



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE – UFRN
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO – PPGEEC

A new approach to optimizing aeration using XGB-Bi-LSTM via the online monitoring of oxygen transfer efficiency and oxygen uptake rate

AUTORES: ZIDING AO, HAO LI, JIABO CHEN, JUNLI YUAN, ZHIHENG XIA, JINSEN ZHANG, HUILING CHEN, HONGCHEN WANG, GUOHUA LIU, LU QI,

DISCIPLINA: REDES NEURAIIS E DEEP LEARNING - PPGEEC2321

PROFESSOR: ADRIÃO DUARTE DORIA NETO

ALUNO: WILSON FRANCELINO DE MORAIS JÚNIOR

Introdução

Contexto e Motivação

- A taxa de aeração é essencial para o fornecimento de oxigênio aos microrganismos em Estações de Tratamento de Efluentes (ETE).
- O estudo propõe uma abordagem baseada em **XGB-Bi-LSTM** com monitoramento online de:
 - Eficiência de transferência de oxigênio (OTE)
 - Taxa de consumo de oxigênio (OUR)

Metodologia Proposta

- XGB: identifica os parâmetros mais importantes com base no ganho.
- Bi-LSTM: realiza previsões usando apenas os parâmetros selecionados.
- Modelos comparados: XGB, Random Forest, LightGBM, Gradient Boosting e LSTM.
- Dados operacionais usados são de uma ETE em Suzhou, China.

Estação de Tratamento de Efluentes

- A ETE de Liuhe, com capacidade de 20.000 m³ por dia, localizado em Taicang, Suzhou, China.
- O conjunto de dados para o presente estudo incluiu os dados operacionais das 12h do dia 19 de agosto de 2022 às 12h do dia 20 de outubro de 2022, abrangendo 1.462 leituras horárias.

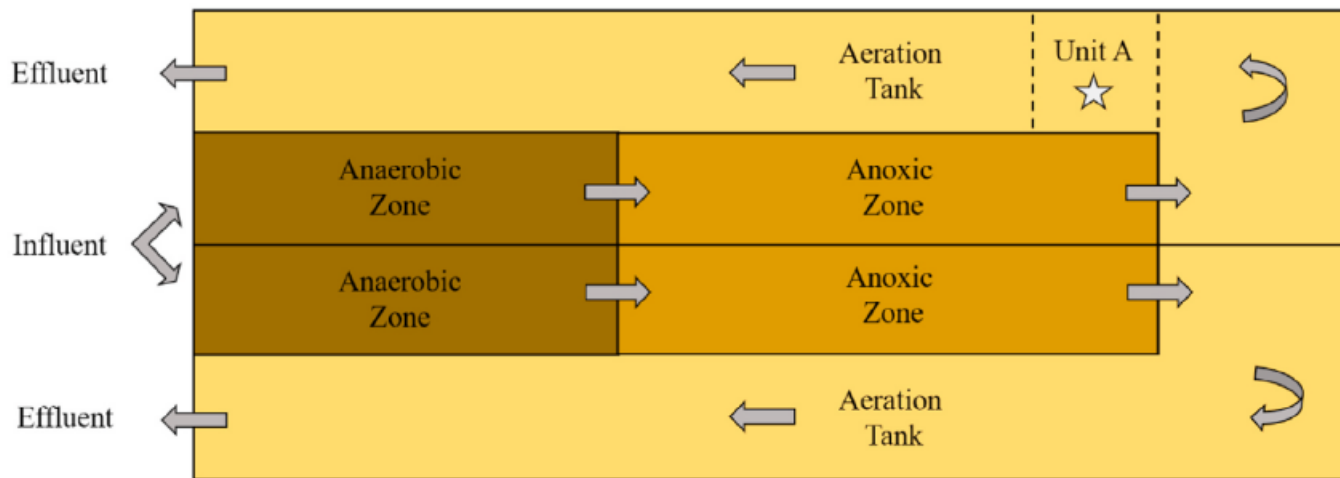


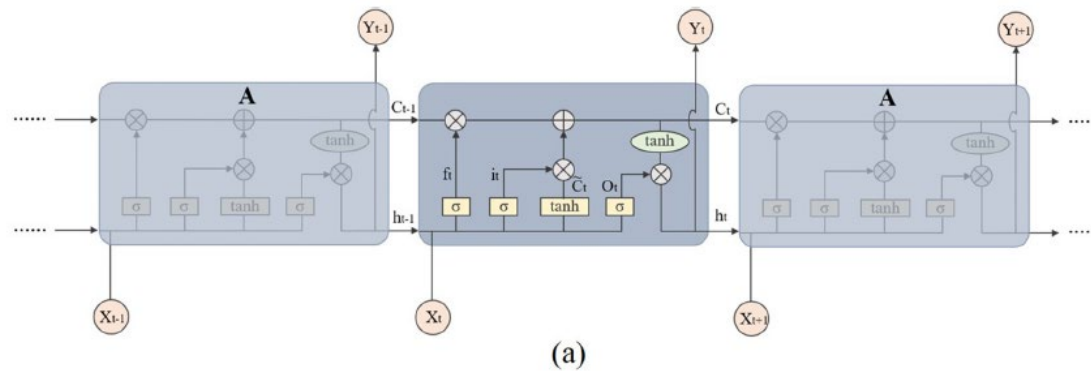
Fig. 2. AAO process diagram.

Extreme Gradient Boosting - XGB

- Algoritmo amplamente utilizado para aumento de gradiente, que pode ser usado para várias tarefas de aprendizado de máquina.
- Um de seus principais recursos é a capacidade de determinar a importância relativa de cada *feature* em um conjunto de dados.
- O XGB calcula a importância do *feature* avaliando o ganho de cada *feature* em cada ponto de divisão em uma árvore de decisão.
- A importância do *feature* i é a proporção de ganhos divididos em todos os nós que usam recurso i .

$$Gain_i = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma$$
$$Importance_i = \frac{\sum_{j \in I_i} Gain_j}{\sum_j Gain_j}$$

Long Short-Term Memory and Bidirectional Long Short-Term Memory



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

$$\begin{cases} i_t = \sigma(W_{it} \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \end{cases}$$

$$\begin{cases} o_t = \sigma(W_{ot} \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \times \tanh(C_t) \end{cases}$$

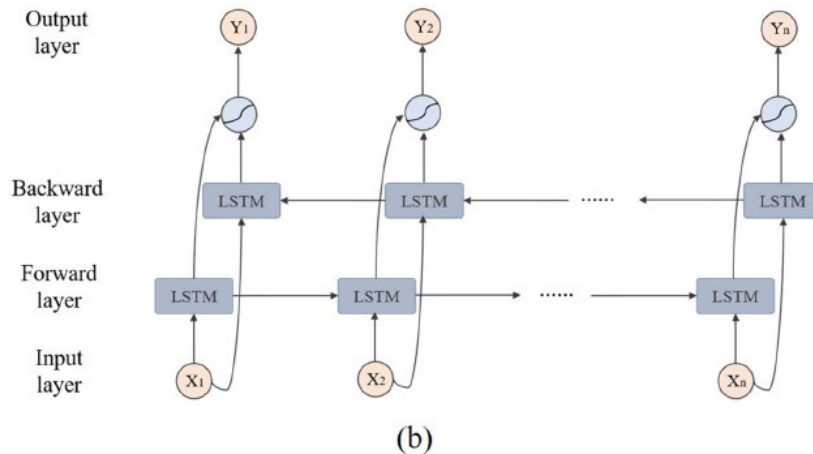


Fig. 1. (a) Cells for LSTM and (b) architecture of Bi-LSTM.

Dataset

The statistical information of the dataset.

	Q ^a	COD	NH ₄ ⁺ -N	TP	pH	OUR	OTE	DO	MLSS	T	TDS	QAIR_tank ^b	Qoff-gas ^c
unit	m ³ h ⁻¹	mg L ⁻¹	mg L ⁻¹	mg L ⁻¹	–	mg L ⁻¹ h ⁻¹	%	mg L ⁻¹	mg L ⁻¹	°C	mg L ⁻¹	m ³ h ⁻¹	m ³ h ⁻¹
count	1462	1462	1462	1462	1462	1462	1462	1462	1462	1462	1462	1462	1462
mean	800.87	151.54	28.41	3.13	6.56	19.14	19.12	2.16	5092.32	26.33	568.32	795.64	1.66
std	52.52	36.66	6.40	0.66	0.23	5.62	2.72	0.55	837.62	0.83	63.61	174.56	0.33
min	662.07	55.63	10.38	1.34	6.09	5.32	14.01	0.82	3557.22	24.59	439.44	340.87	0.93
25%	772.93	126.88	24.69	2.67	6.36	15.52	17.01	1.83	4381.87	25.70	519.99	686.90	1.48
50%	811.27	151.25	28.47	3.09	6.58	19.57	19.01	2.19	4995.84	26.22	553.65	813.40	1.62
75%	838.94	176.25	31.88	3.53	6.73	22.19	21.16	2.49	5559.17	26.63	610.58	919.05	1.84
max	927.55	250.00	45.44	4.94	7.22	32.51	25.00	3.59	7468.90	28.22	741.66	1274.78	2.36

^a Q = total influent flow rate.

^b QAIR_tank = total AFR.

^c Qoff-gas = off-gas flow.

- Q - air flow rate in an aeration unit (m³ h⁻¹)
- COD - Chemical oxygen demand (mg L⁻¹)
- NH₄⁺-N - nitrogénio amónio
- TP - Total Phosphorus (mg L⁻¹)
- pH of the influent
- OUR - Oxygen uptake rate (mg L⁻¹ h⁻¹)
- OTE - Oxygen Transfer Efficiency
- DO - Dissolved Oxygen (mg L⁻¹)
- MLSS - Mixed Liquor Suspended Solids (mg L⁻¹)
- T – Temperature (°C)
- TDS - Total Dissolved Solids (mg L⁻¹)
- QAIR_tank – Total AFR
- Qoff-gas – off-gas flow

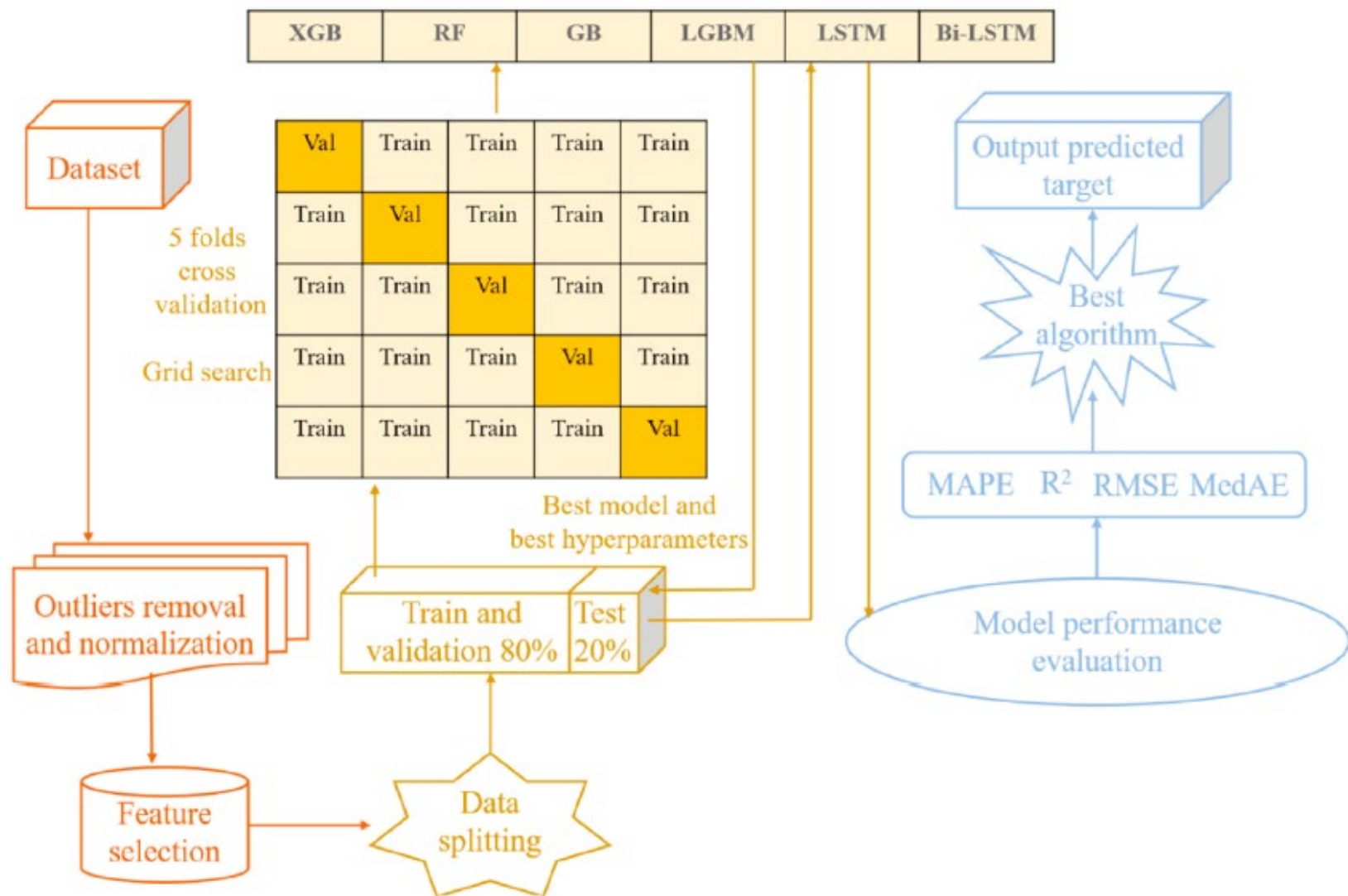


Fig. 3. Modeling workflow.

Avaliação de desempenho do modelo

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2}$$

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

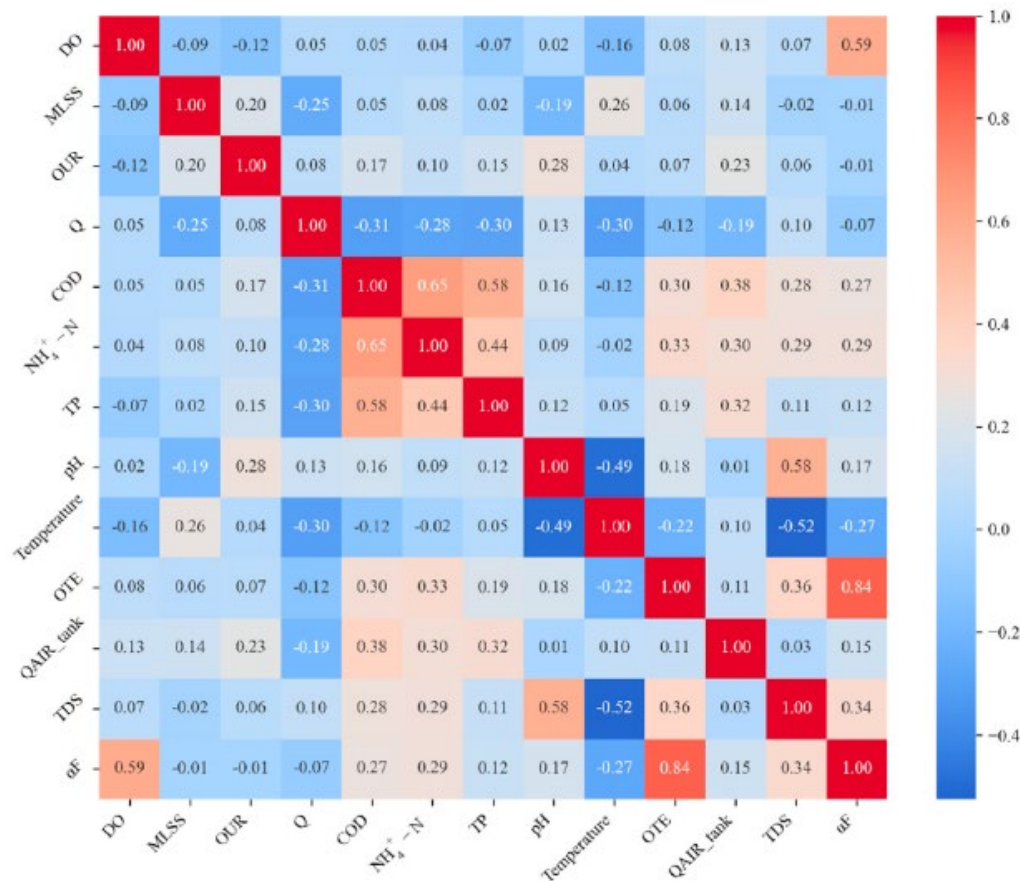
$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$

$$MedAE = median(|\hat{y}_i - y_i|)$$

- R^2 (Coeficiente de Determinação)
- RMSE (Root Mean Square Error – Raiz do Erro Quadrático Médio)
- MAPE (Mean Absolute Percentage Error – Erro Percentual Absoluto Médio)
- MedAE (Median Absolute Error – Erro Absoluto Mediano)

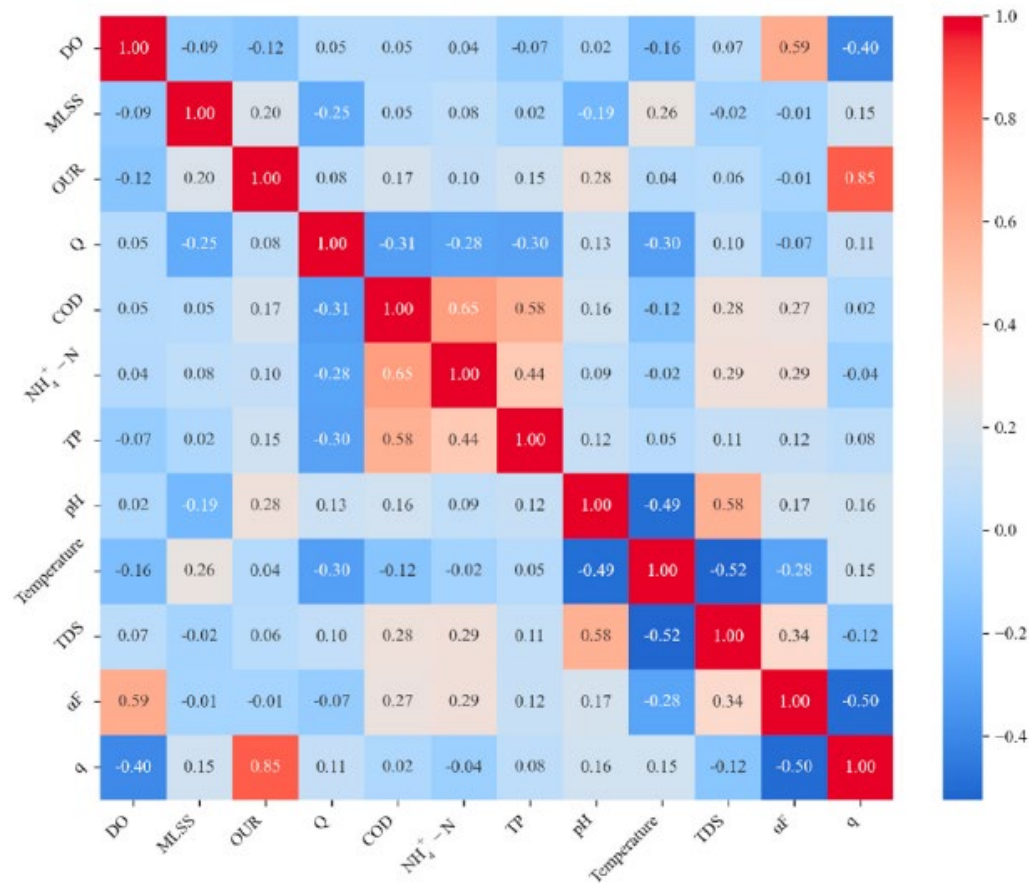
Resultados

- Matriz do coeficiente de correlação de Pearson com aF como alvo.



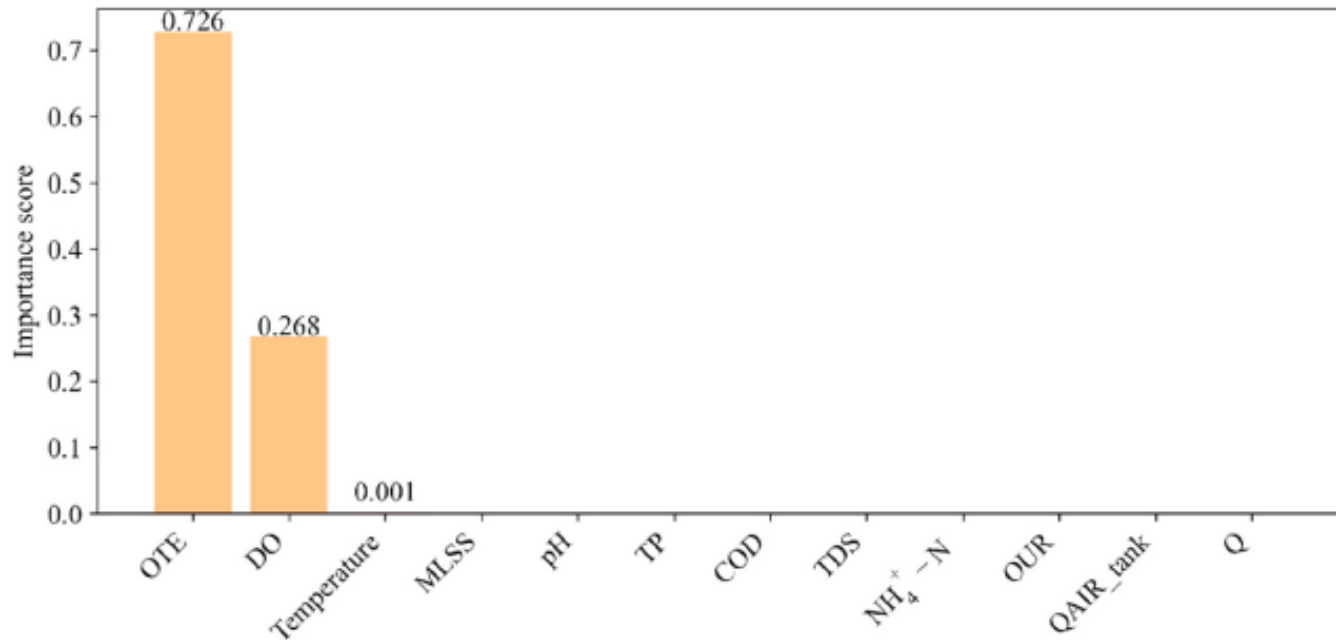
Resultados

- Matriz do coeficiente de correlação de Pearson com q como alvo.



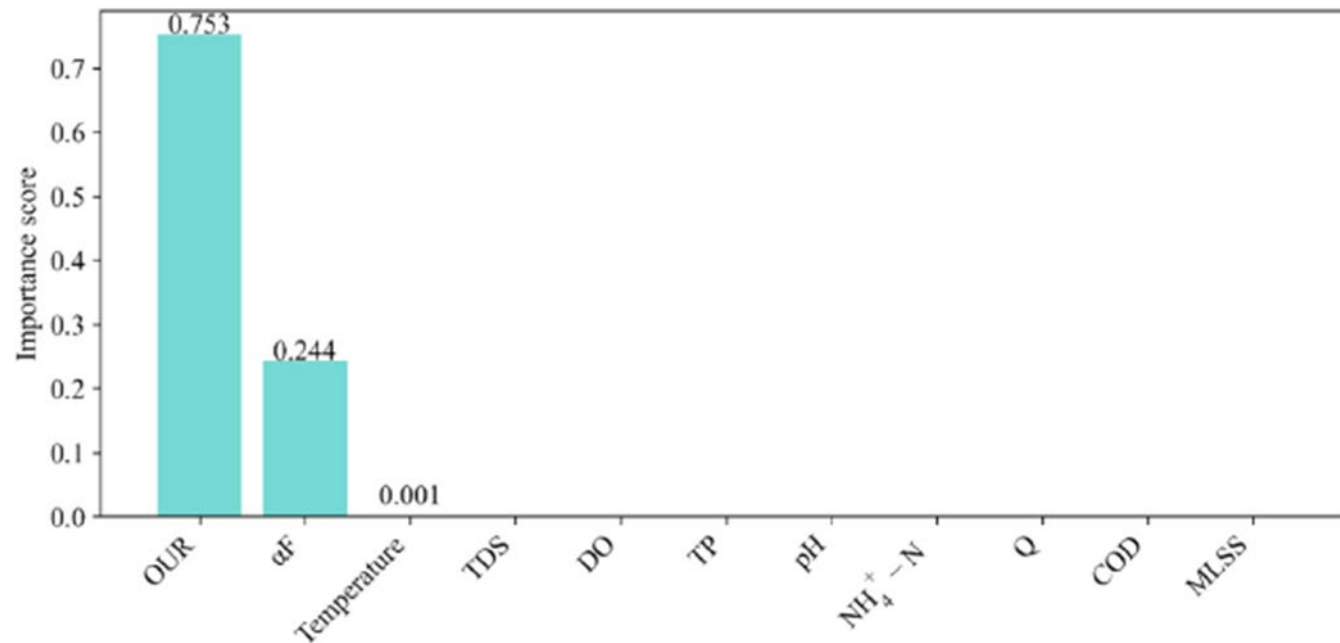
Resultados

- Classificação das pontuações de importância do *feature* para a previsão de aF .



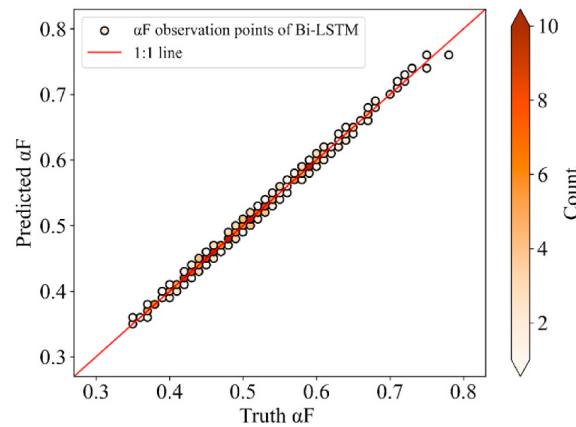
Resultados

- Classificação das pontuações de importância do *feature* para a previsão de q .

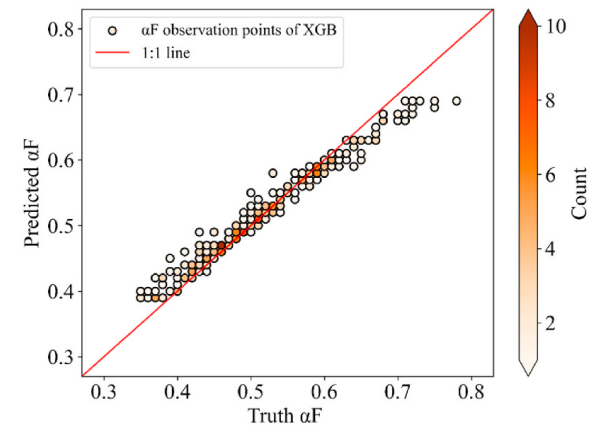


Resultados

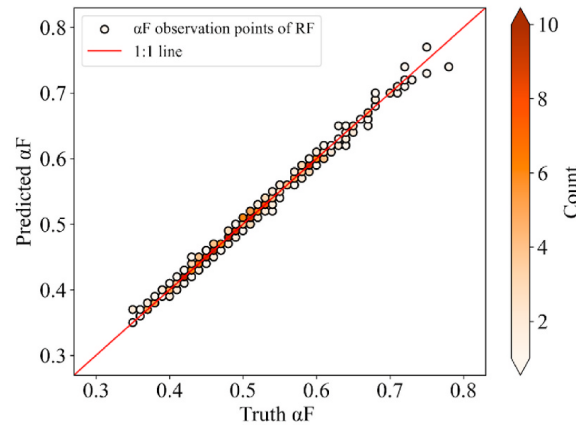
- Gráfico de dispersão de densidade de αF no conjunto de dados de teste.



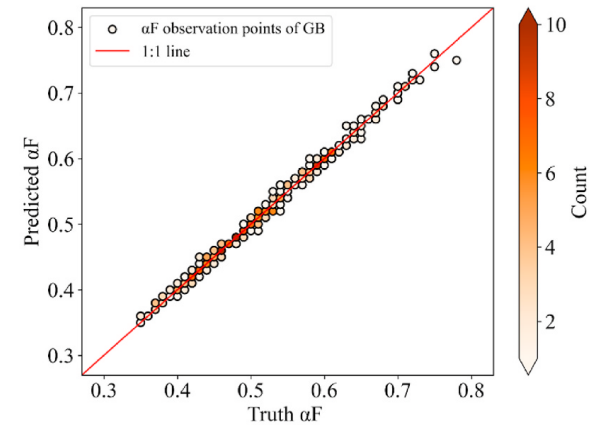
(a)



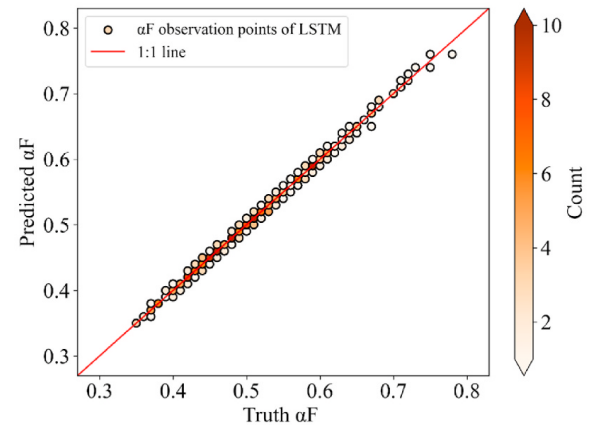
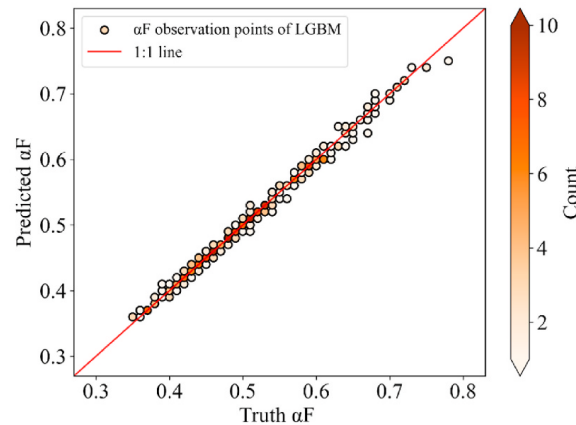
(b)



(c)



(d)



Resultados

- Comparação de desempenho de modelos para previsão de αF no conjunto de dados de teste.

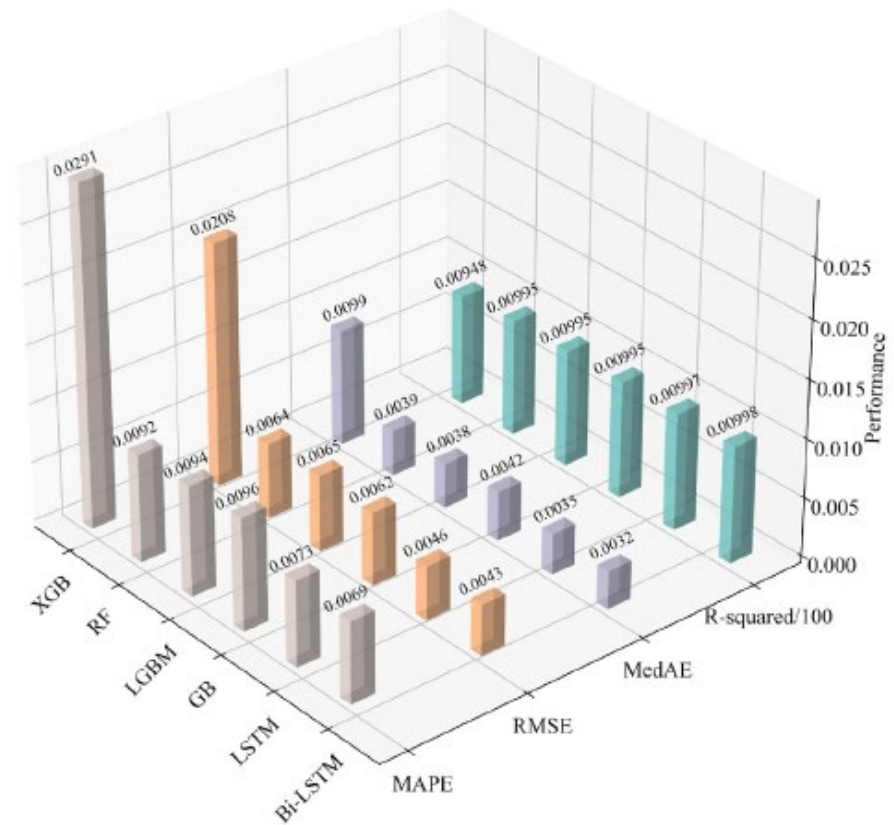


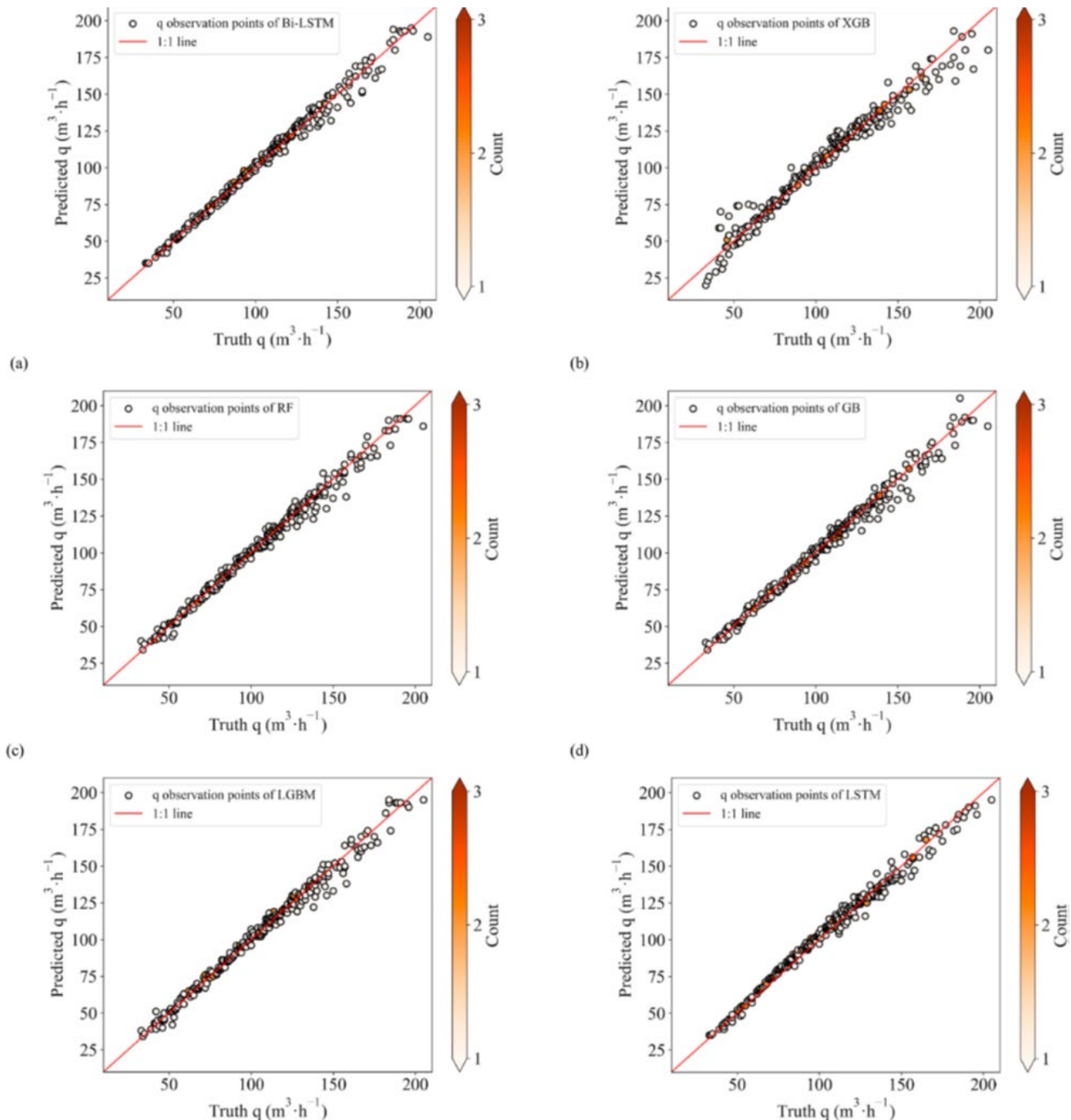
Fig. 5. Performance comparison of models for the prediction of αF on the test dataset.

Statistics of performance metrics for predicting αF .

Metrics	Dataset	Bi-LSTM	XGB	RF	GB	LGBM	LSTM
R ²	Training	0.9979	0.9440	0.9992	0.9985	0.9990	0.9968
	Validation	0.9979	0.9358	0.9951	0.9949	0.9943	0.9968
	Test	0.9977	0.9479	0.9950	0.9954	0.9948	0.9975
RMSE	Training	0.0043	0.0222	0.0026	0.0036	0.0030	0.0053
	Validation	0.0043	0.0238	0.0066	0.0067	0.0071	0.0053
	Test	0.0043	0.0208	0.0064	0.0062	0.0065	0.0046
MAPE	Training	0.0066	0.0302	0.0037	0.0054	0.0040	0.0080
	Validation	0.0065	0.0328	0.0093	0.0099	0.0101	0.0081
	Test	0.0069	0.0291	0.0092	0.0096	0.0094	0.0073
MedAE	Training	0.0030	0.0104	0.0016	0.0023	0.0016	0.0037
	Validation	0.0031	0.0119	0.0038	0.0044	0.0041	0.0036
	Test	0.0032	0.0099	0.0039	0.0042	0.0038	0.0035

Resultados

- Gráfico de dispersão de densidade de q no conjunto de dados de teste.



Resultados

- Comparação de desempenho de modelos para previsão de q no conjunto de dados de teste.

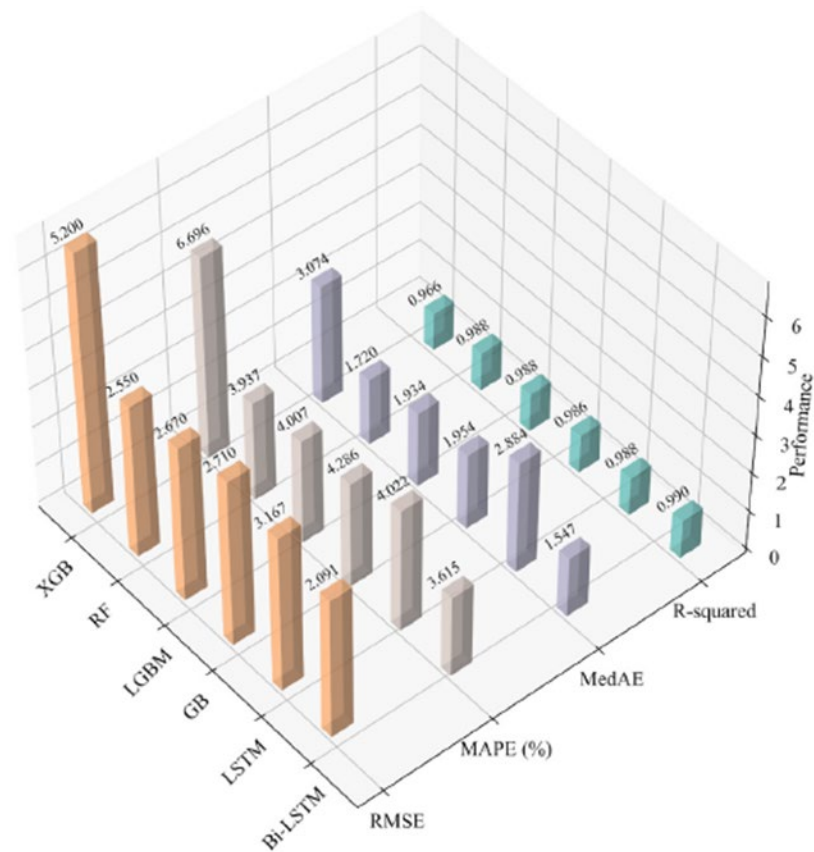


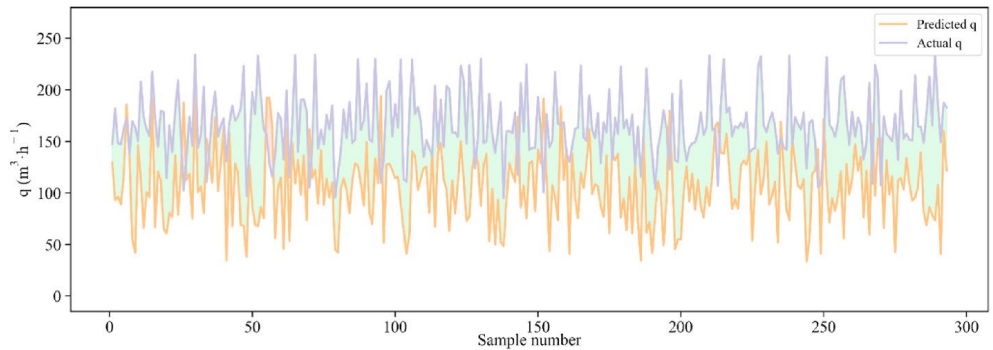
Fig. 7. Performance comparison of models for the prediction of q on the test dataset.

Statistics of performance metrics for predicting q .

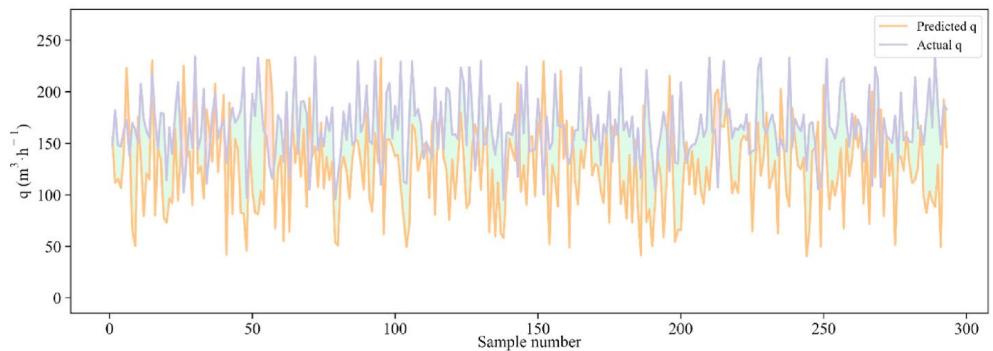
Metrics	Dataset	Bi-LSTM	XGB	RF	GB	LGBM	LSTM
R^2	Training	0.9900	0.9838	0.9975	0.9957	0.9962	0.9897
	Validation	0.9899	0.9680	0.9849	0.9868	0.9846	0.9870
	Test	0.9901	0.9661	0.9883	0.9861	0.9879	0.9878
RMSE	Training	3.7375	4.7498	1.8641	2.4467	2.2742	3.8048
	Validation	3.7509	6.6519	4.5742	4.2905	4.6292	4.5051
	Test	3.6150	6.6963	3.9371	4.2857	4.0073	4.0218
MAPE	Training	0.0229	0.0368	0.0119	0.0170	0.0132	0.0228
	Validation	0.0232	0.0470	0.0292	0.0290	0.0291	0.0268
	Test	0.0209	0.0520	0.0255	0.0271	0.0267	0.0317
MedAE	Training	1.6816	2.4227	0.8450	1.3133	0.8747	1.6683
	Validation	1.7546	2.8243	2.0469	2.1111	2.0194	1.9916
	Test	1.5472	3.0736	1.7199	1.9545	1.9342	2.8843

Resultados

- Comparação da $q_{\text{operação}}$ com o q_{previsto} usando o Bi-LSTM, (a) antes e (b) após aplicação do fator de segurança de 1,2.
- O valor médio do método de otimização de aeração é $106 \text{ m}^3\text{h}^{-1}$, enquanto a média real é $165 \text{ m}^3\text{h}^{-1}$, representando uma redução de aprox. 36%.
- O valor médio de saída da abordagem de otimização com fator de segurança aplicado é $127 \text{ m}^3\text{h}^{-1}$, o que resulta numa poupança de energia de aprox. 23%.



(a)



(b)

Conclusões

- Classificação das pontuações de importância do *feature* para a previsão de q .
- Redução do número de *feature* de entrada necessários para o treinamento.
- Obtenção de um modelo otimizado capaz de reduzir o consumo de energia pode ser reduzido em 23%.

Referências

- ZIDING AO, HAO LI, JIABO CHEN, JUNLI YUAN, ZHIHENG XIA, JINSEN ZHANG, HUILING CHEN, HONGCHEN WANG, GUOHUA LIU, LU QI, ***“A new approach to optimizing aeration using XGB-Bi-LSTM via the online monitoring of oxygen transfer efficiency and oxygen uptake rate”***, Environmental Research, Volume 238, Part 1, 2023, 117142, ISSN 0013-9351.