

LSTMs Encoder-Decoders na Geração de Meteogramas Horários Multivariados para Brasília

Thomas Alexandre da Silva, Geraldo P. Rocha Filho, Andre L. M. Serrano
Fábio L. L. de Mendonça, Rodolfo I. Meneguette, Vinícius P. Gonçalves

Resumo—Diversos setores se beneficiam de previsões meteorológicas, como a produção agrícola, o transporte e o turismo, evidenciando seu valor socioeconômico. Nesta conjuntura, e considerando a crescente popularização das Redes Neurais Artificiais, é natural que se explore novos modelos preditivos. Destarte, este estudo propõe redes LSTM Encoder-Decoder para geração de previsões multivariadas em 24 horas para Brasília. Outrossim, visando performance ótima, testam-se variações hiperparamétricas em profundidade, dropout, largura de janela e bidirecionalidade. Dentre as variáveis previstas, estão: temperatura, umidade relativa, pressão atmosférica e irradiância solar.

Palavras-Chave—Meteorologia, LSTM, Seq2Seq, EncDec, Brasília.

Abstract—Several sectors benefit from weather forecasts, such as agricultural production, transport and tourism, highlighting its socioeconomic value. At this juncture, and considering the increasing popularization of Artificial Neural Networks, it is natural to explore new predictive models. Therefore, this study proposes LSTM Encoder-Decoder networks to generate 24-hour multivariate forecasts for Brasília. Furthermore, aiming for optimal performance, hyperparametric variations in depth, dropout, windowing width and bidirectionality are tested. Among the predicted variables are: temperature, relative humidity, atmospheric pressure and solar irradiance.

Keywords—Meteorology, LSTM, Seq2Seq, EncDec, Brasília.

I. INTRODUÇÃO

Tendo em vista a importância socioeconômica dos prognósticos meteorológicos, constata-se o interesse histórico em métodos para gerá-los. Essa prática remonta aos primórdios de diversas civilizações, as quais a exerciam com base na sazonalidade dos eventos climáticos [14]. Dentre diversos interesses, destacam-se impactos na produção agrícola e a prevenção de desastres ambientais, como queimadas em biomas que lidam com períodos de estiagem, ao exemplo do cerrado brasileiro.

Hodiernamente, é comum que se empregue modelos físicos de simulação atmosférica, a exemplo do INMET (Instituto Nacional de Meteorologia), que utiliza o COSMO (Consortium for Small-scale Modeling) em previsões não hidrostáticas limitadas em área (2.8x2.8 ou 7x7 km) [7]. Por outro lado, ao considerar a estocasticidade sequencial dos fenômenos meteorológicos, diversos trabalhos têm sido continuamente publicados para previsões envolvendo redes LSTM (Long Short-Term Memory) neste meio. Percebe-se, portanto, aos exemplos de [14], [5], [12], [11], [2], [9] e [8], uma crescente busca por soluções meteorológicas alternativas envolvendo métodos de Aprendizagem Profunda.

Temperatura, umidade, pressão e pluviometria são grandezas comuns em meteogramas. A irradiância solar também

compõe uma variável de interesse, podendo auxiliar na estimativa de produtividade para painéis fotovoltaicos [12] e sugerindo a nebulosidade ocorrida. Por fim, espera-se que uma estação meteorológica automática possa fornecer tais dados e estes corroborem com o treino de modelos sequenciais capazes de prever, com base em dados recentes, um meteograma horário do dia seguinte. Para esta estrutura Seq2Seq (Sequence-to-Sequence) de múltiplas variáveis correlatas em que se admite variações no tamanho das sequências, é comum a utilização de LSTMs EncDecs (Encoder-Decoders), como em [2] e [14]. Entretanto, conforme apontado por [8], a maioria dos estudos produz previsões de médias diárias, mensais ou anuais, com poucos se atendo a precisão horária.

É natural, ainda, que se considere a pluviometria nestas predições. Redes LSTM são capazes de fornecer saídas multivariadas temporalmente distribuídas, mas, para uma resolução horária com funções de erro comuns, ao exemplo do MAE (Mean Average Error) e MSE (Mean Squared Error), correr-se-ia o risco de punir excessivamente um prognóstico pluviométrico por localizar eventos com uma única hora de defasagem do ocorrido, mesmo que na milimetragem correta. Este erro não deveria implicar em tamanha perda, evidenciando que, tão importante quanto os valores previstos, é a avaliação de sua distância temporal. Por fim, observa-se que a pluviometria, embora relevante e pertinente como entrada, compõe uma saída que seria melhor abordada em modelos a parte, com estrutura de dados e função de erro próprias.

Este trabalho almeja, portanto, previsões horárias de temperatura, umidade, pressão e irradiância solar, baseado nos mesmos acrescidos da pluviometria. Para tal, são adotadas redes LSTM no formato Encoder-Decoder a fim de flexibilizar o tamanho das janelas de análise e permitir prognósticos diários baseados nas últimas 24 e 48 horas. Adjunto a tal, tem-se que diversos hiperparâmetros devem ser testados almejando resultados ótimos e evidenciando a eficiência do modelo multivariado em questão frente a alternativas de saída única. Para as bases de treinamento, validação e testes, são utilizados os registros públicos do INMET para Brasília (DF).

Por fim, o presente documento se divide nas seguintes partes: Sessão II, que expõe a literatura vigente sobre o tema; III, com definições da arquitetura proposta; IV, com a análise dos dados e definições de testes; V, com os relatórios dos modelos; e VI, com conclusões, limitações e lacunas.

II. TRABALHOS RELACIONADOS

No que se refere a previsão de séries meteorológicas multivariadas Seq2Seq por redes neurais, é de praxe que se observe

modelos LSTM performarem com boa acurácia, conforme já exposto. Neste contexto, o modelo de [14] opera previsões horárias de temperatura, umidade e velocidade eólica. Este propõe um Encoder-Decoder de duas redes LSTMs, uma rede densa entre elas e outra na saída (LSTM-Dense-LSTM-Dense). A estrutura dos dados compõe, neste, janelamentos de 24 ou 72h nas variáveis de temperatura, umidade e velocidade eólica. Aplicou-se uma divisão de 80% da base de dados para treinamento, 10% para validação e o restante para testes, com nenhum dropout citado. Tal experimentação preferiu o MSE enquanto função de perda e um escalonamento próprio de mínimos e máximos, mas não justificou tais escolhas.

Por outro lado, [2] propõe um estudo para previsões diárias de temperatura, umidade e horas de luz solar utilizando um empilhamento mais simples: duas redes LSTM (uma encoder e outra decoder) adjunta de uma densamente conectada na saída (LSTM-LSTM-Dense). São, ainda, destinadas 64% das amostras para treino, 16% para a validação e 20% para testes (dados de 1977 a 2022) e, apesar de apresentar diversas métricas de testes, não explicita a função de erro utilizada no aprendizado supervisionado ou o uso de dropout.

Distanciando-se dos modelos EncDecs e se atendo a um leque mais abrangente de LSTMs e derivados, encontram-se alguns trabalhos de ponta no ramo da meteorologia horária monovariada. Dentre tais, destacam-se [8], [9], [5], [11] e [12]. Vide a Tabela I, comparando o escopo do presente estudo com as produções expostas.

TABELA I

COMPARAÇÃO DO PRESENTE TRABALHO COM PRODUÇÕES ANTERIORES

	Jingwei (Conv LSTM)	Kreuzer (Conv LSTM)	Ozbek (LSTM)	Qing (LSTM)	Alakkari (EncDec LSTM)	Zaytar (EncDec LSTM)	Este trabalho (EncDec LSTM)
Considera Pluviometria		✓					✓
Prevê	Temperatura	✓			✓	✓	✓
	Umidade			✓	✓	✓	✓
	Pressão						✓
	Irradiância			✓			✓
Analisa Outliers	✓	✓		✓	✓		✓
	Profundidades						✓
Testa	Dropouts						✓
	Bidirecionalidade						✓
	Janelamentos				✓	✓	✓
	Precisão Horária	✓	✓	✓	✓	✓	✓

III. MODELO PROPOSTO

Para estimar um meteograma de 24 horas com base em dados recentes, é necessário operar uma estrutura Seq2Seq multivariada – isto é, receber uma sequência entrante e projetar temporalmente uma sequência-resposta, ambas em múltiplas variáveis. São estas: temperatura (°C); umidade relativa (%); pressão atmosférica (hPa) e irradiância solar (kJ/mh), com a pluviometria (mm/h) presente apenas nos dados entrantes (vide Sessão I). Considere, ainda, B_X e B_Y como bases de entrada e saída filtradas por suas grandezas. A partir de sua segmentação temporal, geram-se as janelas traseiras e frontais que compõem, respectivamente, os conjuntos X e Y .

O modelo deve receber, como entrada, a i -ésima amostra X_i do conjunto de janelas traseiras X . Cada X_i contém as cinco submatrizes Tmp_X , Hr_X , Pr_X , Rd_X e Plv_X , referentes aos dados já escalonados de temperatura, umidade, pressão, irradiância e pluviometria, cada uma com n_X passos horários. Diz-se, portanto, que a largura de atributos da janela X_i é $w_X = 5$

e seu comprimento é de n_X horas. Vide a formalização de X_i na Expressão 1, com os índices de submatrizes superescritos.

$$X_i = \begin{bmatrix} Tmp_X^1 & Hr_X^1 & Pr_X^1 & Rd_X^1 & Plv_X^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Tmp_X^{n_X} & Hr_X^{n_X} & Pr_X^{n_X} & Rd_X^{n_X} & Plv_X^{n_X} \end{bmatrix} \quad (1)$$

Analogamente, a saída Y possui $w_Y = 4$ por excetuar a pluviometria de sua largura. É fixado $n_Y = 24$, uma vez que se almeja um prognóstico diário discretizado horariamente. Vide a Expressão 2 abaixo, bem como a representação do janelamento na Figura 1. Nesta, cada janela caminha adiante em passos horários e compõe uma grandeza amostrada de X_i , se traseira, ou de Y_i , se frontal.

$$Y_i = \begin{bmatrix} Tmp_Y^1 & Hr_Y^1 & Pr_Y^1 & Rd_Y^1 \\ \vdots & \vdots & \vdots & \vdots \\ Tmp_Y^{24} & Hr_Y^{24} & Pr_Y^{24} & Rd_Y^{24} \end{bmatrix} \quad (2)$$

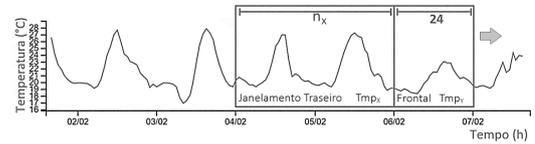


Fig. 1. Janelamentos deslizantes $Tmp_X \in X_i$, $Tmp_Y \in Y_i$ sobre uma série de temperatura. Janeiro de 2020, Estação 001 - Brasília [4].

Destarte, os modelos de hipótese devem aproximar Y com base em X e nos hiperparâmetros variáveis h de uma transformação multivariada $T\{\cdot\}$ qualquer. Existem ainda, os hiperparâmetros fixos h_f , sendo estes previamente estabelecidos por haver aporte teórico embasando seu valor ou simplesmente por viabilização prática do estudo. Desprende-se que os hiperparâmetros variáveis ótimos (\hat{h}) do modelo devem ser tais que façam $T\{\cdot\}$ gerar \hat{Y} com a função de erro $E(Y, \hat{Y})$ retornando valores mínimos. Vide a Expressão 3.

$$\hat{h} = \underset{h}{\operatorname{argmin}}(E(Y, \hat{Y})) \quad | \quad \hat{Y} = T\{X, h, h_f\} \quad (3)$$

Para evitar sobreajustes comuns a métodos de aprendizado supervisionado, vale considerar uma regularização por dropout em cada amostra $X_i \in X$ durante a fase de treino. Para operá-lo em X_i , deseja-se substituir aleatoriamente elementos por zeros na proporção $0 \leq p_{dpout} < 1$. Para tal, considera-se \odot como o operador de Hadamard e se efetua a atribuição $X_i \leftarrow X_i \odot DP$, tal que DP é a matriz de dropout de mesma ordem (n_X, w_X) que X_i . Vide suas definições formais na Expressão 4, sendo $p_{j,k}$ um variável estocástica binária iterada pelos índices j e k em DP . Deve ser percebido que este dropout é um processo matricial sobre as amostras, válido para qualquer que seja o método supervisionado $T\{\cdot\}$.

$$DP = \begin{bmatrix} p_{1,1} & \cdots & p_{1,w_X} \\ \vdots & \ddots & \vdots \\ p_{n_X,1} & \cdots & p_{n_X,w_X} \end{bmatrix} \quad \forall p_{j,k} \in \{0, 1\} \quad (4)$$

$$p_{dpout} = 1 - \frac{1}{n_X w_X} \sum_{j=1}^{n_X} \sum_{k=1}^{w_X} p_{j,k} \quad \forall p_{j,k} \in DP$$

Por fim, é proposta uma arquitetura LSTM EncDec para operar $T\{\cdot\}$, tendo em vista a performance destes modelos em problemas Seq2Seq (vide Sessão II) e como ela permite variar o tamanho das sequências entrantes a partir da codificação destas em uma matriz intermediária de tamanho fixo. Dentre as várias formas de se organizar esta arquitetura, optou-se pela mais simples e prática: a LSTM-LSTM-Dense.

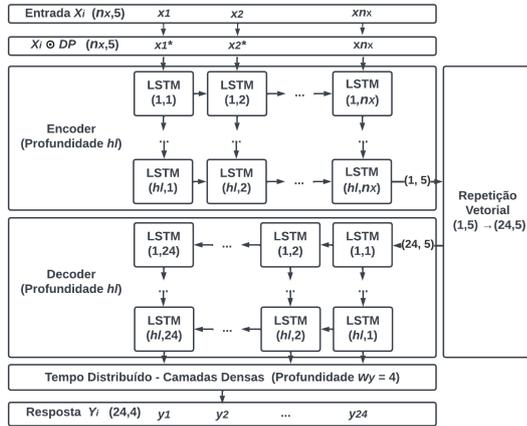


Fig. 2. Arquitetura LSTM-LSTM-Dense proposta pelo autor.

Considerando os valores iniciais $w_X = 5$, $w_Y = 4$ e $n_Y = 24$, vide o diagrama arquitetural da Figura 2. Neste, a primeira LSTM codifica a entrada X_i (n_X , 5) em tamanho fixo (1, 5), sendo posteriormente repetido 24 vezes para gerar uma matriz bidimensional (24, 5). Esta é, então, decodificada na rede LSTM conseguinte e processada por uma rede densamente conectada. A saída desta, agora de ordem (24, 4), compreende uma matriz-resposta Y_i . Percebe-se, portanto, que o arranjo apresentado permitiu transformar a sequência dinâmica (n_X , 5) em uma resposta de tamanho fixo (24, 4).

IV. CENÁRIO DE APLICAÇÃO

A. Análise dos Dados

Trabalhar com previsões através de quaisquer métodos de Machine Learning requer boa compreensão acerca dos dados, sua distribuição e origem. Almejando alta confiabilidade e acessibilidade acerca de dados do cerrado brasileiro, escolheu-se o repositório público da Estação Meteorológica Automática 001, em Brasília – DF, disponibilizados através do [4], de janeiro de 2014 a janeiro de 2020.

Filtrando-se as grandezas de interesse para B , são observadas as distribuições da Figura 3. Nesta, há um comportamento similar a $N(\mu, \sigma)$ para os dados de pressão, abreviados como "pres". No entanto, tal gaussianidade não ocorre tão caracteristicamente para os valores de temperatura, umidade, irradiância ou pluviometria ("temp", "hum", "rad" e "pluv", respectivamente), estes últimos apresentando diversos outliers.

Métodos de escalonamento são famosos por ajudar na convergência de redes neurais, sendo anteriores à etapa de janelamento e redistribuição de valores de B . Estes retêm a função de escalonamento $Esc(\cdot)$ e sua inversa, ambas parametrizadas pelas estatísticas dos dados iniciais para a uniformidade das

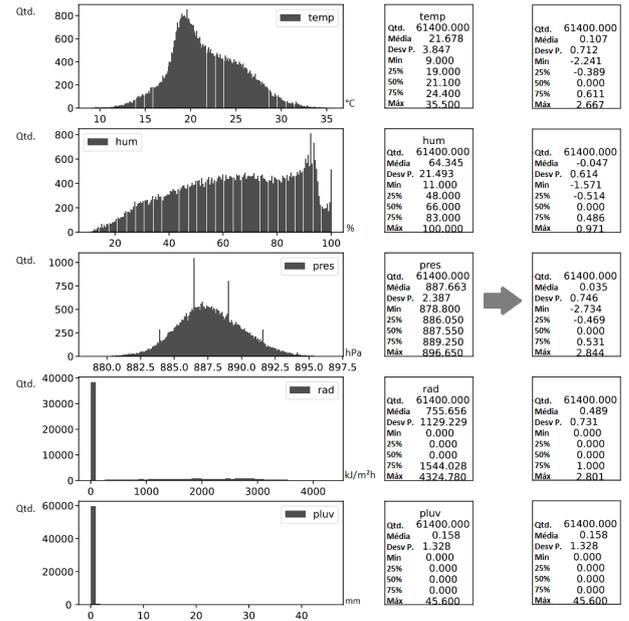


Fig. 3. Distribuições dos dados com sumários antes e pós escalonamento.

transformações futuras. Este processo requer, no entanto, uma análise de distribuições. Considere, portanto, B_g como o conjunto dos elementos ordenados e filtrados pela g -ésima grandeza de B . Dentre várias $Esc(\cdot)$ possíveis, escolhas comuns são a padronização $[Pad(\cdot)]$, para distribuições gaussianas, o escalonamento robusto $[Rob(\cdot)]$ e o método de mínimos e máximos $[MM(\cdot)]$. Vide-os na Expressão 5 a seguir, sendo Q o valor de cada quartil, σ o desvio padrão e \bar{B}_g sua média.

$$\begin{aligned}
 Pad(B_g) &= \frac{b_g - \bar{B}_g}{\sigma} \\
 Rob(B_g) &= \frac{b_g - Q_1(B_g)}{Q_3(B_g) - Q_1(B_g)} \\
 MM(B_g) &= \frac{b_g - \min(B_g)}{\max(B_g) - \min(B_g)}
 \end{aligned} \quad (5)$$

Considerando-se a elevada quantidade de outliers, bem como a sensibilidade da padronização e da transformação por mínimos e máximos, é de se esperar a preferência pela opção robusta $Rob(\cdot)$. Vide, ainda na Figura 3, o sumário estatístico de cada grandeza antes do escalonamento e após este processo. Devido à alta quantidade de zeros nos dados pluviométricos, seu intervalo interquartil e mediana são zerados e a transformação robusta não é aplicada a fim de evitar indefinições matemáticas, resultado em um conjunto inalterado. Tem-se, então, que todos os conjuntos $B_g \in B$ são transformados e redistribuídos em B com exceção da pluviometria. Por não estar presente nos janelamentos de Y e, por conseguinte, nas previsões \hat{Y} , tem-se que a inalteração pluviométrica não influi negativamente no cálculo de erro e retropropagação. Explicita-se que, após o treinamento da rede neural sobre uma base de dados transformada, quaisquer novas submatrizes $X_{ig} \in X_i$ deverão estar igualmente escalonadas e produzirão resultados \hat{Y}_{ig} nesta mesma escala. Destarte, é possível operar a inversa $\hat{Y}_{ig} \leftarrow Rob^{-1}(\hat{Y}_{ig})$ a fim de se recuperar a interpretabilidade.

Outrossim, os outliers observados também influem na escolha de $E(Y, \hat{Y})$. Para este, métricas quadráticas implicam uma penalização muito forte, distorcendo as regressões em direção a tais pontos. É, portanto, mais cabida a utilização de outras funções, como o próprio MAE, que não apresenta hipersensibilidade e melhor preserva a interpretabilidade dos erros. Vide suas definições formais na Expressão 6, onde n denota o comprimento das sequências Y e \hat{Y} .

$$\begin{aligned} MAE(Y, \hat{Y}) &= \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \\ MSE(Y, \hat{Y}) &= \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \end{aligned} \quad (6)$$

B. Casos de Teste

Em termos práticos, a arquitetura exposta na Sessão III pode ser implementada em linguagem Python 3 [13] com auxílio da biblioteca Tensorflow 2 [1] a fim de abstrair operações internas dos tensores na rede.

Não sendo viável avaliar todos os hiperparâmetros que otimizem a geração de prognósticos horários \hat{Y} , conjuga-se a teoria com a necessidade material a fim de fixar alguns elementos de h_f e diminuir o escopo de testes. Destarte, seguem as seguintes condições:

- Utiliza-se o MAE enquanto função de erro.
- Quando houver validação ótima em uma determinada época, recuperaram-se os pesos desta iteração em detrimento da época de parada - isto é, implementa-se um model checkpointer [1].
- Caso o erro de aumento por P épocas consecutivas, o treinamento é interrompido e recupera-se o modelo do checkpointer. Isto é, aplica-se um early stopper [1].
- Utiliza-se o NADAM como gradiente de convergência [6].
- Aplica-se treinamento por mini lotes de 32 unidades, sendo este o padrão da biblioteca Tensorflow e defendido como um dos valores ótimos por [10].
- Dos sete anos utilizados como base, destinam-se cinco ao treinamento, um à validação local e um aos testes.

Fixados estes hiperparâmetros, é prudente questionar o quanto a variação dos demais afeta a performance. Visando \hat{h} , executam-se as 48 combinações de testes na Tabela II.

TABELA II
VARIÁÇÕES HIPERPARAMÉTRICAS A SEREM TESTADAS.

hiperparâmetros h	Valores Testados
Profundidade das redes LSTMs (hl)	10, 50, 100, 150
Proporção de dropout (p_{dpout})	0, 0.25, 0.5
Janelamento traseiro (n_X)	24, 48
Bidirecionalidade das camadas LSTM (bid)	0, 1

V. RESULTADOS

Mediante compilação de todas as redes, foram gerados relatórios de erro para cada variação testada - vide Figura 4. Observou-se aumento do MAE conforme camadas ocultas foram adicionadas, explicitando que a solução ótima coincide com o caso de maior simplicidade e levando à preferência por $hl = 10$. Notou-se ainda que, apesar de $n_X = 48$ apresentar

pequena vantagem para $hl = 10$, o janelamento $n_X = 48$ performou melhor em todos os demais, acentuando sua vantagem conforme hl cresce. Em todos os casos, a presença de qualquer dropout foi prejudicial à performance, sugerindo não haver sobreajuste. Por fim, modelos unidirecionais obtiveram pior performance para a maioria dos cenários, exceto em $hl = 50$.

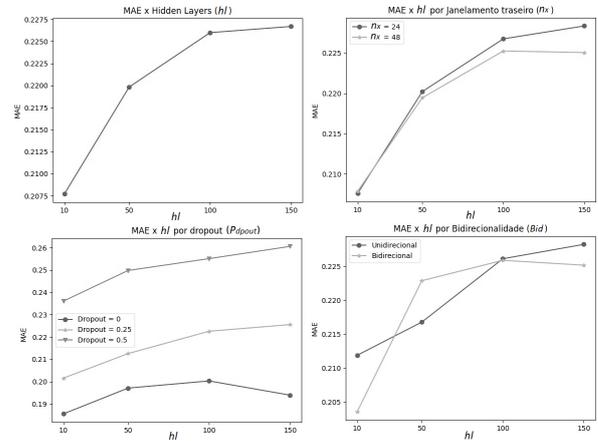


Fig. 4. MAE geral (escalonado) em função dos hiperparâmetros variáveis.

Vide a relação dos dez melhores modelos compilados na Tabela III, ordenados de forma crescente por erro médio absoluto. Tal métrica, nesta tabela, denota uma avaliação média geral da compilação com dados ainda escalonados, sendo adimensional e não distinguindo grandezas. Faz-se necessário reparar, portanto, que pequenas variações sobre os dados transformados influem em grandes discrepâncias na escala original. Outrossim, o MAE unificado possibilita ordenar mais facilmente os modelos por performance. Nota-se, por fim, que a acurácia ótima ocorreu para $\hat{h}_1 = \{hl = 10, p_{dpout} = 0, n_X = 24, bid = 1\}$.

TABELA III
LISTA DE TESTES HIPERPARAMÉTRICOS ORDENADA POR MAE.

	bid	p_{dpout}	hl	n_X	MAE de teste (escalonado)
1	1	0	10	24	0.1748
2	1	0	10	48	0.1805
3	1	0	150	48	0.1881
4	0	0	50	24	0.1917
5	0	0	50	48	0.1919
6	0	0	10	48	0.1920
7	0	0	150	48	0.1945
8	0	0	10	24	0.1948
9	1	0.25	10	48	0.1954
10	1	0	150	24	0.1958

Ao analisar a Figura 5, observa-se que a curva de validação obteve maior MAE em relação à de aprendizado, ambas decrescentes em função das épocas. Esta relação confirma o não sobreajuste e sugere o motivo pelo qual valores de $p_{dpout} > 0$ foram prejudiciais. Faz-se mister considerar que este resultado de dropout é influenciado pela quantidade de dados fornecidos, podendo performar de forma diferente em bases de dados reduzidas. Ainda em tal análise, seu sumário de erros traz um MAE de 1.32 °C, 7.14%, 0.629 hPa e 204.4 kJ/m²h (56.7 W/m²) para as previsões diárias de temperatura, umidade, pressão e irradiância solar. O RMSE, por sua vez, foi de 1.7°C, 9.6%, 0.827 hPa e 421 kJ/m²h (117 W/m²).

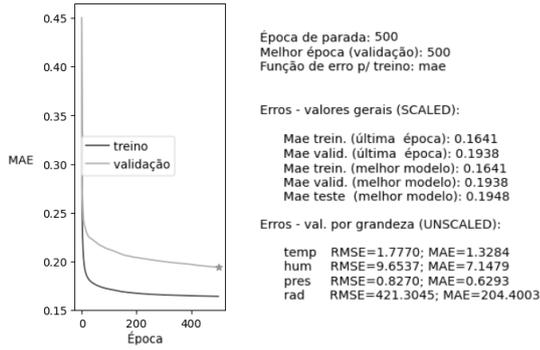


Fig. 5. MAE em função das épocas e sumário de erros para o modelo ótimo.

VI. CONCLUSÃO

Embora [14] seja o trabalho com maior similaridade arquitetural para comparação de resultados, tem-se que suas métricas foram aplicadas aos dados ainda escalonados, dificultando a comparação com trabalhos cujos dados foram escalonados por outras funções em outras bases de dados. Similarmente, [2] apresenta previsões discretizadas diariamente, incompatibilizando a equiparação numérica.

TABELA IV

RMSE COMPARATIVO DE TRABALHOS BASEADOS EM LSTM HORÁRIAS.

		<i>Tmp</i> (°C)	<i>Pr</i> (hPa)	<i>Rd</i> (W/m ²)	<i>Hr</i> (%)
Kreuzer	(ConvLSTM)	2.10	-	-	-
Jingwei	(ConvLSTM)	1.97	-	-	-
Qing	(LSTM)	-	-	122.7	-
Ozbek	(LSTM)	-	-	-	7.51
Este Trabalho	(ED-LSTM)	1.77	0.827	117	9.65

Analisando outros trabalhos de ponta que implementam LSTMs para meteorologia em resolução horária, porém fora do formato Encoder-Decoder, notam-se as ConvLSTM (Convolutional LSTM) de [8] e [9], e as LSTMs de [5], [12] e [11]. Destarte, [8] explicita RMSE de 1.97 °C para previsão horária do próximo dia e [9] de 2.10 °C, confirmando a qualidade do valor obtido pelo presente documento (1.77 °C). O erro mensurado na irradiância (117 W/m²) é comparável ao de 122.7 W/m² encontrado pelos estudos de [12]. O RMSE barométrico obtido (0.827 hPa) também apresentou boa acurácia, mas possui difícil comparação com os estudos de [5], cuja métrica pontuou 0.29 hPa. Embora apresente discretização horária, suas previsões são no escopo de uma única hora adiante, contrastando com o horizonte de 24 passos dado pelo estudo presente. Por fim, a única grandeza aquém dos trabalhos analisados foi a umidade relativa, no qual [11] apresenta RMSE de 7.51%, inferior aos 9.65% aqui obtidos. Vide a Tabela IV, que sumariza os dados supracitados por cada autor excetuando-se [5], devido a diferenças de escopo.

Faz-se mister notar que, enquanto os referidos trabalhos propõem redes dedicadas a cada grandeza, este as prevê de forma conjunta no escopo de Brasília (DF), utilizando um valor significativamente baixo de camadas ocultas ($hl = 10$). Destarte, esta produção apresenta rede única de baixo custo computacional para previsão multivariada, comprovando não apenas sua eficiência, mas sobretudo sua aplicabilidade neste

exemplar do cerrado brasileiro - localização inédita para estudos de prognósticos climáticos baseados em LSTM.

Por fim, o código desenvolvido pode ser publicamente acessado pelo repositório [3]. A este modelo, deu-se o nome de "MZDN-HF", acrônimo para Meteorologic Zone-Delimited Neural Networks - Hourly Forecaster (Redes Neurais Meteorológicas Delimitadas por Zona - Preditor Horário). Nesta, consta não apenas o algoritmo para geração das redes, mas também os seus arquivos de pesos e formato, facilitando integrações externas. A compilação do modelo disponibilizada pode ser diretamente reaplicada a futuras evidências da Estação 001 [4]. Para outros pontos de Brasília e suas proximidades, esta compilação pode ser utilizada mediante regulação barométrica (devido a diferenças de altitude) e menores ajustes de equivalência microclimática entre os dados da localização treinada e os que se desejam inserir e prever. Por outro lado, para localidades consideravelmente distantes, uma nova compilação desta arquitetura com dados próprios é sugerida, bem como novas análises de performance e aplicabilidade.

Considera-se, para iterações futuras deste trabalho, a inclusão de mais grandezas entrantes ($w_X > 5$), tais como dados que explicitem a sazonalidade anual, ou até mesmo o treinamento com dados de diversas estações próximas para abstrair as variações microclimáticas entre elas. De tal forma, é possível que a querela da barometria seja mitigada pela própria rede neural. Não obstante, há evidente abertura para contornar o problema de *P_{lv}* ser incluso como grandeza-submatriz de *Y* (vide Sessões I e III) ou tratá-la em modelo próprio.

REFERÊNCIAS

- [1] Abadi, M. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. (2015), <https://www.tensorflow.org/>, Software available from tensorflow.org
- [2] Alakkari Modelling Weather Conditions Using Encoder-Decoder and Attention Based on LSTM Deep Regression Model. *International Journal Of Advances In Applied Computational Intelligence*. **1** pp. 08-29 (2022,1)
- [3] Alek, T. THOM-32-DATALAB. *GitHub Repository*. (2024), <https://github.com/redstaralek/THOM-32-DATALAB/tree/018cac93ded9e3e1489a30d9fee4b2c35dd759e1>
- [4] BDMEP Banco de Dados Meteorológicos do INMET. (<https://bdmep.inmet.gov.br/2024>), Acessado em 14/01/2024
- [5] Bilgin, O., Mağa, P., Vergutz, T. & Mehrkanoon, S. TENT: Tensorized Encoder Transformer for Temperature Forecasting. (2021,6)
- [6] Dozat, T. Incorporating Nesterov Momentum into Adam. *Proceedings Of The 4th International Conference On Learning Representations*. pp. 1-4 (2016)
- [7] INMET Previsão Numérica - Modelo. (<https://portal.inmet.gov.br/servicos/previs%C3%A3o-num%C3%A9rica-modelo,2024>), Acessado em 14/01/2024
- [8] Jingwei Prediction of hourly air temperature based on CNN-LSTM. *Geomatics, Natural Hazards And Risk*. **13** pp. 1962-1986 (2022,12)
- [9] Kreuzer, D., Munz, M. & Schlueter, S. Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany. *Machine Learning With Applications*. **2** pp. 100007 (2020,12)
- [10] Masters, D. & Luschi, C. Revisiting Small Batch Training for Deep Neural Networks. (2018,4)
- [11] Ozbek, A., Ünal, Ş. & Bilgili, M. Daily average relative humidity forecasting with LSTM neural network and ANFIS approaches. (2022,3)
- [12] Qing, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*. **148** (2018,2)
- [13] Van Rossum, G. & Drake, F. Python 3 Reference Manual. (CreateSpace,2009)
- [14] Zaytar, M. Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. *International Journal Of Computer Applications*. (2016)