Avaliação da Capacidade de Generalização de Arquiteturas de Aprendizado Profundo para Previsão de Séries Temporais Meteorológicas

Arthur Henrique R. dos Santos *, **. Matheus Henrique D. M. Ribeiro**
Roberto Z. Freire**

** Grupo Volvo do Brasil (Volvo Group Connected Solutions - VGCS), Curitiba, PR 81260-900, Brasil (Tel:+5541992119904; e-mail: arthurroseira95@gmail.com).*

***Programa de Pós-Graduação em Engenharia de Produção e Sistemas (PPGEPS), Universidade Tecnológica Federal do Paraná (UTFPR), Pato Branco, PR 81280-340, Brasil (e-mails: mribeiro@utfpr.edu.br; robertofreire@utfpr.edu.br).*

Abstract: The objective of climate forecasting, through time series, is to provide the most accurate results possible to assist in determining future atmospheric states. In the context of time series forecasting, the state-of-the-art focuses on deep learning algorithms and the different architectures that have emerged in recent years. However, even though complex architectures have demonstrated interesting results, the amount of work that evaluates the generalization capacity of these architectures is still tiny. Based on this, in this work, a comparative study is proposed between three neural network architectures adapted for the prediction of time series: Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN), Gated Recurrent Unit (GRU) with Attention mechanism, and U-Net convolutional architecture), aiming to predict the maximum air temperature with a one-step-ahead forecast horizon using data from Meteorology National Institute (INMET) collected across Brazil. The results showed that, although new neural network architectures are emerging, the Ensemble Learning methodology still presents better results when evaluated in the prediction of new application data, surpassing other architectures by 25% in the coefficient of determination (R^2) and reducing it by up to 2.45°C of Root Mean-Square Error ($RMSE$).

Resumo: O objetivo da previsão climática, por meio de séries temporais, é fornecer os resultados mais precisos possíveis para auxiliar na determinação de estados atmosféricos futuros. O estado da arte destas áreas está voltado aos algoritmos de aprendizado profundo e as diferentes arquiteturas que surgiram nos últimos anos. Porém, por mais que arquiteturas complexas venham apresentando resultados interessantes, ainda é pequena a quantidade de trabalhos que avaliam a capacidade de generalização destas arquiteturas. Nesse sentido, este trabalho propõe um estudo comparativo entre três arquiteturas de redes neurais adaptadas para a previsão de séries temporais: comitê de máquinas *Long Short-Term Memory* (LSTM) *Temporal Convolutional Network* (TCN), *Gated Recurrent Unit* (GRU) com mecanismos de atenção e rede arquitetura convolucional U-Net. O objetivo é prever a temperatura máxima do ar com um horizonte de previsão de um passo à frente utilizando dados do Instituto Nacional de Meteorologia (INMET) coletados em todo o Brasil. Os resultados mostraram que, apesar de novas arquiteturas de redes neurais estarem surgindo, a metodologia de comitê de máquinas ainda apresenta melhores resultados quando avaliadas na previsão de novos dados, superandos as demais arquiteturas em 25% no coeficiente de determinação ($\boldsymbol{R}^2$) e reduzindo em até 2,45°C a Raiz do Erro Médio Quadrático ($\boldsymbol{RMSE}$).

Keywords: artificial neural network; ensemble forecasting; machine learning; model generalization capacity; deep learning; time series forecasting; weather forecasting.

Palavras-chaves: aprendizado de máquina; aprendizado profundo; capacidade de generalização do modelo; previsão de séries temporais; previsão meteorológica, rede neural artificial.

1. INTRODUÇÃO

A previsão meteorológica desempenha um papel importante na compreensão e mitigação de eventos climáticos extremos, na gestão de recursos naturais e tomada de decisão em diversos setores, como na agricultura, energia e transporte. Neste contexto, é comum a representação de análises meteorológicas ser realizada por meio de séries temporais (Krzyszczak et al., 2017), devido principalmente à necessidade de acompanhamento de longos ciclos para a determinação das características climáticas.

No âmbito da meteorologia, os modelos numéricos de previsão (do inglês, *Numerical Weather Prediction*, NWP) são predominantes. Isto se deve ao fato de que estes modelos baseiam-se em representações matemáticas da dinâmica atmosférica (Dongjin Cho e Cheolhee Yoo, 2020). Porém, em

decorrência das imperfeições nessas representações, estes modelos podem gerar tendências em função das sistemáticas atmosféricas locais e globais, limitando sua capacidade de generalização. Desta forma, fez-se necessário o avanço no estudo de metodologias alternativas a estes modelos, onde é possível encontrar aplicações de modelos estatísticos na previsão de séries temporais meteorológicas, como é o caso de algoritmos como o ARIMA e suas derivações como os modelos SARIMA (Dimri et al., 2020).

As redes neurais recorrentes (RNNs) foram projetadas especificamente para lidar com dados sequenciais visando obter a capacidade de armazenar informações sobre entradas anteriores para realizar previsões. Já as redes neurais convolucionais (CNNs), apesar de terem sido criadas para processamento de dados espaciais, sofrem menos com problemas de desvanecimento do gradiente em comparação com as RNNs, sendo esta uma vantagem na previsão de séries temporais (Benidis, 2022). Outras variantes mais avançadas, como as redes *Long Short-Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU), também foram desenvolvidas para lidar com dependências de longo prazo, evitando o problema do desvanecimento do gradiente (Lin et al., 2021).

Neste contexto, a presente pesquisa apresenta um estudo comparativo onde diferentes tipos de arquiteturas de aprendizado profundo são aplicadas para solucionar um problema de previsão climática. O principal fator avaliado é a capacidade de generalização dos modelos, de forma a verificar se estes são capazes de manter a consistência nos resultados de previsão sobre novos dados, ou seja, em séries meteorológicas coletadas em diferentes regiões da qual modelo foi treinado.

Na seção 2 apresenta-se um levantamento dos estudos recentes, a partir de 2018, envolvendo comitê de máquinas e aprendizado profundo para séries temporais. Na seção 3 descreve-se o problema de previsão, bem como a base de dados e os modelos utilizados. Os resultados foram dispostos na seção 4, comparando as diferentes arquiteturas selecionas em uma base de teste não utilizada na fase de treinamento, referente a uma região com características meteorológicas distintas. Por fim, a conclusão é apresentada na seção 5 em conjunto com sugestões para trabalhos futuros.

2. REVISÃO DE LITERATURA

O objetivo desta seção é realizar uma análise das pesquisas recentes com o foco em previsão de séries temporais por meio da utilização da metodologia de comitê de máquinas e aprendizado profundo, dando ênfase nas diferentes arquiteturas de redes neurais. Por se tratar de um tema amplo, este levantamento visa apresentar as diferentes metodologias de aplicação destes conceitos e sua respectiva relevância dentro do tema, a fim de justificar a escolha do formato da implementação apresentada na seção 3. As principais métricas de avaliação também foram levantadas por meio desta revisão.

Na pesquisa proposta por Larrea et al. (2021), direcionada à previsão do consumo de energia elétrica, foi proposta a utilização de redes neurais de aprendizado extremo (ELM) agregadas em um comitê de máquinas através da distribuição de pesos por meio de otimização via algoritmo de enxame de partículas (PSO). A metodologia foi avaliada pelas métricas $RMSE$ e Erro Médio Quadrático Normalizado ($NMSE$). Os resultados mostraram que o algoritmo PSO melhorou em 10% os resultados de previsão do comitê em comparação com a agregação dos modelos utilizando a média dos seus resultados.

Para a previsão de volume de tráfego em Israel, Ma et al. (2020) propuseram um comitê de máquinas onde as séries residuais resultantes da previsão de uma rede neural *Multi-Layer Perceptron* (MLP) foram aplicadas em um modelo ARIMA, gerando o resultado de previsão. Foi avaliado de forma individual o modelo MLP um passo à frente e o comitê híbrido denominado NN-ARIMA por meio do Erro Médio Quadrático (MSE). O método proposto conseguiu superar o algoritmo estatístico, apresentando uma melhora de 14,65%.

Phyo e Byun (2021) propuseram um comitê de redes neurais LSTM e CNN formando um conjunto híbrido denominado CNN-LSTM. Os modelos foram implementados na previsão da demanda de energia elétrica na Coreia do Sul em um horizonte de previsão de um passo à frente (uma hora à frente, uma semana à frente e um mês à frente). Os modelos foram avaliados utilizando as métricas MSE, MAE e $MAPE$, apresentando uma superioridade do modelo conjunto em relação à aplicação dos modelos de maneira individual, fornecendo um $MAPE$ mínimo de 0,75%.

Tabela 1. Revisão dos modelos de *Deep Ensemble Learning* para previsão de séries temporais.

Referência	Ano	Pré-Processamento	Métodos	Otimização	Horizonte Previsão (passos à frente)	Métricas
Larrea et al.	2019	-	ELM	PSO	1	$RMSE$ e $NMSE$
Ma et al.	2020	-	NN-ARIMA	*Stacked Ensemble*	1	MSE
Phyo et al.	2021	-	LSTM e CNN	*Stacked Ensemble*	1	MSE, MAE e $MAPE$
Tekin et al.	2021	-	Atenção LSTM/GRU	*Stacked Ensemble*	1	$RMSE$, MAE e $MAPE$
Shukla e Marlin	2021	-	Atenção GRU	*Stacked Ensemble*	N	MSE
Benson et al.	2021	-	LSTM	-	27	$MAPE$
Guo et al.	2022	EMD	LSTM e TCN	-	1	$MAPE$ e R^2
Li Z. et al.	2023	-	LSTM, TCN e XGBoost	WB	1	MSE e MAE
Li Q. et al.	2023	CEEMDAN	LSTM e WGAN	-	1	MAE, $MAPE$ e $RMSE$

Visando a utilização de arquiteturas de redes neurais de maior complexidade em conjunto com modelos NWP, Tekin et al. (2021) desenvolveram um estudo comparativo considerando o modelo numérico ERA5 em conjunto com redes CNN e LSTM em diferentes arquiteturas, como a estrutura convolucional U-Net e a aplicação de mecanismos de atenção na arquitetura TrajGRU. Os modelos foram avaliados utilizando as métricas $RMSE$, MAE e $MAPE$ e os resultados mostraram que o modelo NWP se sobressai sobre os modelos de aprendizado profundo em intervalos curtos de previsão, mas perde eficiência em intervalos maiores, aumentando o $RMSE$ em até 0,5°C.

Shukla e Marlin (2021) propuseram uma arquitetura de redes GRU com múltiplas camadas com mecanismos de atenção nomeada como *Multi-Time Attention Network* (mTAND) para previsão de séries temporais multivariadas com amostragem não uniforme em uma aplicação de interpolação de dados faltantes ou de baixa qualidade. Avaliando a arquitetura proposta em relação à outras arquiteturas de redes RNN e CNN por meio da métrica MSE, os resultados mostram que o modelo proposto reduz em 33% o erro de interpolação em relação às demais arquiteturas para esta aplicação.

Fazendo o uso de comitês de máquinas de aprendizado profundo e utilizando redes neurais do tipo LSTM, Benson et al. (2021) desenvolveram um estudo para prever o fluxo de radiação solar acoplando o uso de séries temporais e imagens durante um período de 27 dias. Os resultados demonstraram uma melhoria de 30 a 40% no $MAPE$ ao combinar dados de imagens solares com dados de séries temporais em comparação com o uso exclusivo de dados de séries temporais.

Um estudo comparativo entre comitês homogêneos de redes LSTM e redes neurais temporais convolucionais (TCN), em conjunto com um algoritmo de decomposição de séries temporais chamado de *Empirical Mode Decomposition* (EMD) foi proposto por Guo et al. (2023). Os comitês foram comparados através de três bases de dados multivariadas distintas. Como resultados, os autores mostraram que em o comitê de redes TCN juntamente com o algoritmo EMD conseguiu se sobressair em relação ao conjunto LSTM em até 17% no $MAPE$ e em 8% no R^2.

Z. Li et al. (2024) propuseram um comitê de máquinas heterogêneo formado por redes LSTM e TCN, juntamente com o algoritmo XGBoost, para previsão de demanda de energia elétrica no formato de série temporal univariada. Os modelos realizam previsões de forma paralela e o resultado foi obtido através de um algoritmo de balanceamento de pesos. No trabalho, o comitê de máquinas superou os algoritmos de forma individual em 49% no MSE e 33% no MAE.

No contexto de previsão meteorológica para produção de energia solar, o trabalho proposto por Q. Li et al. (2023) descreve a aplicação de um comitê formado por redes LSTM e *Wasserstein Generative Adversarial Network* (WGAN). No estudo foi avaliado a utilização do algoritmo *Ensemble Empirical Mode Decomposition with Adaptive Noise* (CEEMDAN) no pré-processamento das séries temporais. Os experimentos mostraram que, em comparação com muitos métodos tradicionais de previsão e modelos de integração de decomposição, o modelo desenvolvido minimizou os critérios de avaliação MAE, $MAPE$ e $RMSE$ das quatro estações, os quais diminuíram 3,51%, 6,11% e 2,25%, respectivamente.

A Tabela 1 descreve um resumo da revisão de literatura desta seção, onde é possível notar que prevalece a utilização das redes LSTM, CNN e GRU combinadas de duas formas: no formato de pilha (*Stacked*) e de comitês (*Ensemble*s). Este fato é um ponto importante para a definição das arquiteturas implementadas neste trabalho, onde os modelos apresentados na seção 4 foram baseadas nessas redes.

3. DESCRIÇÃO DO PROBLEMA

O instituto nacional de meteorologia do Brasil (INMET) é um órgão federal responsável em prover informações de monitoramento atmosférico e realizar previsões climáticas através do Modelo Numérico Brasileiro de Alta Resolução (MBAR) (Ministério da Agricultura e Pecuária, 2024). Em todo o território do Brasil, o INMET administra 567 estações meteorológicas automáticas e fornece dados publicamente desde o ano 2000 com amostragem horária.

Na Fig. 1 é possível verificar a cobertura das estações automáticas de coleta de dados gerenciada pelo INMET em todo o território nacional. Já a Tabela 2 possui um descritivo de todas as variáveis disponibilizadas nas bases de dados públicas referentes à todas as estações, ao todo são 567 bases de dados diferentes coletadas no período de maio de 2000 até julho de 2023. Ao todo, a base de dados totaliza aproximadamente 610.000 amostras das 17 variáveis para todas as estações do território brasileiro. De forma prática, estes modelos serão avaliados realizando previsões para uma determinada região, sendo esta diferente daquela na qual os dados de treinamento foram coletados.

Assim, será avaliada a capacidade dos algoritmos de aprenderem características das séries temporais meteorologias independentemente da região de coleta dos dados, visando validar a utilização de um algoritmo na previsão do clima em diferentes regiões auxiliando previsões em regiões onde não há dados históricos suficientes para o treinamento de um algoritmo de aprendizado de máquina, ou desenvolvimento de um modelo numérico específico para aquela região.

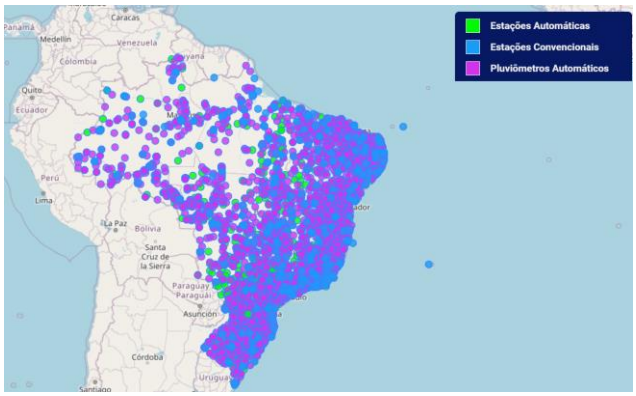


Fig. 1. Mapa estações automáticas INMET (Ministério da Agricultura e Pecuária, 2024).

Tabela 2. Variáveis da base de dados INMET.

Nomenclatura	Unidade
Precipitação total, horário	mm
Pressão atmosférica ao nível da estação, horária	mbar
Pressão atmosférica máx. na hora anterior	mbar
Pressão atmosférica min. Na hora anterior	mbar
Radiação global	kJ/m²
Temperatura do ar - bulbo seco, horária	°C
Temperatura do ponto de orvalho	°C
Temperatura máxima na hora anterior	°C
Temperatura mínima na hora anterior	°C
Temperatura orvalho máx. na hora anterior	°C
Temperatura orvalho min. na hora anterior	°C
Umidade relativa máx. na hora anterior	%
Umidade relativa. min. na hora anterior	%
Umidade relativa do ar, horária	%
Vento, direção horária	°
Vento, rajada máxima	m/s
Vento, velocidade horária	m/s

4. MODELOS DE PREVISÃO

Nesta seção são descritas as 3 arquiteturas de redes neurais implementadas neste estudo, estas determinadas de acordo com a análise dos trabalhos apresentados seção 2.

4.1 LSTM/TCN Ensemble

Combinar redes LSTM (*Long Short-Term Memory*) e TCN (*Temporal Convolutional Network*) em um comitê de máquinas é uma estratégia que visa unificar as qualidades de ambas as arquiteturas para melhorar o desempenho do modelo. A rede LSTM é um tipo de rede neural recorrente (RNN) adequada para modelar sequências de dados (Hochreiter e Schmidhuber, 1997), enquanto a TCN é um tipo de rede neural convolucional projetada para capturar dependências de longo alcance em séries de dados (Lea et al., 2016).

Embora os métodos de *ensemble* sejam eficientes na previsão de series temporais por meio de modelos de aprendizado de máquina, sua implementação enfrenta desafios, principalmente na definição do número de modelos e na combinação dos resultados. Para favorecer a generalização, as redes neurais foram treinadas usando bases de dados distintas, onde cada modelo é treinado com dados de estações automáticas selecionadas aleatoriamente. Para criação, seleção e combinação de preditores individuais, foi desenvolvido um algoritmo genético de otimização binário que seleciona as arquiteturas treinadas que irão compor o modelo final.

Cada possível solução do algoritmo genético desenvolvido é formada por uma lista binária onde cada elemento é um modelo específico dentre todos os que foram treinados. A função custo do problema de otimização foi determinada como sendo a métrica de decomposição de ambiguidade de perda de erro quadrático (Krogh e Vedelsby, 1994). A decomposição da ambiguidade de perda de erro quadrático para M modelos $(f_1, f_2, \dots f_M)$ pode ser vista na equação (1).

$$(\bar{f} - \mathrm{y})^2 = \frac{1}{M}\sum_{i=1}^{M}(y - f_\mathrm{i})^2 - \frac{1}{M}\sum_{i=1}^{M}(\bar{f} - f_i)^2 \qquad (1)$$

O erro quadrático é calculado pelo termo $(y - f_\mathrm{i})^2$, onde y representa os valores reais da variável objetivo. A parcela de ambiguidade é calculada pelo termo $(\bar{f} - f_i)^2$, representando a diferença da combinação por meio do cálculo da média aritmética $\bar{f}$ dos resultados de cada modelo e de seus respectivos valores preditos.

Foram definidos 28 subconjuntos de dados para treinamento dos modelos do comitê e, para cada um, foi definido uma rede neural LSTN ou TCN aleatoriamente. Para promover maior diversidade entre os modelos, os hiperparâmetros foram definidos de forma aleatória, sendo que o número de camadas ocultas variou de 1 a 5, e cada camada pode conter de 16 a 256 neurônios. As funções de ativação variaram entre a função ReLu e Sigmoide. O treinamento do modelo foi realizado com o otimizador Adam (Kingma e Ba, 2015), conhecido por sua eficiência em ajustar os pesos durante o treinamento.

4.2 Rede GRU Encoder-Decoder com Mecanismo de Atenção

O *Gated Recurrent Unit* (GRU), introduzido por Cho et al. (2014), foi projetado para superar limitações das RNNs na captura de dependências de longo prazo. No estudo, um modelo de previsão de séries temporais é implementado usando uma estrutura de codificação-decodificação que integra GRUs e um mecanismo de auto atenção. O modelo é composto por três partes: codificação, decodificação e módulos de atenção. Na codificação, GRUs extraem recursos das séries temporais, enquanto na decodificação, o mecanismo de auto atenção, calcula pontuações de atenção entre as saídas do codificador, permitindo que o modelo se concentre em diferentes partes da série para melhorar as previsões.

A arquitetura GRU *Encoder-Decoder* descrita utiliza camadas GRU com 8 unidades em ambas as fases de codificação e decodificação. O inicializador de pesos adotado é o Normal He, conforme descrito por He et al. (2015), que é projetado para melhorar a convergência em redes profundas. Assim como no comitê de máquinas LSTM/TCN o algoritmo otimizador selecionado foi o Adam. O mecanismo de atenção empregado é o *self-attention*, atenção por produto escalar, que calcula a relevância entre diferentes partes da sequência de entrada para ajustar os pesos das previsões com base na importância relativa das partes (Vaswani et al., 2017).

4.3 Rede Convolucional U-Net

A arquitetura U-Net é formada por camadas totalmente convolucionais (CNN) (Ronneberger et al., 2015). Este tipo de algoritmo é considerado uma estrutura *Encoder-Decoder*, caracterizado por sua estrutura em forma de U, que consiste em um caminho de contração seguido por um caminho de expansão. Para esta aplicação, definiu-se uma arquitetura U-Net para dados de saída unidimensionais, que é comumente utilizada para tarefas de análise de séries temporais.

De acordo com esta arquitetura, a rede U-Net implementada possui 3 camadas de contração com 64, 128 e 256 filtros e um *kernel* de tamanho 3. A camada de *bottleneck*, que tem função de reduzir a dimensão espacial dos dados, possui 512 filtros e tamanho 3. A rede também possui 3 camadas de expansão com filtros de tamanho 256, 128 e 64 e uma camada de saída unidimensional. Todas as camadas possuem função de ReLu.

4.4 Critérios de Avaliação de Desempenho

Como métodos de avaliação foram usados o $RMSE$ e o R^2, apresentados em (2) e (3), respectivamente.

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

$$R^2 = \frac{\sum_1^n (\hat{y}_i - \bar{y})^2}{\sum_1^n (y_i - \bar{y})^2} \quad (3)$$

Os termos apresentados $\hat{y}_i$ em (2) e (3) correspondem aos valores preditos, já o termo y_i correspondem aos valores reais, por fim, o termo $\bar{y}$ presente em (2), corresponde à média dos valores reais ao longo do período de previsão.

5. RESULTADOS

Nesta seção são apresentados os resultados obtidos por meio da implementação das três arquiteturas descritas na seção 4. Os resultados para comparação consideraram dois cenários, os quais consideraram as métricas $RMSE$ e R^2. As etapas de treinamento e validação constituíram o primeiro cenário de coleta de resultados, onde os dados de todas as estações meteorológicas dos estados do Rio Grande do Sul, São Paulo, Minas Gerais, Rio de Janeiro foram utilizadas para treinar e validar as arquiteturas, em uma divisão de 80% (anos de 2000 até 2018) e 20% (anos de 2018 até 2023) dos dados, respectivamente. Já no segundo cenário, os dados das estações meteorológicas do estado do Paraná são utilizados para uma nova seção de previsões. A série temporal utilizada teve início na mesma data de início da base de validação (2018 até 2023).

Para garantir uma avaliação consistente, foi utilizada a validação cruzada na primeira etapa. Os dados foram divididos em 5 subconjuntos de treino que variaram de tamanho com 23.772, 47.542, 71.312, 95.082 e 1.188.252 amostras. Para todos os conjuntos de treinamento foi utilizada a mesma base de teste com 23.770 amostras. Conforme a Tabela 3, todas as arquiteturas propostas conseguiram se adaptar às características das séries temporais utilizadas. O melhor resultado foi alcançado através da validação cruzada, por meio do comitê LSTN/TCN, que apresentou R^2 de 0,87±0,04 e 2,60±1,00 para o $RMSE$. Com relação ao comitê de máquinas, ao todo foram treinados 28 modelos, sendo que a aplicação do algoritmo de otimização finalizou por selecionar apenas 5. Os valores de R^2 de 0,88 e $RMSE$ de 3,01°C foram obtidos da combinação dos resultados destes 5 modelos através da média aritmética. Os resultados apresentaram variância de 0,002 para R^2 e 1,87°C para o $RMSE$ dentre os modelos selecionados.

Tabela 3. Resultados referentes à base de validação.

	Temperatura Máxima	
	R^2	$RMSE$
LSTM/TCN	**0,87±0,04**	**2,60±1,00**
U-Net	0,83±0,06	3,13±1,16
GRU ATM	0,80±0,07	3,60±1,22

Na segunda etapa de avaliação dos resultados, foi utilizado um conjunto de dados de teste originado a partir de estações automáticas não incluídas nas bases de treinamento e validação. A Tabela 4 apresenta os dados desta etapa.

Tabela 4. Resultados referentes à base de teste.

	Temperatura Máxima	
	R^2	$RMSE$
LSTM/TCN	**0,88**	**2,58**
U-Net	0,77	5,03
GRU ATM	0,66	7,22

Avaliando a Tabela 4, percebe-se uma diferença acentuada entre os desempenhos de previsão dos 3 modelos. Tanto o modelo GRU ATM quando o U-Net apresentaram uma queda brusca na previsão quando confrontados com dados de séries temporais diferentes daquelas utilizadas no treinamento e validação, ou seja, dados totalmente novos e possivelmente com leves diferenças de suas características climáticas. Já o modelo de comitê de máquinas LSTM/TCN conseguiu manter seu desempenho e até melhorar seu índice $RMSE$ em 15%. Por meio da análise dos gráficos (Fig. 2), é possível perceber que o modelo GRU gerou resultados com picos de temperatura muito acima dos demais modelos, ultrapassando a faixa dos 32°C, apresentando o pior resultado de previsão. Já o comitê de máquinas se mostrou mais aderente aos dados de validação.

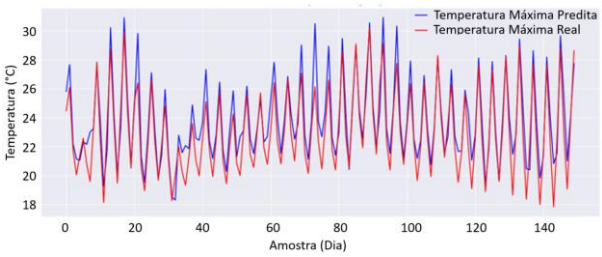
6. CONCLUSÕES

Este trabalho teve como premissa o estudo da capacidade de generalização de algoritmos de aprendizado profundo sobre séries temporais meteorológicas multivariadas, com o objetivo comparar diferentes arquiteturas de redes neurais utilizando dados que não foram utilizados na fase de treinamento dos modelos. O trabalho incluiu a aplicação da de comitê de máquinas com modelos LSTM e TCN, a aplicação de redes neurais GRU com mecanismos de atenção e redes neurais convolucionais dispostas na arquitetura U-Net.

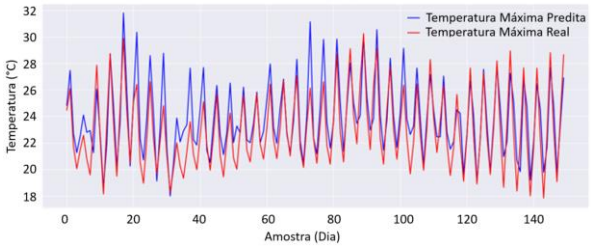
Na primeira etapa, denominada de validação, foi possível constatar que os modelos apresentaram desempenho similar. Contudo, na segunda etapa, foi notada uma diferença expressiva na capacidade de generalização destes modelos na previsão de novas séries temporais, onde foi visível a melhor capacidade de armazenar diferentes características das séries temporais no algoritmo de *Ensemble Learning*.

Como trabalhos futuros, sugere-se a inclusão de redes para processamento de imagens meteorológicas ao modelo e o

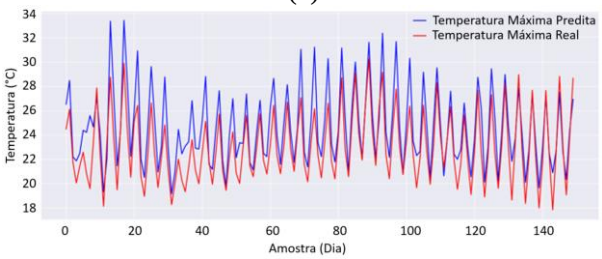
comparativo de diferentes comitês de máquinas envolvendo outras arquiteturas e horizontes de previsão.



(a)



(b)



(c)

Fig. 2. Representação gráfica dos resultados de teste: (a) LSTM/TCN; (b) U-Net; (c) GRU ATM.

AGRADECIMENTOS

Os autores agradecem o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo: 312688/2021-0, pelo financiamento desta pesquisa.

REFERÊNCIAS

Benson, B., Brown, E., Bonasera, S., Acciarini, G., Pérez-Hernández, J. A., Sutton, E., Jah, M. K., Bridges, C., Jin, M., Güne¸s, A., & Baydin, G. (2021). Simultaneous Multivariate Forecast of Space Weather Indices using Deep Neural Network Ensembles. *Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021)*, 1–6. https://lasp.colorado.edu/lisird/data/

Dimri, T., Ahmad, S., & Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. Journal of Earth System Science, 129(1), 149.

Dongjin Cho, & Cheolhee Yoo. (2020, February 18). *UCI Machine Learning Repository: Bias correction of numerical prediction model temperature forecast Data Set*. https://archive.ics.uci.edu/ml/datasets/Bias+correction+of+numerical+prediction+model+temperature+forecast

Guo, C., Kang, X., Xiong, J., & Wu, J. (2023). A New Time Series Forecasting Model Based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Temporal Convolutional Network. *Neural Processing Letters*, 55(4), 4397–4417.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.

Kingma D. P., & Ba J. (2015). Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations.*

Krogh, A., & Vedelsby, J. (1994). Neural Network Ensembles, Cross Validation, and Active Learning. In Advances in *Neural Information Processing Systems* 7 (NIPS 1994).

Krzyszczak, J., Baranowski, P., Hoffmann, H., Zubik, M., & Sławiński, C. (2017). Analysis of Climate Dynamics Across a European Transect Using a Multifractal Method (pp. 103–116).

Larrea, M., Porto, A., Irigoyen, E., Barragán, A. J., & Andújar, J. M. (2021). Extreme learning machine ensemble model for time series forecasting boosted by PSO: Application to an electric consumption problem. *Neurocomputing*, 452, 465–472.

Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal Convolutional Networks: A Unified Approach to Action Segmentation (pp. 47–54).

Li, Q., Zhang, D., & Yan, K. (2023). A Solar Irradiance Forecasting Framework Based on the CEE-WGAN-LSTM Model. *Sensors*, 23(5), 2799.

Li, Z., Qian, X., Li, L., & Xia, Z. (2024). Time series prediction model based on autoregression weight network. *Engineering Reports*, 6(4).

Lin, W.-H., Wang, P., Chao, K.-M., Lin, H.-C., Yang, Z.-Y., & Lai, Y.-H. (2021). Wind Power Forecasting with Deep Learning Networks: Time-Series Forecasting. *Applied Sciences*, 11(21), 10335.

Ministério da Agricultura e Pecuária. (2024). Instituto Nacional de Meteorologia. INMET, Governo Federal. https://portal.inmet.gov.br/

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science* (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351, 234–241.

Shukla, S. N., & Marlin, B. M. (2021). Multi-Time Attention Networks for Irregularly Sampled Time Series. 9th *International Conference on Learning Representations*, 1–15.

Tekin, S. F., Fazla, A., & Kozat, S. S. (2021). Numerical Weather Forecasting using Convolutional-LSTM with Attention and Context Matcher Mechanisms. (Submitted to) *IEEE Transactions on Geoscience and Remote Sensing*, 1–11.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.