ET_o$  calculada pelo método de Hargreaves, a pressão atmosférica e a velocidade do vento como variáveis de entrada. A Tabela 17 mostra as métricas das configurações testadas neste trabalho em comparação ao trabalho base.

Tabela 17 Comparação das configurações 13 e 14 com o trabalho de Arca, Beniscasa e Vincenzi (2001)

Métrica	ANN 1	RNA <sub>13</sub> (C <sub>1</sub> )	RNA <sub>13</sub> (C <sub>2</sub> )	RNA <sub>14</sub> (C <sub>1</sub> )	RNA <sub>14</sub> (C <sub>2</sub> )
MAE (mm/dia)	0,19	0,25	0,24	0,02	0,02
RMSE (mm/dia)	0,27	0,62	0,62	0,33	0,33
R <sup>2</sup>	0,97	0,88	0,89	0,99	0,99

Fonte: A autora

A RNA<sub>14</sub> destacou-se como a segunda melhor configuração avaliada neste estudo, evidenciando um desempenho significativamente superior em termos de MAE e R<sup>2</sup> em comparação com a ANN 1. A eficácia da RNA<sub>14</sub> é ainda mais interessante considerando que ela utiliza a radiação estimada. Em contraste, a RNA<sub>13</sub>, que faz uso da radiação da estação, apresentou um desempenho inferior à ANN 1, embora tenha alcançado a sexta posição entre todas as configurações testadas neste trabalho. Esses resultados ressaltam a importância da escolha apropriada das variáveis de entrada, evidenciando que a qualidade dos dados de entrada pode ter um impacto substancial no desempenho das RNAs no contexto da estimativa da  $ET_o$ .

### 5.3 Análise das 5 melhores configurações

Após completar as fases de treinamento e validação, cada uma realizada cinco vezes, as cinco configurações de RNA mais bem-sucedidas, considerando os dois otimizadores, SGD e Adam, foram submetidas a um conjunto de teste composto por 4.478 amostras, equivalentes a 20% do total dos dados disponíveis no conjunto de dados do INMET. Essa seleção visava garantir que as configurações escolhidas apresentassem robustez e generalização diante de

dados não utilizados durante o treinamento. A Tabela 18 apresenta as métricas resultantes desses testes, destacando não apenas o desempenho em termos de métricas de avaliação, como MAE e RMSE, mas também detalhes cruciais como o otimizador empregado, o número de neurônios de entrada (correspondente ao número de variáveis de entrada), a configuração de neurônios na camada oculta e o número de épocas utilizado durante o treinamento. Essas informações detalhadas fornecem uma visão abrangente do desempenho das RNA sob diferentes configurações, proporcionando uma base sólida para análises comparativas e futuras decisões de implementação.

Tabela 18 Sumarização das 5 melhores configurações

Config.	Otim.	Variáveis Entrada	Neur. Camada Oculta	Num. Épocas	MAE	MSE	RMSE	R <sup>2</sup>
RNA <sub>9</sub>	Adam	8	16	111	0,01	0,06	0,24	1,00
RNA <sub>9</sub>	SGD	8	16	196	0,01	0,08	0,28	0,99
RNA <sub>14</sub>	SGD	5	10	211	0,02	0,10	0,32	0,99
RNA <sub>14</sub>	Adam	5	10	90	0,02	0,10	0,32	0,99
RNA <sub>10</sub>	SGD	4	8	80	0,03	0,12	0,35	0,99
RNA <sub>10</sub>	Adam	4	8	96	0,03	0,13	0,35	0,99
RNA <sub>7</sub>	Adam	6	12	107	0,03	0,12	0,34	0,99
RNA <sub>7</sub>	SGD	6	12	98	0,04	0,13	0,36	0,98
RNA <sub>8</sub>	SGD	7	14	94	0,15	0,26	0,51	0,93
RNA <sub>8</sub>	Adam	7	14	75	0,16	0,26	0,51	0,93

Fonte: A autora

Os 10 testes foram organizados em ordem decrescente de desempenho, da melhor (configuração RNA<sub>9</sub>, com o otimizador Adam (C<sub>2</sub>)) para a pior (configuração RNA<sub>8</sub>, também com o otimizador Adam (C<sub>2</sub>)). Todas as configurações alcançaram um ótimo desempenho, com valores de R<sup>2</sup> próximos a 1, indicando uma boa capacidade de ajuste aos dados. Mas duas configurações destacam-se, a RNA<sub>9</sub> (Adam) e RNA<sub>14</sub> (ambos os otimizadores), mostrando métricas bastante promissoras.

A melhor arquitetura foi a configuração com 8 variáveis de entrada (RNA<sub>9</sub>), o que já era esperado, visto que quanto maior o número de variáveis disponíveis, maior será a precisão da predição (Ferraz, 2014). Avaliando as métricas de erro desta arquitetura, podemos verificar que os valores das 4 métricas são consideradas muito boas. O MAE nos mostra que a dife-

rença entre o valor predito e o valor real é de apenas 0,01 mm/dia com os dois otimizadores, enquanto o  $R^2$  nos mostra que os valores preditos se ajustam quase que perfeitamente aos valores reais. Essa informação poderia indicar que a RNA “decorou” (*overfitting*) os dados, mas quando a RNA é confrontada com dados não avaliados durante o treinamento, validação e testes (dados do produtor rural, que serão descritos na Seção 5.4), seu  $R^2$  continua com o valor de 0,99, mostrando que a configuração RNA<sub>9</sub> se ajustou muito bem.

Quando consideramos arquiteturas com um menor número de variáveis de entrada, a melhor foi a configuração com 4 variáveis (RNA<sub>10</sub>), com o MAE mostrando que a diferença entre o valor predito e o valor real é de apenas 0,03 mm/dia com os dois otimizadores, enquanto o  $R^2$  nos mostra que os valores preditos se ajustam quase que perfeitamente aos valores reais. Para garantir que não houve *overfitting*, esta arquitetura também foi confrontada com os dados do produtor (Seção 5.4) e seu  $R^2$  continua com um valor alto, de 0,98.

A Tabela 19 mostra as variáveis de entrada utilizadas em cada configuração<sup>3</sup>. Ao analisar as variáveis comuns às 5 melhores configurações, é possível perceber que a umidade relativa (Ur) aparece em todas as configurações, sendo acompanhada por suas variações U<sub>max</sub> e U<sub>min</sub> na configuração RNA<sub>7</sub>. Similarmente, as variáveis relacionadas à temperatura (T, T<sub>min</sub>, T<sub>max</sub>) estão presentes em todas as configurações, indicando sua importância na modelagem. A variável de velocidade do vento (Vento) está presente em quatro das cinco configurações, sendo ausente apenas na configuração RNA<sub>8</sub>, o que pode indicar a importância desta variável para a previsão da  $ET_o$ , visto que é a única que não utiliza essa variável e foi a que obteve menor desempenho.

Tabela 19 Variáveis utilizadas pelas 5 melhores configurações

Configuração	Var. Entrada	Variáveis
RNA <sub>9</sub>	8	DiaJ, T <sub>max</sub> , T <sub>min</sub> , Ur, Vento, Lat, Long, Alt
RNA <sub>14</sub>	5	T, Ur, Pressao, Vento, Rs
RNA <sub>10</sub>	4	T, Ur, Vento, Rs
RNA <sub>7</sub>	6	T <sub>max</sub> , T <sub>min</sub> , U <sub>max</sub> , U <sub>min</sub> , Vento, Rs
RNA <sub>8</sub>	7	DiaJ, T <sub>max</sub> , T <sub>min</sub> , Ur, Lat, Long, Alt

Fonte: A autora

Essa análise sugere que as variáveis fundamentais para o desempenho do modelo

<sup>3</sup> A descrição de cada variável utilizada, juntamente com o nome de cada uma delas no código implementado pode ser vista na Tabela 6.

estão relacionadas à umidade relativa e à temperatura. A presença ou ausência de outras variáveis, como a velocidade do vento, parece ter impacto nas diferentes configurações.

Agora, munidos dessas cinco melhores configurações, prosseguimos com o próximo passo da análise, que consiste em testá-las com os dados do produtor. Essa etapa foi importante para avaliar se as configurações selecionadas conseguiram generalizar eficazmente as previsões da  $ET_o$  para um conjunto de dados até então não apresentado às RNAs. Esse teste realista permitiu uma avaliação mais completa da capacidade de adaptação e precisão das redes neurais artificiais desenvolvidas, fornecendo informações valiosas para aplicações práticas.

#### 5.4 Análise das configurações com os dados do produtor

Na análise dos testes com os dados do produtor (Seção 4.1.2), das cinco configurações ( $RNA_7$ ,  $RNA_8$ ,  $RNA_9$ ,  $RNA_{10}$  e  $RNA_{14}$ ), a configuração  $RNA_7$  não pôde ser avaliada devido ao uso das variáveis  $U_{max}$  e  $U_{min}$ , não disponíveis no conjunto de dados do produtor.

Os testes foram conduzidos com as 1.144 amostras disponíveis. A Tabela 20 apresenta um resumo das métricas obtidas nas análises das configurações de RNAs utilizando diferentes otimizadores. A ordem está classificada da melhor para a pior configuração, considerando o  $R^2$  e o MAE. Podemos verificar que o otimizador Adam saiu-se melhor em todas as configurações. Assim como nos dados INMET, a  $RNA_9$  manteve-se como a melhor, e a  $RNA_8$ , como a menos eficaz.

Tabela 20 Sumarização das configurações com os dados do produtor

Configuração	Otimizador	Variáveis Entrada	MAE	MSE	RMSE	$R^2$
$RNA_9$	Adam	8	0,01	0,06	0,25	1,00
$RNA_9$	SGD	8	0,01	0,08	0,29	0,99
$RNA_{14}$	Adam	5	0,01	0,08	0,29	0,99
$RNA_{14}$	SGD	5	0,02	0,10	0,32	0,99
$RNA_{10}$	Adam	4	0,02	0,11	0,33	0,99
$RNA_{10}$	SGD	4	0,02	0,12	0,35	0,98
$RNA_8$	Adam	7	0,14	0,27	0,52	0,91
$RNA_8$	SGD	7	0,22	0,38	0,61	0,85

Fonte: A autora

A  $RNA_9$ , com otimizador Adam ( $C_2$ ), obteve o melhor desempenho, apresentando MAE

de 0,01 mm/dia e o  $R^2$  de 1,00, mostrando que os valores preditos se ajustam muito bem aos valores reais. Na análise realizada na Seção 5.3, destaca-se o  $R^2$  observado na configuração RNA<sub>9</sub>, que atingiu o valor máximo de 1. Esse resultado inicialmente levantou a possibilidade de *overfitting*, indicando que a RNA poderia ter memorizado os dados de treinamento. Contudo, ao realizar os testes com os dados do produtor, que não foram utilizados durante o treinamento, a RNA<sub>9</sub> mostrou um ótimo desempenho na previsão da  $ET_o$ . Isso sugere que o modelo foi capaz de generalizar eficazmente novos conjuntos de dados, reforçando sua capacidade de adaptação e precisão na estimativa da  $ET_o$ . Esse comportamento ressalta a robustez da configuração RNA<sub>9</sub>, sendo um indicativo positivo de sua aplicabilidade em cenários práticos.

Quando consideramos arquiteturas com um menor número de variáveis de entrada, a melhor foi novamente a configuração com 4 variáveis (RNA<sub>10</sub>), com o MAE mostrando que a diferença entre o valor predito e o valor real é de apenas 0,02 mm/dia com os dois otimizadores, enquanto o  $R^2$  nos mostra que os valores preditos se ajustam quase que perfeitamente aos valores reais.

A fim de ilustrar as diferenças entre os valores reais e os previstos, foram elaboradas duas tabelas destacando as menores e maiores diferenças entre esses valores, considerando os dois otimizadores. Estas tabelas incluem as variáveis de entrada, os valores reais (calculados pelo método de Penman-Monteith) e os valores previstos para os otimizadores SGD e Adam.

Para a RNA<sub>9</sub> (Tabela 21), identificada como a configuração mais eficiente, as menores diferenças variam de 0,000 a 0,001 mm/dia (para o otimizador SGD - C<sub>1</sub>) e de 0,002 a 0,112 mm/dia (para o otimizador Adam - C<sub>2</sub>). As maiores diferenças oscilam entre 0,270 a 0,548 mm/dia (para C<sub>1</sub>) e 0,073 a 0,329 mm/dia (para C<sub>2</sub>).

No caso da RNA<sub>10</sub> (Tabela 22), considerada a mais eficaz quando se leva em conta o menor número de variáveis, as menores diferenças variam de 0,000 a 0,002 mm/dia (para C<sub>1</sub>) e de 0,001 a 0,162 mm/dia (para C<sub>2</sub>). Já as maiores diferenças situam-se entre 0,392 a 0,477 mm/dia (para C<sub>1</sub>) e 0,175 a 0,553 mm/dia (para C<sub>2</sub>). Esses dados mostram uma visão detalhada das capacidades das RNAs em ajustar-se aos valores reais, diferenciando-se conforme o otimizador utilizado.

Na próxima seção é realizada uma avaliação das duas configurações mais promissoras, RNA<sub>9</sub> e RNA<sub>10</sub>. Esta avaliação foi realizada empregando as quatro métricas utilizadas na avaliação das configurações (MAE, MSE, RMSE e  $R^2$ ), além do Teste Kolmogorov-Smirnov (KS) e o valor-p associado. O KS será empregado para examinar a aderência das previsões

ao comportamento estatístico dos dados reais, enquanto o valor-p fornecerá uma medida estatística da significância dessa aderência. Essa análise mais abrangente permitirá uma compreensão mais aprofundada da capacidade de generalização dessas RNAs em cenários práticos de previsão da  $ET_o$ , proporcionando uma visão mais completa sobre o desempenho e a confiabilidade desses modelos.

Tabela 21 Menores e maiores diferenças entre os valores reais e preditos para a configuração RNA<sub>9</sub>

Lat	Long	Alt	DiaJ	Tmin	Tmax	Ur	Vento	$ET_o$	$ET_o (C_1)$	$ET_o (C_2)$
-25,14	-53,46	524	140	14,20	28,20	60,09	1,47	2,69	2,69	2,65
-25,14	-53,46	524	95	7,70	27,80	71,11	0,66	3,13	3,13	3,13
-25,14	-53,46	524	257	16,20	33,40	71,89	1,04	3,90	3,90	3,88
-25,14	-53,46	524	302	21,10	32,90	67,89	2,67	5,15	5,15	5,21
-25,14	-53,46	524	229	17,40	30,10	48,65	2,08	3,98	3,98	4,04
-25,14	-53,46	524	85	14,20	30,70	66,27	0,89	3,71	3,71	3,68
-25,14	-53,46	524	190	9,80	17,60	93,58	1,45	1,14	1,14	1,14
-25,14	-53,46	524	211	6,30	17,30	85,24	1,98	1,51	1,51	1,54
-25,14	-53,46	524	308	13,80	28,90	51,95	1,19	4,69	4,69	4,80
-25,14	-53,46	524	330	18,60	34,70	57,26	2,02	6,02	6,02	5,97
(...)										
-25,14	-53,46	524	133	12,60	19,30	94,27	0,76	1,35	1,08	1,43
-25,14	-53,46	524	284	16,40	35,50	50,66	1,89	5,69	5,42	5,53
-25,14	-53,46	524	203	6,10	28,60	55,77	0,66	2,16	2,43	2,37
-25,14	-53,46	524	122	14,20	28,60	84,04	0,63	2,39	2,66	2,49
-25,14	-53,46	524	91	18,40	30,50	77,46	0,87	3,26	3,53	3,34
-25,14	-53,46	524	302	9,90	30,90	65,10	1,04	4,78	4,49	4,62
-25,14	-53,46	524	285	11,40	30,80	60,96	1,00	4,45	4,15	4,32
-25,14	-53,46	524	276	17,40	18,40	97,08	0,95	1,17	1,47	1,45
-25,14	-53,46	524	340	19,20	21,30	97,90	1,04	1,81	2,32	2,13
-25,14	-53,46	524	350	18,70	21,10	95,51	1,07	1,95	2,49	2,27

Fonte: A autora

Tabela 22 Menores e maiores diferenças entre os valores reais e preditos para a configuração RNA<sub>10</sub>

<b>Rs</b>	<b>T</b>	<b>Ur</b>	<b>Vento</b>	<i>ET<sub>o</sub></i>	<i>ET<sub>o</sub> (C<sub>1</sub>)</i>	<i>ET<sub>o</sub> (C<sub>2</sub>)</i>
19,34	25,95	67,79	1,47	4,38	4,38	4,37
13,84	17,20	86,33	0,68	2,12	2,12	2,12
18,87	23,85	85,83	1,81	3,72	3,72	3,72
13,67	22,75	91,02	1,16	2,70	2,70	2,68
13,54	12,45	76,75	0,74	1,49	1,49	1,65
13,02	22,70	94,67	1,02	2,55	2,55	2,58
15,47	22,25	90,63	0,70	2,95	2,94	2,95
15,55	21,40	72,11	1,43	3,04	3,04	3,20
21,44	25,20	84,93	0,94	4,30	4,30	4,31
20,40	23,80	87,89	0,66	3,95	3,95	4,00
(...)						
11,45	18,10	88,82	2,55	1,66	2,05	1,84
14,28	19,10	82,78	0,98	2,09	2,49	2,41
11,36	18,85	91,29	1,85	1,66	2,06	1,89
16,14	17,95	75,68	0,44	2,08	2,49	2,36
9,71	24,05	92,22	2,13	1,76	2,18	2,11
15,21	30,35	41,76	4,30	7,32	6,90	7,12
12,05	15,55	91,54	2,21	1,43	1,86	1,74
12,45	18,35	85,42	1,50	1,77	2,22	2,01
18,49	18,35	73,19	1,36	2,75	3,22	3,30
21,14	23,05	74,30	0,87	3,68	4,16	4,12

Fonte: A autora

### 5.5 Análise das métricas e do teste Kolmogorov-Smirnov

Nesta seção, são apresentados e analisados os resultados obtidos pela avaliação das configurações RNA<sub>9</sub> e RNA<sub>10</sub> nos dados do produtor, considerando métricas essenciais como MAE, MSE, RMSE e R<sup>2</sup>. Além disso, a inclusão do teste de Kolmogorov-Smirnov (KS) e seus valores-p ampliam a compreensão sobre quão bem as distribuições dos dados observados se harmonizam com as previsões geradas pelas RNAs. Esses valores mostram o desempenho preditivo dessas RNAs e são utilizados em uma análise aprofundada das métricas específicas

e do teste KS, fornecendo assim uma compreensão mais completa da adequação do modelo às complexidades dos dados de evapotranspiração ( $ET_o$ ). A Tabela 23 apresenta os resultados destas métricas, que serão analisados para cada configuração.

Tabela 23 Comparação das 4 métricas utilizadas com o valor do KS e Valor-p

<b>Configuração</b>	<b>Otimizador</b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>	<b>Teste KS</b>	<b>Valor-p</b>
RNA <sub>9</sub>	SGD	0,01	0,08	0,29	0,99	0,04	0,42
RNA <sub>9</sub>	Adam	0,01	0,06	0,25	1,00	0,02	0,93
RNA <sub>10</sub>	SGD	0,02	0,12	0,35	0,98	0,04	0,24
RNA <sub>10</sub>	Adam	0,02	0,11	0,33	0,99	0,04	0,31

Fonte: A autora

A análise da configuração RNA<sub>9</sub> com ambos otimizadores (SGD e Adam) revela resultados vantajosos em relação à estimativa da  $ET_o$  para o estado do Paraná. Ambas as variantes apresentaram métricas de desempenho muito boas, com valores baixos para MAE (0,01 mm/dia), MSE (0,06 a 0,08), RMSE (0,25 a 0,29 mm/dia) e alto para R<sup>2</sup> (0,99 a 1,00). Essas métricas sugerem uma correspondência robusta entre as previsões da RNA<sub>9</sub> e os valores reais de  $ET_o$ .

Ao avaliar o teste KS, observa-se que as estatísticas de KS são relativamente baixas (0,02 a 0,04), indicando uma discrepância limitada entre a função de distribuição acumulativa empírica (ECDF) das previsões da RNA e a ECDF dos dados reais. Com base na interpretação dos valores-p obtidos no teste KS, que são superiores a 0,05, e considerando a análise realizada em uma amostra de 1.144 observações do produtor, é estatisticamente indicativo que o modelo neural representado pela configuração RNA<sub>9</sub> possui uma concordância aceitável com a distribuição dos dados reais de  $ET_o$  para a região Oeste do estado do Paraná. Essa conclusão sugere que o modelo tem uma capacidade razoável de estimar a  $ET_o$  com um alto grau de confiabilidade estatística, considerando o conjunto de dados avaliado.

Os resultados da avaliação da configuração RNA<sub>10</sub> com os otimizadores SGD e Adam apresentam métricas consistentemente favoráveis, indicando um desempenho robusto da RNA na estimativa da  $ET_o$  para o estado do Paraná. Tanto o MAE quanto o MSE são baixos, evidenciando uma precisão satisfatória nas previsões. Além disso, as métricas RMSE e R<sup>2</sup> confirmam a qualidade do modelo na representação dos padrões de  $ET_o$ . A análise do teste KS mostra valores iguais a 0,04 para ambos os otimizadores, sugerindo discrepâncias limitadas entre a ECDF das previsões da RNA e a ECDF dos dados reais. O valor-p superior a 0,05, em con-

junto com as demais métricas, indica que a configuração RNA<sub>10</sub>, com suas quatro variáveis de entrada, apresenta uma capacidade plausível de representar com precisão e generalizar os padrões de  $ET_o$  na região estudada.

Esses resultados respaldam a eficácia do modelo neural artificial no contexto da estimativa da  $ET_o$ , proporcionando uma ferramenta valiosa para aplicações práticas relacionadas à gestão hídrica e agrícola na região. Dessa forma, considerando a eficiência estatística e a capacidade de generalização, ambas as configurações revelam-se como opções robustas para a estimativa de  $ET_o$  no contexto específico do Paraná, permitindo escolhas com base nas necessidades específicas do modelo e dos recursos disponíveis.

## **5.6 Divergências nas estimativas da $ET_o$ entre o método de Penman-Monteith e os dados observados**

Como descrito na Seção 4.1.1, as estações do INMET não possuíam o valor da  $ET_o$ . Por esse motivo, calculou-se esse valor para todas as amostras das estações utilizando o método de Penman-Monteith (Equação 1). Já a estação do produtor possuía o valor da  $ET_o$ , mas como discutido na Seção 4.1.2, não existiam informações sobre qual o método de estimativa da  $ET_o$  a estação utilizava. Para que a comparação fosse justa entre os dados do treinamento e os de validação, optou-se por calcular o valor da  $ET_o$  para todas as amostras da estação do produtor.

Após uma análise abrangente das configurações propostas, considerando os trabalhos relacionados e a escolha das melhores configurações para o teste com os dados do produtor, procedeu-se à validação do modelo. Durante esse processo, observou-se uma diferença significativa entre os valores de  $ET_o$  disponíveis na estação e os valores de  $ET_o$  calculados pelo método de Penman-Monteith. Essa divergência é destacada pelas métricas de desempenho, onde a diferença entre os dados observados na estação e os dados calculados apresentam os seguintes valores: MAE = 0,57 mm/dia, MSE = 0,49, RMSE = 0,70 mm/dia e  $R^2 = 0,76$ . Essas métricas indicam uma variação entre os dois conjuntos de dados, apontando para a possibilidade de a estação do produtor empregar um método de estimativa de  $ET_o$  diferente.

A análise prosseguiu ao comparar os resultados das RNAs treinadas com dados do método de Penman-Monteith, com os valores calculados pela mesma abordagem. As métricas obtidas para as melhores configurações das RNAs (Tabela 24), quando confrontadas com os dados da estação do produtor, espelham a disparidade observada entre os valores calculados e os da estação. Essa congruência nas métricas reforça hipótese de que a estação do pro-

ductor utiliza um método diferente do Penman-Monteith para estimar a  $ET_o$ , e isso influenciou diretamente a eficácia das RNAs em reproduzir os padrões observado.

Tabela 24 Métricas que mostram as divergências entre o valor da  $ET_o$  da estação do produtor e os valores calculados e preditos

<b>Tipo <math>ET_o</math></b>	<b>MAE</b>	<b>MSE</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
$ET_o$ calculada	0,57	0,49	0,70	0,76
RNA <sub>9</sub> , SGD	0,57	0,50	0,71	0,75
RNA <sub>9</sub> , Adam	0,58	0,51	0,72	0,74
RNA <sub>10</sub> , SGD	0,57	0,50	0,71	0,75
RNA <sub>10</sub> , Adam	0,56	0,49	0,70	0,75

Fonte: A autora

Importante ressaltar que, embora as métricas revelem diferenças consideráveis, essa disparidade não necessariamente desabona a análise realizada com as RNAs. O treinamento dessas redes foi conduzido com base nos valores calculados pelo método de Penman-Monteith, o que justifica a consistência nas métricas quando comparadas aos dados da estação do produtor. No entanto, é importante levar em consideração essa divergência na interpretação dos resultados, reconhecendo a possibilidade de diferentes métodos de cálculo de  $ET_o$  empregados entre as estações.

Este capítulo proporcionou uma análise detalhada das configurações das RNAs para a estimativa da  $ET_o$  no contexto do manejo de irrigação. As avaliações abrangeram métricas de desempenho, como MAE, MSE, RMSE e R<sup>2</sup>, além do teste de Kolmogorov-Smirnov (KS) e seus valores-p, proporcionando uma visão abrangente do desempenho das RNAs em diversos aspectos. A interpretação desses resultados foi importante para compreender a capacidade das RNAs em representar e prever padrões meteorológicos complexos, essenciais para a gestão eficaz da irrigação. Com base nessas análises, o próximo capítulo é dedicado às conclusões do trabalho, destacando os principais achados. Em seguida, são indicadas limitações e possíveis direções futuras. Esse desfecho visa consolidar as descobertas obtidas ao longo desta pesquisa e fornecer uma perspectiva mais abrangente sobre o papel das RNAs na estimativa da  $ET_o$  e seu impacto potencial no manejo sustentável da irrigação.

## 6 CONCLUSÕES

1. O modelo de previsão de evapotranspiração de referência ( $ET_o$ ) baseado em Redes Neurais Artificiais (RNAs) foi desenvolvido com sucesso, alcançando o objetivo de simplificar o cálculo da  $ET_o$  ao empregar um conjunto reduzido de variáveis climáticas.
2. A organização dos dados meteorológicos provenientes das estações do Instituto Nacional de Meteorologia (INMET) foi essencial para o treinamento eficaz das configurações de RNAs, contribuindo para a obtenção de estimativas precisas de  $ET_o$ .
3. As configurações de RNAs foram desenvolvidas e treinadas utilizando variáveis meteorológicas relevantes como entrada, demonstrando sua capacidade de aprendizado e adaptação aos padrões climáticos.
4. A avaliação das variáveis de entrada viabilizou a identificação da configuração ótima para a RNA (configuração RNA<sub>9</sub>) e um segundo modelo, a configuração RNA<sub>10</sub>, que mostrou um equilíbrio entre precisão e simplicidade na estimativa da  $ET_o$ , o que facilita sua utilização por agricultores.
5. A avaliação de desempenho das configurações, especialmente a RNA<sub>9</sub> e a RNA<sub>10</sub>, por meio de métricas estatísticas como Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e Coeficiente de Determinação ( $R^2$ ), evidenciou a eficácia das RNAs na estimativa precisa da  $ET_o$ .
6. A validação das configurações RNA<sub>9</sub> e RNA<sub>10</sub>, com dados de  $ET_o$  de um produtor na região oeste do Paraná reforçou a confiabilidade das estimativas, indicando uma concordância satisfatória entre as previsões geradas pelos modelos e os valores observados.

## 7 CONSIDERAÇÕES FINAIS

O presente estudo, voltado para a estimativa de  $ET_o$  no contexto agrícola, proporcionou valiosas percepções sobre a aplicação de modelos baseados em RNAs para a previsão de  $ET_o$ . No entanto, como em todo trabalho científico, a pesquisa não está isenta de limitações, que merecem ser reconhecidas. Estas limitações destacam-se como áreas potenciais para futuras investigações. Nesta seção, serão discutidas essas limitações, ao mesmo tempo em que serão delineados possíveis caminhos para pesquisas subsequentes, visando enriquecer e aprimorar o entendimento nesta área específica.

### 7.1 Limitações do trabalho

Esta seção visa oferecer uma análise crítica e transparente sobre os aspectos que podem ter impactado o escopo e os resultados deste estudo. Reconhecendo que todo processo de pesquisa possui suas próprias restrições, esta seção procura identificar e discutir as limitações metodológicas e contextuais que podem ter influenciado os resultados alcançados. Essa abordagem reflexiva é fundamental para contextualizar as conclusões e percepções obtidas, além de fornecer diretrizes claras para pesquisas futuras na área de estimativa de  $ET_o$  em contextos agrícolas.

- **Variação na configuração da RNA:**

- *Limitação:* embora tenha sido realizada uma análise de várias configurações das RNAs, utilizou-se apenas uma arquitetura e dois otimizadores em todas elas.
- *Sugestão:* pesquisas futuras podem explorar uma variedade mais ampla de arquiteturas de RNAs e técnicas de otimização para determinar configurações mais robustas.

- **Validação com dados de apenas um produtor:**

- *Limitação:* a validação do modelo foi conduzida com dados provenientes de um

único produtor na região oeste do Paraná. Tal abordagem pode restringir a aplicabilidade dos resultados a outras localidades ou contextos.

- *Sugestão*: pesquisas adicionais poderiam incluir dados de uma variedade mais ampla de produtores ou explorar diferentes regiões, permitindo uma compreensão mais abrangente da eficácia do modelo em contextos diversos.

- **Sazonalidade e variações climáticas locais:**

- *Limitação*: a influência da sazonalidade e variações climáticas específicas da região pode não ter sido totalmente considerada.
- *Sugestão*: pesquisas futuras podem explorar como o modelo responde a diferentes condições sazonais e climáticas locais para aprimorar sua robustez.

## 7.2 Trabalhos futuros

Nesta seção, foram identificadas oportunidades para a expansão e aprimoramento da pesquisa em estimativa de  $ET_o$ . Destacam-se caminhos promissores para pesquisas futuras, levando em consideração as lacunas e desafios identificados durante o curso deste estudo. Projetos já em andamento, desdobramentos diretos deste trabalho, foram indicados, juntamente com outras sugestões, proporcionando orientações valiosas para pesquisadores e profissionais interessados em avançar no entendimento e na aplicação de modelos baseados em RNAs na previsão de  $ET_o$  em contextos agrícolas.

A partir dos resultados deste trabalho, dois projetos já estão em andamento. No primeiro, temos um projeto que visa aprimorar a acessibilidade e utilidade dos dados meteorológicos provenientes do Instituto Nacional de Meteorologia (INMET). Dada a complexidade envolvida na coleta, pré-processamento e organização desses dados, o projeto busca desenvolver um sistema integrado que automatize esses processos. A implementação desse sistema tem o potencial de simplificar significativamente a obtenção de dados meteorológicos, tornando-os mais acessíveis para análises futuras e facilitando a expansão e replicação deste estudo em diferentes contextos.

O segundo projeto em andamento, considerando o resultado bem-sucedido na estimativa da  $ET_o$  por meio de modelos baseados em RNAs, é o desenvolvimento de um aplicativo destinado ao manejo da irrigação. Os modelos mais eficientes, especificamente a  $RNA_9$  e a  $RNA_{10}$ , serão testados diretamente com irrigantes em condições práticas. Esse aplicativo visa fornecer aos agricultores uma ferramenta prática e eficaz para otimizar a irrigação, promovendo

uma abordagem sustentável e eficiente no uso dos recursos hídricos. Essas iniciativas representam passos importantes na aplicação prática dos resultados alcançados neste trabalho, contribuindo para a inovação no campo do manejo hídrico agrícola.

Além dos dois projetos em andamento, podemos citar algumas sugestões adicionais para trabalhos futuros:

**Análise regionalizada:** explorar a possibilidade de uma análise mais regionalizada, considerando diferentes zonas climáticas dentro do estado do Paraná. Isso permitiria uma adaptação mais precisa do modelo às variações climáticas locais.

**Integração de dados adicionais:** investigar a inclusão de dados adicionais, como características específicas do solo, tipo de cultura e práticas agrícolas, para aprimorar ainda mais a precisão das estimativas de  $ET_o$ .

**Validação em outras regiões:** estender a validação do modelo para outras regiões geográficas com características climáticas distintas, a fim de avaliar a capacidade de generalização do modelo para diferentes ambientes.

**Avaliação econômica do manejo da irrigação:** realizar uma análise econômica do manejo da irrigação com base nas estimativas de evapotranspiração geradas pelos modelos, considerando custos e benefícios associados à implementação das práticas sugeridas.

## REFERÊNCIAS

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. **FAO, Rome**, v. 300, n. 9, p. D05109, 1998.

ARCA, B.; BENISCASA, F.; VINCENZI, M. D. Evaluation of neural network techniques for estimating evapotranspiration. **Evolving Solution with Neural Networks**, p. 62–97, 2001.

BACHOUR, R.; MASLOVA, I.; TICLAVILCA, A. M.; WALKER, W. R.; MCKEE, M. Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration. **Stochastic environmental research and risk assessment**, Springer, v. 30, n. 1, p. 103–117, 2016.

BRAGA, A. de P.; CARVALHO, A. C. F. de; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e aplicações**. 2. ed. Rio de Janeiro: LTC, 2012.

CARMO, C. R. S.; SILVA, J. R. de M. Aprendizado de máquina e prestação de serviços de armazenamento de dados: métricas para análise e validação de algoritmos previsores. **Revista GeTeC**, v. 12, n. 38, p. 123–144, 2023.

CARVALHO, L. G. d.; RIOS, G. F. A.; MIRANDA, W. L.; NETO, P. C. Evapotranspiração de referência: uma abordagem atual de diferentes métodos de estimativa. **Pesquisa Agropecuária Tropical**, SciELO Brasil, v. 41, n. 3, p. 456–465, 2011.

CARVALHO, R. L. da S.; DELGADO, A. R. S. Comparação de modelos perceptron multicamadas na estimava da evapotranspiração de referência. **Revista Inova Ciência Tecnologia**, v. 8, n. 1, p. e0221213, 2022.

CASTANEDO, F. **Data preparation in the Big Data era best practices for data integration**. 1. ed. Boston: O'Reilly, 2015.

CECCON, D. **Funções de ativação: definição, características, e quando usar cada uma**. 2020. Disponível em: <[https://iaexpert.academy/2020/05/25/funcoes-de-ativacao-definicao-caracteristicas-e-quando-usar-cada-uma/?doing\\_wp\\_cron=1706512055.9711220264434814453125](https://iaexpert.academy/2020/05/25/funcoes-de-ativacao-definicao-caracteristicas-e-quando-usar-cada-uma/?doing_wp_cron=1706512055.9711220264434814453125)>. Acesso em: 29 jan 2024.

CONCEIÇÃO, M. A. F. Roteiro de cálculo da evapotranspiração de referência pelo método de Penman-Monteith-FAO. **Embrapa Uva e Vinho. Circular Técnica**, Bento Gonçalves: Embrapa Uva e Vinho, 2006.

Data Science Academy. **Deep Learning Book**. 2019. Disponível em: <<https://www.deeplearningbook.com.br/>>. Acesso em: 29 jan 2024.

DAVIS INSTRUMENTS. **User Manual: Vantage Pro2 GroWeather**. Hayward, 2017.

DAVIS INSTRUMENTS. **User Manual Console For Vantage Pro2 and Vantage Pro2 Plus Weather Stations**. Hayward, 2023. Disponível em: <[https://cdn.shopify.com/s/files/1/0515/5992/3873/files/07395\\\_234\\\_Manual\\\_VP2\\\_Console\\\_RevZ\\\_web.pdf?v=1656098534](https://cdn.shopify.com/s/files/1/0515/5992/3873/files/07395\_234\_Manual\_VP2\_Console\_RevZ\_web.pdf?v=1656098534)>. Acesso em: 29 jan 2024.

FAUSETT, L. **Fundamentals of Neural Networks – Architectures, Algorithms, and Applications**. 1. ed. New Jersey: Prentice-Hall, 1994.

FERRAZ, R. C. Estimativa de evapotranspiração de referência utilizando redes neurais artificiais para o estado do Rio Grande do Sul. **Revista Tecnológica**, v. 23, n. 1, p. 25–31, 2014.

FIORIN, D. V.; MARTINS, F. R.; SCHUCH, N. J.; PEREIRA, E. B. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 33, n. 1, p. 01–20, 2011.

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

GUTIERRES, M. I.; NEVES, E. A importância do monitoramento da umidade do solo através de sensores para otimizar a irrigação nas culturas. **Revista Enciclopédia Biosfera**, v. 18, n. 35, p. 1–16, 2021. Disponível em: <<https://conhecer.org.br/ojs/index.php/biosfera/article/view/3819>>. Acesso em: 29 jan 2024.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. New York: Springer, 2009. v. 2.

HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. New Jersey: Prentice-Hall, 2009.

HEATON, J. **Introduction to Neural Networks with Java**. 2. ed. St. Louis: Heaton Research, Inc, 2008.

HONNEF, B. G. **Redes neurais para previsão do coeficiente de condutância global de refrigeradores domésticos**. 2020. Dissertação (Bacharelado em Ciência da Computação) — Universidade Federal de Santa Catarina, Florianópolis-SC, 2020.

IRIGOYEN, A. I. **Modelagem da evapotranspiração de referência e da evapotranspiração de limeira acida com aplicação de técnicas de regressão e redes neurais artificiais**. 2010. Tese (Doutorado em Ciências) — Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba-SP, 2010.

JIMENEZ, A.-F.; ORTIZ, B. V.; BONDESAN, L.; MORATA, G.; DAMIANIDIS, D. Long short-term memory neural network for irrigation management: a case study from southern alabama, usa. **Precision Agriculture**, Springer, v. 22, n. 2, p. 475–492, 2021.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *In*: BENGIO, Y.; LECUN, Y. (ed.). **Electronic annals [...]**. San Diego: ICLR, 2015.

KRISHNA, P. A. Evapotranspiration and agriculture - a review. **Agricultural Reviews**, v. 40, n. 1, p. 1–11, 2019.

KUMAR, M.; RAGHUWANSHI, N. S.; SINGH, R.; WALLENDER, W. W.; PRUITT, W. O. Estimating evapotranspiration using artificial neural network. **Journal of Irrigation and Drainage Engineering**, American Society of Civil Engineers, v. 128, n. 4, p. 224–233, 2002.

- LEIVAS, J.; GUSSO, A.; FONTANA, D. C.; BERLATO, M. Estimativa do balanço de radiação na superfície a partir de imagens do satélite ASTER. *In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, 13., 2007, Floprianópolis. **Anais eletrônicos [...]**. Floprianópolis: INPE, 2007. p. 255–262.
- LIMA, J. G. A.; VIANA, P. C.; SOBRINHO, J. E.; COUTO, J. P. C. Comparação de métodos de estimativa de  $ET_o$  e análise de sensibilidade para diferentes climas brasileiros. **Irriga, Botucatu**, v. 24, n. 3, p. 538–551, 2019.
- LOUZADA, D. R. **Deteção e caracterização de danos estruturais através de sensores a rede de Bragg e Redes Neurais Artificiais**. 2013. Tese (Doutorado em Engenharia Mecânica) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro-RJ, 2013.
- LUCAS, P. de Oliveira e. **Previsão de Séries Temporais de Evapotranspiração de Referência com Redes Neurais Convolucionais**. 2019. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal de Minas Gerais, Belo Horizonte-MG, 2019.
- LUIZ, A. J. B. *et al.* Application of the Kolmogorov-Smirnov test to compare greenhouse gas emissions over time. **Brazilian Journal of Biometrics**, v. 39, n. 1, p. 60–70, 2021.
- MARIÑO, M. A.; TRACY, J. C.; TAGHAVI, S. A. Forecasting of reference crop evapotranspiration. **Agricultural water management**, Elsevier, v. 24, n. 3, p. 163–187, 1993.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.
- MONTEITH, J. L. Evaporation and environment. *In: SYMPOSIA OF THE SOCIETY FOR EXPERIMENTAL BIOLOGY*, 19., 1965, Cambridge. **Annals [...]**. Cambridge: Cambridge University Press (CUP) Cambridge, 1965. p. 205–234.
- ODHIAMBO, L. O.; YODER, R. E.; YODER, D. C.; HINES, J. W. Optimization of fuzzy evapotranspiration model through neural training with input–output examples. **Transactions of the ASAE**, American Society of Agricultural and Biological Engineers, v. 44, n. 6, p. 1625, 2001.
- ORSSATTO, A. J. **Aplicação de modelos estatísticos e de aprendizagem de máquina na predição de casos de dengue em Cascavel - PR**. 2021. Dissertação (Bacharelado em Ciência da Computação) — Universidade Estadual do Oeste do Paraná, Cascavel-PR, 2021.
- PENMAN, H. L. Natural evaporation from open water, bare soil and grass. **Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences**, The Royal Society London, v. 193, n. 1032, p. 120–145, 1948.
- PEREA, R. G.; POYATO, E. C.; MONTESINOS, P.; DÍAZ, J. R. Irrigation demand forecasting using artificial neuro-genetic networks. **Water Resources Management**, Springer, v. 29, n. 15, p. 5551–5567, 2015.
- REIS, C. H. **Otimização de Hiperparâmetros em Redes Neurais Profundas**. 2018. Dissertação (Bacharelado em Ciência da Computação) — Universidade Federal de Itajubá, Itajubá-MG, 2018.
- RIBEIRO, A. M.; JUNIOR, F. de P. S. A. Um estudo comparativo entre cinco métodos de otimização aplicados em uma RNC voltada ao diagnóstico do glaucoma. **Revista de Sistemas e Computação-RSC**, Unifacs, v. 10, n. 1, p. 122–130, 2020.

SANTOS, D. da S. **Aplicação de Redes Neurais MLP na estimativa de Evapotranspiração de Referência nos municípios de Mendes e Teresópolis**. 2020. Dissertação (Mestrado em Ciências) — Universidade Federal Rural do Rio de Janeiro, Seropédica-RJ, 2020.

SANTOS, S. A.; MOREIRA, A. G.; JUNIOR, I. C. P. Análise comparativa da influência de otimizadores no desempenho de uma cnn para detecção do câncer de mama. *In: IX ESCOLA REGIONAL DE COMPUTAÇÃO DO CEARÁ, MARANHÃO E PIAUÍ*, 9., 2021, Ceará. **Anais eletrônicos [...]**. Ceará: SBC, 2021. p. 1–8.

SEDIYAMA, G. C. Estimativa da evapotranspiração: histórico, evolução e análise crítica. **Revista Brasileira de Agrometeorologia**, v. 4, n. 1, p. 1–12, 1996.

SENAR (ed.). **Irrigação: gestão e manejo**. Brasília: Serviço Nacional de Aprendizagem Rural, 2019.

SILVA, A. F. da. **Previsão da evapotranspiração de referência utilizando redes neurais**. 2002. Dissertação (Mestrado em Engenharia Agrícola) — Universidade Federal de Viçosa, Viçosa-MG, 2002.

SILVA, S. C. da. **Custo do bombeamento de água em sistema de irrigação sob diferentes modalidades tarifárias**. 2020. Dissertação (Mestrado em Conservação de Recursos Naturais do Cerrado) — Instituto Federal de Educação, Ciência e Tecnologia Goiano, Urutaí-GO, 2020.

SILVEIRA, P. M. da; STONE, L. F. **Manejo da irrigação do feijoeiro: uso do tensiômetro e avaliação do desempenho do pivô central**. Brasília: EMBRAPA-SPI, 1994.

SOUSA, V. F. de; MAROUELLI, W. A.; COELHO, E. F.; PINTO, J. M.; FILHO, M. A. C. (ed.). **Irrigação e fertirrigação em fruteiras e hortaliças**. Brasília-DF: Embrapa Informação Tecnológica, 2011.

SOWMYA, M.; KUMAR, M. S.; AMBAT, S. K. Comparison of deep neural networks for reference evapotranspiration prediction using minimal meteorological data. *In: 2020 ADVANCED COMPUTING AND COMMUNICATION TECHNOLOGIES FOR HIGH PERFORMANCE APPLICATIONS (ACCTHPA)*, Cochin. **Electronic Annals [...]**. India, 2020. p. 27–33.

SUDHEER, K.; GOSAIN, A.; RAMASASTRI, K. Estimating actual evapotranspiration from limited climatic data using neural computing technique. **Journal of irrigation and drainage engineering**, American Society of Civil Engineers, v. 129, n. 3, p. 214–218, 2003.

SUTSKEVER, I.; MARTENS, J.; DAHL, G.; HINTON, G. On the importance of initialization and momentum in deep learning. *In: DASGUPTA, S.; MCALLESTER, D. (ed.)*. **Electronic annals [...]**. Atlanta: PMLR, 2013. v. 28, n. 3, p. 1139–1147.

TEIXEIRA, L. F. H. de S. de B.; GOMES, C. F. S.; SANTOS, M. dos. Revisitando os testes de aderência à curva normal: Contribuições para a engenharia de produção. *In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO*, 40., 2020, Foz do Iguaçu. **Anais eletrônicos [...]**. Foz do Iguaçu: Enegep, 2020. p. 1–16.

THORNTHWAITE, C. W. An approach toward a rational classification of climate. **Geographical review**, JSTOR, v. 38, n. 1, p. 55–94, 1948.

TRAJKOVIC, S.; TODOROVIC, B.; STANKOVIC, M. Forecasting of reference evapotranspiration by artificial neural networks. **Journal of Irrigation and Drainage Engineering**, American Society of Civil Engineers, v. 129, n. 6, p. 454–457, 2003.

WALLS, S.; BINNS, A. D.; LEVISON, J.; MACRITCHIE, S. Prediction of actual evapotranspiration by artificial neural network models using data from a bowen ratio energy balance station. **Neural Computing and Applications**, Springer, v. 32, p. 14001–14018, 2020.

WREGE, M. S.; CARLOS, S. S.; JÚNIOR, R.; ALMEIDA, I. R. de (ed.). **Atlas Climático da Região Sul do Brasil: estados do Paraná, Santa Catarina e Rio Grande do Sul**. 2. ed. Brasília-DF: Embrapa Clima Temperado; Embrapa Florestas, 2012.

XAVIER, A. C.; JERSZURKI, D.; SOUZA, J. L. M. de; BARATTO, R. L. Coeficiente de proporcionalidade KRs da equação de hargreaves e samani para o brasil. *In*: CONGRESSO VIRTUAL DE AGRONOMIA, 6., 2017, São Paulo. **Anais eletrônicos [...]**. São Paulo: Convibra, 2017. p. 13.

ZANETTI, S.; SOUSA, E.; OLIVEIRA, V.; ALMEIDA, F.; BERNARDO, S. Estimating evapotranspiration using artificial neural network and minimum climatological data. **Journal of Irrigation and Drainage Engineering**, American Society of Civil Engineers, v. 133, n. 2, p. 83–89, 2007.

ZANETTI, S. S.; SOUSA, E. F.; CARVALHO, D. F. d.; BERNARDO, S. Estimação da evapotranspiração de referência no Estado do Rio de Janeiro usando redes neurais artificiais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, SciELO Brasil, v. 12, n. 2, p. 174–180, 2008.

ZHAO, T.; WANG, Q. J.; SCHEPEN, A.; GRIFFITHS, M. Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs. **Agricultural and forest meteorology**, Elsevier, v. 264, p. 114–124, 2019.