



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIA
DEPARTAMENTO DE ENGENHARIA MECÂNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

DJAYR ALVES BISPO JUNIOR

**PREVISÃO DE POTÊNCIA PARA AVALIAÇÃO DE CONDIÇÕES
OPERACIONAIS DE TURBINAS EÓLICAS USANDO REDES DO TIPO
TRANSFORMERS**

Recife

2025

DJAYR ALVES BISPO JUNIOR

**PREVISÃO DE POTÊNCIA PARA AVALIAÇÃO DE CONDIÇÕES
OPERACIONAIS DE TURBINAS EÓLICAS USANDO REDES DO TIPO
TRANSFORMERS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Mecânica do Departamento de Engenharia Mecânica, Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como parte dos requisitos parciais para obtenção do título de doutor em Engenharia Mecânica.

Área de concentração: Energia.

Orientador: Profº. Dr. Alvaro Antonio Ochoa Villa

Coorientador: Profº. Dr. Gustavo de Novaes Pires Leite

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Bispo Junior, Djayr Alves.

Previsão de potência para avaliação de condições operacionais de turbinas eólicas usando redes do tipo Transformers / Djayr Alves Bispo Junior. - Recife, 2025.

130f.: il.

Tese (Doutorado) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Programa de Pós-Graduação em Engenharia Mecânica, 2025.

Orientação: Alvaro Antonio Ochoa Villa.

Coorientação: Gustavo de Novaes Pires Leite.

Inclui referências e apêndices.

1. Energia Eólica; 2. Previsão de Potência; 3. Transformer;
4. Time2Vec; 5. FlowAttention; 6. FlashAttention. I. Ochoa Villa, Alvaro Antonio. II. Leite, Gustavo de Novaes Pires. III. Título.

UFPE-Biblioteca Central

DJAYR ALVES BISPO JUNIOR

**PREVISÃO DE POTÊNCIA PARA AVALIAÇÃO DE CONDIÇÕES
OPERACIONAIS DE TURBINAS EÓLICAS USANDO REDES DO TIPO
TRANSFORMERS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Mecânica do Departamento de Engenharia Mecânica, Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como parte dos requisitos parciais para obtenção do título de doutor em Engenharia Mecânica.

Área de concentração: Energia.

Aprovada em: 04/12/2025

BANCA EXAMINADORA

Documento assinado digitalmente

 ALVARO ANTONIO OCHOA VILLA
Data: 23/12/2025 17:41:52-0300
Verifique em <https://validar.itii.gov.br>

Prof. Dr. Alvaro Antonio Ochoa Villa (Orientador)

Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco - IFPE

Documento assinado digitalmente

 JOSE ANGELO PEIXOTO DA COSTA
Data: 19/12/2025 15:30:24-0300
Verifique em <https://validar.itii.gov.br>

Prof. Dr. José Ângelo Peixoto da Costa (Examinador Interno)

Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco - IFPE

Documento assinado digitalmente

 CARMELO JOSE ALBANEZ BASTOS FILHO
Data: 23/12/2025 17:28:26-0300
Verifique em <https://validar.itii.gov.br>

Prof. Dr. Carmelo Jose Albanez Bastos Filho (Examinador Externo)

Universidade de Pernambuco - UPE

Documento assinado digitalmente

 PEDRO ANDRE CARVALHO ROSAS
Data: 22/12/2025 19:09:37-0300
Verifique em <https://validar.itii.gov.br>

Prof. Dr. Pedro André Carvalho Rosas (Examinador Externo)

Universidade Federal de Pernambuco - UFPE

Documento assinado digitalmente

 LEANDRO MACIEL ALMEIDA
Data: 23/12/2025 11:03:16-0300
Verifique em <https://validar.itii.gov.br>

Prof. Dr. Leandro Maciel Almeida (Examinador Externo)

Universidade Federal de Pernambuco - UFPE

AGRADECIMENTOS

Primeiramente, agradeço a Deus por todos os desafios que enfrentei ao longo da minha vida. Sou profundamente grato por cada conquista e por cada degrau que pude subir.

A toda a minha família — em especial ao meu pai, Djair Alves Bispo; à minha mãe, Maria de Fátima Ribeiro Bispo; ao meu irmão, Thyciano Ribeiro Bispo; à minha cunhada, Carla de Araújo; e à minha sobrinha, Laura de Araújo Ribeiro Bispo — deixo minha eterna gratidão por todo o investimento, suporte, paciência, incentivo, dedicação, confiança e amor ao longo de toda a minha trajetória pessoal e profissional.

A todos os meus amigos de vida.

Ao meu orientador, Alvaro Antonio Ochoa Villa, e ao meu coorientador, Gustavo de Novaes Pires Leite, agradeço por todos os ensinamentos, conselhos, apoio, confiança, compreensão, cuidado e paciência.

Ao professor Enrique Lopez Drogue, da UCLA, por me receber e me oferecer a oportunidade de realizar um intercâmbio nos Estados Unidos durante o doutorado.

Aos professores da banca; a todos os professores, colegas e funcionários do PPGEM e do CER da UFPE; aos meus orientadores de mestrado, professor Raphael Abrahão e professora Monica Carvalho, da UFPB; e a todos os professores do PPGEM da UFCG.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).

“Em memória do meu amado pai, Djair Alves Bispo”.

RESUMO

A natureza intermitente e imprevisível dos ventos impõe desafios significativos à operação, despacho e manutenção de parques eólicos, motivo pelo qual métodos de previsão precisos tornam-se fundamentais para ganhos operacionais, como redução de custos, maior disponibilidade das turbinas e integração segura da energia eólica na matriz elétrica brasileira. Partindo desse problema, esta tese propõe novos modelos de previsão de potência eólica de curto prazo baseados em arquiteturas Transformer. Foram desenvolvidos três modelos originais — T2V-Transformer, T2V-Flowformer e T2V-Flashformer — que integram a codificação temporal Time2Vec e empregam mecanismos alternativos de atenção (FlowAttention e FlashAttention) para reduzir a complexidade computacional do FullAttention, mantendo ou ampliando a precisão das previsões. A pesquisa foi conduzida com dados reais de turbinas eólicas localizadas no Nordeste do Brasil, considerando diferentes condições sazonais e horizontes de previsão de até 12 horas. Os resultados evidenciaram ganhos consistentes de desempenho, com os modelos propostos superando os métodos de referência (ARIMA, MLP, LSTM e DLinear) em grande parte dos cenários avaliados, apresentando maior precisão e desempenho nas métricas de avaliação utilizadas. A análise de sensibilidade mostrou que a aplicação do Time2Vec na entrada do codificador gera impactos positivos expressivos, especialmente nas variantes Flowformer e Flashformer. Adicionalmente, observou-se que melhorias aparentemente pequenas se tornam relevantes quando acumuladas em longos períodos de operação, resultando em benefícios concretos à eficiência energética e à confiabilidade operacional. Do ponto de vista metodológico, este trabalho se destaca por apresentar a primeira aplicação do mecanismo FlashAttention à previsão de energia eólica, além da primeira integração conjunta de Time2Vec com múltiplos mecanismos de atenção nesse contexto. A abordagem proposta combina robustez preditiva e viabilidade computacional, oferecendo um procedimento metodológico que avança o estado da arte em previsão de séries temporais eólicas e fornece subsídios práticos para uma gestão mais inteligente e sustentável de parques eólicos.

Palavras-chave: Energia Eólica; Previsão de Potência; Transformer; Time2Vec; FlowAttention; FlashAttention; Séries Temporais.

ABSTRACT

The intermittent and unpredictable nature of wind imposes significant challenges on the operation, dispatch, and maintenance of wind farms, making accurate forecasting methods essential for achieving operational gains such as cost reduction, increased turbine availability, and the secure integration of wind energy into the Brazilian electricity grid. Addressing this problem, this thesis proposes new short-term wind power forecasting models based on Transformer architectures. Three original models were developed—T2V-Transformer, T2V-Flowformer, and T2V-Flashformer—which integrate the Time2Vec temporal encoding and employ alternative attention mechanisms (FlowAttention and FlashAttention) to reduce the computational complexity of FullAttention while maintaining or enhancing predictive accuracy. The research was conducted using real data from wind turbines located in Northeastern Brazil, considering different seasonal conditions and forecasting horizons of up to 12 hours ahead. The results demonstrated consistent performance gains, with the proposed models outperforming reference methods (ARIMA, MLP, LSTM, and DLinear) in most evaluated scenarios, achieving higher accuracy and superior performance across the assessment metrics. The sensitivity analysis revealed that applying Time2Vec at the encoder input yields substantial positive impacts, particularly in the Flowformer and Flashformer variants. Additionally, seemingly small improvements were shown to become highly relevant when accumulated over long periods of continuous operation, resulting in tangible benefits for energy efficiency and operational reliability. From a methodological perspective, this work stands out for presenting the first application of the FlashAttention mechanism to wind power forecasting, as well as the first joint integration of Time2Vec with multiple attention mechanisms in this context. The proposed approach combines predictive robustness and computational feasibility, offering a methodological framework that advances the state of the art in wind time-series forecasting and provides practical support for more intelligent and sustainable wind farm management.

Keywords: Wind Energy; Power Forecast; Transformer; Time2Vec; FlowAttention; FlashAttention; Time Series.

LISTA DE FIGURAS

Figura 1 - Desenvolvimento histórico do total de instalações (GW)	11
Figura 2 - Matriz Energética Brasileira	12
Figura 3 - Arquitetura do Transformer Vanilla	19
Figura 4 - Visão da rede de fluxo para atenção	25
Figura 5 - Mecanismo FlashAttention	27
Figura 6 - Representação do Time2Vec	30
Figura 7 - Componentes principais de TEs para uso industrial.....	33
Figura 8 - Análise de dados na energia eólica	36
Figura 9 - Estrutura de uma MLP	44
Figura 10 - Estrutura de uma LSTM	45
Figura 11 - Ilustração básica de um modelo linear.....	46
Figura 12 - Fluxograma do passo a passo metodológico da revisão de literatura	50
Figura 13 - Documentos por área temática da Etapa 1	51
Figura 14 - Número de estudos desenvolvidos a partir da Etapa 1 (2017-2025).....	52
Figura 15 - Classificação dos 90 artigos incluídos na revisão sistemática de acordo com o desafio investigado: potência eólica, velocidade do vento, detecção de anomalias e desafios combinados.	53
Figura 16 - Representação do procedimento metodológico utilizado neste estudo	65
Figura 17 - Análise de sensibilidade na integração da camada Time2Vec na arquitetura Transformer	66
Figura 18 - Representação para o Estudo de Caso 1 (Cenário A) das dimensões geradas pela camada Time2Vec para uma janela de entrada de 60 horas, utilizada como incorporação temporal na entrada do Transformer	67
Figura 19 - Representação para o Estudo de Caso 2 das dimensões geradas pela camada Time2Vec para uma janela de entrada de 60 horas, utilizada como incorporação temporal na entrada do Transformer	68
Figura 20 - Representações de características Time2Vec aprendidas para a série temporal de energia eólica do Estudo de Caso 2	69
Figura 21 - Arquitetura dos modelos propostos.....	70
Figura 22 - Séries temporais da potência ativa para a turbina eólica do Estudo de Caso 1: Cenário A (acima) e Cenário B (abaixo)	74
Figura 23 - Série temporal da potência ativa da usina eólica do Estudo de Caso 2	76
Figura 24 - Visualização das métricas de avaliação para diferentes horizontes de previsão da potência eólica, para cada modelo avaliado neste estudo. MAE na parte superior, RMSE na parte inferior (Cenário A)	90
Figura 25 - Visualização das métricas de avaliação para diferentes horizontes de previsão da potência eólica, para cada modelo avaliado neste estudo. MAE na parte superior, RMSE na parte inferior (Cenário B)	91
Figura 26 - Potência eólica observada e prevista em uma janela de 24 horas do Estudo de Caso 1 (Cenário A), entre 3 e 4 de março de 2020	93

Figura 27 - Potência eólica observada e prevista em uma janela de 24 horas do Estudo de Caso 1 (Cenário B), entre 25 e 26 de outubro de 2020	93
Figura 28 - Visualização das métricas de avaliação para diferentes horizontes de previsão de potência eólica, para cada modelo avaliado neste estudo. MAE na parte superior, RMSE na parte inferior (Estudo de Caso 2)	99
Figura 29 - Potência eólica observada e prevista em uma janela de 24 horas do Estudo de Caso 2, entre 27 e 28 de setembro de 2020.....	101

LISTA DE TABELAS

Tabela 1 - Resumo e visão geral dos X-formers	28
Tabela 2 - Componentes que influenciam a produção de uma turbina eólica.....	34
Tabela 3 - Alguns dos artigos identificados na pesquisa	54
Tabela 4 - Modelos propostos para previsão de energia eólica	56
Tabela 5 - Resultados dos modelos propostos para a previsão de energia eólica	57
Tabela 6 - Visão geral dos conjuntos de treinamento, validação e teste para cada cenário do Estudo de Caso 1	73
Tabela 7 - Visão geral dos conjuntos de treinamento, validação e teste	75
Tabela 8 - Resultados finais da busca de hiperparâmetros para todos os modelos avaliados no Estudo de Caso 1 (Cenário A)	79
Tabela 9 - Resultados finais da busca de hiperparâmetros para todos os modelos avaliados no Estudo de Caso 1 (Cenário B)	80
Tabela 10 - Resultados finais da busca de hiperparâmetros para todos os modelos avaliados no Estudo de Caso 2	82
Tabela 11 - Comparação do desempenho do modelo para diferentes horizontes de previsão (Cenário A)	86
Tabela 12 - Comparação do desempenho do modelo para diferentes horizontes de previsão (Cenário B)	88
Tabela 13 - Melhorias com a adição do Time2Vec	96
Tabela 14 - Comparação do desempenho dos modelos para diferentes horizontes de previsão (Estudo de Caso 2)	98
Tabela 15 - Melhorias com a adição do Time2Vec (Estudo de Caso 2)	102
Tabela 16 - Valores de nMAE e nRMSE para os modelos avaliados no Estudo de Caso 1 (Cenário A e B) e Estudo de Caso 2, considerando o horizonte de previsão de 12 horas	104
Tabela 17 - Média ± desvio padrão do IoR-RMSE obtidos a partir de 10 execuções independentes para os modelos avaliados nos Estudos de Caso 1 (Cenários A e B) e Estudo de Caso 2, considerando horizonte de 12 h	106
Tabela 18 - Valores de p obtidos nos testes estatísticos t-test e Wilcoxon aplicados à métrica RMSE para os modelos avaliados nos Estudos de Caso 1 (Cenários A e B) e Estudo de Caso 2, considerando o horizonte de previsão de 12 horas.....	107
Tabela 19 - Critérios de avaliação comparativa dos modelos considerando 100 ensaios. O tempo unitário de inferência da melhor configuração é apresentado nas Tabelas 8, 9 e 10	110

LISTA DE SIGLAS E ABREVIATURAS

1D – Unidimensional

ANN – Artificial Neural Network (Rede Neural Artificial)

ARIMA – AutoRegressive Integrated Moving Average (Modelo Autorregressivo Integrado de Média Móvel)

ARMA – AutoRegressive Moving Average (Modelo Autorregressivo de Média Móvel)

BPNN – Back Propagation Neural Network (Rede Neural com Retropropagação)

BiLSTM – Bi-directional Long Short-Term Memory (LSTM Bidirecional)

CAGR – Compound Annual Growth Rate (Taxa Média de Crescimento Anual)

CEEMDAN – Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (Decomposição Empírica em Modo por Conjunto Completo com Ruído Adaptativo)

CGH – Centrais Geradoras Hidrelétricas

CMS – Condition Monitoring System (Sistema de Monitoramento de Condição)

CNN – Convolutional Neural Network (Rede Neural Convolucional)

ELU – Exponential Linear Unit (Unidade Linear Exponencial)

EMD – Empirical Mode Decomposition (Decomposição Empírica em Modo)

f-ARIMA – filtered Autoregressive Integrated Moving Average (Modelo Autorregressivo Integrado de Médias Móveis filtrado)

FFN – Feed-Forward Network

FFT – Fast Fourier Transform (Transformada Rápida de Fourier)

FV – Fotovoltaica

GAT – Graph Attention Network (Rede de Atenção em Grafos)

GCN – Graph Convolutional Network (Rede Convolucional de Grafos)

GNN – Graph Neural Network (Rede Neural de Grafos)

GPU – Graphics Processing Unit (Unidade de Processamento Gráfico)

GRU – Gated Recurrent Unit (Unidade Recorrente com Portões)

GW – Gigawatt

GWEC – Global Wind Energy Council

HBM – High Bandwidth Memory (Memória de Alta Largura de Banda)

IA – Inteligência Artificial

IoR – Improvement over Reference (Melhoria em Relação à Persistência)

kW – Kilowatt

LSTM – Long Short-Term Memory (Memória de Longo e Curto Prazo)

MAE – Mean Absolute Error (Erro Absoluto Médio)

MLP – Multi-Layer Perceptron (Perceptron Multicamadas)

MSE – Mean Squared Error (Erro Quadrático Médio)

MW – Megawatt

NLP – Natural Language Processing (Processamento de Linguagem Natural)

NWP – Numerical Weather Prediction (Previsão Numérica do Tempo)

PCC – Pearson Correlation Coefficient (Coeficiente de Correlação de Pearson)

PCH – Pequenas Centrais Hidrelétricas

ReLU – Rectified Linear Unit (Unidade Linear Retificada)

RMSprop – Root Mean Square Propagation

RMSE – Root Mean Squared Error (Raiz do Erro Quadrático Médio)

RNN – Recurrent Neural Network (Rede Neural Recorrente)

RUL – Remaining Useful Life (Vida Útil Remanescente)

seq len – Sequence Length (Comprimento da Sequência)

SCADA – Supervisory Control and Data Acquisition (Supervisão, Controle e Aquisição de Dados)

SGD – Stochastic Gradient Descent (Gradiente Descendente Estocástico)

SRAM – Static Random Access Memory (Memória Estática de Acesso Aleatório)

STG – Spatio-Temporal Graph (Grafo Espaço-Temporal)

SVR – Support Vector Regression (Regressão por Vetores de Suporte)

T2V – Time2Vec

TE – Turbinas Eólicas

Time2Vec – Time to Vector (Mecanismo de Codificação Temporal)

TW – Terawatt

VMD – Variational Mode Decomposition (Decomposição por Modo Variacional)

VRAM – Video Random Access Memory (Memória de Vídeo de Acesso Aleatório)

WRF – Weather Research and Forecasting Model

X-formers – Transformer e Derivados

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 Desafios a serem enfrentados	12
1.2 Justificativa da Tese	13
1.3 Potencial de Inovação.....	15
1.4 Objetivo Geral	16
1.4.1 Objetivos Específicos.....	16
1.5 Estrutura da Tese	17
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 Redes Transformers	18
2.1.1 Flowformer.....	24
2.1.2 Flashformer	26
2.1.3 Visão geral dos X-formers utilizados neste estudo	27
2.2 Time2Vec: Aprendendo uma representação vetorial do tempo.....	29
3 ESTADO DA ARTE	32
3.1 Turbinas Eólicas	32
3.2 Análise de Dados na Energia Eólica	34
3.3 Séries Temporais.....	37
3.4 Modelos de Previsão.....	40
3.4.1 Modelos Físicos	40
3.4.2 Modelos Estatísticos	41
3.4.3 Modelos de Inteligência Artificial (IA).....	42
3.5 Aplicação das redes Transformers.....	48
3.5.1 Aplicações Gerais	50
3.5.2 Aplicações na Energia Eólica	52
3.5.3 Discussão e Limitações.....	59
3.6 Lacunas Identificadas na Literatura	60
4 METODOLOGIA	62
4.1 Motivação e Formalização da Proposta.....	62
4.2 Visão Geral da Metodologia	62
4.3 Análise de Sensibilidade na Integração Time2Vec	65
4.4 Modelos Propostos	69
4.5 Estudos de Caso	70

4.5.1 Estudo de Caso 1	71
4.5.2 Estudo de Caso 2	74
4.6 Análise Experimental	76
4.6.1 Estudo de Caso 1.....	77
4.6.2 Estudo de Caso 2	81
5 RESULTADOS E DISCUSSÃO	84
5.1 Estudo de Caso 1.....	85
5.1.1 Discussão Física dos Erros de Previsão.....	92
5.1.2 Impacto da integração do Time2Vec no desempenho dos modelos (Estudo de Caso 1)	94
5.2 Estudo de Caso 2.....	97
5.2.1 Discussão Física dos Erros de Previsão.....	100
5.2.2 Impacto da integração do Time2Vec no desempenho dos modelos (Estudo de Caso 2)	101
5.3 Análise comparativa entre o Estudos de Caso 1 e 2	103
5.4 Análise comparativa de desempenho dos modelos, custo computacional e escalabilidade.....	107
6 CONCLUSÃO.....	111
7 PERSPECTIVAS FUTURAS.....	114
REFERÊNCIAS BIBLIOGRÁFICAS.....	115
APÊNDICE A	124
APÊNDICE B.....	125

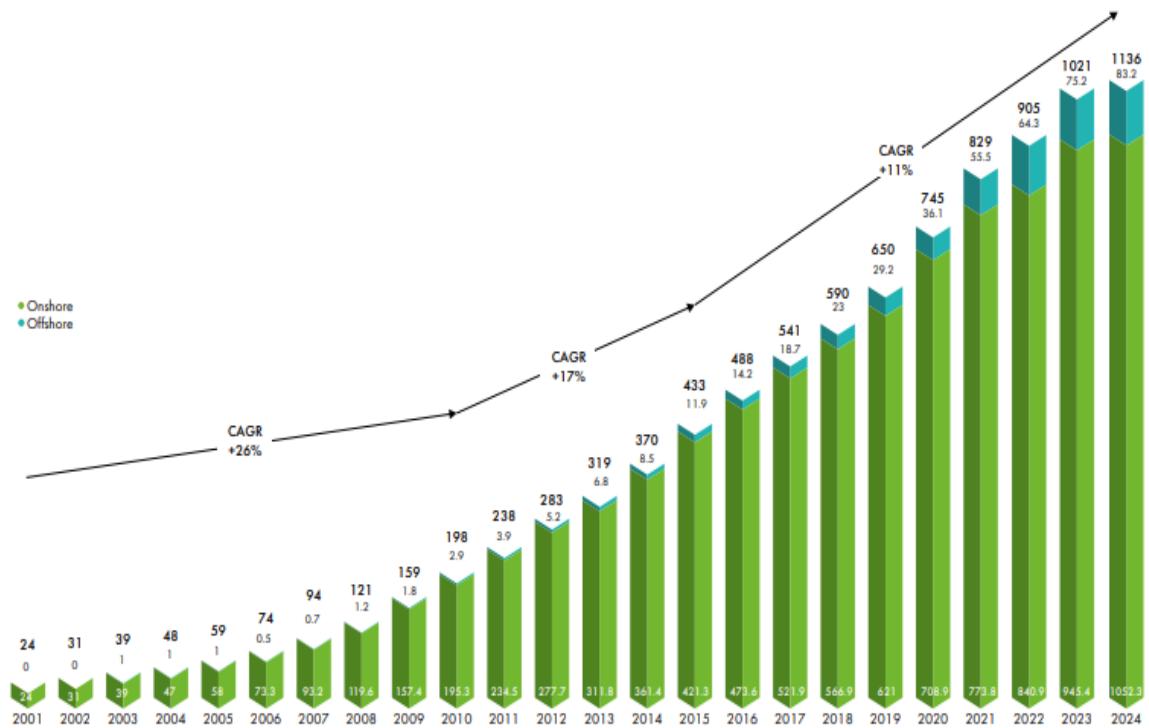
1 INTRODUÇÃO

A energia eólica consolidou-se como um dos pilares centrais da transição da matriz energética global para fontes renováveis, em virtude de seu crescimento acelerado e contínuo ao longo das últimas décadas. Ao configurar-se como alternativa limpa e sustentável aos combustíveis fósseis, desempenha papel estratégico na mitigação das emissões de gases de efeito estufa e na redução dos impactos ambientais associados à geração de eletricidade. A expansão de sua utilização evidencia um movimento global direcionado tanto ao enfrentamento das mudanças climáticas quanto à promoção da segurança energética em longo prazo (ELLABBAN; ABU-RUB; BLAABJERG, 2014; GIELEN *et al.*, 2019; LYDIA *et al.*, 2014).

De acordo com o relatório mais recente da Global Wind Energy Council (GWEC, 2025), a capacidade global de energia eólica atingiu 283 GW em 2012 e 319 GW em 2013. Em 2017 e 2018, essa capacidade já havia se elevado para aproximadamente 540 GW e 591 GW, respectivamente. Essa expansão tornou-se ainda mais expressiva nos últimos anos, alcançando cerca de 830 GW em 2021, 906 GW em 2022, 1.021 GW em 2023 e 1.136 GW em 2024. Projeções futuras indicam que a tendência de crescimento persistirá, impulsionada por investimentos em novas instalações e políticas energéticas de apoio. A Figura 1 ilustra a evolução histórica do total de instalações (GW) de energia eólica ao longo dos anos, onde CAGR significa Taxa de Crescimento Anual Composta.

A expectativa é que esse crescimento atinja 3 TW de capacidade eólica acumulada até 2030. Esse tipo de energia renovável se apresenta como uma solução fundamental para atender às altas demandas por energia elétrica e mitigar os impactos ambientais do planeta. Os cinco principais mercados globais para novas instalações em 2024 foram: China, EUA, Brasil, Índia e Alemanha.

Figura 1 - Desenvolvimento histórico do total de instalações (GW)

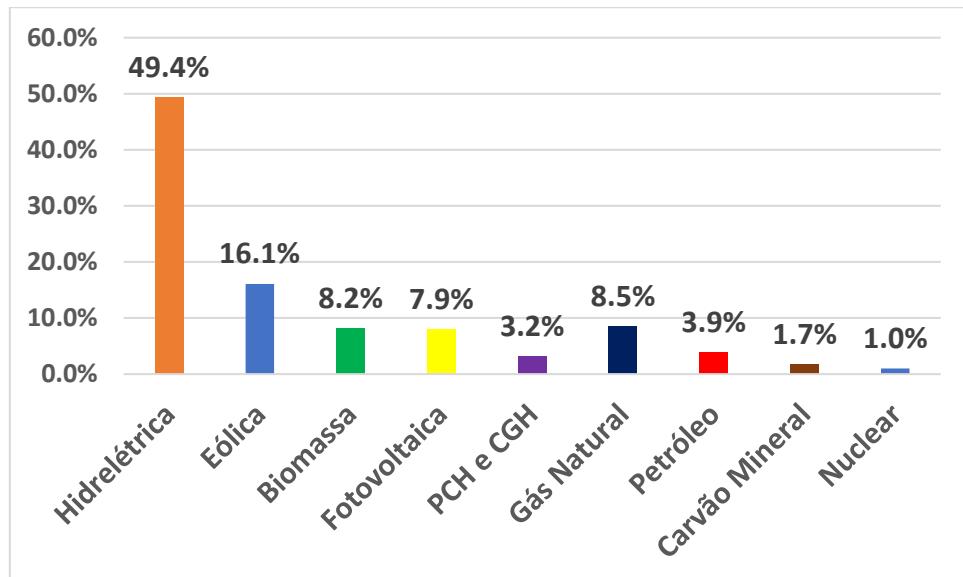


Fonte: (GWEC, 2025).

Em 2024, o Brasil ocupou o quinto lugar no mundo em capacidade instalada total de energia eólica (GWEC, 2025). Durante o mesmo ano, foram instalados 3,3 GW de potência, e a eólica foi a segunda fonte que mais cresceu, representando 10,8% da nova capacidade instalada no ano. A nova capacidade eólica instalada em 2024 fez a fonte eólica atingir uma participação de 16,1% (33,73 GW) da matriz elétrica brasileira, conforme ilustrado na Figura 2, que apresenta a participação de todas as fontes de geração na matriz elétrica brasileira no fim de 2024, de acordo com Associação Brasileira de Energia Eólica E Novas Tecnologias (ABEEÓLICA, 2024). A hidrelétrica representa aproximadamente 49,4% (103,20 GW); biomassa 8,2% (17,1 GW); fotovoltaica 7,9% (16,60 GW); gás natural 8,5% (17,81 GW); petróleo 3,9% (8,13 GW); carvão mineral 1,7% (3,46 GW); nuclear 1,0% (1,99 GW). As fontes PCH e CGH representam respectivamente Pequenas Centrais Hidrelétricas e Centrais Geradoras Hidrelétricas, com aproximadamente 3,2% (6,70 GW).

Com relação a fonte eólica, somente a região Nordeste representou aproximadamente 92,2% de toda a energia eólica produzida na matriz energética nacional (ABEEÓLICA, 2024).

Figura 2 - Matriz Energética Brasileira



Fonte: Adaptado de ABEEÓLICA (2024).

A energia eólica é caracterizada pela conversão de energia cinética, proveniente do vento, em eletricidade. A força dos ventos impulsiona as pás da turbina eólica, que giram a uma determinada velocidade, e o gerador converte essa energia mecânica em energia elétrica (HEIER, 2014). No nível industrial, esse processo de conversão ocorre em larga escala, em parques eólicos *onshore* (em terra) e *offshore* (no mar).

1.1 Desafios a serem enfrentados

A geração de energia eólica envolve diversos desafios técnicos, operacionais e científicos que ainda precisam ser superados, entre os quais destacam-se:

1. Intermitência e Variabilidade: A produção de energia eólica é altamente dependente da intensidade e constância do vento, que apresentam comportamentos variáveis e imprevisíveis ao longo do tempo. Essas flutuações podem resultar em períodos de baixa geração, representando um

- desafio para a previsão precisa da produção e para a operação eficiente da rede elétrica, especialmente em momentos de transição rápida entre diferentes níveis de geração.
2. Custo Inicial e Investimento: O custo inicial de instalação de parques eólicos, incluindo turbinas, infraestrutura e conexão à rede elétrica, pode ser alto.
 3. Durabilidade e Manutenção: As turbinas eólicas exigem manutenção regular e sua durabilidade pode ser afetada por condições climáticas adversas, como ventos fortes e corrosão salina em regiões costeiras.
 4. Impactos Ambientais e Ecológicos: Embora considerada uma fonte de energia limpa, a instalação de parques eólicos pode impactar a vida selvagem (especialmente aves e morcegos) e alterar os ecossistemas locais.
 5. Riscos de Mercado e Regulatórios: A energia eólica pode ser influenciada por mudanças nas políticas de subsídios, regulamentações ambientais e flutuações nos preços de mercado.

Para enfrentar estes desafios é essencial o uso de previsões precisas para otimizar a produção de energia em parques eólicos, bem como o diagnóstico eficiente de falhas em componentes críticos do sistema, a fim de promover os melhores planos de operação e manutenção (LEITE; ARAÚJO; ROSAS, 2018; VEERS *et al.*, 2019; YANG, W.; COURT; JIANG, J., 2013). Esses desafios estão relacionados à análise de dados para previsão e estão frequentemente presentes em parques eólicos *onshore* e *offshore*.

Para fazer previsões precisas sobre a velocidade do vento, o potencial eólico a ser explorado e a detecção de anomalias em componentes críticos de turbinas eólicas (TEs), diversas técnicas, incluindo modelos de aprendizado profundo, podem ser empregadas (este tópico será discutido com mais detalhes ao longo deste trabalho). A análise de dados é essencial para maximizar o desempenho e a confiabilidade de TEs, minimizando os custos operacionais e os impactos ambientais. Com essa abordagem, torna-se possível tomar decisões importantes e garantir a produção eficiente de energia.

1.2 Justificativa da Tese

A análise de dados é um elemento central para a previsão do potencial eólico em parques eólicos, uma vez que a produção de energia está diretamente associada a

fenômenos atmosféricos altamente dinâmicos e não estacionários. Nesse contexto, métodos tradicionais de previsão frequentemente apresentam limitações na captura de padrões temporais complexos, variações abruptas e mudanças de regime características das séries de geração eólica. Assim, a escolha deste tema é justificada pela necessidade de desenvolver e investigar modelos capazes de representar de forma mais eficiente as dependências temporais de curto prazo presentes nos dados de geração eólica, visando aumentar a precisão e a robustez das previsões em cenários reais de operação.

Neste trabalho, o foco está na previsão de curto prazo da potência eólica, em horizontes que variam de poucas horas à frente, os quais são particularmente relevantes para a operação de parques eólicos e para a integração da energia eólica ao sistema elétrico.

Este estudo fundamenta-se na aplicação de modelos de referência amplamente discutidos na literatura científica, com foco principal em arquiteturas baseadas no Transformer. Apesar das limitações associadas ao alto custo computacional, à complexidade do mecanismo de atenção e à representação limitada de padrões multiescalares, os Transformers têm apresentado resultados consistentemente superiores em tarefas de previsão de séries temporais, incluindo aplicações em energia eólica. Essa evidência, somada à sua capacidade de capturar dependências de longo alcance e padrões não lineares, justifica a escolha dessas arquiteturas como base para o desenvolvimento e aprimoramento dos modelos propostos neste trabalho.

Uma dessas limitações está relacionada às codificações posicional senoidais fixas, que podem não capturar de forma adequada características típicas da periodicidade do vento, tais como:

- o ciclo diurno, no qual a velocidade do vento varia de maneira sistemática ao longo do tempo;
- os padrões sazonais irregulares, influenciados por frentes meteorológicas e mudanças atmosféricas de maior escala;
- as flutuações multiescalares, nas quais ciclos curtos e longos interagem de maneira não linear, dificultando sua representação por funções senoidais estáticas.

Além disso, o elevado custo computacional dos Transformers pode resultar em tempos de processamento que, dependendo da configuração, variam de alguns minutos até várias horas por experimento — o que se torna particularmente crítico quando a periodicidade operacional exige atualizações frequentes das previsões. Nesses cenários, um modelo cujo tempo de inferência ou treinamento é maior que a janela de atualização

compromete sua aplicabilidade prática em ambientes industriais. Soma-se a isso o fato de que o Transformer e suas derivações, por se tratarem de modelos relativamente recentes, podem funcionar como uma “caixa-preta”, dificultando a interpretação dos resultados em aplicações reais.

Por fim, um detalhamento técnico aprofundado de todas as limitações mencionadas — incluindo mecanismos de atenção, codificações posicionais e custos computacionais — é apresentado na Seção 2 (Fundamentação Teórica).

Dessa forma, este trabalho busca mitigar essas limitações, com o objetivo principal de aprimorar a precisão da previsão da potência gerada por turbinas eólicas e, simultaneamente, avaliar o desempenho de diferentes abordagens em comparação a modelos clássicos. Ademais, este estudo pretende fornecer uma referência para pesquisadores e engenheiros quanto ao desempenho e potencial dessas arquiteturas modernas em cenários reais de previsão.

Em última análise, ao elevar a acurácia das previsões, pretende-se contribuir para um planejamento operacional e de manutenção mais eficiente, otimizar a produção de energia e maximizar os retornos econômicos, promovendo uma gestão de parques eólicos mais confiável e sustentável.

1.3 Potencial de Inovação

Este estudo investiga a aplicação de modelos ainda não explorados para a previsão de potência em sistemas de geração eólica, adaptando-os para tarefas de séries temporais com o objetivo de superar o desempenho de modelos consagrados na literatura científica.

Dessa forma, a pesquisa se caracteriza como uma das pioneiras na utilização desses modelos para previsão de potência eólica. Ademais, trata-se do primeiro trabalho que, além de propor novos modelos, apresenta uma revisão abrangente sobre a aplicação de arquiteturas baseadas em Transformers no contexto da energia eólica.

Os principais aspectos inovadores desta tese incluem:

1. A utilização dos mecanismos de atenção FlowAttention e Flashformer na previsão de potência de turbinas eólicas, com busca sistemática de hiperparâmetros visando à obtenção da melhor configuração preditiva possível;

2. Este estudo apresenta uma contribuição inédita na literatura consultada ao integrar o Time2Vec à arquitetura Transformer, além da realização de uma análise de sensibilidade com diferentes arranjos;
3. A avaliação do desempenho do Time2Vec combinado a distintos mecanismos de atenção, com o objetivo de aprimorar a precisão e a eficiência preditiva nas tarefas de previsão;
4. A realização de um estudo de revisão comparativo entre distintas derivações do Transformer aplicadas à previsão de séries temporais, com análise quantitativa e qualitativa dos resultados por meio de métricas de avaliação.

1.4 Objetivo Geral

O objetivo geral da tese é maximizar a eficiência produtiva de parques eólicos, gerando previsões precisas de oferta e disponibilidade de energia eólica, permitindo um melhor planejamento da operação e manutenção. O estudo se concentra em modelos de referência, com ênfase em arquiteturas baseadas em Transformers, e utiliza a análise de séries temporais para estimar horizontes de previsão.

1.4.1 Objetivos Específicos

Os objetivos específicos da tese incluem:

1. Realizar uma revisão qualitativa e quantitativa de modelos baseados em Transformers aplicados à área de energia eólica, com o intuito de analisar os resultados alcançados e avaliar os impactos efetivos desses modelos em tais aplicações.
2. Propor modelos baseados em Transformers, testando diferentes mecanismos de atenção voltados à previsão de potência eólica. O objetivo é identificar quais mecanismos proporcionam melhor desempenho preditivo, comparando-os com modelos de referência e conduzindo uma otimização de hiperparâmetros para determinar a configuração ideal de cada modelo.

3. Conduzir uma análise de sensibilidade sobre a integração do Time2Vec na arquitetura Transformer, a fim de avaliar os arranjos que mais contribuem para o aprimoramento preditivo em tarefas de previsão de séries temporais, com foco específico na previsão de potência eólica.
4. Fornecer uma referência estruturada e orientadora para pesquisadores e profissionais interessados no desenvolvimento e aplicação de modelos baseados em Transformers voltados à previsão de energia eólica.

1.5 Estrutura da Tese

Esta tese está estruturada nos seguintes capítulos:

- No Capítulo 1, é apresentada introdução, juntamente com a justificativa da tese, potencial de inovação, além do objetivo geral e específicos para o desenvolvimento deste estudo;
- No Capítulo 2, é apresentada a fundamentação teórica, com foco principal nas redes Transformers e na técnica Time2Vec;
- No Capítulo 3, é apresentado o estado da arte, abordando os tópicos de turbinas eólicas; análise de dados; séries temporais; e modelos de previsão;
- No Capítulo 4, é detalhada a metodologia adotada neste estudo, abordando a descrição do problema; visão geral da metodologia; análise de sensibilidade da integração Time2Vec; modelos propostos; Estudos de Caso 1 e 2; e análise experimental;
- No Capítulo 5, são apresentados os resultados e discussão da tese, abordando os Estudos de Caso 1 e 2; o impacto da integração Time2Vec no desempenho dos modelos e o custo computacional;
- No Capítulo 6, são expostas as conclusões finais do trabalho, ressaltando os principais resultados alcançados.
- No Capítulo 7, são apresentadas as perspectivas futuras, voltadas à expansão e ao aprimoramento dos modelos propostos neste estudo, com aplicações voltadas à previsão de médio e longo prazo, bem como a outras áreas da literatura científica.

2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica é essencial para contextualizar e embasar o desenvolvimento metodológico desta pesquisa. Nesta seção, apresentam-se os conceitos, técnicas e procedimentos que sustentam os modelos propostos, permitindo compreender os princípios por trás das escolhas adotadas. Dessa forma, a fundamentação teórica não apenas explica as ferramentas e métodos utilizados, mas também estabelece o suporte conceitual necessário para interpretar os resultados e justificar cada decisão metodológica adotada ao longo deste estudo.

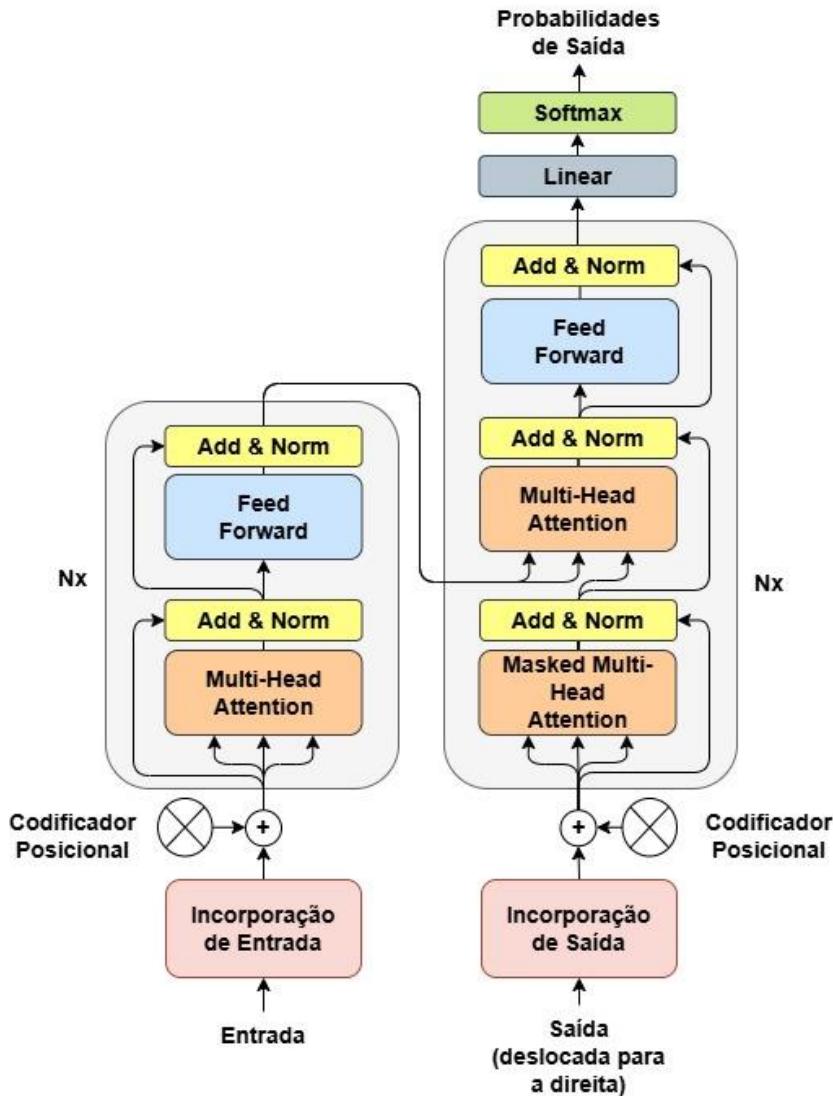
2.1 Redes Transformers

O Transformer é um modelo inicialmente desenvolvido para tarefas de Processamento de Linguagem Natural (*Natural Language Processing* – NLP) (VASWANI *et al.*, 2017), mas tem sido recentemente adaptada para problemas de previsão de séries temporais. O Transformer apresenta uma estrutura de camadas codificador-decodificador, incorporando um mecanismo de auto atenção e camadas totalmente conectadas. Conforme ilustrado na Figura 3, o lado esquerdo do diagrama representa a pilha do codificador, enquanto o lado direito representa a pilha do decodificador. Cada codificador consiste em duas subcamadas principais: (I) um mecanismo de auto atenção multicabeças e (II) uma rede neural *feed-forward* posicional. Ambas as subcamadas são seguidas por conexões residuais e normalização de camadas ('Add & Norm'). As incorporações de entrada são combinadas com codificações posicionais para reter informações de sequência antes de serem alimentadas no codificador.

O decodificador, no lado direito da Figura 3, inclui três subcamadas: (I) uma camada mascarada de auto atenção multicabeças que impede o decodificador de atender a posições futuras, (II) uma camada de atenção multicabeças sobre a saída do codificador (permitindo a interação entre o codificador e o decodificador) e (III) uma rede neural *feed-forward*. Da mesma forma, conexões residuais e normalização são aplicadas após cada subcamada. As incorporações de saída também são combinadas com codificações posicionais e deslocados para a direita para garantir a decodificação autorregressiva. Por

fim, a saída do decodificador passa por uma transformação linear e uma camada *softmax*, a qual normaliza os valores produzidos pelo modelo e os converte em uma distribuição de probabilidades, permitindo a interpretação probabilística das saídas e a seleção da predição mais provável. Na literatura científica, esse modelo também é conhecido como Transformer Vanilla, o primeiro Transformer introduzido na literatura.

Figura 3 - Arquitetura do Transformer Vanilla



Fonte: Adaptado de Vaswani *et al.* (2017).

O mecanismo de auto atenção no modelo Transformer é baseado nos vetores Consulta (Q), Chave (K) e Valor (V), que apresenta representações de *tokens* compactados da sequência de entrada. Esses vetores são organizados nas matrizes $Q \in \mathbb{R}^{N \times D_k}$, $K \in \mathbb{R}^{N \times D_k}$, and $V \in \mathbb{R}^{N \times D_V}$, Onde N denota o comprimento da

sequência, D_k representa as dimensões de Q e de K , e D_V corresponde a dimensão de V . A atenção do produto escalar empregada pelo Transformer é definida de acordo com a Equação (1):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_K}}\right)V \quad (1)$$

O fator de escala $\sqrt{D_K}$ é introduzido para evitar a saturação da função *softmax*, estabilizando os gradientes durante o treinamento ao reduzir a variância dos valores do produto escalar entre Q e K . Em vez de aplicar uma única função de atenção, o Transformer emprega um mecanismo de atenção multicabeça, que permite ao modelo capturar diversos padrões relacionais entre *tokens*. Para atingir esse objetivo, as matrizes originais Q, K, V são projetadas em subespaços de dimensões inferiores usando matrizes de peso treináveis $W^Q \in \mathbb{R}^{D_m \times D_k}$, $W^K \in \mathbb{R}^{D_m \times D_k}$, e $W^V \in \mathbb{R}^{D_m \times D_V}$. O cálculo da atenção para cada uma das cabeças H é então dado pela Equação (2):

$$\text{head}_1 = \text{Attention}(QW_i^Q, KW_i^K \text{ e } VW_i^V) \quad (2)$$

Posteriormente, as saídas de todas as cabeças de atenção H são concatenadas e projetadas de volta à dimensão do recurso original D_m usando uma matriz de projeção final $W^0 \in \mathbb{R}^{HD_V \times D_m}$, resultando na Equação (3):

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^0 \quad (3)$$

Este mecanismo permite que o modelo aprenda múltiplas representações da entrada, aprimorando sua capacidade de capturar dependências contextuais entre *tokens*. Alguns estudos mostram em detalhes a taxonomia das redes Transformers, bem como possíveis modificações que podem ser feitas na arquitetura. Também é possível verificar o pré-treinamento dos modelos e suas respectivas aplicações (LIN, T. et al., 2022).

Além das subcamadas de atenção, cada camada do codificador e do

decodificador contém também uma subcamada de rede totalmente conectada, conhecida como *Position-wise Feed-Forward Network* (FFN). Essa rede é aplicada de forma independente e idêntica a cada posição, consistindo em duas transformações lineares separadas por uma ativação não linear ReLU, conforme a Equação (4):

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (4)$$

As transformações lineares são compartilhadas entre as posições, mas possuem parâmetros distintos entre as camadas. Em termos equivalentes, essa operação pode ser interpretada como duas convoluções com tamanho de *kernel* igual a 1. No Transformer original, a dimensionalidade de entrada e saída é $d_{model}=512$, enquanto a camada interna possui dimensionalidade $d_{ff}=2048$, valores sugeridos no artigo “*Attention Is All You Need*” (VASWANI *et al.*, 2017). Esses valores podem ser ajustados de acordo com a aplicação ou disponibilidade de recursos computacionais, permitindo maior flexibilidade na configuração do modelo. O parâmetro d_{model} representa a dimensão do vetor de representação do modelo, isto é, o tamanho da incorporação em cada posição da sequência. Enquanto d_{ff} corresponde à dimensão da camada intermediária da rede *feed-forward*, responsável por expandir e projetar as representações em um espaço de maior capacidade antes de retorná-las à dimensão original.

Além do mecanismo de atenção, o Transformer original emprega codificações posicionais senoidais fixas (*fixed sinusoidal positional encodings*) para representar a ordem dos elementos na sequência de entrada, como demonstrado na Figura 3. Essas codificações são determinísticas e não aprendíveis, sendo definidas por funções seno e cosseno de diferentes frequências, conforme as Equações 5 e 6:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (5)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (6)$$

onde pos é a posição e i é a dimensão. Cada dimensão da codificação posicional corresponde a uma senoide, cujos comprimentos de onda formam uma progressão geométrica de 2π a $10000 \cdot 2\pi$. Segundo os autores, essa função foi adotada para permitir que o modelo aprenda facilmente a atender por posições relativas, uma vez que, para qualquer deslocamento fixo k , PE_{pos+k} pode ser representado como uma função linear

de PE_{pos} (VASWANI *et al.*, 2017).

Embora as codificações posicionais senoidais fixas sejam adequadas para capturar padrões temporais regulares, séries temporais de potência eólica apresentam estruturas muito mais complexas, incluindo periodicidade diurna moderada, sazonalidades irregulares influenciadas por frentes meteorológicas e flutuações multiescalares decorrentes da interação entre ciclos de minutos, horas e dias. Essas características tornam a representação exclusivamente senoidal menos expressiva para capturar a variabilidade inerente do vento, o que motiva o uso de incorporações aprendíveis e, particularmente, a adoção de funções temporais parametrizáveis, como o Time2Vec.

No contexto desta tese, o termo incorporação refere-se à representação vetorial densa aprendível que transforma a entrada contínua da série temporal em um vetor compatível com a dimensão do modelo d_{model} . Seguindo a formulação adotada na literatura de Transformers para séries temporais (ZERVEAS *et al.*, 2021; ZHOU *et al.*, 2021; WU *et al.*, 2021), cada amostra temporal $x_t \in R^{d_{in}}$ que é projetada para o espaço interno do modelo por meio de uma camada linear aprendível representado pela Equação (7):

$$e_t = W_e x_t + b_e \quad (7)$$

onde $W_e \in R^{d_{model} \times d_{in}}$ e $b_e \in R^{d_{model}}$ são parâmetros ajustados durante o treinamento. Formalmente, essa operação define uma aplicação linear $\phi : R^{d_{in}} \rightarrow R^{d_{model}}$, que projeta cada observação da série temporal para o espaço vetorial interno utilizado pelo Transformer. No caso particular de séries temporais univariadas ($d_{in} = 1$), essa transformação reduz-se a uma projeção direta de $R \rightarrow R^{d_{model}}$. O vetor e_t representa, portanto, a incorporação resultante dessa projeção; enquanto x_t é a observação no instante t ; W_e é a matriz de projeção que mapeia a entrada para o espaço interno do modelo; b_e representa o viés da projeção. Diferentemente das aplicações em NLP, em que incorporações representam relações semânticas entre *tokens*, para séries temporais eles funcionam como uma transformação numérica que captura escala, correlações e padrões estatísticos das variáveis físicas de entrada (por exemplo, potência eólica, velocidade do vento e variáveis meteorológicas). A codificação posicional senoidal é então adicionada a essa incorporação aprendível,

compondo o vetor final de entrada do Transformer, representado pela Equação (8):

$$z_t = e_t + PE(pos_t) \quad (8)$$

onde pos_t indica a posição temporal correspondente ao instante t . A Eq. (8) incorpora simultaneamente informações sobre o valor da série e sua posição temporal. Esse procedimento segue a prática padrão em modelos Transformer aplicados à previsão de séries temporais contínuas (ZERVEAS et al., 2021; ZHOU et al., 2021).

De acordo com a Equação (1), a atenção resulta do produto escalar $(QK)^T$, produzindo uma matriz de pontuação de tamanho $N \times N$ com custo computacional de $O(N^2d)$. A função *softmax* aplicada a $O(N^2)$ amplifica o problema da complexidade quadrática. À medida que N aumenta, o número de operações necessárias cresce quadraticamente, tornando o processamento de sequências longas dispendioso em termos de tempo, computação e memória. Isso se torna ainda mais crítico em aplicações de séries com limitação de tempo, particularmente para sequências de entrada longas.

Dentre essas propostas, destaca-se o Informer (ZHOU, Haoyi et al., 2021), desenvolvido para o processamento de séries temporais de longo alcance. O Informer adota uma atenção probabilística esparsa, que amostra de forma seletiva os *tokens* mais informativos da sequência, reduzindo a complexidade computacional da operação de atenção de $O(N^2)$ para $O(N \log N)$. Além disso, o modelo implementa técnicas de compressão de sequência e camadas convolucionais unidimensionais (1D), responsáveis por realçar padrões locais relevantes nos dados. Para manter a representatividade das informações mais expressivas, o Informer utiliza o *MaxPooling* como mecanismo de seleção, preservando os valores mais significativos e contribuindo para um aprendizado mais eficiente e robusto.

Outro exemplo é o Autoformer (WU, H. et al., 2021). Modelo desenvolvido especificamente para tarefas de previsão de séries temporais, com ênfase na captura de padrões sazonais e tendências de longo prazo. Diferentemente das abordagens baseadas na atenção tradicional, o Autoformer introduz o mecanismo de Auto-Correlação, capaz de identificar diretamente as dependências relevantes nas séries temporais, substituindo o cálculo explícito de atenção entre todos os pares de *tokens*. Essa estratégia permite ao modelo aproveitar de forma mais eficiente as repetições e periodicidades presentes nos dados. Além disso, o Autoformer adota uma arquitetura codificador-decodificador livre

de conexões residuais, projetada para reduzir redundâncias e melhorar a eficiência do treinamento. Graças a essas modificações estruturais, o modelo alcança uma complexidade computacional reduzida, na ordem de $O(N \log N)$.

Na literatura científica, além do Informer e do Autoformer, existem uma infinidade de modelos desenvolvidos para trabalhar com aplicações de séries temporais, como Pyraformer (LIU, S. *et al.*, 2022); Crossformer (ZHANG, Yunhao; YAN, Junchi, 2023); FEDformer (ZHOU, T. *et al.*, 2022); LogSparse Transformer (LI, Shiyang *et al.*, 2019); Flowformer (WU, H. *et al.*, 2022); entre outros. Cada um desses modelos foi projetado para abordar limitações específicas do Transformer Vanilla, demonstrando desempenho competitivo ou superior em tarefas de previsão de séries temporais. Na literatura científica, as variantes dos modelos Transformer são reconhecidas sob o termo X-formers (TAY *et al.*, 2023).

2.1.1 Flowformer

O Flowformer (WU, H. *et al.*, 2022) foi desenvolvido para reduzir a complexidade quadrática do mecanismo FullAttention. O modelo propôs uma reformulação da atenção baseada em modelagem de fluxo contínuo, mecanismo denominado FlowAttention. Diferentemente da atenção padrão, que calcula pesos discretos por meio de produtos escalares entre vetores Q e K , o mecanismo FlowAttention interpreta o processo de atenção como um campo vetorial contínuo de transporte de informações, no qual cada *token* atua como uma fonte que envia informação e um sumidouro que a recebe, conforme ilustrado na Figura 4.

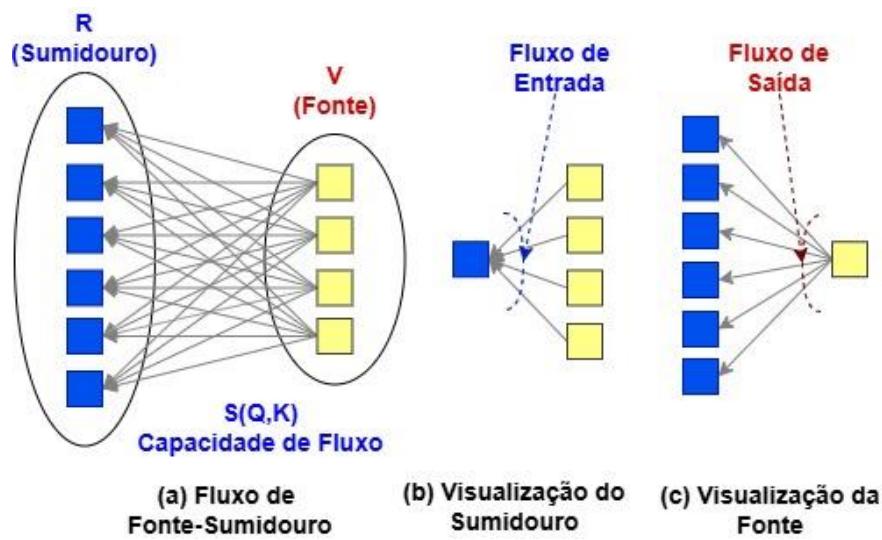
Nessa esquema, o fluxo de atenção é representado como a capacidade de transporte entre fontes e sumidouros, modelada pela função $S(Q, K)$, que define a intensidade do fluxo entre *tokens* com base em suas representações Q e K . O diagrama evidencia duas perspectivas complementares: (b) a Visualização do Sumidouro, que mostra como cada elemento receptor (R) agrupa informações das fontes (V), e (c) a Visualização da Fonte, que descreve como cada elemento emissor distribui sua contribuição para os receptores.

A atenção é formulada como um problema de transporte, onde Q e K são tratados como distribuições de probabilidade. A solução ótima para este problema define o

mecanismo de atenção, denominado FlowAttention. Por fim, o produto escalar $S(Q, K)$ no Transformador Vanilla é substituído por $S = \phi(Q)\phi(K^T)$, onde $\phi(\cdot)$ é uma projeção não linear aplicada elemento a elemento. Devido às propriedades das redes de fluxo, $\phi(\cdot)$ é escolhida como uma função não negativa, de modo a garantir capacidades de fluxo positivas e permitir a interpretação do mecanismo de atenção sob a perspectiva de redes de fluxo. Uma escolha comum na literatura é $\phi(\cdot) = ELU(x) + 1$, onde ELU denota a função de ativação *Exponential Linear Unit* e x representa o vetor de entrada da projeção não linear, conforme adotado no Linear Transformer (KATHAROPOULOS et al., 2020), embora outras projeções não negativas também sejam admissíveis.

Essa reformulação melhora a eficiência computacional e a escalabilidade para sequências longas, além de preservar de forma mais consistente as dependências globais entre os elementos da sequência. O Flowformer demonstrou desempenho competitivo e, em muitos casos, superior ao Transformer tradicional em tarefas de previsão de séries temporais e processamento dinâmico de dados. Entretanto, sua aplicação em domínios energéticos, como a previsão de potência eólica, ainda é pouco explorada, representando um campo promissor para estudos futuros.

Figura 4 - Visão da rede de fluxo para atenção



Fonte: Adaptado de Wu *et al.* (2022).

2.1.2 Flashformer

O modelo é otimizado para previsão de séries temporais e incorpora um mecanismo de atenção eficiente em memória, com consciência de operações de entrada e saída (I/O *awareness*). O uso do algoritmo FlashAttention (DAO *et al.*, 2022) minimiza o número de operações de leitura e escrita entre a *High Bandwidth Memory* (HBM), conhecida como memória de alta largura de banda, e a *Static Random Access Memory* (SRAM), denominada memória estática de acesso aleatório, presentes no chip da GPU.

Como ilustrado na Figura 5, o FlashAttention processa os *tokens* dentro de uma janela deslizante, permitindo capturar dependências locais, essenciais em séries temporais. O algoritmo adota uma estratégia de divisão em blocos (*tiling*) para evitar a materialização explícita da matriz de atenção completa $N \times N$ na HBM, que possui acesso relativamente mais lento. No *loop* externo (setas vermelhas), o FlashAttention itera sobre blocos das matrizes K e V , carregando-os na SRAM rápida *on-chip*. Dentro de cada bloco, ocorre o *loop* interno (setas azuis), no qual segmentos da matriz Q são carregados para a SRAM, e os resultados da atenção são posteriormente gravados na HBM.

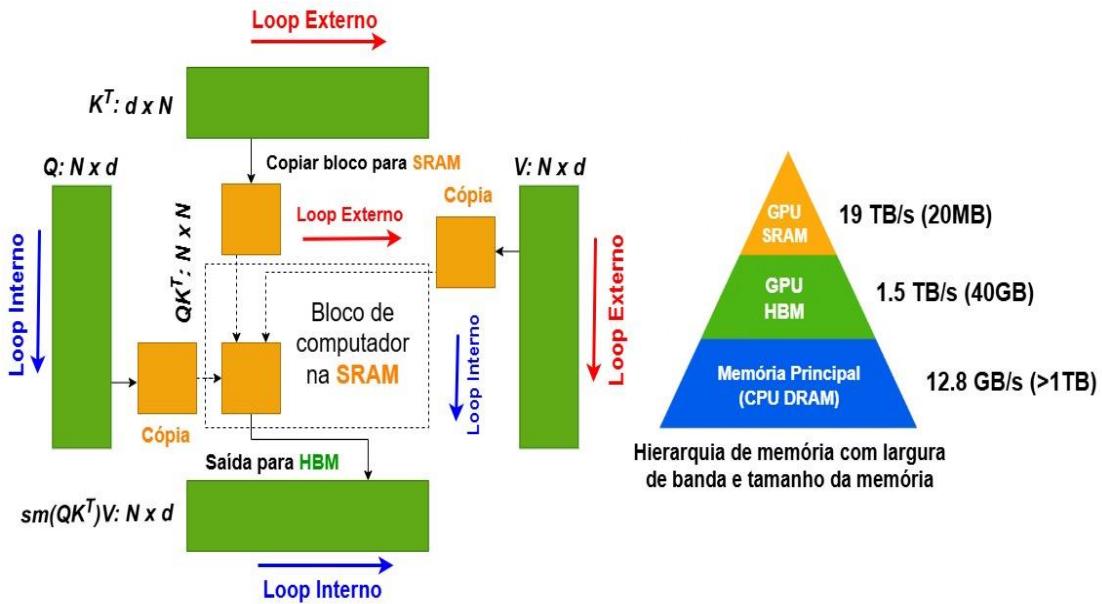
Embora a complexidade aritmética permaneça $O(N^2)$, o FlashAttention reduz substancialmente a complexidade de I/O, limitando o tráfego de memória entre HBM e SRAM. Heuristicamente, essa redução pode ser aproximada por $O(N^2/M)$, onde M representa a capacidade efetiva da memória *on-chip*. O ganho de eficiência exato depende do tamanho dos blocos e da configuração de *hardware*.

Do ponto de vista prático para previsão de potência eólica, essa otimização implica:

1. Menor latência na inferência, favorecendo aplicações que exigem atualização em escalas de minutos;
2. Maior estabilidade numérica e menor necessidade de memória, permitindo treinar modelos de maior profundidade ou sequências mais longas;
3. Capacidade de execução em *hardware* mais acessível, diminuindo o custo computacional de operação em centros de controle de parques eólicos; e
4. Maior escalabilidade quando múltiplas turbinas e janelas temporais precisam ser processadas simultaneamente

O FlashAttention pode substituir diretamente o mecanismo padrão de FullAttention no Transformer Vanilla. Neste trabalho, a variante do Transformer que utiliza o FlashAttention é denominada Flashformer.

Figura 5 - Mecanismo FlashAttention



Fonte: Adaptado de Dao *et al.* (2022).

2.1.3 Visão geral dos X-formers utilizados neste estudo

Conforme discutido anteriormente, os modelos analisados apresentam características distintas, principalmente em relação ao mecanismo de atenção e à complexidade computacional. A Tabela 1 resume brevemente essas diferenças. O Transformer Vanilla apresenta complexidade quadrática com relação ao tamanho da sequência, o que limita sua escalabilidade para séries temporais muito longas. O Flowformer, por sua vez, utiliza o mecanismo FlowAttention para evitar a materialização da matriz de atenção completa, resultando em uma complexidade linear em N , ainda que com constantes computacionais maiores devido às operações adicionais de normalização contínua. Já o Flashformer emprega o algoritmo FlashAttention, que mantém complexidade aritmética quadrática, porém reduz significativamente os custos de I/O,

oferecendo maior eficiência prática — principalmente em GPUs que exploram bem hierarquias de memória e paralelismo. Assim, a escolha do modelo mais adequado deve considerar as características da aplicação, o volume de dados e os recursos computacionais disponíveis, equilibrando desempenho, custo e precisão.

Tabela 1 - Resumo e visão geral dos X-formers

Modelo	Mecanismo de atenção	Complexidade
Transformer	FullAttention	$O(N^2)$
Flowformer	FlowAttention	$O(N)$
Flashformer	FlashAttention	$O(N^2/M)^1$

¹Complexidade aproximada, mais detalhes em Dao *et al.* (2022).

Fonte: Autoria própria (2025).

Na energia eólica, existem diversos estudos que utilizam modelos baseados na arquitetura Transformer, os quais são sistematicamente discutidos e referenciados na Seção 3.5.2. Em (ZHANG, K.; LI, X.; SU, 2022), os autores propuseram o VMD-Transformer (VMD-TF), modelo combinado com a Decomposição de Modos Variacionais (*Variational Mode Decomposition – VMD*), para mitigar os efeitos da não estacionariedade da velocidade do vento, decompondo os sinais em modos estáveis. Os resultados demonstraram que o VMD-TF superou modelos como o VMD-ARIMA e o VMD-LSTM na previsão de curto prazo. Outro modelo proposto foi o FFTransformer (BENTSEN *et al.*, 2023a), que incorpora a decomposição de sinais por meio de dois fluxos para analisar tendências e componentes periódicos, ao mesmo tempo em que captura relações espaço-temporais. O FFTransformer superou o LSTM e o MLP na previsão de velocidade e potência do vento em curto prazo.

Mais recentemente, o mecanismo FlowAttention foi utilizado em (DONG, Z. *et al.*, 2025), sendo proposto o modelo Wind-Mambaformer. O Mamba (GU; DAO, 2024) é uma alternativa ao decodificador Transformer tradicional, projetado para minimizar o acúmulo de erros e aprimorar a adaptabilidade do modelo sob condições variáveis. O Wind-Mambaformer demonstrou desempenho superior na previsão de curto prazo de energia de turbinas eólicas, em comparação com modelos de referência como CNN, LSTM e outros modelos. Com relação ao FlashAttention, ainda não existe na literatura científica

aplicação para previsão de séries temporais na energia eólica, sendo esta tese o primeiro trabalho a tratar desse tema.

2.2 Time2Vec: Aprendendo uma representação vetorial do tempo

O Aprendizado de Características (*Featuring Learning*) visa extrair automaticamente representações informativas de dados brutos, aprimorando o desempenho do modelo por meio da captura de estruturas e dependências subjacentes. No contexto da previsão de séries temporais, o aprendizado de representação temporal desempenha um papel crucial ao permitir que os modelos entendam a periodicidade e a dinâmica temporal de forma eficaz. Dentre as abordagens existentes, o Time2Vec (KAZEMI *et al.*, 2019) se destaca como uma técnica simples, porém poderosa, para codificação de informações relacionadas ao tempo. Ele fornece uma maneira sistemática de representar componentes periódicos e não periódicos de dados temporais, oferecendo uma incorporação temporal mais rica e interpretável para arquiteturas de redes neurais. O Time2Vec adota três propriedades principais:

1. Periodicidade: Captura padrões periódicos e não periódicos nos dados.
2. Invariância na escala temporal: A representação permanece consistente independentemente das variações na escala temporal.
3. Simplicidade: A representação temporal foi projetada para ser simples o suficiente para integração em vários modelos e arquiteturas.

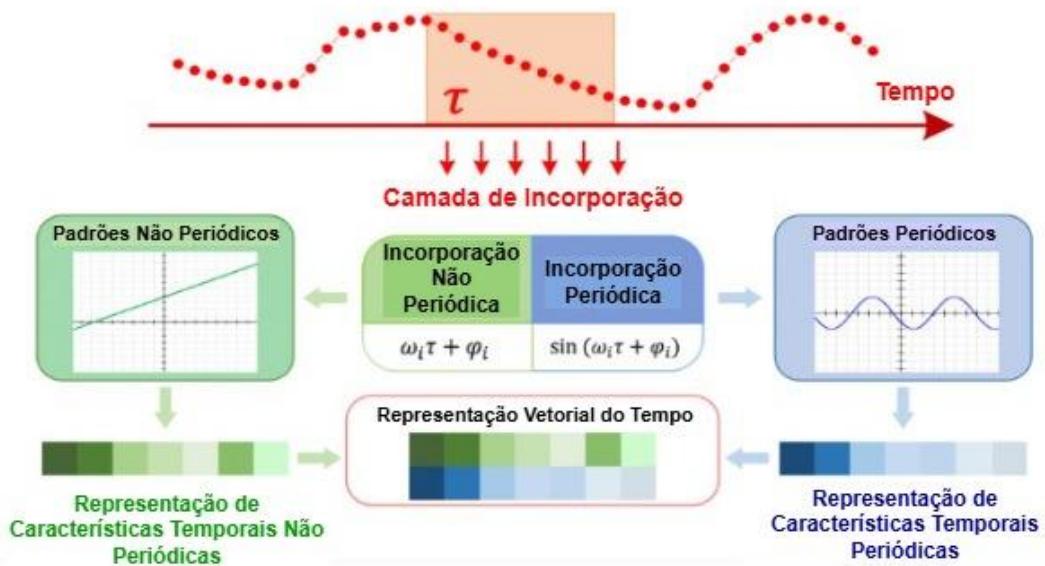
Assim, em vez de aplicar o conjunto de dados diretamente ao modelo, os autores propõem que a série temporal original seja transformada usando a seguinte representação, conforme a Equação (9):

$$t2v(\tau)[i] = \begin{cases} \omega_i \cdot \tau + \phi_i, & \text{if } i = 0 \\ F(\omega_i \cdot \tau + \phi_i), & \text{if } 1 \leq i \leq k \end{cases} \quad (9)$$

Onde k denota a dimensão da função Time2Vec, τ é uma série temporal bruta, F denota uma função de ativação periódica e ω e ϕ denotam um conjunto de parâmetros aprendíveis. F é uma função seno ou cosseno que permite ao algoritmo escolhido detectar padrões periódicos nos dados. Simultaneamente, o termo linear indica a progressão do

tempo e pode ser usado para capturar padrões não periódicos na entrada que são dependentes do tempo. De acordo com a Figura 6, o Time2Vec facilita a representação de comportamentos específicos dentro de uma série temporal, com incorporações periódicas capturando padrões periódicos e incorporações não periódicas abordando padrões não periódicos.

Figura 6 - Representação do Time2Vec



Fonte: Adaptado de Kazemi *et al.* (2019).

Time2Vec é uma técnica poderosa que aprimora modelos de previsão, especialmente em problemas com variáveis temporais complexas. Sua principal vantagem é a forma como representa o tempo, permitindo que os modelos capturem padrões sazonais e periódicos de forma eficaz. Em vez de usar uma representação de tempo simples ou linear, o Time2Vec utiliza funções trigonométricas para criar um vetor que captura as nuances de periodicidade e sazonalidade nos dados. Outra característica importante do Time2Vec é sua capacidade de expandir a entrada temporal, gerando múltiplas características que representam o tempo em diferentes escalas. Isso proporciona ao modelo uma compreensão mais detalhada do contexto temporal, aprimorando seu poder preditivo.

Ao integrar o Time2Vec, os modelos podem representar explicitamente a informação temporal como vetores aprendíveis, capturando padrões sazonais e periódicos

de forma mais expressiva do que codificações temporais tradicionais (como seno e cosseno fixos). Em (GENG; WANG, B.; GAO, Q., 2023), essa representação foi combinada com uma Rede Neural Convolucional Profunda (WDCNN) — responsável por extrair características locais e não lineares dos dados — e com uma BiLSTM, que modela as dependências temporais bidirecionais. Essa arquitetura híbrida mostrou que o Time2Vec potencializa a capacidade do modelo em aprender relações temporais complexas, resultando em previsões mais precisas de potência em sistemas híbridos fotovoltaico-eólicos.

Em (COSTA, R. *et al.*, 2023), os autores propuseram um modelo de previsão de energia fotovoltaica (FV) de curto prazo que integra o Time2Vec à arquitetura do Transformer, bem como às linhas de base MLP e LSTM. Seus resultados mostraram melhorias superiores a 20% em certos horizontes de previsão.

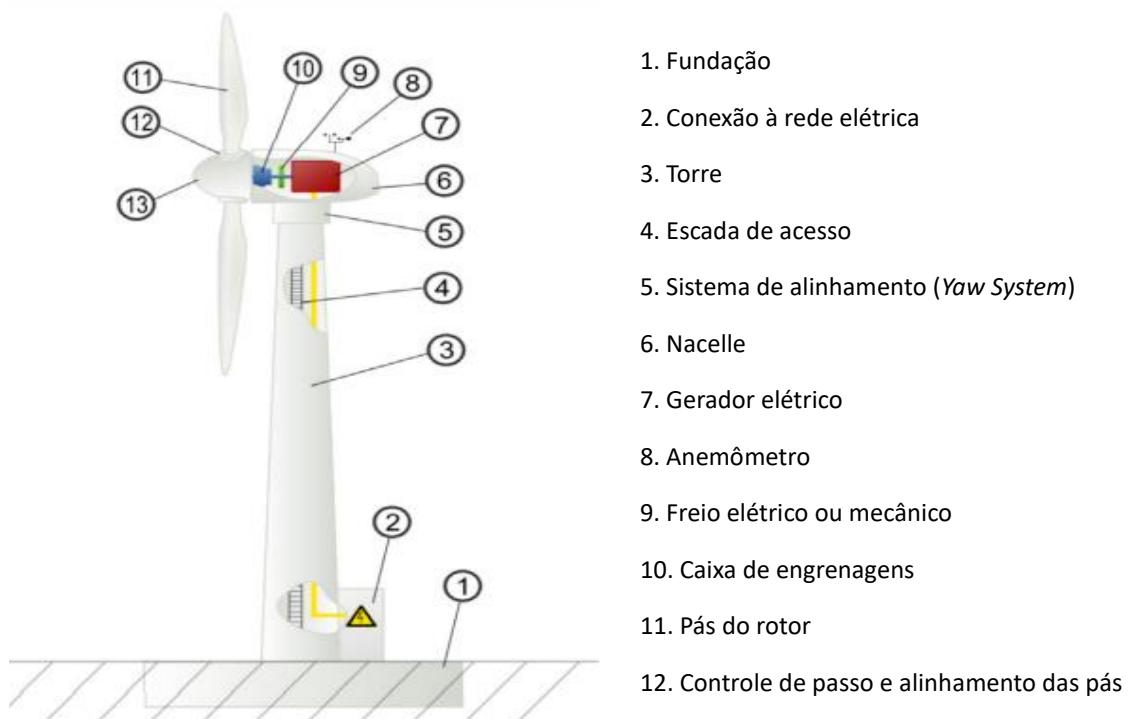
3 ESTADO DA ARTE

Esta seção apresenta uma revisão crítica dos principais conceitos, métodos e avanços relevantes para previsão de potência eólica orientada por dados. Primeiro descrevemos as características e componentes das turbinas eólicas e as fontes de dados mais comuns; em seguida discutimos propriedades das séries temporais na eólica e técnicas de pré-processamento; por fim revisamos abordagens de previsão — desde modelos físicos e estatísticos até métodos de Inteligência Artificial (IA) e variantes de Transformers. O objetivo é identificar lacunas metodológicas e motivar as escolhas técnicas adotadas nesta tese.

3.1 Turbinas Eólicas

As TEs estão disponíveis em diversos modelos e tipos, diferindo em tamanho, orientação do eixo (vertical ou horizontal), número de pás, aerodinâmica e outras características. O tipo mais comumente utilizado em parques eólicos apresenta um eixo horizontal com três pás, conforme ilustrado na Figura 7. Essas turbinas são particularmente adequadas para geração de eletricidade em larga escala devido à sua eficiência e estabilidade superiores. Suas dimensões e capacidades de produção variam significativamente, com diâmetros de rotor variando de 50 a 100 metros, alturas de torre entre 80 e 120 metros e capacidades de produção de 1 a 3 MW (KHUDRI JOHARI; AZIM A JALIL; FAIZAL MOHD SHARIFF, 2018; MCKENNA; OSTMAN V.D. LEYE; FICHTNER, 2016). A Figura 7 ilustra os principais componentes das TEs, enquanto a Tabela 2 explica os principais componentes que influenciam o potencial eólico e a capacidade de produção de energia.

Figura 7 - Componentes principais de TEs para uso industrial



Fonte: (MCKENNA; OSTMAN V.D. LEYE; FICHTNER, 2016).

Tabela 2 - Componentes que influenciam a produção de uma turbina eólica

Componente	Função
Sistema de alinhamento (Yaw System)	Alinha a direção da turbina com a direção do vento, maximizando sua eficiência. A rotação ocorre por meio de um sistema automatizado, com engrenagens que ajustam a velocidade de rotação. Possibilidade de proteger a turbina em condições climáticas adversas, orientando as pás de forma a minimizar a carga exercida sobre a estrutura.
Gerador elétrico	Converte a energia mecânica fornecida pelo rotor em energia elétrica. Geralmente, utiliza-se um gerador síncrono ou assíncrono, que opera com base nos princípios da indução eletromagnética.
Anemômetro	Instrumento usado para medir a velocidade do vento.
Caixa de engrenagens	Projetado para transferir energia mecânica do eixo de transmissão para o gerador, amplificando a velocidade de rotação. Fornece uma interface adequada para absorver choques e vibrações, garantindo uma operação suave e estável.
Pás	Perfis aerodinâmicos responsáveis pela interação com o vento, convertendo a energia cinética do vento em energia mecânica no eixo de transmissão.
Sistemas de passo	Controlar o ângulo de inclinação das pás, permitindo a captura máxima da quantidade de energia cinética do vento.

Fonte: Autoria própria (2025).

Cada um desses componentes envolve variáveis que podem ser monitoradas por meio de diversos métodos de análise de dados. Dados de aplicação de energia eólica podem ser coletados de diversas fontes em múltiplos domínios.

3.2 Análise de Dados na Energia Eólica

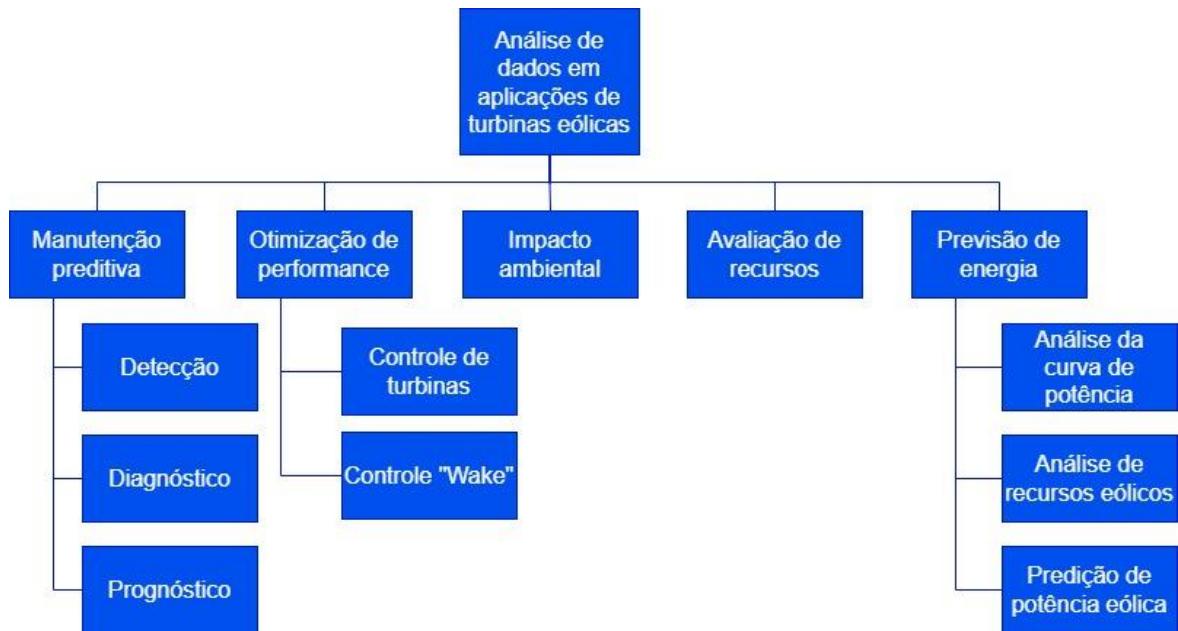
A análise de dados é essencial para maximizar o desempenho e a confiabilidade das TEs, minimizando os custos operacionais e os impactos ambientais. A partir desse ponto, decisões importantes podem ser tomadas, garantindo a produção sustentável e eficiente da energia eólica. Nesse contexto, a análise de dados pode ser aplicada em diferentes etapas do ciclo de vida das TEs, abrangendo desde a operação até a manutenção. A Figura 8 apresenta alguns possíveis usos da análise de dados em

aplicações de energia eólica. A manutenção preditiva baseia-se na detecção, diagnóstico e prognóstico dos componentes das TEs. O diagnóstico está diretamente relacionado à detecção, isolamento e identificação de uma condição de falha iminente ou incipiente, enquanto o prognóstico refere-se à capacidade de prever a vida útil restante (Remaining Useful Life – RUL) de um componente e seus limites de confiança ou nível de incerteza associados (LEITE; ARAÚJO; ROSAS, 2018).

A previsão de energia é baseada na análise da curva de potência, que é diretamente influenciada pelos recursos eólicos, permitindo a previsão de potência de TEs (LYDIA *et al.*, 2014). A otimização de performance busca a máxima extração do recurso eólico disponível, ocorrendo em dois casos: I) Controle de TEs, como o controle do sistema de passo, que regula o ângulo de ataque das pás do rotor, permitindo a resposta aerodinâmica das TEs de acordo com as condições do vento (TIWARI; BABU, 2016); E II) Controle “Wake”, que diz respeito ao controle de TEs ao campo de fluxo previamente perturbado, buscando minimizar os efeitos de fluxos turbulentos (ITALLO RIBEIRO DIAS DA SILVA *et al.*, 2024; NASH; NOURI; VASEL-BE-HAGH, 2021).

No âmbito da análise de dados em aplicações de turbinas eólicas, os impactos ambientais podem ser avaliados por meio da análise das vantagens e limitações da energia eólica, auxiliando na tomada de decisão para a mitigação de efeitos adversos (DAI *et al.*, 2015). Além disso, a avaliação dos recursos incluem condições geográficas, como topografia, obstáculos circundantes, orografia e rugosidade da superfície (MURTHY; RAHI, 2017).

Figura 8 - Análise de dados na energia eólica



Fonte: Autoria própria (2025).

Os dados utilizados para o desenvolvimento de sistemas orientados a dados em aplicações de energia eólica podem ser categorizados com base em suas origens, características e finalidades pretendidas. Este trabalho concentra-se em dados meteorológicos e de recursos eólicos, bem como em dados de parques eólicos e turbinas. Os dados meteorológicos e de recursos eólicos são subdivididos em três categorias: I) Mastros meteorológicos, utilizados em estações para medir temperatura (termômetro), velocidade do vento (anemômetro), pressão atmosférica (barômetro), entre outros. (MÖHRLÉN *et al.*, 2022); II) Dados de sensoriamento remoto, que consiste na coleta de dados e informações de uma região específica da superfície terrestre sem a necessidade de contato direto (HASAGER *et al.*, 2008); e III) Dados meteorológicos históricos que consistem em condições e características de variáveis, como vento, temperatura e pressão, durante um determinado período.

Os dados de usinas eólicas são frequentemente implementados com métodos de diagnóstico de falhas de componentes diretamente relacionados à detecção, isolamento e identificação de uma condição de falha iminente, que inclui: I) Controle de Supervisão e Aquisição de Dados (SCADA), que monitora de forma abrangente o status operacional de TEs (BANGALORE; PATRIKSSON, 2018; TAUTZ-WEINERT; WATSON, 2017; YANG, W.; COURT; JIANG, J., 2013); II) Sensores de Monitoramento de Condição

(CMS), que monitoram mais de perto componentes específicos (CRABTREE; ZAPPALÁ; TAVNER, 2011; GARCÍA MÁRQUEZ *et al.*, 2012; QIAO; LU, D., 2015a, 2015b); E III) Os registros de manutenção referem-se à exploração de métodos de engenharia de confiabilidade, que buscam economizar custos de forma impactante na Operação e Manutenção (O&M) de TEs (SHENG; O'CONNOR, 2023).

Para melhorar a eficiência produtiva dos sistemas de energia eólica, especialmente no contexto da previsão de séries temporais, é crucial superar os desafios associados à complexidade dos dados e às condições operacionais dinâmicas. Nesse contexto, o desenvolvimento de modelos analíticos robustos é essencial para garantir sua confiabilidade, eficiência e sustentabilidade. Esse cenário destaca a necessidade urgente de modelos preditivos avançados capazes de extrair *insights* significativos de conjuntos de dados complexos, particularmente em tarefas de previsão e detecção de anomalias. Para enfrentar esses desafios, além dos modelos estatísticos, inúmeras abordagens baseadas em IA foram propostas na literatura.

Um dos elementos fundamentais na análise de dados de energia eólica é o tratamento de séries temporais, que representam medições coletadas ao longo do tempo de diversas fontes, como sistemas SCADA, sensores meteorológicos e unidades de monitoramento de condições.

3.3 Séries Temporais

A previsão de séries temporais consiste em estimar valores futuros a partir de observações históricas, as quais são interpretadas como realizações de um processo estocástico subjacente. Essas observações podem ser representadas em um espaço de estados indexado por um parâmetro discreto ou contínuo, conforme a natureza temporal dos dados.

A adoção de hipóteses como a estacionariedade fraca — na qual a média permanece constante e a auto covariância depende apenas do intervalo temporal entre as observações — permite simplificar a modelagem estatística e viabilizar o desenvolvimento de métodos preditivos mais robustos (HORVÁTH; KOKOSZKA; RICE, 2014).

Na previsão da geração eólica, são utilizadas técnicas para prever a geração futura de energia por meio da amostragem de um sinal contínuo $p(t)$ de transdutores,

convertendo em uma série temporal discreta p_t por meio de médias móveis em cada passo de tempo Δ_t . Na modelagem de séries temporais, os dados históricos são utilizados como entradas ou regressores, sendo estruturados em intervalos de tempo definidos (ou múltiplos destes) entre cada observação. O intervalo de amostragem é uma característica intrínseca da série, enquanto o passo de previsão constitui um parâmetro metodológico associado à estratégia preditiva adotada.

Outro parâmetro essencial é o horizonte de previsão, que determina o ponto futuro — em número de intervalos de tempo — para o qual se deseja estimar o valor da variável de interesse. No contexto da energia eólica, as previsões de curto prazo geralmente abrangem escalas de minutos a horas, as de médio prazo estendem-se de dias a semanas, e as de longo prazo correspondem a horizontes de meses a anos.

A abordagem estatística para previsão fundamenta-se na modelagem de relações empíricas entre variáveis, utilizando informações históricas da variável de interesse e suas características autorregressivas para estimar o comportamento futuro em um horizonte de previsão k . A Equação (10) expressa esse modelo por meio de uma função que pode assumir forma de um modelo linear autorregressivo, em que p_i indica o valor da variável p observada no enésimo instante de tempo; f_u indica a função empírica desconhecida e ε_i corresponde ao termo de erro associado à predição.

$$p_{t+k} = f_u(p_t, p_{t-1}, p_{t-2}, \dots) + \varepsilon_{t+k} \quad (10)$$

Em geral, a função de previsão da geração eólica consiste em uma relação não linear e pode ser estimada por modelos mais sofisticados que podem ser uma função de variáveis exógenas. Tal modelo pode ser representado pela função na Equação (11). Essa função tem como entradas os dados observados de um SCADA que compõe a série temporal de energia; dados de modelos de Previsão Numérica do Tempo (*Numerical Weather Prediction – NWP*), como previsões de velocidade e direção do vento, temperatura e pressão; e um conjunto de parâmetros Θ , normalmente aproximados por meio de técnicas de aprendizado ou regressão (GALLEGO CASTILLO, 2013).

$$p_{t+k} = f(\text{SCADA}, \text{NWP}, \Theta) \quad (11)$$

A decomposição é uma técnica amplamente utilizada na análise de séries temporais para aprimorar o processamento de dados. Normalmente, as séries temporais são divididas em três componentes: tendência, sazonalidade e ruído. A tendência representa os valores esperados a longo prazo, a sazonalidade captura padrões repetitivos a curto prazo e o ruído se refere a variações inesperadas e aleatórias. Os principais tipos estão representados abaixo, com sua denominação em inglês:

- 1. Classic Decomposition:** Envolve a separação de uma série temporal em componentes de tendência, sazonalidade e ruído, normalmente usando médias móveis ou filtros (CLEVELAND, R. B. *et al.*, 1990);
- 2. Variational Mode Decomposition (VMD):** Decompõe adaptativamente um sinal em modos intrínsecos com base em princípios variacionais, capturando as oscilações do sinal (DRAGOMIRETSKIY; ZOSO, 2014);
- 3. Wavelet Transform:** Analisa sinais em múltiplas escalas ou resoluções, usando wavelets para capturar frequência e tempo, tornando-o ideal para dados não estacionários (RHIF *et al.*, 2019);
- 4. Fast Fourier Transform (FFT):** Converte um sinal do domínio do tempo para o domínio da frequência, identificando frequências dominantes (COOLEY; LEWIS; WELCH, P. D., 1969);
- 5. Empirical Modal Decomposition (EMD):** Método que decompõe um sinal em um conjunto de funções de modo intrínseco (IMFs) e um resíduo, permitindo isolar os modos oscilatórios presentes em sinais não estacionários e não lineares, sem a necessidade de uma base de funções predefinida (TANAKA; MANDIC, 2007);
- 6. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN):** Uma derivação aprimorada do EMD, amplamente utilizada por utilizar múltiplas instâncias de ruído branco adaptativo para melhorar a precisão da decomposição do sinal, reduzindo problemas de ruído. É especialmente eficaz para sinais não lineares e não estacionários. (HE, Yingying *et al.*, 2024).

Para realizar previsões de séries temporais, diversos modelos estão disponíveis na literatura científica. Esses modelos são discutidos na seção a seguir.

3.4 Modelos de Previsão

Para enfrentar os desafios descritos na Seção 1.1, é crucial fazer previsões precisas sobre o potencial de geração de energia eólica. A literatura sobre previsão de potência eólica tradicionalmente organiza os métodos em três grandes classes:

1. modelos físicos, baseados na dinâmica da atmosfera;
2. modelos estatísticos, que descrevem relações matemáticas entre observações passadas e futuras; e
3. modelos baseados em Inteligência Artificial (IA), capazes de capturar padrões não lineares complexos.

Essa taxonomia é amplamente adotada tanto na indústria quanto na comunidade científica, pois cada classe atende a horizontes de previsão, custos computacionais e níveis de granularidade distintos. A seguir, apresentam-se essas três classes de modelos de forma estruturada, destacando suas características, vantagens e limitações para previsão de curto prazo.

3.4.1 Modelos Físicos

Os modelos físicos utilizados em previsão eólica baseiam-se em modelos de NWP (*Numerical Weather Prediction*), entre os quais o WRF (*Weather Research and Forecasting*) é o mais difundido na academia e na indústria (CHAWLA *et al.*, 2018; VOYANT *et al.*, 2012). Esses modelos constituem o padrão operacional utilizado por centros meteorológicos, empresas de energia e operadores do sistema elétrico, devido à sua capacidade de representar a dinâmica atmosférica em grande escala.

Essas abordagens baseiam-se na solução de equações diferenciais parciais que descrevem os processos físicos e dinâmicos da atmosfera, como conservação de massa, momento e energia. Para isso, integram variáveis meteorológicas fundamentais, incluindo pressão atmosférica, umidade relativa, temperatura, radiação solar e interações superfície-atmosfera (ZHAO, J. *et al.*, 2021a). Esses modelos são particularmente eficazes em previsões de médio e longo prazo, pois capturam adequadamente a evolução de sistemas meteorológicos de grande escala (CHANG, 2014). No entanto, apresentam limitações importantes para previsão local e de curto prazo em parques eólicos:

1. Resolução espacial restrita, que suaviza efeitos topográficos e microclimáticos;
2. Resolução temporal limitada, geralmente de minutos a horas;
3. Alto custo computacional, decorrente do processamento de grandes domínios atmosféricos; e
4. Sensibilidade às condições iniciais, que pode amplificar incertezas em regiões de elevada variabilidade do vento.

Essas características podem estar associadas as condições climáticas e meteorológicas do Nordeste brasileiro. Em função dessas limitações, estudos recentes têm explorado abordagens híbridas, seja com modelos estatísticos ou com modelos de IA, capazes de complementar ou substituir parcialmente as previsões físicas, oferecendo maior eficiência computacional e adaptação a condições locais específicas.

3.4.2 Modelos Estatísticos

Os modelos estatísticos mais comumente usados para previsão na eólica incluem média móvel autorregressiva (ARMA) (ERDEM; SHI, J., 2011), média móvel integrada autorregressiva (ARIMA) (AASIM; SINGH, S. N.; MOHAPATRA, 2019) e ARIMA fracionário (f-ARIMA) (KAVASSERI; SEETHARAMAN, 2009). Essas abordagens baseiam-se em relações lineares entre observações passadas e futuras, assumindo que o comportamento histórico da variável contém informação suficiente para descrever sua evolução temporal.

Os modelos AR (AutoRegressive) representam o valor corrente de uma série como uma combinação linear de seus valores defasados, enquanto os modelos MA (Moving Average) utilizam combinações lineares de erros passados. Assim, o modelo ARMA combina ambas as estruturas, sendo adequado para séries aproximadamente estacionárias. O modelo ARIMA estende essa formulação ao introduzir um operador de diferenciação, permitindo lidar com tendências e eliminar não estacionaridades. Já o modelo f-ARIMA emprega diferenciação fracionária, possibilitando capturar dependências de longo alcance com maior flexibilidade matemática.

Uma das formas de representar matematicamente o modelo ARIMA é de acordo com a Equação (12):

$$(1 - B)^d y_t = c + \sum_{i=1}^p \phi_i (1 - B)^d y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \quad (12)$$

onde B representa o operador de defasagem, tal que $By_t = y_{t-1}$; d é o grau de diferenciação aplicado para remover tendência; p e q são, respectivamente, as ordens autorregressivas (AR) e de média móvel (MA); c é um termo constante (*drift*), responsável por introduzir uma tendência linear residual mesmo após a diferenciação; ϕ_i e θ_j são os coeficientes AR e MA; e ϵ_t é um termo de ruído branco com média zero e variância constante.

Apesar de sua eficiência em determinadas condições, esses modelos apresentam limitações estruturais significativas. Por dependerem de suposições de linearidade e estacionariedade, têm desempenho restrito quando aplicados a séries de velocidade do vento, que frequentemente exibem comportamentos não lineares, sazonalidades complexas e alta variabilidade estocástica (Zhao et al., 2021b). Além disso, sua capacidade de generalização é limitada em contextos dominados por múltiplos fatores meteorológicos e espaciais, que introduzem dinâmicas não lineares difíceis de serem capturadas por modelos lineares.

3.4.3 Modelos de Inteligência Artificial (IA)

Para superar essas limitações, a literatura recente tem apontado para a necessidade de modelos de IA, capazes de representar de forma mais realista a dinâmica intrínseca do vento e suas interações com o ambiente atmosférico. Nos últimos anos, técnicas de IA têm desempenhado um papel central no aprimoramento da previsão de potência eólica, especialmente devido à sua capacidade de lidar com relações não lineares, padrões complexos e dependências temporais nos dados.

Modelos clássicos, como o Multi-Layer Perceptron (MLP) (MARVUGLIA; MESSINEO, 2012) e as Redes Neurais Artificiais (ANN) (HE, Yaoyao; LI, Haiyan, 2018), foram amplamente utilizados nas primeiras abordagens baseadas em IA, principalmente pela capacidade de modelar relações não lineares entre variáveis meteorológicas e potência gerada. Entretanto, por não possuírem mecanismos internos que tratem sequências temporais, esses modelos dependem de janelas fixas de entrada e apresentam limitações na captura de dinâmicas temporais mais longas.

Para lidar diretamente com dados sequenciais, surgiram as Recurrent Neural

Networks (RNN) (CAO, Q.; EWING; THOMPSON, M. A., 2012), capazes de representar dependências temporais por meio de estados recorrentes. No entanto, problemas como gradientes explosivos e desvanecentes restringiam seu desempenho em séries longas.

Para mitigar essas limitações, arquiteturas avançadas como Long Short-Term Memory (LSTM) (ZHANG, Zhendong *et al.*, 2019) e Gated Recurrent Unit (GRU) (SONG, J.; WANG, Jianzhou; LU, H., 2018) foram desenvolvidas, tornando-se amplamente aplicadas em previsão eólica devido à sua maior capacidade de capturar dependências de médio e longo prazo. Extensões como o BiLSTM (PENG, T. *et al.*, 2021) ampliam essa capacidade ao considerar dependências temporais em ambas as direções.

Outras abordagens relevantes incluem Support Vector Regression (SVR) (HE et al., 2021), Convolutional Neural Networks (CNN) (HARBOLA; COORS, 2019), que extraem padrões espaciais ou spatio-temporais, e redes com treinamento baseado em Backpropagation (BPNN) (SONG; WANG; LU, 2018).

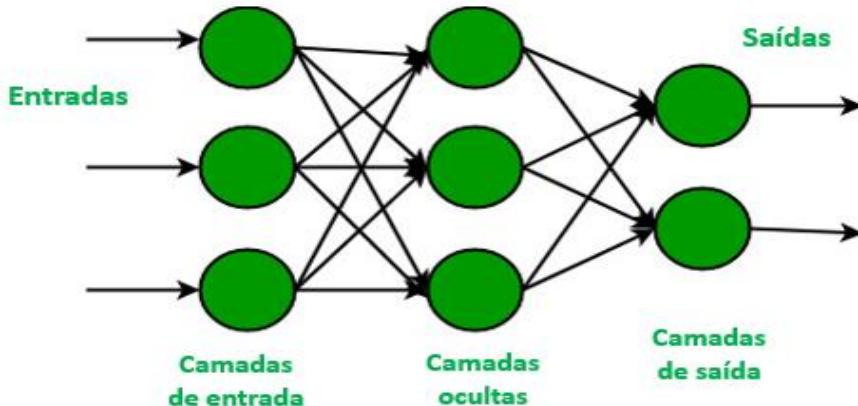
- **MLP: modelos *feed-forward* e suas limitações temporais**

MLPs são modelos de aprendizado profundo que consistem em múltiplas camadas de nós (neurônios). Essas camadas são redes de *feed-forward* que aprendem pesos Θ e mapeiam a entrada para a saída $y \approx f(x|\Theta)$ (Como ilustrado na Figura 9). A saída gera uma estrutura em cadeia, onde múltiplas camadas são empilhadas, dando profundidade ao modelo. Portanto, a saída é caracterizada pela Equação (13) abaixo:

$$\hat{y} = f^{(n+1)}(f^{(n)}(\dots f^{(2)}(f^{(1)}(x|\Theta_1)|\Theta_2) \dots |\Theta_n)|\Theta_{n+1}) \quad (13)$$

Onde $f^{(1)}$ representa a transformação aplicada pela primeira camada oculta, com peso Θ_1 , $f^{(2)}$ pela segunda camada oculta, com peso Θ_2 . $f^{(n)}$ pela enésima camada oculta, com peso Θ_n . E $f^{(n+1)}$ para a última camada oculta, com peso Θ_{n+1} . A Equação (13) representa como a entrada x é progressivamente transformada através das n camadas ocultas e, finalmente, mapeada para a camada de saída \hat{y} (TAUD; MAS, 2018).

Figura 9 - Estrutura de uma MLP

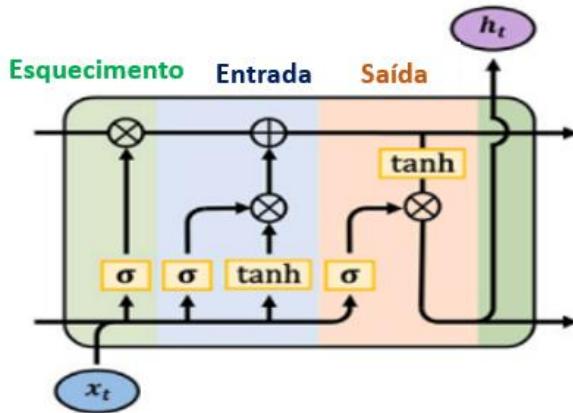


Fonte: Adaptado de Taud e Mas (2018).

- **LSTM e RNNs: Aprendizado sequencial e dependências de longo prazo**

LSTM é uma arquitetura de RNN projetada para aprendizado de sequências. Diferentemente das RNNs tradicionais, as LSTMs são capazes de capturar dependências de longo prazo, superando o problema do desvanecimento do gradiente (*vanishing gradient*), que compromete o aprendizado eficaz durante o processo de retropropagação. Essa capacidade advém do uso de um mecanismo interno de portas de controle, que regula o fluxo de informações ao longo do tempo. A porta de entrada determina quais informações são incorporadas ao estado de memória; a porta de esquecimento define quais informações devem ser descartadas; e a porta de saída seleciona os conteúdos relevantes a serem transmitidos para a próxima etapa de processamento. Essa estrutura permite que a LSTM mantenha e atualize informações de forma seletiva, otimizando a eficiência e a estabilidade do modelo durante o treinamento (YU, Y. et al., 2019). A Figura 10 ilustra a estrutura típica de uma célula LSTM, em que x_t representa a entrada no instante de tempo t (por exemplo, dados de uma série temporal), enquanto h_t corresponde ao estado oculto, que transporta informações processadas de etapas anteriores da sequência.

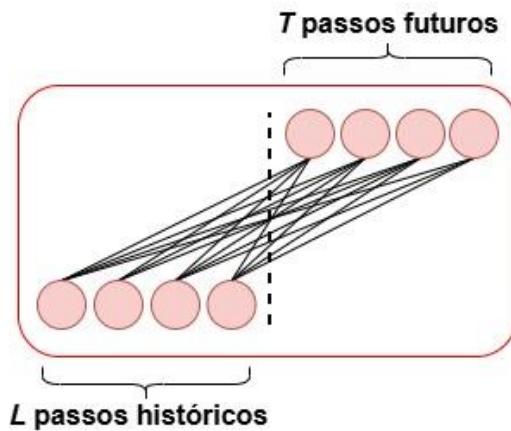
Figura 10 - Estrutura de uma LSTM

Fonte: Adaptado de Bentsen *et al.* (2023a)

- **Modelos lineares profundos: NLinear e DLinear**

Diversos modelos baseados em aprendizado profundo são discutidos na literatura científica, entre eles o DLinear (ZENG *et al.*, 2023). Nesse estudo, os autores questionam a eficácia das arquiteturas Transformer na previsão de séries temporais, destacando limitações como alto custo computacional, ineficiência e propensão ao sobre ajuste (*overfitting*), especialmente em séries de longa duração. Com o objetivo de investigar se a complexidade dos Transformers é realmente necessária, os autores propuseram duas variantes simplificadas — NLinear e DLinear — que mantêm a estrutura de redes neurais profundas, mas substituem o mecanismo de atenção por camadas lineares aplicadas ao longo do eixo temporal. Esses modelos realizam a regressão da série temporal por meio de uma operação de soma ponderada, conforme ilustrado na Figura 11. A formulação matemática é dada por $\hat{X} = WX_i$, em que $W \in R^{T \times L}$ representa a camada linear aplicada na dimensão temporal, enquanto X_i e \hat{X}_i correspondem, respectivamente, à entrada e à previsão da enésima variável, respectivamente.

Figura 11 - Ilustração básica de um modelo linear



Fonte: Adaptado de Zeng *et al.* (2023).

A principal diferença entre os dois modelos está na forma como tratam a série de entrada. O NLinear aplica diretamente uma transformação linear sobre a sequência original, sem qualquer pré-processamento adicional. Já o DLinear realiza uma decomposição explícita da série temporal em dois componentes — tendência e sazonalidade — e aplica uma camada linear independente a cada um deles. As saídas dessas duas camadas são então combinadas para formar a previsão final. Essa decomposição permite ao DLinear capturar padrões de variação lenta e rápida de forma mais interpretável, ao mesmo tempo em que mantém a simplicidade e eficiência computacional do modelo linear.

- **Limitações gerais e motivação para abordagens mais avançadas**

Embora esses modelos tenham demonstrado sucesso na aprendizagem de não linearidades complexas e padrões temporais, cada abordagem apresenta limitações inerentes. As MLPs, por exemplo, são modelos essencialmente estáticos, incapazes de capturar dependências temporais de forma direta, exigindo engenharia manual de atributos para incorporar informações históricas. Os modelos baseados em LSTM, apesar de sua capacidade de manipular sequências, enfrentam dificuldades para representar dependências de longo prazo e estão sujeitos a altos custos computacionais, além de demandarem ajuste cuidadoso de hiperparâmetros. Essa complexidade os torna menos eficientes quando aplicados a grandes volumes de dados.

Por outro lado, abordagens lineares como o DLinear oferecem alta eficiência

computacional e se mostram eficazes em cenários nos quais os padrões temporais apresentam comportamento predominantemente linear. Entretanto, tais modelos possuem capacidade representacional limitada, o que restringe seu desempenho na modelagem de relações não lineares complexas — características frequentemente observadas em séries temporais de energia eólica, que envolvem interações multivariadas e variabilidade estocástica

Essas limitações motivaram o desenvolvimento de mecanismos capazes de:

1. Capturar relações de curto e longo prazo de forma simultânea;
2. Operar de maneira paralela;
3. Identificar automaticamente quais partes da sequência são mais relevantes.

Essas limitações levaram ao desenvolvimento do mecanismo de atenção, ponto de transição entre modelos recorrentes e arquiteturas totalmente paralelas.

- **Evolução até o mecanismo de atenção e transição para o Transformer.**

Os mecanismos de atenção foram introduzidos como alternativa aos modelos recorrentes, permitindo que o modelo atribuísse pesos relativos aos elementos mais relevantes de uma sequência sem depender de processamento sequencial. A formulação básica da atenção escalonada *dot-product* é apresentada na Seção 2.1 deste trabalho

A consolidação da atenção levou ao desenvolvimento do Transformer (VASWANI et al., 2017), que elimina a recorrência e avalia todas as relações temporais de forma paralela, capturando dependências de curto e longo prazo com maior eficiência. O uso de múltiplas cabeças de atenção (ver Equação 3 na Seção 2.1) aprimora a capacidade de identificar diferentes padrões temporais. Por não possuir recorrência nativa, o Transformer utiliza codificadores posicionais para introduzir informações de ordem temporal, apresentados na Figura 3 e formalizados pelas Equações 5 e 6 da seção anterior. Com o intuito de ampliar essa representação, este trabalho adota o Time2Vec, conforme descrito na Seção 2.2.

Historicamente, a evolução ocorre em duas etapas:

1. A atenção é introduzida como componente adicional em modelos recorrentes;
2. Torna-se a base estrutural de uma arquitetura completa — o Transformer.

Essa progressão cronológica fundamenta o uso de Transformers em previsão de séries temporais e contextualiza sua aplicação neste trabalho.

3.5 Aplicação das redes Transformers

Para uma revisão detalhada das aplicações das redes Transformer, realizou-se um estudo de revisão sistemático, fundamentado em uma metodologia estruturada que assegura transparência, reproduzibilidade e rigor científico. Neste estudo, adotou-se uma abordagem quantitativa e qualitativa combinada, seguindo as melhores práticas para revisões sistemáticas e alinhando-nos às diretrizes PRISMA quando aplicável (PAGE *et al.*, 2021). O componente quantitativo consistiu em uma análise bibliométrica, que emprega técnicas estatísticas e matemáticas para descobrir tendências e padrões de pesquisa dentro de um domínio (DONTHU *et al.*, 2021). Optou-se por esse método porque é possível identificar a dinâmica de publicação dos estudos, tópicos emergentes e clusters de pesquisa no crescente campo de aplicações do Transformer. O componente qualitativo envolveu uma análise de conteúdo aprofundada dos estudos selecionados (SOARES *et al.*, 2018). Esta etapa foi essencial para avaliar criticamente as arquiteturas do modelo, conjuntos de dados, procedimentos de avaliação e resultados.

A combinação dessas duas abordagens fornece uma ampla visão geral da evolução da pesquisa e uma compreensão detalhada do desempenho e das limitações do modelo. A Figura 12 ilustra a metodologia de pesquisa dessa revisão sistemática, destacando o refinamento progressivo do conjunto de dados ao longo das três etapas (Planejamento; Coleta e identificação dos artigos; Relatório). Essa representação reforça a natureza sistemática e transparente do processo de revisão, estruturada da seguinte forma:

Na etapa de planejamento, definiu-se o escopo da pesquisa e as palavras-chave a serem utilizadas. Foram empregadas as expressões “*Transformers Networks*” e seus sinônimos correspondentes, com o objetivo de abranger o maior conjunto possível de estudos relevantes. A base de dados Scopus foi selecionada por sua ampla cobertura interdisciplinar, pelos recursos avançados de busca e pela extensa indexação de periódicos revisados por pares, assegurando uma identificação abrangente da literatura pertinente.

Na etapa de coleta e identificação, restringiu-se a busca por artigos de periódicos revisados por pares, redigidos em inglês e publicados entre 2017 e 2025, considerando o ano de introdução do modelo Transformer. Foram excluídos resumos de conferências, livros, teses e outras fontes não associadas a periódicos científicos. A busca inicial retornou 16.236 artigos (Etapa 1). Em seguida, foram incorporadas palavras-chave

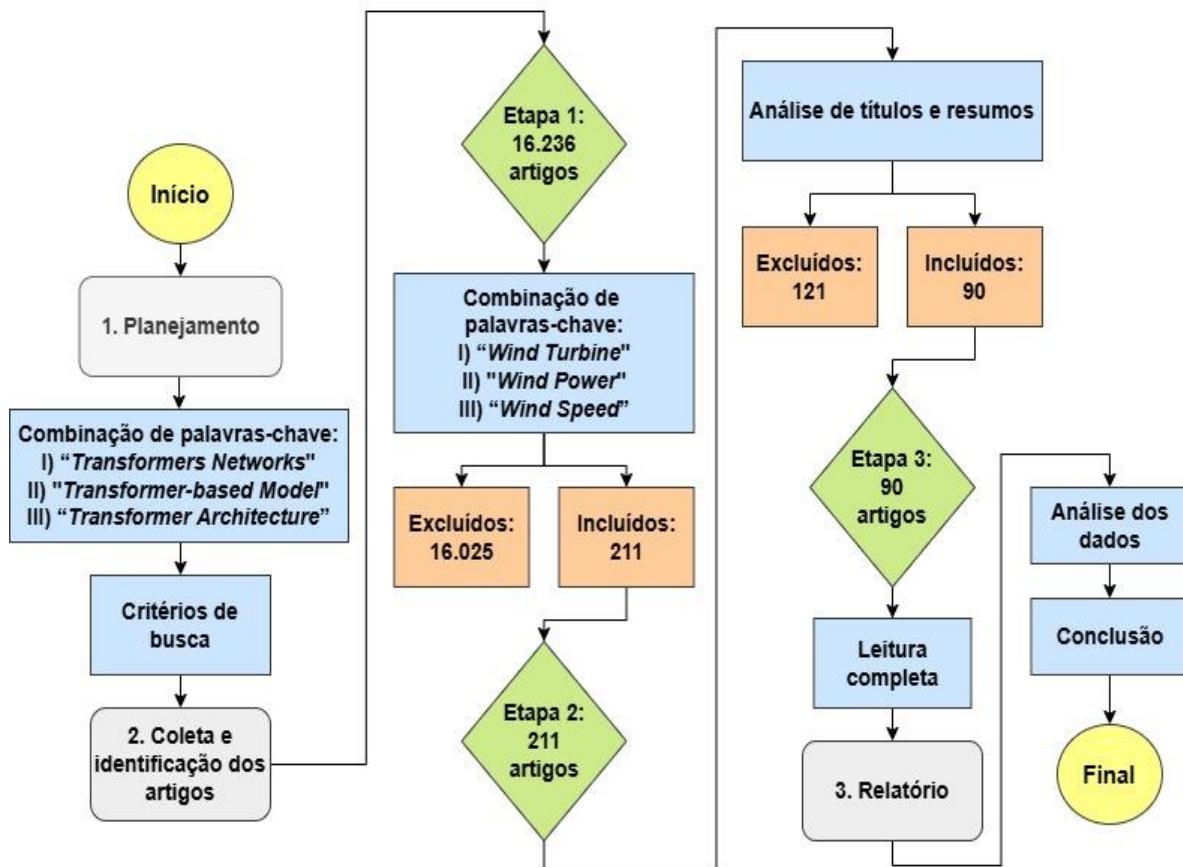
específicas do domínio, como “*Wind Turbine*”, “*Wind Energy*” e “*Wind Speed*”, bem como seus sinônimos, resultando em um conjunto reduzido de 211 artigos (Etapa 2).

As *strings* completas utilizadas nas Etapas 1 e 2 encontram-se apresentadas no Apêndice A, garantindo transparência, rastreabilidade e reproduzibilidade do processo de busca.

Por fim, na etapa de relatório, os títulos e resumos foram analisados de forma criteriosa. Excluíram-se os trabalhos que: (1) não apresentavam avaliação objetiva de desempenho de modelos baseados em Transformers; (2) tinham foco principal em sistemas híbridos solar-eólicos ou em temas mais amplos de energia; e (3) não demonstravam transparência metodológica. Após essa triagem, 90 artigos permaneceram para análise aprofundada durante a avaliação qualitativa.

A abordagem metodológica qualitativa foi empregada para conduzir uma análise explicativa do estudo, que foi sistematicamente integrada aos 90 artigos identificados durante a Etapa 3 (Fig. 12). Esse procedimento analítico teve como objetivo reforçar a avaliação dos dados na fase de relatório e discutir os resultados da pesquisa com mais detalhes, levando, em última análise, a conclusões mais detalhadas sobre alguns dos estudos identificados.

Figura 12 - Fluxograma do passo a passo metodológico da revisão de literatura



Fonte: Autoria própria (2025).

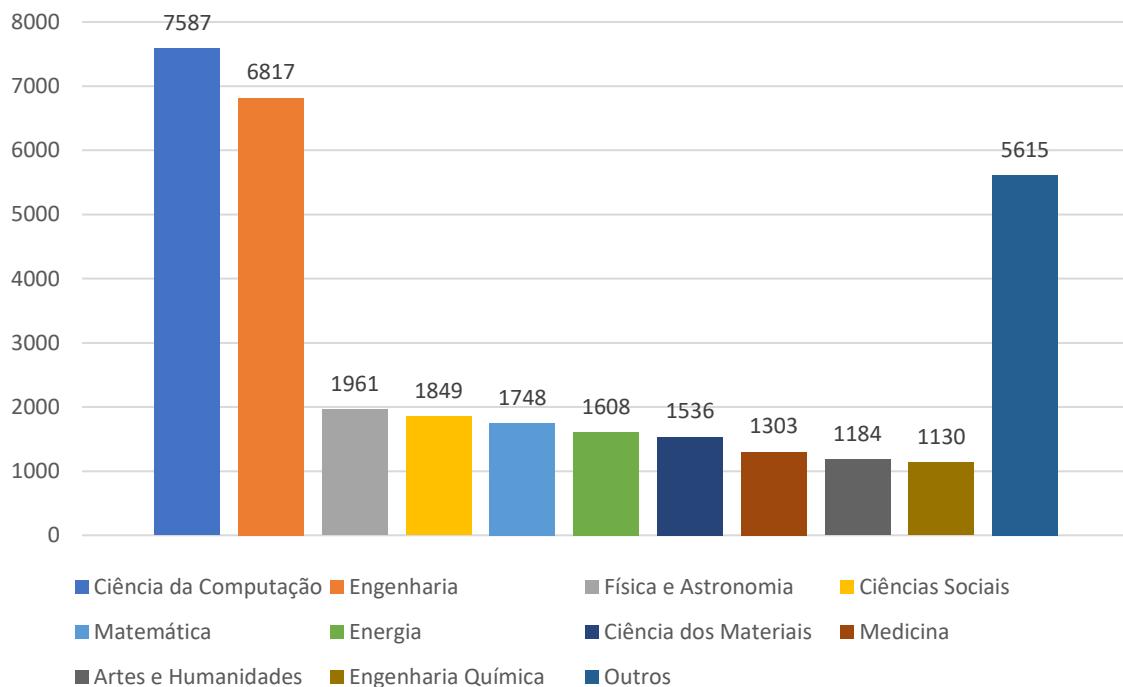
Uma breve visão das aplicações gerais foi realizada para analisar o domínio potencial da análise de dados em TEs usando redes Transformer. Como resultado, é possível identificar os potenciais benefícios dessa rede neural em diversas áreas da sociedade.

3.5.1 Aplicações Gerais

A estratégia de busca realizada na Etapa 1 resultou em 16.236 artigos. A Figura 13 apresenta uma visão geral das múltiplas aplicações das redes Transformer em diferentes domínios da sociedade, evidenciando as áreas de estudo que adotaram essa arquitetura em suas pesquisas. Ressalta-se que um mesmo artigo pode estar associado a mais de uma área de aplicação. Em (HU, Zhaoyu *et al.*, 2023), foi desenvolvido um modelo para detecção de tumores cerebrais com base em imagens tridimensionais de

ressonância magnética. O estudo abrange áreas como medicina e ciência da computação. Em (YU, Z. *et al.*, 2022), os autores proporam um modelo para reidentificação de veículos em ambientes urbanos, abrangendo áreas como engenharia, ciências da computação e ciências sociais.

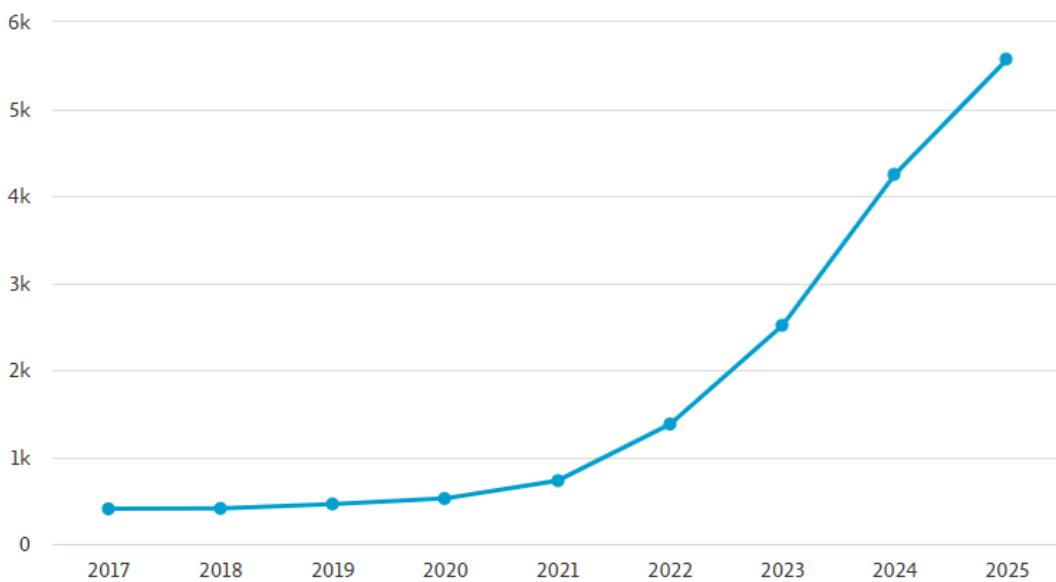
Figura 13 - Documentos por área temática da Etapa 1



Fonte: Adaptado de SCOPUS (2025).

A Figura 14 ilustra o uso crescente de redes Transformer ao longo do tempo, com um aumento notável entre 2021 e 2025. Essa tendência indica uma preferência crescente pelo emprego dessa arquitetura de rede neural para solucionar desafios de análise preditiva de dados. Além disso, é possível observar a notável relevância das redes Transformer em diversas aplicações. Essas aplicações incluem classificação de imagens, detecção e segmentação de objetos, geração de imagens e linguagem, processamento de dados multimodais (incluindo texto, fala e imagem) e análise de dados de séries temporais (KHAN, S. *et al.*, 2022).

Figura 14 - Número de estudos desenvolvidos a partir da Etapa 1 (2017-2025)



Fonte: Adaptado de SCOPUS (2025).

3.5.2 Aplicações na Energia Eólica

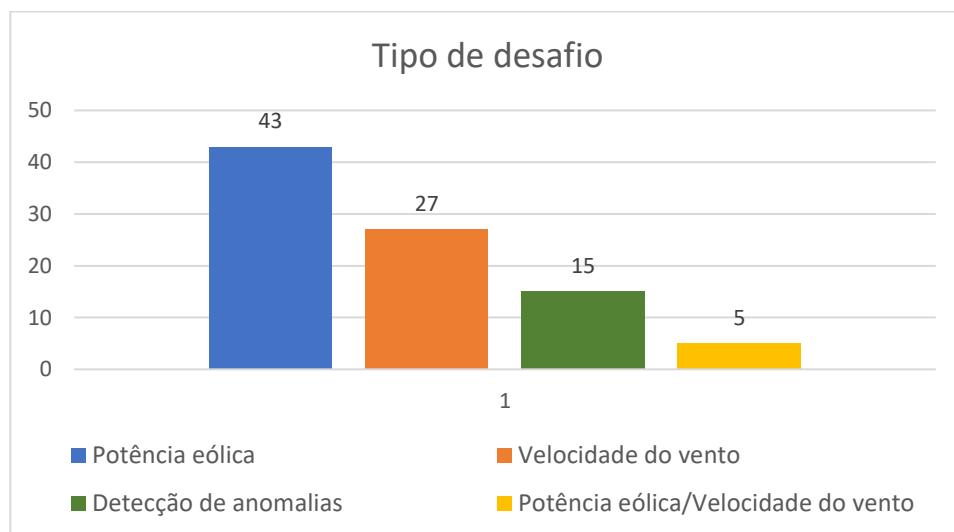
A estratégia de busca adotada nesta pesquisa resultou em 90 artigos na etapa final de seleção. A Figura 15 apresenta a distribuição desses trabalhos por tipo de desafio identificado. A Tabela 3 apresenta uma amostra desses trabalhos, organizados em ordem cronológica. Cada artigo aborda diferentes tipos de desafios, classificados em: (I) previsão da velocidade do vento; (II) previsão da potência eólica; e (III) detecção de anomalias. Alguns estudos contemplam simultaneamente a previsão da velocidade do vento e da potência eólica. Todos os trabalhos analisados empregam metodologias baseadas na coleta de dados de turbinas reais e em estudos experimentais. Em alguns casos, os autores especificaram as fontes dos dados de entrada, obtidos por meio de vibrações (registradas por acelerômetros), sistemas SCADA ou anemômetros. Nos estudos em que essa informação não foi explicitada, os dados foram considerados como operacionais. Por fim, a última coluna da Tabela 3 indica o local de coleta dos dados utilizados em cada pesquisa.

Durante a análise dos estudos identificados, observou-se que a arquitetura do Transformer frequentemente requer modificações para lidar de forma eficaz com a previsão de séries temporais. Consequentemente, cada trabalho propõe um modelo

personalizado com validação experimental de sua abordagem. Um aprimoramento comum envolve a integração de técnicas de decomposição, conforme comentado na Seção 3.3 desse estudo, para isolar componentes de tendência e sazonais, melhorando assim a qualidade das entradas do modelo.

Outra abordagem recorrente consiste na utilização de técnicas híbridas que integram modelos baseados em grafos, tais como Grafos Espaço-Temporais (STGs), Redes Neurais de Grafos (GNNs), Redes Convolucionais de Grafos (GCNs) e Redes de Atenção em Grafos (GATs). Os STGs possibilitam a representação conjunta das dimensões espaciais e temporais dos dados em uma estrutura de grafo, preservando a topologia e a dinâmica das relações entre os nós. As GNNs, por sua vez, são projetadas para processar esses dados estruturados, enquanto as GCNs se especializam em capturar correlações espaciais locais por meio de operações de convolução sobre o grafo. As GATs aprimoram esse processo incorporando mecanismos de atenção que atribuem pesos diferenciados aos vizinhos mais relevantes de cada nó. Ao integrar as dependências espaciais extraídas pelos modelos baseados em grafos com a capacidade dos Transformers de modelar relações temporais de longo alcance, as arquiteturas híbridas resultantes oferecem uma estrutura robusta e eficiente para tarefas de previsão espaço-temporal, como a estimativa de fluxo de tráfego e o monitoramento de redes de sensores.

Figura 15 - Classificação dos 90 artigos incluídos na revisão sistemática de acordo com o desafio investigado: potência eólica, velocidade do vento, detecção de anomalias e desafios combinados.



Fonte: Autoria própria (2025).

Tabela 3 - Alguns dos artigos identificados na pesquisa

Nº	Fonte	Tipo de desafio	Dados de entrada	Local de coleta dos dados
1	(WANG, H.-K.; SONG, K.; CHENG, 2022)	Potência eólica	SCADA	China
2	(ZHANG, K.; LI, X.; SU, 2022)	Velocidade do vento	Dados operacionais	China
3	(WANG, Lei; HE, Yigang; LI, Lie; <i>et al.</i> , 2022)	Potência eólica	Dados operacionais	China
4	(ZHOU, Haoxuan <i>et al.</i> , 2022)	Detecção de anomalias	Dados de vibração	Canada
5	(QU, K. <i>et al.</i> , 2022a)	Potência eólica	Dados operacionais	China
6	(HUANG, Xiaohan; JIANG, A., 2022)	Potência eólica	Dados operacionais	Espanha
7	(PAN <i>et al.</i> , 2022)	Velocidade do vento	Dados operacionais	Dinamarca
8	(WANG, Lei; HE, Yigang; LIU, Xiaoyan; <i>et al.</i> , 2022)	Potência eólica	Dados operacionais	China
9	(ZHA <i>et al.</i> , 2023a)	Velocidade do vento/ Potência eólica	SCADA	China
10	(TANG, W.; LIU, C.; ZHANG, B., 2023)	Detecção de anomalias	SCADA	China
11	(HUANG, S.; YAN, Chang; QU, Y., 2023a)	Potência eólica	SCADA	China
12	(BOMMIDI; TEEPARTHI; KOSANA, 2023)	Velocidade do vento/ Potência eólica	Dados operacionais	EUA
13	(BENTSEN <i>et al.</i> , 2023a)	Velocidade do vento/ Potência eólica	Dados operacionais	Mar do Norte (Noruega)
14	(WANG, Yun <i>et al.</i> , 2023)	Velocidade do vento	Dados operacionais	Groelândia e EUA
15	(XIAO; HE, X.; LI, C., 2023)	Potência eólica	Dados operacionais	Bélgica
16	(TIAN <i>et al.</i> , 2023)	Potência eólica	Dados operacionais	China e Espanha
17	(ZHENG, H. <i>et al.</i> , 2023)	Potência eólica	Dados operacionais	China
18	(NASCIMENTO; MELO, DE; MOREIRA, 2023)	Velocidade do vento	Anemômetros	Brasil
19	(XINXIN <i>et al.</i> , 2023)	Velocidade do vento	Dados operacionais	Não especificado
20	(HU, J. <i>et al.</i> , 2023)	Potência eólica	Dados operacionais	Espanha e Austrália
21	(WANG, Hai-Kun <i>et al.</i> , 2023)	Potência eólica	Dados operacionais	China
22	(BENTSEN <i>et al.</i> , 2023b)	Velocidade do vento/ Potência eólica	Dados operacionais	Mar do Norte (Noruega)
23	(YU, Chengqing <i>et al.</i> , 2023)	Velocidade do vento	Dados operacionais	China
24	(GONG, M. <i>et al.</i> , 2023a)	Potência eólica	Dados operacionais	Página virtual
25	(WEI, H.; WANG, W.; KAO, 2023)	Potência eólica	SCADA	Turquia
26	(LI, N. <i>et al.</i> , 2023)	Potência eólica	Dados operacionais	China
27	(CHEN, Yaoran <i>et al.</i> , 2024)	Velocidade do vento	Dados operacionais	Oceano Pacífico (EUA)
28	(BAN <i>et al.</i> , 2024)	Velocidade do vento	Dados operacionais	China e páginas virtuais
29	(LIN, S. <i>et al.</i> , 2024)	Velocidade do vento	Dados operacionais	China
30	(TAO <i>et al.</i> , 2024)	Detecção de anomalias	CMS	China
31	(SHI, Z. <i>et al.</i> , 2024).	Velocidade do vento	Dados operacionais	EUA, Groelândia e Antártida
32	(BOMMIDI; TEEPARTHI; DULLA MALLESHAM, 2024)	Velocidade do vento	Dados operacionais	EUA
33	(MO <i>et al.</i> , 2024)	Potência eólica	Dados operacionais	China
34	(WAN <i>et al.</i> , 2024)	Potência eólica	SCADA	China
35	(GAO, Y. <i>et al.</i> , 2024)	Velocidade do vento	Dados operacionais	China
36	(ZHENG, Y. <i>et al.</i> , 2024)	Detecção de anomalias	SCADA	China
37	(PARRI; TEEPARTHI, 2024)	Velocidade do vento	Dados operacionais	EUA
38	(JIN, Z. <i>et al.</i> , 2025)	Velocidade do vento	Dados operacionais	China
39	(DONG, Z. <i>et al.</i> , 2025)	Potência eólica	Dados operacionais	China
40	(WANG, Zhongrui <i>et al.</i> , 2025)	Potência eólica	Dados operacionais	Estreito de Dover (Reino Unido e França)
41	(LENG <i>et al.</i> , 2025)	Velocidade do vento	Dados operacionais	EUA
42	(HONG <i>et al.</i> , 2025)	Potência eólica	Dados operacionais	China

Fonte: Autoria própria (2025).

Outra abordagem aplicada em modelos híbridos envolve o uso de aprendizado multitarefa e multimodal, que permite explorar simultaneamente diferentes tarefas relacionadas e integrar múltiplas fontes de dados. O aprendizado multitarefa melhora a generalização e reduz o erro ao compartilhar informações entre tarefas principais e auxiliares, enquanto o aprendizado multimodal combina variáveis meteorológicas e séries temporais de potência, enriquecendo as representações e aumentando a precisão preditiva (WANG, Lei; HE, Yigang; LIU, Xiaoyan; *et al.*, 2022; WANG, Zhongrui *et al.*, 2025).

Além das dependências temporais, a geração eólica também apresenta dependências espaciais, uma vez que parques eólicos geograficamente próximos tendem a ser influenciados por sistemas meteorológicos semelhantes. Para modelar essas interações espaciais dinâmicas, alguns estudos adotam representações em grafo, nas quais a similaridade entre parques é explicitamente quantificada. Nesse contexto, o kernel gaussiano e o coeficiente de correlação de Pearson dinâmico (PCC) são utilizados para capturar as relações espaciais dinâmicas entre os parques eólicos, em que o PCC mede a correlação temporal entre locais e o kernel gaussiano gera a matriz de similaridade usada na construção do grafo espacial do modelo (WANG, Zhongrui *et al.*, 2025).

Conforme apresentado na Tabela 4, as aplicações do Transformer e de suas variantes na literatura abrangem tarefas como previsão de velocidade do vento, previsão de potência eólica e detecção de anomalias. Considerando que o foco central deste trabalho é a previsão de potência eólica, optou-se por aprofundar a discussão nesse domínio. A Tabela 4 sintetiza alguns dos estudos identificados que empregam arquiteturas baseadas em Transformers ou em variantes híbridas, apresentando para cada um o modelo proposto, as técnicas e componentes utilizados, o objetivo principal e suas características-chave.

Os modelos apresentados na Tabela 4 foram desenvolvidos com o objetivo de aprimorar o desempenho da arquitetura Transformer na previsão de potência eólica. Esses estudos também realizaram comparações entre os modelos propostos e abordagens clássicas amplamente utilizadas na literatura, empregando métricas de avaliação preditiva, tais como MAE, MSE e RMSE (detalhadas na Seção 5 deste estudo). A Tabela 5 apresenta os resultados obtidos para alguns dos modelos propostos, em comparação com os modelos de referência, incluindo os horizontes de previsão considerados e as tabelas de referência dos estudos de onde esses resultados foram extraídos.

Tabela 4 - Modelos propostos para previsão de energia eólica

Modelo	Técnicas /Componentes	Objetivo principal	Características-chave
Transformer (QU, K. et al., 2022b)	Transformer padrão	Melhorar a previsão de curto prazo para vários parques eólicos com dependências complexas	Captura longas dependências e correlações espaciais entre parques eólicos
CNN-Informer (WANG, H.-K.; SONG, K.; CHENG, 2022)	CNN + Informer	Aprimorar a previsão de energia eólica por meio da extração de características espaço-temporais	Combina CNN para extração espacial e Informer para modelagem temporal de longo alcance
Transformer (WANG, Lei; HE, Yigang; LI, Lie; et al., 2022)	Transformer padrão	Melhorar a precisão e a eficiência de previsões em múltiplos estágios	Ajuste de hiperparâmetros; equilibra precisão e eficiência computacional
M2TNet (WANG, Lei; HE, Yigang; LIU, Xiaoyan; et al., 2022)	Aprendizado multitarefa e multiorigem baseado em Transformer	Prever potência eólica em horizonte ultracurto utilizando dados de múltiplas fontes	Integra camadas de extração de características, fusão e predição; utiliza aprendizado multimodal e multitarefa
MSIN (HUANG, Xiaohan; JIANG, A., 2022)	Informer + Retropropagação truncada	Melhorar a precisão das previsões de médio e longo prazo	Utiliza variáveis meteorológicas e treinamento truncado para aprimorar interpretabilidade e desempenho
Transformer (HUANG, S.; YAN, Chang; QU, Y., 2023b)	Transformer padrão	Melhorar a capacidade de generalização do modelo em previsões	Estende o Transformer para lidar com entradas generalizadas
VMD-CAT (ZHENG, H. et al., 2023)	VMD + Transformer modificado	Capturar correlações entre padrões de flutuação e segmentos históricos	Modelo híbrido de decomposição e atenção que aprimora a representação temporal
IVMD-FE-Ad-Informer (TIAN et al., 2023)	VMD aprimorado + Entropia Fuzzy + Informer (com função de perda adaptativa)	Reduzir a complexidade na previsão de séries temporais	Utiliza decomposição, entropia e perda adaptativa para aprimorar o desempenho do Informer
IFORNLD (HU, J. et al., 2023)	Informer + Regressão Ordinal + Diversidade de Rótulos	Prever eventos de variação abrupta de potência eólica (<i>ramp events</i>)	Modelo de múltiplas saídas para longas sequências, com melhor desempenho em classificação
GCNInformer (WANG, Hai-Kun et al., 2023)	GCN + Informer	Aprimorar a modelagem de correlações entre turbinas	Utiliza GCN para dependências espaciais e Informer para dependências temporais
TCNInformer (GONG, M. et al., 2023b)	TCN + Informer	Melhorar a extração de características temporais	TCN captura dependências locais; Informer captura padrões de longo alcance
EMD-CCTransformer (LI, N. et al., 2023)	EMD + Convolução Causal + Transformer	Melhorar a previsão de curto prazo de potência eólica	Combina decomposição com atenção causal para modelagem de padrões não estacionários
Powerformer (MO et al., 2024)	Arquitetura Transformer modificada	Aprimorar a extração de características temporais e reduzir a complexidade computacional	Integra variáveis meteorológicas como velocidade e direção do vento, temperatura e pressão
VM-MSI-GTTS (WAN et al., 2024)	VMD + Entrada em múltiplas escalas temporais + GRU + Transformer + CNN + Rede Neural Totalmente Conectada (FCN)	Melhorar previsões multietapas capturando dependências de curto e longo prazo	Utiliza decomposição baseada em VMD; GRU para dependências de curto prazo e Transformer para longo prazo; CNN e FCN para predição final
Wind-Mambaformer (DONG, Z. et al., 2025)	Estrutura do Transformer + Mamba + mecanismo FlowAttention	Melhorar previsão de curto prazo de potência eólica	Utiliza o Mamba para capturar dependências de longo prazo e o mecanismo FlowAttention para reduzir a complexidade computacional
M3STIN (WANG, Zhongrui et al., 2025)	Informer + GAT + Fusão multimodal + Aprendizagem multitarefa + Kernel gaussiano + PCC	Realizar previsão de potência eólica offshore de curto prazo em múltiplas localizações, considerando dependências espaciais e temporais	Integra GAT e Informer; explora correlações espaciais; usa tarefas auxiliares e fusão multimodal para maior precisão e eficiência.
DPPFformer-MEC (HONG et al., 2025)	Transformer + Mamba + FFT + Função de perda de frequência Kernel (FK), Correção de erro ponderada por coeficiente (MEC)	Melhorar a previsão de potência eólica por meio da extração de características em múltiplas escalas e do uso de informações no domínio da frequência	Integra Mamba e Transformer para eficiência e modelagem global; decompõe séries via FFT; usa FK para aprendizado de alta frequência e MEC para correção de erro multivariável

Fonte: Autoria própria (2025).

Tabela 5 - Resultados dos modelos propostos para a previsão de energia eólica

Fonte	Modelos	MAE	MSE	RMSE	Horizonte de previsão
(QU, K. <i>et al.</i> , 2022b)	LSTM	110,2			Curto prazo
	Transformer	81,0			
(WANG, H.-K.; SONG, K.; CHENG, 2022)	LSTM	0,815	1,156	10,748	Curto prazo
	CNN-Informer	0,0064	0,007	0,0086	
(WANG, Lei; HE, Yigang; LI, Lie; <i>et al.</i> , 2022)	SLSTM	7,08		10,15	Curto
	Transformer	4,33		7,51	
(WANG, Lei; HE, Yigang; LIU, Xiaoyan; <i>et al.</i> , 2022)	SVR	15,34		18,48	Curto prazo
	M2TNet	8,33		11,66	
(HUANG, Xiaohan; JIANG, A., 2022)	LSTM	137,019		727,754	Curto prazo
	Informer	61,600		436,576	
	MSIN	41,701		307,553	
(HUANG, S.; YAN, Chang; QU, Y., 2023b)	GRU	99,28		168,20	Curto prazo
	LSTM	86,20		149,91	
	Transformer	10,96		18,70	
(ZHENG, H. <i>et al.</i> , 2023)	ARIMA	1,61		1,94	Curto prazo
	VMD-CAT	0,79		0,93	
(TIAN <i>et al.</i> , 2023)	LSTM	207,879		278,866	Curto prazo
	MLP	361,476		478,394	
	IVMD-FE-Ad-Informer	183,962		252,690	
(HU, J. <i>et al.</i> , 2023)	LSTM-random	918,261		1,376	Curto prazo
	IFORNLD-random	789,224		1,273	
(WANG, Hai-Kun <i>et al.</i> , 2023)	LSTM	0,445	0,592	0,769	Curto prazo
	GCNInformer	0,104	0,024	0,154	
(GONG, M. <i>et al.</i> , 2023b)	GRU	0,399	0,298	0,546	Curto prazo
	LSTM	0,405	0,342	0,569	
	TCN-Informer	0,229	0,108	0,329	
(LI, N. <i>et al.</i> , 2023)	ARIMA	0,502	0,162	0,402	Curto prazo
	EMD-CCTransformer	0,089	0,013	0,113	
(MO <i>et al.</i> , 2024)	LSTM	0,751	0,739	0,859	Curto prazo
	Transformer	0,819	1,042	1,021	
	Powerformer	0,314	0,241	0,491	
(WAN <i>et al.</i> , 2024)	LSTM	0,109	0,030	0,174	Curto prazo
	Informer	0,085	0,019	0,137	
	VM-MSI-GTTS	0,062	0,011	0,104	
(WANG, Zhongrui <i>et al.</i> , 2025)	SVR	11,39		16,08	Curto prazo
	M3STIN	7,97		12,67	
(HONG <i>et al.</i> , 2025)	LSTM	2,361		3,336	Curto prazo
	DPFMformer-MEC	0,2958		0,3218	

Fonte: Autoria própria (2025).

Nota: As métricas apresentadas na Tabela 5 não devem ser comparadas entre estudos distintos, uma vez que cada trabalho adota conjuntos de dados, escalas, frequências temporais, metodologias de pré-processamento e horizontes de previsão próprios. Dessa forma, a Tabela 5 não tem como objetivo estabelecer uma comparação direta entre valores absolutos de erro, mas sim sintetizar as comparações internas realizadas por cada autor, evidenciando a relação de desempenho entre os modelos avaliados sob as mesmas condições experimentais em cada estudo.

De acordo com a Tabela 5, os modelos propostos apresentaram, de modo geral, desempenho superior aos modelos de referência em todos os experimentos. Esse resultado é evidenciado pelas métricas de avaliação, uma vez que menores valores de MAE, MSE e RMSE indicam melhor desempenho preditivo. Outra informação importante é que os erros apresentados na Tabela 5 estão exatamente no formato coletado em cada estudo correspondente. Portanto, os erros podem estar em formato absoluto ou percentual. É importante destacar que os erros apresentados na Tabela 5 são utilizados exatamente no formato disponibilizado pelos estudos originais — podendo estar expressos em valores absolutos ou percentuais — a fim de preservar a comparabilidade metodológica entre as diferentes abordagens.

Para o Transformer, os modelos propostos performaram melhor que modelos como LSTM, GRU, ARIMA, entre outros (QU, K. *et al.*, 2022b) (HUANG, S.; YAN, Chang; QU, Y., 2023b; QU, K. *et al.*, 2022b; WANG, Lei; HE, Yigang; LI, Lie; *et al.*, 2022). Nos estudos que propuseram modelos híbridos, as abordagens desenvolvidas também apresentaram desempenho superior aos modelos de referência, incluindo ARIMA, SVR, GRU, MLP, LSTM e até mesmo o Transformer padrão (GONG, M. *et al.*, 2023b; LI, N. *et al.*, 2023; MO *et al.*, 2024; ZHENG, H. *et al.*, 2023). Já para os estudos que utilizaram o Informer, os modelos propostos também superaram os modelos de referência (GONG, M. *et al.*, 2023b; HU, J. *et al.*, 2023; HUANG, Xiaohan; JIANG, A., 2022; TIAN *et al.*, 2023; WAN *et al.*, 2024; WANG, Hai-Kun *et al.*, 2023; WANG, H.-K.; SONG, K.; CHENG, 2022; WANG, Zhongrui *et al.*, 2025).

Os estudos analisados evidenciam que os modelos baseados na arquitetura Transformer apresentam desempenho notável na previsão de potência eólica a curto prazo, mostrando-se altamente adequados para essa finalidade. Verifica-se, ainda, que aprimoramentos estruturais na arquitetura desses modelos contribuem significativamente para o aumento da capacidade preditiva. Além disso, o modelo Informer destacou-se como uma alternativa promissora para séries temporais com grandes volumes de dados, demonstrando, em alguns casos, desempenho superior ao Transformer padrão, conforme observado nos trabalhos revisados. Algumas derivações do Transformer padrão são

explicadas detalhadamente na Seção 2.1 deste trabalho, como é o caso do Informer e do Autoformer.

3.5.3 Discussão e Limitações

Os estudos analisados nesta revisão indicam que os modelos baseados na arquitetura Transformer são capazes de atribuir relevância de forma adequada aos dados de séries temporais por meio do mecanismo de atenção, podendo ser adaptados para lidar também com entradas espaço-temporais. Essas características tornam essa arquitetura especialmente promissora para aplicações em energia eólica. No entanto, ainda persistem desafios para o aproveitamento pleno de seu potencial em contextos práticos, sobretudo quando se consideram dados reais e requisitos específicos do domínio. As principais limitações identificadas são apresentadas a seguir:

1. **Custo Computacional:** Os Transformers possuem maior demanda computacional, entendido aqui como maior tempo de processamento, maior consumo de memória e maior exigência de *hardware* durante a inferência. Esses fatores podem dificultar seu uso em aplicações que demandam resposta rápida ou operam sob restrições de processamento, como sistemas de previsão em parques eólicos.
2. **Codificação Temporal:** A codificação posicional tradicional do Transformer nem sempre captura adequadamente padrões sazonais, cíclicos ou dinâmicas temporais não lineares observadas em séries reais, o que limita sua capacidade de representar dependências temporais complexas. Por esse motivo, métodos mais expressivos — como o Time2Vec — tornam-se essenciais para melhorar a modelagem das estruturas temporais.
3. **Riscos de Overfitting:** A alta capacidade dos Transformers pode levar ao sobre ajuste em conjuntos de dados limitados, exigindo o uso de técnicas de regularização e validação rigorosa para garantir robustez.
4. **Interpretabilidade e Aplicação Industrial:** Apesar dos mecanismos de atenção, os Transformers ainda funcionam como caixas-pretas, o que dificulta sua aceitação em contextos industriais que exigem transparência e aplicabilidade prática.

5. **Requisitos de Dados:** Muitos modelos necessitam da disponibilidade de dados extensos e de alta qualidade; contudo, em ambientes reais, como parques eólicos, os dados podem ser ruidosos, incompletos ou limitados.

3.6 Lacunas Identificadas na Literatura

A partir da revisão da literatura apresentada na seção anterior, observa-se que ainda persistem lacunas relevantes quanto ao uso da arquitetura Transformer na previsão de energia eólica. Diante disso, este estudo propõe-se a investigar alguns desses pontos em aberto, os quais também se relacionam às limitações discutidas na Seção 3.5.3 deste trabalho. As principais lacunas abordadas são apresentadas a seguir:

- **Custo Computacional:** Grande parte do custo computacional do Transformer decorre da complexidade quadrática do mecanismo FullAttention presente na arquitetura original (Transformer Vanilla). Assim, este trabalho investiga alternativas capazes de mitigar essa limitação por meio dos mecanismos FlowAttention e FlashAttention, que visam reduzir a demanda de processamento mantendo a eficiência preditiva.
- **Codificação Posicional:** O Transformer Vanilla utiliza codificações posicionais senoidais fixas, frequentemente insuficientes para representar dinâmicas temporais complexas. Com base nisso, este estudo propõe a integração de codificações temporais aprendíveis, capazes de capturar de forma mais eficaz sinais periódicos e comportamentos cíclicos, como ciclos diários e sazonais. Para essa integração, realiza-se uma análise de sensibilidade da adoção da camada temporal na arquitetura Transformer, com o objetivo de identificar o arranjo que melhor favorece o desempenho do modelo.
- **Interpretabilidade:** Apesar do elevado desempenho preditivo, os modelos baseados em Transformer ainda apresentam desafios de interpretabilidade, pois o funcionamento interno dos mecanismos de atenção e das representações latentes é complexo. Essa dificuldade é inerente à própria arquitetura. Embora existam abordagens de interpretabilidade — como análise de importância, visualização de atenção e métodos pós-hoc —, sua aplicação em previsão de energia eólica ainda

é recente. Neste estudo, buscamos contribuir analisando o comportamento dos modelos em dados reais de parques eólicos do Nordeste do Brasil, oferecendo evidências empíricas que podem apoiar investigações futuras e servir de referência para pesquisadores e profissionais da indústria.

Ainda sobre as limitações descritas na Seção 3.5.3, para reduzir o risco de *overfitting* e melhorar a qualidade dos dados, este trabalho adota um conjunto de procedimentos metodológicos de pré-processamento e modelagem. O detalhamento dessas etapas é apresentado na Seção 4.

4 METODOLOGIA

Esta seção apresenta a metodologia adotada neste trabalho, descrevendo as etapas de desenvolvimento, os modelos utilizados, os dados empregados e os procedimentos de avaliação.

4.1 Motivação e Formalização da Proposta

A previsão de energia eólica de curto prazo apresenta desafios devido à natureza estocástica e não estacionária do comportamento do vento. Para abordar essas características, este estudo adota modelos baseados no Transformer com aplicação mais adequada a dados de séries temporais. Conforme discutido ao longo desse trabalho, as arquiteturas convencionais do Transformer enfrentam limitações relacionadas a codificações posicionais fixas, considerável complexidade computacional (especialmente em termos de tempo de processamento), e interpretabilidade limitada. Além disso, essas limitações caracterizam lacunas persistentes na literatura. Para superar essas limitações, os mecanismos FlowAttention e FlashAttention são usados para substituir o FullAttention tradicional, reduzindo o custo computacional e melhorando a escalabilidade. Além disso, uma camada de codificação Time2Vec é incorporada na entrada do modelo, para fornecer uma representação mais rica de padrões temporais. Essas modificações visam aprimorar a precisão das previsões e a eficiência computacional, ao mesmo tempo em que fornecem uma base sólida de resultados que pode servir como referência para pesquisas futuras e aplicações no setor.

4.2 Visão Geral da Metodologia

Este estudo visa prever a energia eólica de curto prazo com base em dados operacionais reais de turbinas eólicas, abordando os desafios da previsão de energia precisa e oportuna. Esta seção detalha o processo metodológico adotado neste estudo,

sendo representado pela Figura 16 que resume o processo metodológico utilizado.

A primeira etapa envolve a coleta de dados de turbinas eólicas em operação por meio do sistema de SCADA. Para garantir a qualidade dos dados, a segunda etapa envolve procedimentos de pré-processamento e filtragem, incluindo limpeza de dados, remoção de outliers e padronização de séries temporais para reduzir o ruído e facilitar o treinamento dos modelos. A remoção de outliers foi realizada por meio de testes de qualidade locais aplicados aos dados de energia eólica observados da turbina analisada. Esses testes hierárquicos visam verificar a consistência física e estatística da variável e detectar comportamentos anômalos de curto prazo, incluindo verificação de intervalo, verificação de persistência e verificação de degrau de curto prazo. Os valores faltantes ou removidos durante o processo foram tratados por interpolação linear, técnica adequada para séries densas e de alta frequência como as geradas por sistemas SCADA. Esse procedimento garante continuidade temporal e preserva a coerência global da série antes do treinamento dos modelos. Essas etapas garantiram que o conjunto de dados estivesse devidamente dimensionado e consistente para a variável potência ativa (variável alvo deste trabalho). Finalmente, os dados processados são validados e preparados para uso na etapa subsequente.

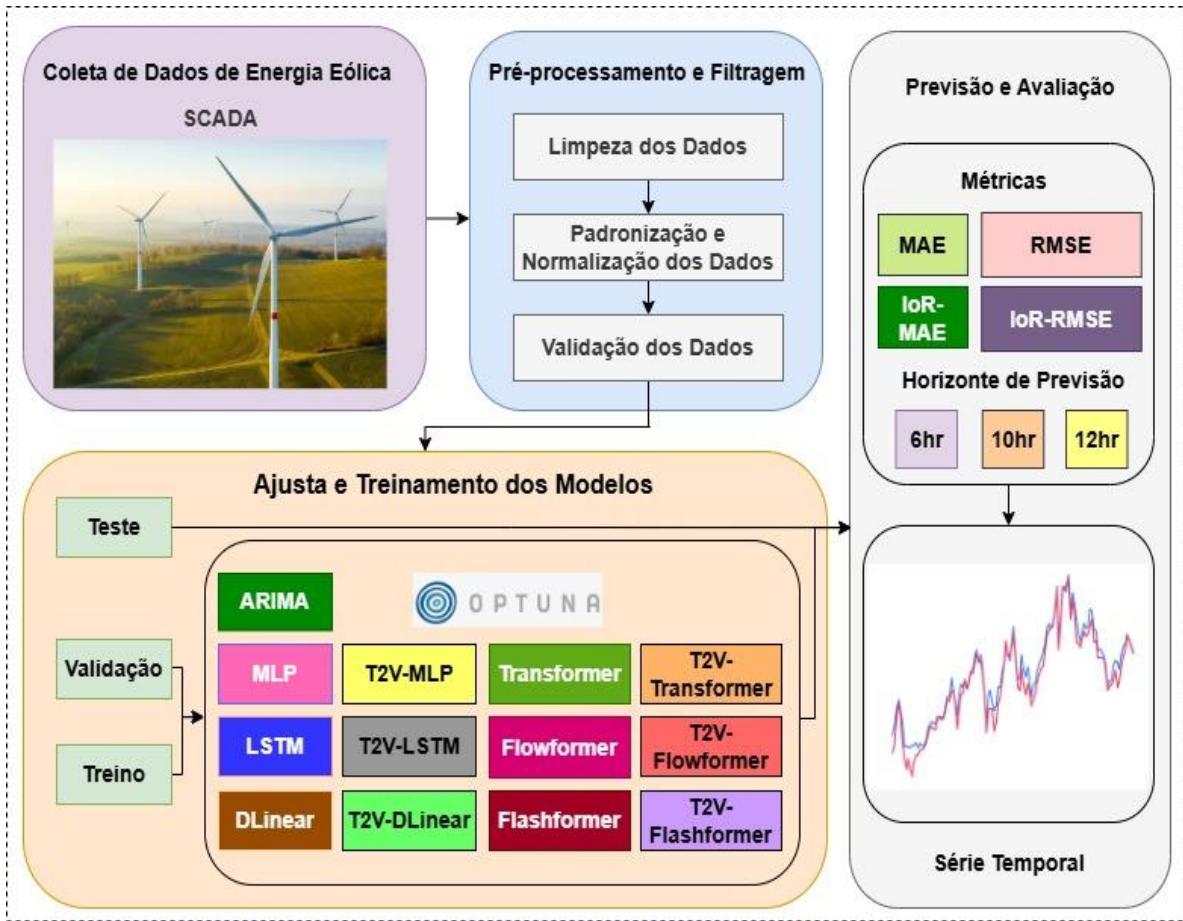
A terceira etapa concentra-se no ajuste e treinamento dos modelos empregados neste estudo. Os modelos utilizados incluem ARIMA, MLP, LSTM, DLinear, T2V-MLP, T2V-LSTM, T2V-DLinear, Transformer, Flowformer e Flashformer, bem como os modelos propostos T2V-Transformer, T2V-Flowformer e T2V-Flashformer. O termo T2V refere-se à incorporação da camada Time2Vec nos respectivos modelos. Além disso, também foi utilizado o modelo de Persistência (DUTTA *et al.*, 2017), frequentemente usado como referência na previsão de séries temporais. Este modelo trivial pressupõe que o valor da variável em um determinado instante t será igual ao valor observado no instante $t - 1$. Em outras palavras, a previsão para o próximo ponto da série é o valor atualmente observado. A terceira etapa envolve a divisão do conjunto de dados em conjuntos de treinamento, validação e teste, conforme ilustrado na Figura 16. Durante a fase de treinamento, os modelos recebem dados e ajustam seus parâmetros por meio da minimização da função de perda Erro Quadrático Médio (MSE), definida pela Equação (14), utilizando o algoritmo de retropropagação do erro (*backpropagation*) para a atualização dos pesos via método de otimização.

$$Loss = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

onde y_i representa o valor real, \hat{y}_i representa o valor prevista, e n é número de amostras no lote. Essa função tem o objetivo de medir a diferença quadrática média entre as previsões e os valores-alvo. Os otimizadores utilizados na fase de treinamento foram Adam (KINGMA; BA, 2014), *Root Mean Square Propagation* (RMSprop) (HINTON; SRIVASTAVA; SWERSKY, 2012) e *Stochastic Gradient Descent* (SGD) (BOTTOU, 2010). O método de *early stopping* foi adotado para encerrar o treinamento assim que não fossem observadas melhorias no desempenho do modelo. Em todas as simulações, utilizou-se um critério de paciência de 5, considerando a evolução das previsões em relação aos valores-alvo.

Durante a fase de validação, a função de perda MSE foi empregada exclusivamente para a avaliação do desempenho do modelo, não havendo retropropagação do erro nem atualização dos parâmetros. Adicionalmente, utilizou-se o Optuna (AKIBA *et al.*, 2019), biblioteca de busca de hiperparâmetros para identificar a melhor configuração possível para cada modelo (a explicação detalhada desta fase encontra-se na Seção 4.6). Na fase de teste, os modelos treinados e otimizados são avaliados usando os dados de teste para analisar seu desempenho. O quarto estágio envolve a previsão e avaliação dos modelos, com base nas métricas descritas na Figura 16 para horizontes de referência de 6, 10 e 12 horas à frente. Para os modelos propostos, uma análise de sensibilidade foi conduzida para determinar a configuração mais eficaz para integrar o mecanismo Time2Vec na arquitetura do Transformer.

Figura 16 - Representação do procedimento metodológico utilizado neste estudo



Fonte: Autoria própria (2025).

4.3 Análise de Sensibilidade na Integração Time2Vec

Para integrar o Time2Vec à arquitetura Transformer, foram avaliados três diferentes arranjos com o objetivo de identificar aquele que promove o melhor desempenho na previsão de energia eólica. A análise de sensibilidade dos arranjos testados é apresentada na Figura 17, e cada configuração é detalhada a seguir.

1. Arranjo I: utiliza tanto o codificador quanto o decodificador, com Time2Vec adicionado exclusivamente à entrada do codificador.

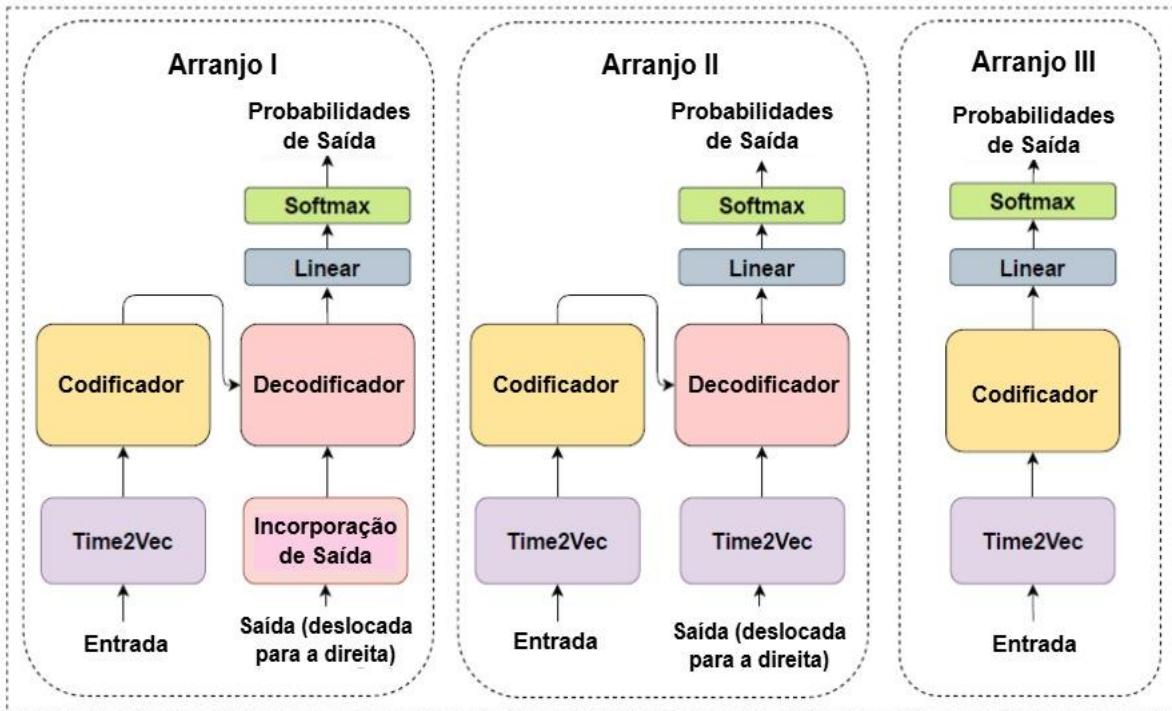
2. Arranjo II: utiliza tanto o codificador quanto o decodificador, com Time2Vec incorporado às entradas do codificador e do decodificador.

3. Arranjo III: utiliza apenas o codificador, sem o decodificador, com Time2Vec aplicado à entrada do codificador.

Até onde a literatura científica indica, esta é a primeira vez que uma análise de

sensibilidade tão específica foi conduzida na integração do Time2Vec na arquitetura do Transformer. Os resultados obtidos nos experimentos, conforme apresentados no Apêndice B, indicam que o Arranjo I proporcionou as condições mais favoráveis para o desempenho do modelo. Esta configuração obteve o maior desempenho de acordo com as métricas de avaliação empregadas neste estudo. Portanto, este foi o arranjo adotado para os modelos propostos neste estudo. A adição do Time2Vec apenas no codificador permitiu que o modelo aprendesse padrões temporais de forma mais eficiente. O decodificador, por sua vez, concentra-se em gerar a saída com base nessas representações, sem a necessidade de incorporar informações temporais novamente. Essa abordagem, portanto, evita complexidade desnecessária, mantendo o desempenho otimizado. No entanto, a remoção do decodificador da arquitetura do modelo comprometeu a capacidade do modelo de gerar previsões adequadamente, visto que o decodificador é crucial para transformar representações codificadas em saídas previsíveis. Portanto, os Arranjos II e III foram desconsiderados neste estudo.

Figura 17 - Análise de sensibilidade na integração da camada Time2Vec na arquitetura Transformer



Fonte: Autoria própria (2025).

A camada Time2Vec substitui a camada original de codificação posicional do Transformer no codificador. Esta camada foi descrita detalhadamente na Seção 2.1 deste

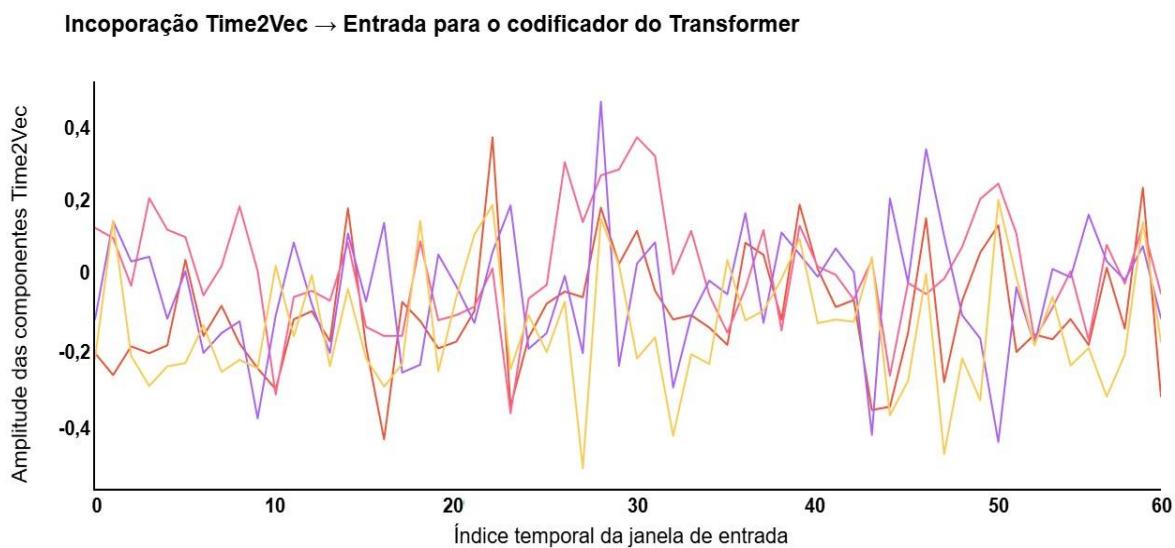
estudo. Assim, conforme o Arranjo I, a codificação posicional original do Transformer é mantida apenas na entrada do decodificador.

Para visualizar como cada série temporal é processada no modelo, as Figuras 18 e 19 apresentam o detalhamento da integração entre a camada Time2Vec e o Transformer conforme os Estudos de Caso utilizados neste trabalho. O conjunto de dados para treino, validação e teste de cada Estudo de Caso encontram-se na Seção 4.5, bem como outras informações importantes. Especificamente, a saída do Time2Vec concatena componentes lineares e periódicas aprendidas pelo Time2Vec ao vetor de entrada na primeira camada do codificador (Arranjo I). Essa fusão enriquece o vetor de entrada encaminhado ao mecanismo de atenção do codificador. As Figura 18 e 19 ilustram explicitamente:

1. A geração do vetor temporal pelo Time2Vec;
2. O fluxo resultante até o primeiro bloco do codificador do Transformer.

Em ambas as Figuras, o eixo x representa o índice temporal da janela de entrada, contendo 60 horas selecionadas dentro do intervalo temporal referente ao conjunto de treinamento de cada Estudo de Caso (ver Tabelas 6 e 7, Seção 4.5). O eixo y representa a amplitude das componentes aprendidas pela camada Time2Vec, isto é, as dimensões do vetor de entrada após a transformação temporal. Dessa forma, as Figuras 18 e 19 documentam de forma clara como o Time2Vec é incorporado a arquitetura do Transformer.

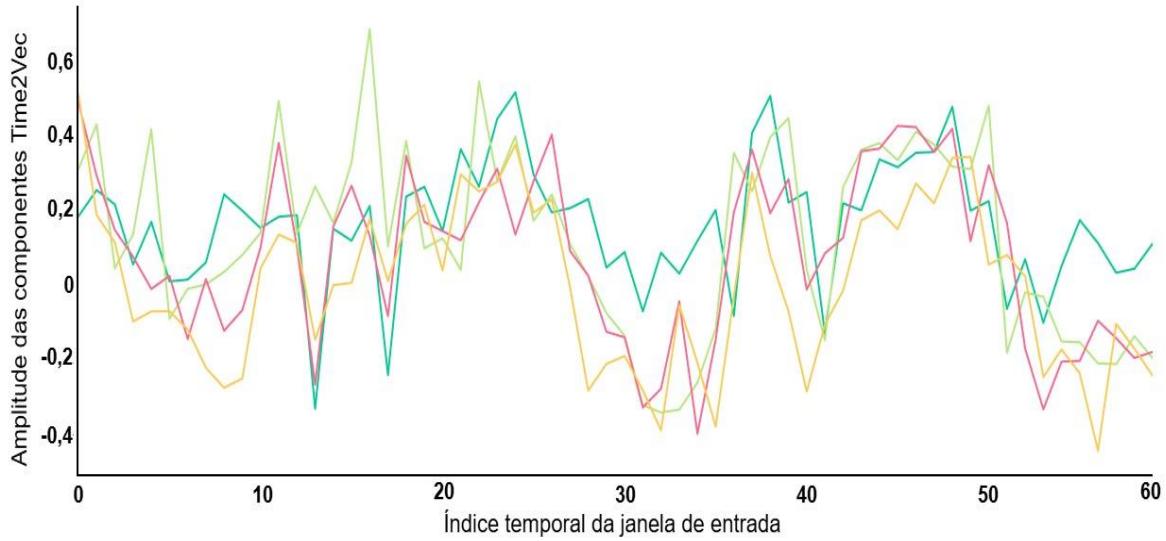
Figura 18 - Representação para o Estudo de Caso 1 (Cenário A) das dimensões geradas pela camada Time2Vec para uma janela de entrada de 60 horas, utilizada como incorporação temporal na entrada do Transformer



Fonte: Autoria própria (2025).

Figura 19 - Representação para o Estudo de Caso 2 das dimensões geradas pela camada Time2Vec para uma janela de entrada de 60 horas, utilizada como incorporação temporal na entrada do Transformer

Incorporação Time2Vec → Entrada para o codificador do Transformer



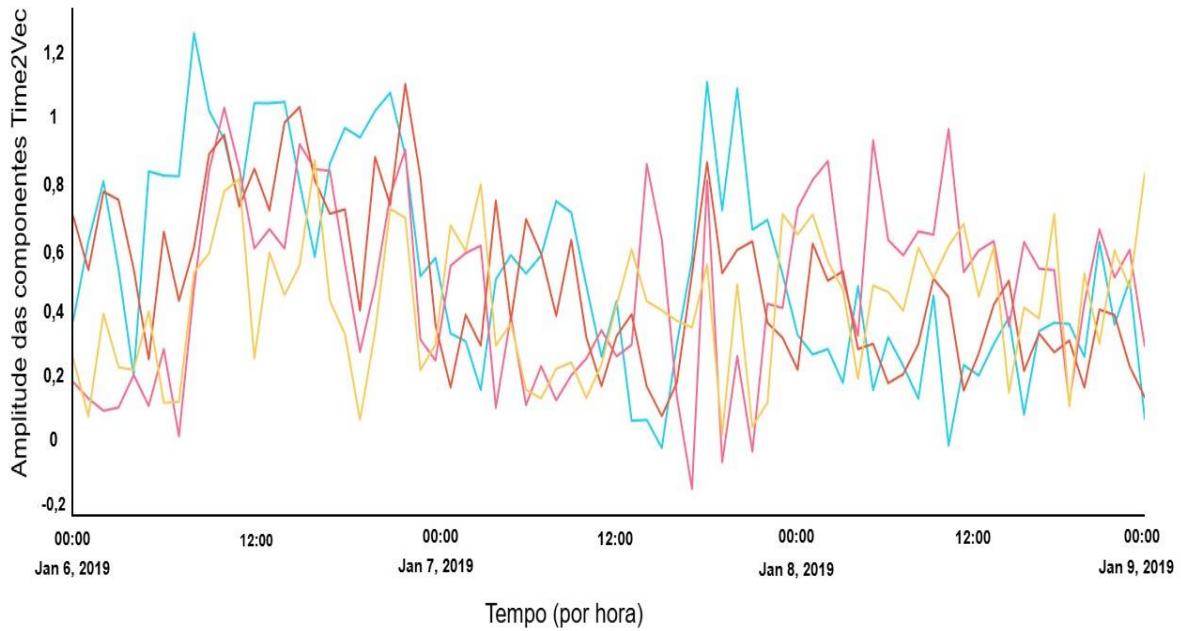
Fonte: Autoria própria (2025).

A camada Time2Vec também foi implementada nos modelos MLP, LSTM e DLinear. Isso foi feito para verificar como a camada se comporta com outras arquiteturas e avaliar seu potencial para melhorar o desempenho em uma variedade de tipos de modelos. O modelo T2V-DLinear, que introduz a camada Time2Vec na arquitetura DLinear, representa uma abordagem inovadora na literatura científica. Embora o foco principal deste estudo seja a integração do Time2Vec em modelos baseados em Transformer, o T2V-DLinear serve como um *benchmark* adicional para demonstrar a versatilidade do Time2Vec em diferentes arquiteturas.

No contexto deste estudo, as componentes aprendidas apresentadas na Figura 20 correspondem às representações temporais produzidas pela camada Time2Vec para cada instante da série, de uma forma mais ampla. Cada curva representa uma dimensão distinta desse vetor temporal, combinando componentes lineares e periódicos capazes de modelar padrões cíclicos, sazonais e tendências presentes nos dados de potência eólica. Embora essas representações não sejam interpretáveis no sentido clássico — como variáveis físicas explícitas —, elas fornecem uma estrutura temporal que o modelo utiliza para identificar relações e dependências ao longo do tempo. A visualização permite observar a variação dessas componentes no horizonte analisado, evidenciando como o Time2Vec

captura oscilações e comportamentos recorrentes relevantes para a previsão. A Figura 20 ilustra um segmento de três dias da série temporal, compreendido entre 6 e 9 de janeiro de 2019, referente a série temporal do Estudo de Caso 2.

Figura 20 - Representações de características Time2Vec aprendidas para a série temporal de energia eólica do Estudo de Caso 2



Fonte: Autoria própria (2025).

4.4 Modelos Propostos

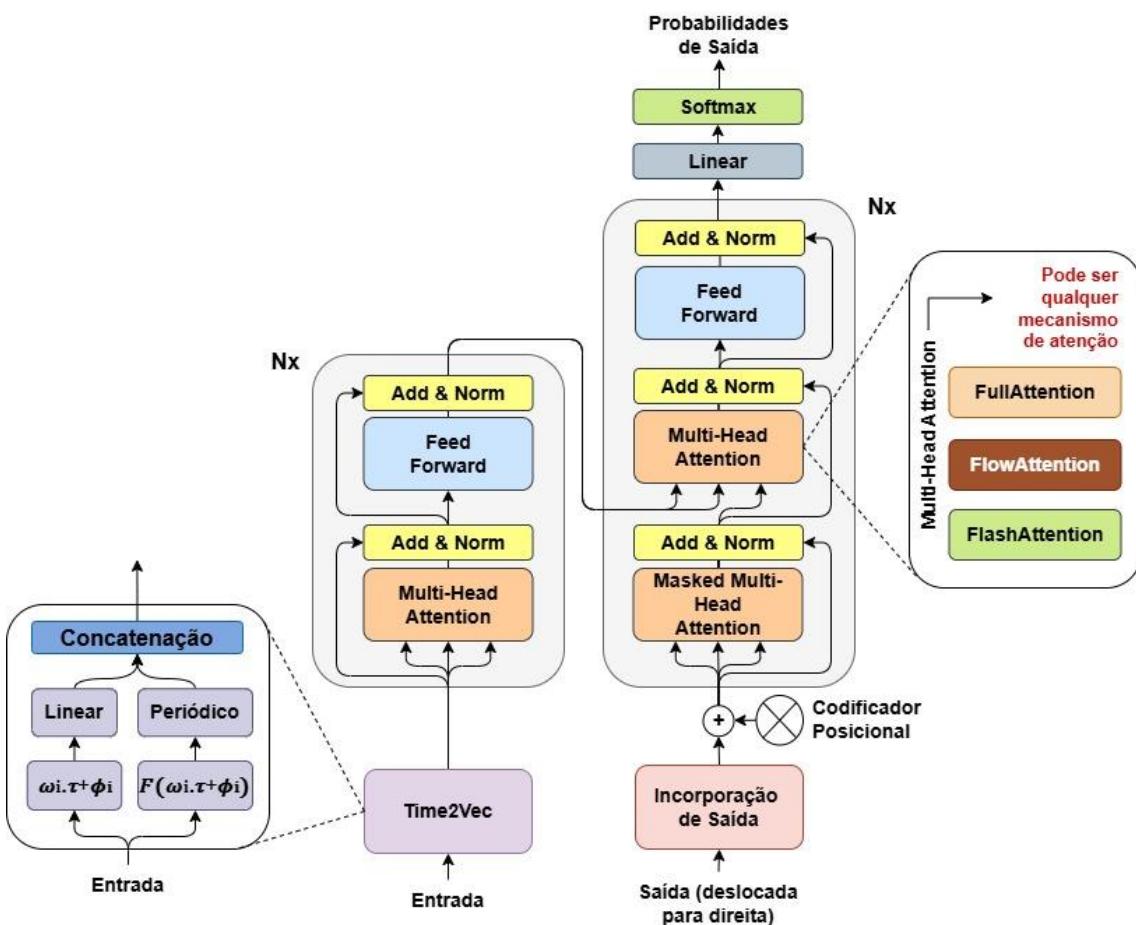
As alterações propostas para a arquitetura do Transformer referem-se ao Arranjo I, mostrado na seção anterior. Além disso, o mecanismo de atenção clássico conhecido como FullAttention foi substituído pelos mecanismos FlowAttention e FlashAttention. O modelo proposto é ilustrado na Figura 21. Qualquer um desses mecanismos de atenção pode ser adotado. Os modelos que usam FullAttention, FlowAttention e FlashAttention neste trabalho são chamados de T2V-Transformer, T2V-Flowformer e T2V-Flashformer, respectivamente. Os modelos propostos cobriram a análise de sensibilidade descrita na Seção 4.3. Eles seguem a arquitetura clássica de codificador-decodificador, com a flexibilidade de modificar o mecanismo de atenção, conforme ilustrado na Figura 21.

Vale destacar que a arquitetura proposta permite a implementação de diferentes mecanismos de atenção, o que reforça sua flexibilidade. Conforme novos mecanismos

são desenvolvidos na literatura científica, eles podem ser incorporados ao modelo sem necessidade de alterações estruturais profundas.

Na literatura científica, a integração do Time2Vec exclusivamente na entrada do codificador do Transformer foi proposta em (VAJIRE *et al.*, 2024), No entanto, esse modelo difere dos propostos neste trabalho devido a modificações aplicadas ao decodificador. Especificamente, os autores empregaram uma camada de Agrupamento Médio Global (*Global Average Pooling*) seguida por uma camada *dropout* e uma camada de saída densa, omitindo completamente o mecanismo de atenção.

Figura 21 - Arquitetura dos modelos propostos



Fonte: Autoria própria (2025).

4.5 Estudos de Caso

Para avaliar a robustez dos modelos propostos, foram conduzidos dois estudos de

caso. O **Estudo de Caso 1** refere-se aos dados operacionais de uma turbina eólica localizada no Nordeste do Brasil, considerando dois cenários distintos. O **Estudo de Caso 2**, por sua vez, baseia-se em dados operacionais de uma central eólica completa, também situada na região Nordeste do Brasil. Ambos os casos são descritos detalhadamente nas Seções 4.5.1 e 4.5.2.

A metodologia adotada neste trabalho inclui múltiplos recortes temporais intencionalmente distintos entre os estudos de caso — dois cenários no Estudo de Caso 1 e um intervalo independente no Estudo de Caso 2 — com o objetivo de avaliar a estabilidade e o desempenho dos modelos em diferentes regimes operacionais e sazonais. Essa abordagem amplia a capacidade de generalização dos experimentos, uma vez que os modelos são analisados em contextos temporais heterogêneos, e não apenas em um único período fixo.

Em tarefas de previsão de séries temporais, a preservação da ordenação cronológica é fundamental para evitar vazamento de informação. Por esse motivo, técnicas de validação cruzada aleatória não foram aplicadas nesta tese. As divisões adotadas para cada conjunto de dados (treinamento, validação e teste) foram definidas de forma a garantir equilíbrio entre quantidade de dados disponível para aprendizado, estabilidade no ajuste de hiperparâmetros e avaliação independente do desempenho. Essas proporções estão de acordo com estudos da literatura, conforme discutido nas Seções 4.5.1 e 4.5.2.

Além disso, o uso de Estudos de Caso com intervalos temporais distintos contribui para atenuar o risco de que o desempenho observado esteja associado a características específicas de um único período. Dessa forma, a análise comparativa entre modelos torna-se mais robusta. Considera-se ainda que o objetivo central desta tese é propor, analisar e comparar arquiteturas derivadas de Transformers em cenários realistas de previsão de curto prazo, e não avaliar diferentes estratégias de validação temporal. Assim, o delineamento experimental adotado mostra-se apropriado para o propósito da pesquisa.

4.5.1 Estudo de Caso 1

Os dados para este estudo foram obtidos de um parque eólico operacional localizado no Nordeste do Brasil, no estado do Rio Grande do Norte. Embora o parque eólico seja composto por múltiplas turbinas eólicas, o Estudo de Caso 1 concentra-se nos

dados de uma única turbina com uma capacidade nominal de potência de aproximadamente 2.300 kW. Os dados foram coletados pelo SCADA, entre janeiro de 2019 e outubro de 2020, totalizando 22 meses de observações com uma frequência de amostragem de uma hora. Inicialmente a série temporal consistia em seis variáveis:

1. Carimbo de data e hora - referência de tempo para cada registro;
2. Velocidade do vento (m/s) — medida no anemômetro da turbina;
3. Potência ativa (kW) — potência elétrica de saída (variável alvo);
4. Velocidade do rotor (rpm) — velocidade de rotação do rotor;
5. Ângulo de inclinação ($^{\circ}$) — posição angular das pás;
6. Posição da nacela ($^{\circ}$) — orientação de guinada da turbina em relação à direção do vento.

Como a variável de interesse desta pesquisa é a potência ativa, todas as demais variáveis foram desconsideradas durante a etapa de pré-processamento e filtragem. Após a padronização da série temporal para frequência horária, o conjunto de dados passou a conter 16.080 linhas, das quais 58 apresentavam valores faltantes. Esses valores ausentes foram tratados por meio do procedimento de interpolação descrito na Seção 4.2, garantindo a continuidade temporal da série antes do treinamento dos modelos.

Para avaliar a robustez preditiva dos modelos utilizados neste estudo, é essencial avaliar seu desempenho em diferentes períodos de tempo. Isso evita que o modelo seja restrito a padrões específicos de uma única estação ou condição climática, aumentando sua capacidade de generalização para novas situações. Ao expor o modelo a variações sazonais e mudanças na dinâmica do vento ao longo do tempo, podemos avaliar melhor sua adaptabilidade e desempenho em cenários do mundo real. Portanto, este estudo considera duas condições temporais distintas. O Cenário A representa a transição do verão para o outono, enquanto o Cenário B corresponde à transição do inverno para a primavera. Com base no calendário sazonal do Brasil, o conjunto de dados foi dividido em três partes: conjunto de treinamento, conjunto de validação e conjunto de teste (ver Tabela 6).

Tabela 6 - Visão geral dos conjuntos de treinamento, validação e teste para cada cenário do Estudo de Caso 1

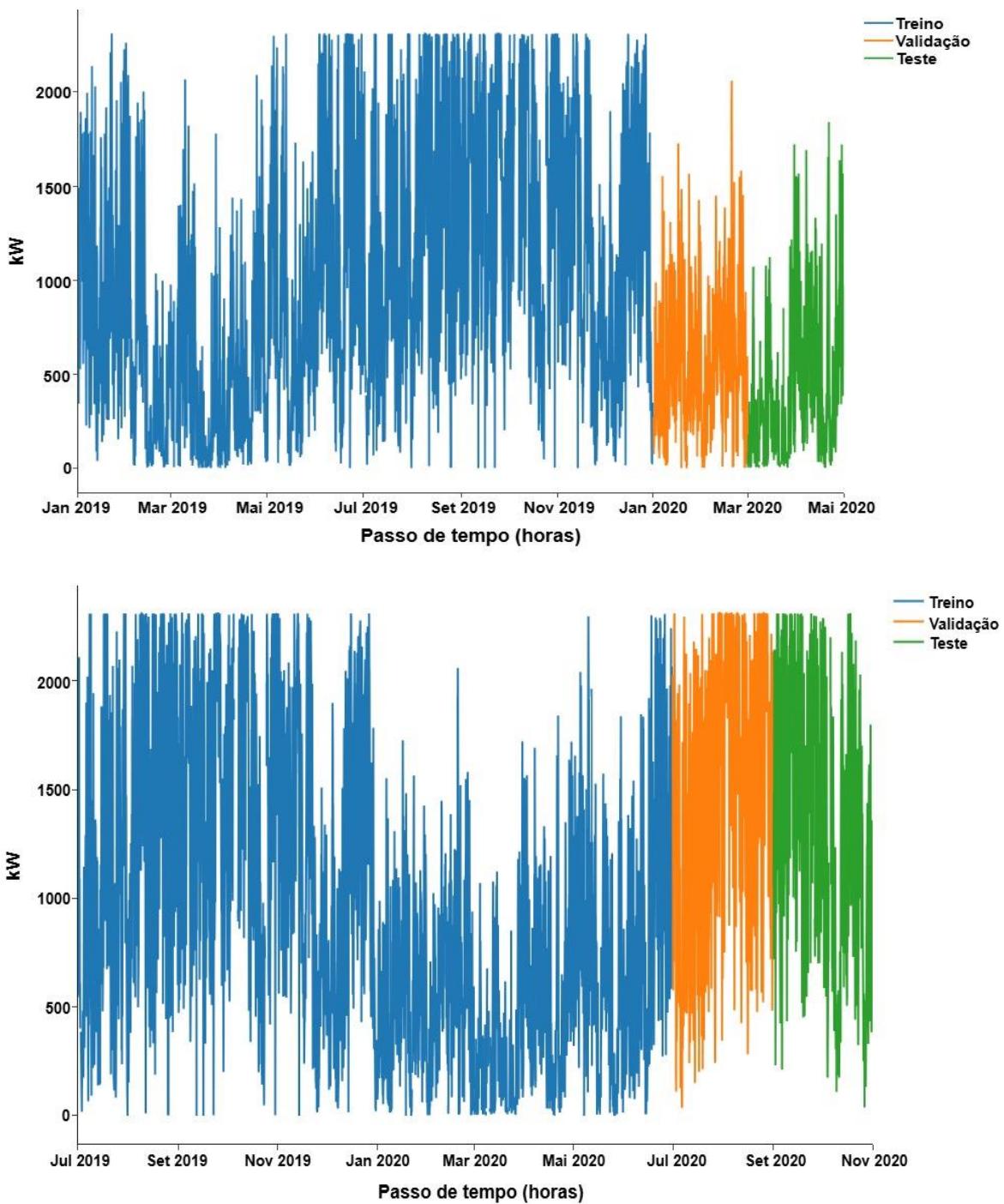
Cenário A	
Conjunto de treinamento	(01/01/2019) - (31/12/2019)
Conjunto de validação	(01/01/2020) – (29/02/2020)
Conjunto de teste	(01/03/2020) – (30/04/2020)
Cenário B	
Conjunto de treinamento	(01/07/2019) - (30/06/2020)
Conjunto de validação	(01/07/2020) - (31/08/2020)
Conjunto de teste	(01/09/2020) - (31/10/2020)

Fonte: Autoria própria (2025).

A prática de alocar mais tempo para o conjunto de treinamento é amplamente adotada na literatura científica. Por exemplo, (ZHA *et al.*, 2023b) utilizou uma proporção de 4:1:1 para os conjuntos de treinamento, validação e teste, enquanto (WANG, Lei; HE, Yigang; LI, Lie; *et al.*, 2022) empregou uma proporção de 3:1:1. Já a proporção de 6:1:1 foi adotada por (FU; WANG; JIN, 2025). Com base nisso, este estudo adota uma proporção de 6:1:1, fornecendo mais dados para o conjunto de treinamento. Essa divisão garante um equilíbrio adequado entre aprendizado, ajuste de hiperparâmetros e avaliação do modelo. Considerando que o horizonte máximo de previsão é de 12 horas à frente e os dados são coletados de hora em hora, o período de 12 meses fornece uma quantidade substancial de dados. Isso permite que os modelos capturem vários padrões sazonais e dinâmicas de séries temporais, tornando o processo de treinamento mais robusto e eficaz. Com a proporção de 6:1:1, tanto o conjunto de validação quanto o de teste contêm 2 meses de dados cada.

A Figura 22 ilustra a potência eólica da turbina eólica em estudo para os dois cenários propostos (A e B). A imagem superior corresponde ao Cenário A, enquanto a imagem inferior corresponde ao Cenário B. Estes são dados reais de uma turbina eólica atualmente em operação. O maior potencial eólico foi observado entre julho e dezembro de 2019 e entre julho e outubro de 2020, enquanto o menor ocorreu entre janeiro e abril de ambos os anos. Portanto, é evidente que os dois cenários capturam condições temporais distintas do parque eólico em estudo.

Figura 22 - Séries temporais da potência ativa para a turbina eólica do Estudo de Caso 1: Cenário A (acima) e Cenário B (abaixo)



Fonte: Autoria própria (2025).

4.5.2 Estudo de Caso 2

Os dados para este estudo foram obtidos de um parque eólico localizado na

Chapada do Araripe, região que se estende entre os estados do Ceará, Pernambuco e Piauí. Portanto, o parque eólico explorado também se localiza no Nordeste do Brasil. Porém, por apresentar-se em região distante do parque eólico do Estudo de Caso 1, apresenta diferenças geográficas e meteorológicas. Para o Estudo de Caso 2, apenas um cenário foi considerado. O parque eólico tem potência nominal de aproximadamente 29,7 MW. Os dados foram coletados também pelo sistema SCADA, entre agosto de 2019 até setembro de 2020, totalizando 14 meses, com uma frequência de amostragem a cada 1 hora.

A base de dados continha apenas a variável de potência ativa, totalizando 10.248 linhas, das quais 120 apresentavam valores ausentes. Esses valores faltantes foram tratados por meio do procedimento de interpolação descrito na Seção 4.2, garantindo a continuidade temporal da série antes do treinamento dos modelos.

A Tabela 7 apresenta a divisão do conjunto de dados adotada neste estudo, composta por 12 meses para o treinamento, 1 mês para a validação e 1 mês para o teste, na proporção de 12:1:1. Diferentemente do Estudo de Caso 1, que utiliza a proporção 6:1:1, essa proporção foi adotada com o objetivo de avaliar a robustez dos modelos frente a diferentes proporções de dados para treinamento, validação e teste. Essa configuração também foi adotada em estudos anteriores, como em Bispo Junior *et al.* (2025). O treinamento foi realizado com dados do período de agosto de 2019 a julho de 2020, enquanto os conjuntos de validação e teste correspondem aos meses de agosto e setembro de 2020, respectivamente. Essa configuração visa assegurar maior disponibilidade de dados para o treinamento dos modelos, preservando períodos mais curtos para a etapa de validação e avaliação final.

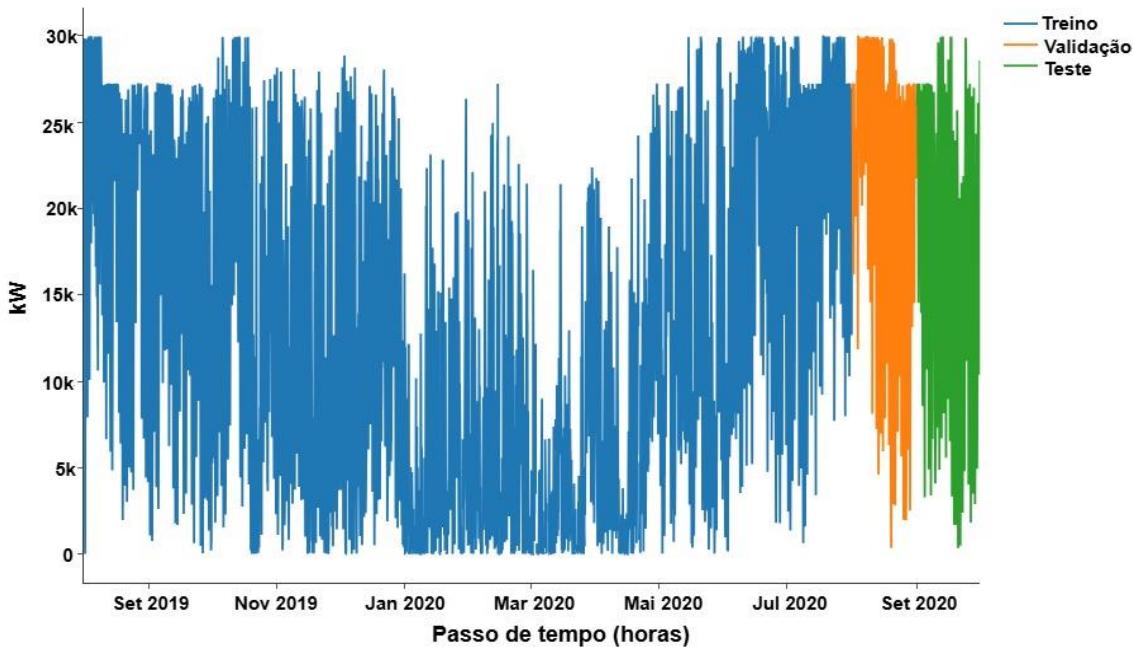
A Figura 23 ilustra a série temporal de geração da usina eólica considerada no Estudo de Caso 2. Observa-se uma redução nos níveis de produção entre os meses de janeiro e abril, ao passo que o intervalo de junho a setembro apresenta os maiores valores de potência nominal gerada. De modo análogo ao Estudo de Caso 1, as cores azul, laranja e verde indicam, respectivamente, os conjuntos de treinamento, validação e teste.

Tabela 7 - Visão geral dos conjuntos de treinamento, validação e teste

Conjunto de treinamento	(01/08/2019) - (31/07/2020)
Conjunto de validação	(01/08/2020) – (31/08/2020)
Conjunto de teste	(01/09/2020) – (30/09/2020)

Fonte: Autoria própria (2025).

Figura 23 - Série temporal da potência ativa da usina eólica do Estudo de Caso 2



Fonte: Autoria própria (2025).

4.6 Análise Experimental

A análise experimental deste estudo foi estruturada em duas etapas distintas, correspondentes aos Estudos de Caso 1 e 2. Todos os experimentos, bem como o processamento dos dados e a implementação dos modelos, foram conduzidos em Python.

Para o Estudo de Caso 1, os experimentos foram realizados em ambiente *PyTorch*, utilizando um sistema equipado com uma GPU Nvidia RTX A4000, de nível profissional, baseada na arquitetura Ampere, com 16 GB de VRAM, otimizada para tarefas de aprendizado profundo e computação de alto desempenho.

No Estudo de Caso 2, os experimentos também foram conduzidos em *PyTorch*, porém em um sistema com GPU Nvidia RTX 4060 Ti, com 8 GB de VRAM, adequada à execução de tarefas de aprendizado profundo em configurações de menor custo computacional.

Essa distinção foi adotada com o objetivo de avaliar o custo computacional e a viabilidade de implementação dos modelos propostos em diferentes contextos experimentais.

Para ambos os casos, foi realizada uma busca por hiperparâmetros para cada

modelo, utilizando o Optuna. O número de ensaios foi definido como 100 neste estudo, com base em um equilíbrio entre o custo computacional e a necessidade de exploração suficiente do espaço de hiperparâmetros. Essa escolha está alinhada à recomendação de (AKIBA *et al.*, 2019), onde 100 ensaios foram utilizados em seu exemplo para a otimização de hiperparâmetros. Esse número permite um bom equilíbrio entre o desempenho do modelo e o tempo disponível para experimentação.

Na análise experimental de ambos os casos o *batch size* foi considerado 16. O comprimento da sequência (*seq len*) foi incluído na busca de hiperparâmetros, com valores variando de 6 a 180. Este intervalo foi escolhido para garantir que o modelo fosse capaz de aprender dependências temporais tanto de curto quanto de longo alcance dentro da série temporal, permitindo que o modelo capturasse tanto variações rápidas quanto padrões mais duradouros nas sequências de dados. A escolha de 6 a 180 como intervalo de *seq len* busca equilibrar o aprendizado de padrões temporais relevantes sem introduzir complexidade desnecessária. Sequências muito curtas poderiam não capturar padrões de longo alcance, enquanto sequências muito longas poderiam aumentar a complexidade computacional sem agregar benefício significativo. Vale ressaltar que, neste contexto, o termo "longo prazo" se refere à capacidade do modelo de capturar dependências temporais mais distantes dentro da janela de entrada, não ao horizonte máximo de previsão, que permanece de 12 horas. O comprimento do rótulo (*label len*), definido como metade do *seq len*, foi mantido constante em todos os cenários para garantir uma proporção equilibrada entre entrada e saída.

Os horizontes de previsão considerados — 6, 10 e 12 horas — estão alinhados com o foco deste estudo em previsões de curto prazo, permitindo que os modelos capturem com maior precisão a dinâmica do futuro próximo da geração de energia eólica. A função de ativação empregada foi a ReLU, que introduz não linearidade nas redes, permitindo que os modelos aprendam padrões complexos nos dados e acelerando o processo de treinamento ao evitar a saturação do gradiente. O *dropout* foi definido com valor de 0,1. Essa técnica realiza a regularização das redes neurais, evitando *overfitting*.

4.6.1 Estudo de Caso 1

Os resultados na busca de hiperparâmetros para o Estudo de Caso 1 são

representados pelas Tabelas 8 e 9, que mostram os parâmetros finais de cada modelo, referenciando as melhores configurações encontradas para o Cenário A e o Cenário B, respectivamente. Os modelos LSTM, MLP e DLinear foram treinados com 50 épocas. Enquanto os modelos propostos e os X-formers foram treinados com 10 épocas. Isso demonstra que os modelos propostos e os X-formers convergem mais rápido do que os modelos de referência, provavelmente devido aos seus mecanismos de atenção eficientes e capacidade superior de capturar padrões complexos. Consequentemente, eles alcançam desempenho ideal em menos épocas, reduzindo o risco de *overfitting*.

De acordo com a Tabela 8, diferentes arquiteturas manipulam dados históricos de maneiras únicas, e o comprimento da sequência varia com base na capacidade de cada modelo de processar e extrair informações relevantes. O MLP tem o maior *seq len*, com um valor de 104, enquanto o T2V-Transformer tem o menor, com um *seq len* de 32. Em relação ao número de camadas, o MLP tem 3, e o T2V-MLP tem 2. Para o LSTM e o T2V-LSTM, o número de camadas é 2 e 1, respectivamente. Ambos os modelos são bidirecionais, o que significa que processam a sequência em duas direções: do passado para o futuro e do futuro para o passado. Essa bidirecionalidade permite que os modelos capturem dependências temporais globais, alavancando informações passadas e futuras, o que é essencial para prever padrões complexos, como na previsão de energia eólica. Em relação aos X-formers e aos modelos propostos, o Transformer e o T2V-Flashformer apresentam um maior número de camadas no codificador (3 e 2, respectivamente) do que no decodificador (1 e 1, respectivamente). Para o T2V-Flowformer e o Flashformer, o número de camadas no codificador foi menor do que no decodificador, com 2 e 3, respectivamente. Para o T2V-Transformer e o Flowformer, o número de camadas foi o mesmo para o codificador e o decodificador. Para todos os modelos de *benchmark*, Adam foi o melhor otimizador. Para os X-formers e os modelos propostos, RMSprop foi o melhor otimizador. O maior d_{model} para os modelos foi 256 para o T2V-Flashformer. Enquanto o menor foi 32, para o T2V-Flowformer e o Flashformer. O maior d_{ff} foi 768 para o Transformer e o Flowformer. O menor foi 64 para o Flashformer. O maior número de cabeças foi 8 para o Flowformer. O menor valor foi 2, para o Flashformer. Quanto maiores esses três parâmetros, melhor o desempenho em tarefas complexas, embora ao custo de maiores recursos computacionais. O oposto é verdadeiro para valores menores.

Como mencionado anteriormente, um total de 100 ensaios foram realizados para cada modelo na busca por hiperparâmetros usando o Optuna. As Tabelas 8 e 9 mostram

o tempo que cada modelo levou para apenas um ensaio. Esses valores indicam que os modelos são aplicáveis e adequados para previsões com 12 horas a frente. É possível observar que o maior tempo foi para o Transformer e o T2V-Transformer, devido à complexidade quadrática do mecanismo FullAttention, conforme explicado na Seção 2.1. Para modelos com adição de Time2Vec, a função de ativação periódica, denotada por F (de acordo com a Equação 9), foi selecionada para ser uma função seno ou cosseno. Portanto, a melhor configuração da função para cada modelo é apresentada nas Tabelas 8 e 9.

Tabela 8 - Resultados finais da busca de hiperparâmetros para todos os modelos avaliados no Estudo de Caso 1 (Cenário A)

Parâmetro	ARIMA	MLP	LSTM	DLinear	T2V-MLP	T2V-LSTM	T2V-DLinear
<i>seq len</i>	50	104	52	46	82	46	65
Camadas	-	3	2	-	2	1	-
Camadas ocultas	-	(44,206,234)	98	-	(246,84)	149	-
Bidirecional	-	-	Sim	-	-	Sim	-
Épocas	-	50	50	50	50	50	50
Otimizador	-	Adam	Adam	Adam	Adam	Adam	Adam
Função	-	-	-	-	cos	sen	sen
Tempo	18 s	30 s	26 s	55 s	35 s	32 s	1 min 5 s

Parâmetro	Transformer	Flowformer	Flashformer	T2V-Transformer	T2V-Flowformer	T2V-Flashformer
<i>seq len</i>	53	58	93	32	49	63
Camadas do codificador	3	2	2	1	2	2
Camadas do decodificador	1	2	3	1	3	1
Épocas	10	10	10	10	10	10
Otimizador	RMSprop	RMSprop	RMSprop	RMSprop	RMSprop	RMSprop
d_{model}	128	128	32	64	32	256
Nº de cabeças	2	8	2	4	4	6
d_{ff}	768	768	64	128	96	512
Função	-	-	-	sen	cos	sen
Tempo	2 min 20 s	2 min 5 s	1 min 30 s	2 min 34 s	2 min 17 s	1 min 50 s

Fonte: Autoria própria (2025).

De acordo com a Tabela 9, o modelo com o maior *seq len* foi o T2V-Flowformer, seguido pelo T2V-Flashformer. Isso significa que os modelos precisaram de mais dados

históricos para fazer previsões para o Cenário B. O número de camadas foi 1 para MLP e T2V-MLP. Para LSTM e T2V-LSTM, o número de camadas foi 1 e 2, respectivamente. Ambos os modelos são bidirecionais, como na Abordagem A. O número de codificadores foi maior que o decodificador para o Transformer e o T2V-Flashformer. Para os outros modelos, o número de camadas para o codificador e o decodificador foi o mesmo. O modelo com a maior complexidade foi o T2V-Flashformer, com d_{model} , número de cabeças e d_{ff} iguais a 256, 8, 1536, respectivamente. Em seguida, veio o Transformer, com d_{model} , heads e d_{ff} iguais a 256, 6 e 1536, respectivamente. O otimizador Adam foi aplicado à maioria dos modelos, enquanto o SGD foi aplicado ao DLinear, e o RMSprop foi aplicado ao Transformer e ao Flowformer. De acordo com as Tabelas 8 e 9, os valores dos parâmetros para os modelos variaram. Isso se explica pelo fato de serem duas abordagens diferentes, considerando duas condições temporais distintas.

Tabela 9 - Resultados finais da busca de hiperparâmetros para todos os modelos avaliados no Estudo de Caso 1 (Cenário B)

Parâmetro	ARIMA	MLP	LSTM	DLinear	T2V-MLP	T2V-LSTM	T2V-DLinear
<i>seq len</i>	48	92	77	76	60	74	104
Camadas	-	1	1	-	1	2	-
Camadas ocultas	-	246	144	-	197	148	-
Bidirecional	-	-	Sim	-	-	Sim	-
Épocas	-	50	50	50	50	50	50
Otimizador	-	Adam	Adam	Adam	Adam	Adam	Adam
Função	-	-	-	-	cos	sen	cos
Tempo	20 s	30 s	26 s	55 s	35 s	32 s	1 min 5 s

Parâmetro	Transformer	Flowformer	Flashformer	T2V-Transformer	T2V-Flowformer	T2V-Flashformer
<i>seq len</i>	76	55	53	44	120	105
Camadas do codificador	3	2	1	1	2	3
Camadas do decodificador	2	2	1	1	2	1
Épocas	10	10	10	10	10	10
Otimizador	RMSprop	RMSprop	Adam	Adam	Adam	Adam
d_{model}	256	64	64	64	128	256
Nº de cabeças	6	8	8	6	2	8
d_{ff}	1536	192	192	384	512	1536
Função	-	-	-	cos	sen	sen
Tempo	2 min 20 s	2 min 5 s	1 min 30 s	2 min 34 s	2 min 17 s	1 min 50 s

Fonte: Autoria própria (2025).

4.6.2 Estudo de Caso 2

Os resultados na busca de hiperparâmetros para o Estudo de Caso 2 estão representados pela Tabela 10, que mostra os parâmetros finais de cada modelo, referenciando as melhores configurações encontradas para este caso. Os modelos LSTM, MLP e DLinear foram treinados com 50 épocas. Enquanto os modelos propostos e os X-formers foram treinados com 10 épocas. Isso demonstra que os modelos propostos e os X-formers convergem mais rápido do que os modelos de referência, reforçando aquilo que foi discutido no Estudo de Caso 1. Consequentemente, eles alcançam desempenho ideal em menos épocas, reduzindo o risco de *overfitting*.

De acordo com a Tabela 10, o modelo com maior valor do *seq len* foi o T2V-Transformer, com 102 unidades, enquanto o menor valor foi observado no LSTM, com 18. Esse resultado indica que modelos com *seq len* mais elevado tendem a apresentar maior capacidade de capturar padrões temporais de longo prazo. Outros modelos com *seq len* relativamente altos foram o T2V-Flowformer e o T2V-Flashformer, com valores de 90 e 72, respectivamente. Em relação ao número de camadas, o modelo MLP apresentou três, ao passo que os modelos LSTM, T2V-MLP e T2V-LSTM contaram com apenas uma. Ademais, tanto o LSTM quanto o T2V-LSTM não são bidirecionais.

Nos X-formers (Transformer, Flowformer e Flashformer), o número de camadas variou entre os módulos codificador e decodificador. O Transformer apresentou 3 camadas em ambos os módulos; o Flowformer, 2 no codificador e 1 no decodificador; e o Flashformer, 2 e 3, respectivamente. Nos modelos propostos, também foram observadas variações: o T2V-Transformer apresentou 2 camadas no codificador e 1 no decodificador; o T2V-Flowformer, 2 e 3; e o T2V-Flashformer, 1 e 2, respectivamente. O otimizador RMSprop foi aplicado aos modelos Transformer e T2V-Transformer, enquanto o DLinear utilizou o SGD. Os demais modelos utilizaram o Adam.

Considerando o parâmetro d_{model} , os modelos Transformer e Flowformer apresentaram o maior valor, igual a 512, enquanto o menor foi observado no T2V-Flowformer, com 32. Para d_{ff} , o Transformer obteve o maior valor (2048), seguido pelo T2V-Transformer (1536), sendo novamente o T2V-Flowformer o modelo com o menor valor. Quanto ao número de cabeças de atenção, o T2V-Transformer apresentou o maior valor (8), e o T2V-Flowformer, o menor (2). Dessa forma, conclui-se que os modelos Transformer e T2V-Transformer demandam maior custo computacional, ao passo que o

T2V-Flowformer se destaca por apresentar a configuração mais eficiente nesse aspecto.

Para a função Time2Vec, os modelos T2V-LSTM, T2V-DLinear e T2V-Transformer apresentaram a função seno, enquanto os modelos T2V-MLP, T2V-Flowformer e T2V-Flashformer apresentaram a função cosseno.

Tabela 10 - Resultados finais da busca de hiperparâmetros para todos os modelos avaliados no Estudo de Caso 2

Parâmetro	ARIMA	MLP	LSTM	DLinear	T2V-MLP	T2V-LSTM	T2V-DLinear
<i>seq len</i>	50	58	18	23	39	28	37
Camadas	-	3	1	-	1	1	-
Camadas ocultas	-	(194,187,130)	106	-	174	88	-
Bidirecional	-	-	Não	-	-	Não	-
Épocas	-	50	50	50	50	50	50
Otimizador	-	Adam	Adam	SGD	Adam	Adam	Adam
Função	-	-	-	-	cos	sen	sen
Tempo	17 s	20 s	23 s	1 min 10 s	36 s	28 s	1 min 24 s
Parâmetro	Transformer	Flowformer	Flashformer	T2V-Transformer	T2V-Flowformer	T2V-Flashformer	
<i>seq len</i>	42	54	54	102	90	102	
Camadas do codificador	3	2	2	2	2	2	
Camadas do decodificador	3	1	3	1	3	2	
Épocas	10	10	10	10	10	10	
Otimizador	RMSprop	Adam	Adam	RMSprop	Adam	Adam	
d_{model}	512	512	64	256	32	64	
Nº de cabeças	6	7	4	8	2	3	
d_{ff}	2048	1024	128	1536	96	192	
Função	-	-	-	sen	cos	sen	
Tempo	1 min 45 s	1 min 30 s	56 s	1 min 50 s	1 min 36 s	1 min 22 s	

Fonte: Autoria própria (2025).

A Tabela 10 apresenta o tempo de processamento de cada modelo para a execução de um único ensaio. Esses resultados indicam que todos os modelos são aplicáveis e adequados para previsões com horizonte de 12 horas à frente. Observa-se que o modelo com maior tempo de processamento foi o T2V-Transformer, seguido pelo Transformer, com aproximadamente 1 minuto e 45 segundos e 1 minuto e 10 segundos, respectivamente. Por outro lado, os modelos Flowformer, Flashformer, T2V-Flowformer e T2V-Flashformer apresentaram tempos inferiores, resultado da adoção dos mecanismos de atenção FlowAttention e FlashAttention, que reduzem a complexidade quadrática

característica do FullAttention do Transformer convencional. Entre os modelos de referência, o menor tempo de execução foi obtido pelo ARIMA, MLP e LSTM, com cerca de 17, 20 e 23 segundos, respectivamente. Esses valores estão em consonância com os resultados do Estudo de Caso 1, no qual os modelos baseados em Transformers e suas variantes propostas também apresentaram maior tempo de treinamento.

5 RESULTADOS E DISCUSSÃO

Neste Capítulo são apresentados os resultados e a discussão dos experimentos computacionais de previsão de potência eólica realizados com os modelos propostos e os modelos de *benchmark*, utilizando dados reais de parques eólicos localizados no Nordeste do Brasil. Os experimentos foram conduzidos em dois casos distintos, denominados Estudo de Caso 1 e Estudo de Caso 2, os quais diferem quanto às condições meteorológicas, configuração dos conjuntos de dados e às estratégias de avaliação adotadas. Inicialmente, são analisados os indicadores de desempenho preditivo dos modelos, seguidos por uma discussão sobre os aspectos computacionais e sobre a capacidade de generalização sob diferentes condições sazonais. Por fim, são destacadas as principais implicações dos resultados obtidos e as limitações observadas.

As métricas utilizadas para avaliar o desempenho dos modelos neste artigo foram erro absoluto médio (*Mean Absolute Error* – MAE), raiz do erro quadrático médio (*Root Mean Squared Error* – RMSE), melhoria em relação ao MAE de referência (*Improvement over Reference Mean Absolute Error* – IoR-MAE) e melhoria em relação ao RMSE de referência (*Improvement over Reference Root Mean Squared Error* – IoR-RMSE). O modelo de referência utilizado no cálculo do IoR-MAE e do IoR-RMSE é o modelo de Persistência. As Equações (15) – (18) apresentam a formulação matemática correspondente a cada uma delas:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

$$IoR - MAE = \left(1 - \frac{MAE_{modelo}}{MAE_{referência}} \right) \times 100\% \quad (17)$$

$$IoR - RMSE = \left(1 - \frac{RMSE_{modelo}}{RMSE_{referência}} \right) \times 100\% \quad (18)$$

Onde n representa o número de observações, y_i são os valores observados e \hat{y}_i são os valores previstos por cada modelo. O MAE e o RMSE de referência correspondem às métricas obtidas do modelo de Persistência. Portanto, valores mais altos de IoR-MAE e IoR-RMSE, juntamente com MAE e RMSE mais baixos, indicam melhor desempenho

do modelo.

5.1 Estudo de Caso 1

De acordo com a Tabela 11, o modelo T2V-DLinear obteve o melhor desempenho para as métricas MAE e IoR-MAE no horizonte de 6 horas, com valores de 213,386 kW e 12,35%, respectivamente, seguido pelo T2V-Transformer, que registrou um MAE de 214,096 kW e um IoR-MAE de 12,06%. No horizonte de 10 horas, o T2V-Transformer superou todos os modelos com um IoR-MAE de 14,56%. Posteriormente, o Flashformer e o T2V-Flashformer alcançaram valores de IoR-MAE de 13,88% e 13,83%, respectivamente. Para o horizonte de 12 horas, os melhores resultados foram alcançados pelo T2V-Transformer e pelo T2V-Flashformer, com valores de IoR-MAE de 13,55% e 13,30%, respectivamente. Os modelos ARIMA, MLP, LSTM e DLinear apresentaram desempenho inferior, com valores de IoR-MAE de 6,20%, 9,22%, 10,43% e 10,52%, respectivamente. No geral, o T2V-Transformer demonstrou a maior consistência e o melhor desempenho em todos os horizontes para as métricas MAE e IoR-MAE, seguido pelo T2V-Flashformer. Em relação às métricas RMSE e IoR-RMSE, os modelos T2V-Transformer e T2V-Flashformer demonstraram desempenho superior.

Como observado na Tabela 11, no horizonte de 6 horas, o T2V-Transformer foi o único modelo a superar um IoR-RMSE de 16%, atingindo 16,03%. O T2V-Flashformer, por sua vez, atingiu um IoR-RMSE de aproximadamente 14,98%. No horizonte de 10 horas, os valores de IoR-RMSE foram de 17,85% para o T2V-Transformer e 16,58% para o T2V-Flashformer, enquanto para o horizonte de 12 horas, os valores de IoR-RMSE foram de 17,73% para o T2V-Transformer e 16,67% para o T2V-Flashformer. Os modelos ARIMA, MLP, LSTM e DLinear demonstraram desempenho inferior, atingindo 13,36%, 15,08%, 13,79% e 14,64%, respectivamente, para a mesma métrica e horizonte de previsão. Em geral, o T2V-Transformer apresentou o melhor desempenho para todos os horizontes, seguido pelo T2V-Flashformer para as métricas RMSE e IoR-RMSE. Analisando os horizontes e as métricas de avaliação apresentados na Tabela 11, os modelos T2V-Transformer e T2V-Flashformer demonstraram consistência e confiabilidade, tornando-os as escolhas mais adequadas para previsão de potência sob esta abordagem. Embora o T2V-DLinear tenha alcançado o melhor desempenho para MAE e IoR-MAE no horizonte de 6 horas, seu desempenho foi inconsistente em outros

horizontes e menos competitivo para as métricas RMSE e IoR-RMSE. Consequentemente, o T2V-DLinear não é tão adequado para o Cenário A, em comparação com os modelos T2V-Flashformer e T2V-Transformer.

Tabela 11 - Comparação do desempenho do modelo para diferentes horizontes de previsão (Cenário A)

Modelo	MAE (kW) / IoR – MAE (%)					
	6 h	10 h	12 h			
Persistência	243,463	0	273,198	0	279,828	0
ARIMA	231,939	4,73	256,666	6,05	262,460	6,20
MLP	222,349	8,67	245,996	9,95	254,032	9,22
T2V-MLP	222,947	8,42	241,690	11,53	249,697	10,76
LSTM	225,592	7,34	248,002	9,22	250,645	10,43
T2V-LSTM	218,730	10,16	249,981	8,49	252,356	9,82
DLinear	218,137	10,40	239,555	12,31	250,375	10,52
T2V-DLinear	213,386	12,35	257,894	5,60	246,195	12,02
Transformer	224,325	7,86	244,365	10,55	254,115	9,18
T2V-Transformer	214,096	12,06	233,430	14,56	241,911	13,55
Flowformer	223,979	8,00	244,266	10,59	249,242	10,93
T2V-Flowformer	219,680	9,77	238,644	12,65	247,056	11,71
Flashformer	215,573	11,45	235,273	13,88	246,213	12,01
T2V-Flashformer	214,368	11,95	235,402	13,83	242,598	13,30

Modelo	RMSE (kW) / IoR – RMSE (%)					
	6 h	10 h	12 h			
Persistência	342,520	0	378,050	0	384,270	0
ARIMA	300,042	12,40	326,909	13,52	332,916	13,36
MLP	294,207	14,10	318,141	15,84	326,291	15,08
T2V-MLP	297,990	12,99	324,162	14,25	330,686	13,94
LSTM	299,486	12,56	325,241	13,96	331,262	13,79
T2V-LSTM	293,626	14,27	320,027	15,35	326,267	15,09
DLinear	297,971	13,00	321,293	15,01	328,022	14,64
T2V-DLinear	298,052	12,98	327,494	13,37	324,737	15,49
Transformer	297,550	13,13	322,429	14,71	330,681	13,94
T2V-Transformer	287,613	16,03	310,545	17,85	316,146	17,73
Flowformer	298,385	12,88	322,098	14,80	326,549	15,02
T2V-Flowformer	296,478	13,44	320,958	15,10	327,351	14,81
Flashformer	295,377	13,76	321,323	15,00	329,660	14,21
T2V-Flashformer	291,201	14,98	315,353	16,58	320,213	16,67

Fonte: Autoria própria (2025).

Observa-se que o modelo de Persistência apresenta valores nulos de IoR, uma vez que este modelo é adotado como referência no cálculo do índice de melhoria (IoR). Assim, o IoR quantifica a melhoria relativa de cada modelo em relação à Persistência, resultando, por definição, em valor zero para o próprio modelo de referência.

De acordo com a Tabela 12, os modelos T2V-Flashformer e T2V-Flowformer alcançaram consistentemente o melhor desempenho em todos os horizontes de previsão e métricas de avaliação. Para as métricas MAE e IoR-MAE, o T2V-Flashformer produziu os melhores resultados, com valores de IoR-MAE de 18,23% para o horizonte de 6 horas e 23,89% para o horizonte de 10 horas, enquanto o T2V-Flowformer seguiu de perto com 17,98% e 23,74% para os mesmos horizontes, respectivamente. No horizonte de 12 horas, o T2V-Flowformer e o T2V-Flashformer registraram valores de IoR-MAE de 24,47% e 24,37%, respectivamente. Notavelmente, esses dois modelos foram os únicos a exceder 23% de IoR-MAE no horizonte de 10 horas e 24% no horizonte de 12 horas. No horizonte de previsão de 12 horas, os modelos ARIMA, MLP, LSTM e DLinear alcançaram valores de IoR-MAE de 21,20%, 20,60%, 19,09% e 20,75%, respectivamente. Para as métricas RMSE e IoR-RMSE, o T2V-Flashformer obteve o melhor desempenho no horizonte de 6 horas, com um IoR-RMSE de 22,88%, seguido pelo T2V-Flowformer com 22,49%. Para os horizontes de 10 e 12 horas, o T2V-Flowformer superou todos os outros modelos, registrando valores de IoR-RMSE de 27,64% e 27,84%, respectivamente, enquanto o T2V-Flashformer obteve 27,34% e 27,45% para os mesmos horizontes. Os modelos ARIMA, MLP, LSTM e DLinear apresentaram pior desempenho, com valores de 24,24%, 23,67%, 23,73% e 23,35% no horizonte de 10 horas, e 23,75%, 23,23%, 23,15% e 23,06% no horizonte de 12 horas, respectivamente. Conforme mostrado na Tabela 12, tanto o T2V-Flowformer quanto o T2V-Flashformer provaram ser os modelos mais adequados para o Cenário B.

De acordo com a Figura 22, no Cenário B, o período de teste apresenta valores de potência eólica mais elevados, com picos mais frequentes e intensos. Isso indica maior variabilidade e magnitude nos dados a serem previstos, aumentando a complexidade da tarefa de previsão. Consequentemente, os modelos apresentam valores maiores de erro absoluto (MAE) e erro quadrático médio (RMSE) neste cenário em comparação ao Cenário A. No entanto, conforme demonstrado nas Tabelas 11 e 12, os valores de IoR-MAE e IoR-RMSE para os modelos avaliados neste estudo são consistentemente maiores no Cenário B. Isso sugere que, apesar do aumento dos erros absolutos devido às condições de teste mais desafiadoras, os modelos propostos superam o modelo de Persistência por

uma margem maior. Portanto, as métricas de IoR mais altas no Cenário B destacam a robustez e a eficácia dos modelos em condições de previsão mais exigentes.

Tabela 12 - Comparação do desempenho do modelo para diferentes horizontes de previsão (Cenário B)

Modelo	MAE (kW) / IoR – MAE (%)					
	6 h	10 h	12 h			
Persistência	472,842	0	534,238	0	539,987	0
ARIMA	391,420	17,22	420,801	21,23	425,504	21,20
MLP	401,633	15,05	422,769	20,86	428,720	20,60
T2V-MLP	399,556	15,06	421,157	21,16	425,582	21,18
LSTM	404,243	14,50	430,225	19,46	436,876	19,09
T2V-LSTM	406,040	14,12	430,061	19,50	427,028	20,91
DLinear	400,527	15,29	423,949	20,64	427,896	20,75
T2V-DLinear	404,430	14,46	427,453	20,54	427,903	20,75
Transformer	402,317	14,91	427,499	19,90	433,263	19,76
T2V-Transformer	397,342	15,97	425,233	20,40	431,457	20,09
Flowformer	405,141	14,32	425,564	20,34	429,519	20,46
T2V-Flowformer	387,832	17,98	407,396	23,74	407,848	24,47
Flashformer	392,619	16,96	422,642	20,89	427,476	20,83
T2V-Flashformer	386,651	18,23	406,608	23,89	408,364	24,37

Modelo	RMSE (kW) / IoR – RMSE (%)					
	6 h	10 h	12 h			
Persistência	610,223	0	674,159	0	676,469	0
ARIMA	483,776	20,72	510,732	24,24	515,762	23,75
MLP	492,236	19,33	514,570	23,67	519,293	23,25
T2V-MLP	496,024	18,71	513,812	23,78	514,389	23,95
LSTM	488,402	19,96	514,152	23,73	519,833	23,15
T2V-LSTM	497,665	18,44	521,738	22,60	516,666	23,62
DLinear	496,257	18,67	516,697	23,35	520,411	23,06
T2V-DLinear	496,896	18,57	514,842	23,63	517,046	23,56
Transformer	492,053	19,36	515,418	23,54	519,182	23,25
T2V-Transformer	484,898	20,54	512,338	24,00	519,001	23,28
Flowformer	492,491	19,29	512,717	23,95	511,864	24,33
T2V-Flowformer	472,943	22,49	487,814	27,64	488,135	27,84
Flashformer	481,754	21,05	511,446	24,13	514,838	23,89
T2V-Flashformer	470,586	22,88	489,812	27,34	490,786	27,45

Fonte: Autoria própria (2025).

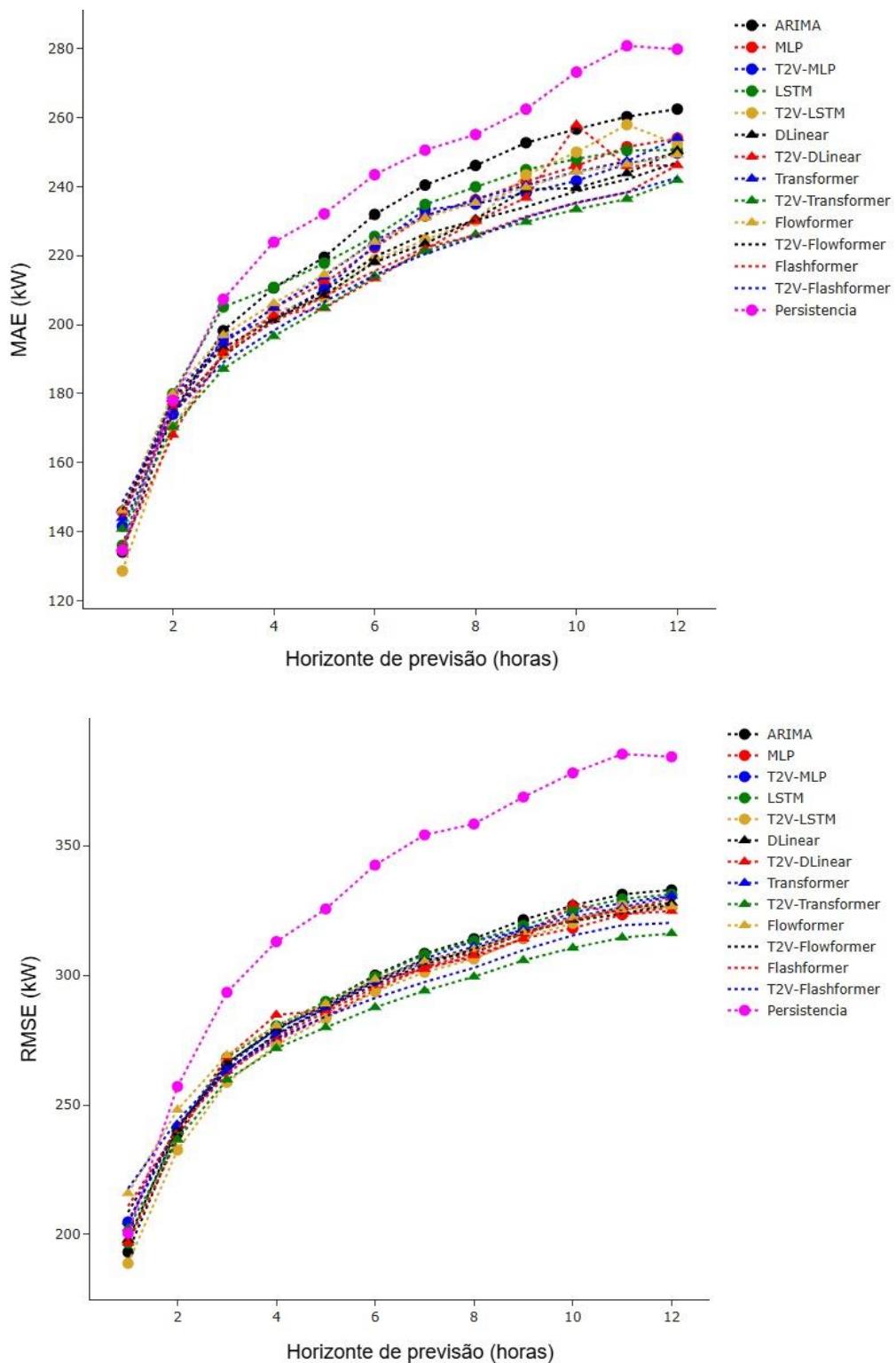
As Figuras 24 e 25 ilustram o desempenho dos modelos em diferentes horizontes

de previsão, oferecendo uma visualização abrangente das métricas MAE e RMSE. De acordo com a Figura 24, para o Cenário A, é evidente que os modelos propostos (especificamente T2V-Transformer e T2V-Flashformer) superaram os modelos de referência, particularmente nos horizontes posteriores. Essa tendência é ainda mais pronunciada no RMSE, onde o T2V-Transformer e o T2V-Flashformer demonstraram consistentemente desempenho superior em quase todos os horizontes de previsão, com a diferença de desempenho se tornando cada vez mais significativa além do horizonte de 4 horas.

De acordo com a Figura 25, para o Cenário B, os modelos propostos — especificamente T2V-Flowformer e T2V-Flashformer — começam a superar significativamente os modelos de *benchmark* após o horizonte de 4 horas. Isso reflete uma melhoria progressiva no desempenho da previsão à medida que o horizonte de previsão aumenta. Embora no curíssimo prazo (por exemplo, horizonte 1) seu desempenho possa inicialmente ficar atrás dos modelos benchmark, como o modelo de persistência, esse comportamento é provavelmente atribuível à sua dependência da codificação temporal via Time2Vec e mecanismos de atenção complexos, que são mais eficazes na captura de dependências temporais latentes em horizontes ligeiramente mais longos. Do horizonte três em diante, no entanto, ambos os modelos exibem uma redução acentuada no erro de previsão e superam consistentemente todos os modelos de base até o horizonte de 12 horas. Essas descobertas sugerem que as arquiteturas propostas são particularmente adequadas para tarefas de previsão de curto prazo envolvendo horizontes de várias horas, como os analisados neste estudo (ou seja, 6, 10 e 12 horas).

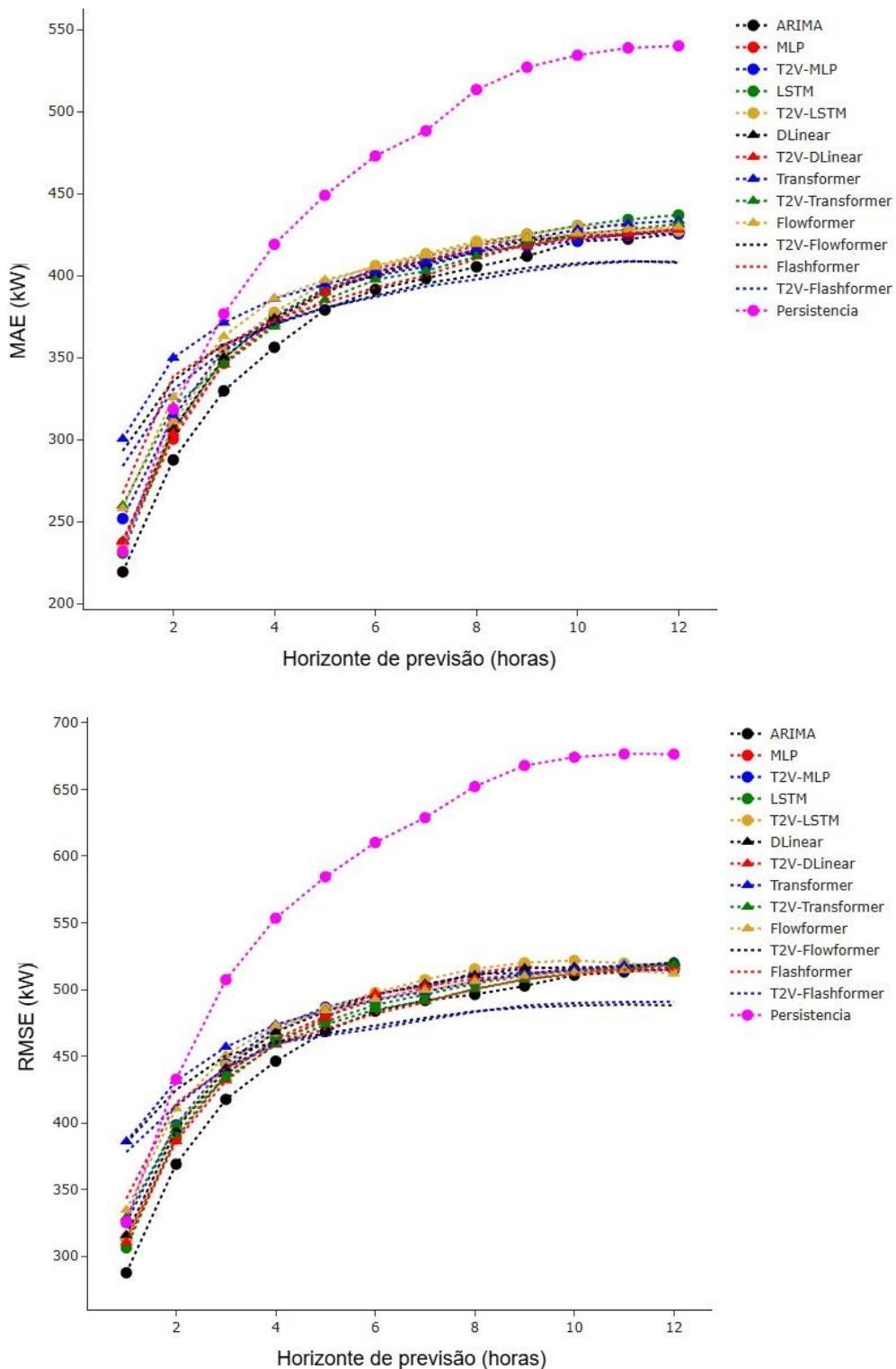
Analizando todos os modelos sem a adição de Time2Vec, observou-se que, para o Cenário A, o Flashformer demonstrou o melhor desempenho em termos da métrica MAE em todos os horizontes. Para a métrica RMSE, o modelo com melhor desempenho nos horizontes de 6, 10 e 12 horas foi o MLP. Flashformer, DLinear e Flowformer foram outros modelos com bons desempenhos nos horizontes de 6, 10 e 12 horas. No Cenário B, o Flashformer superou os demais modelos para a métrica MAE nos horizontes de 6, 10 e 12 horas. Para a mesma métrica, DLinear foi o segundo melhor desempenho nos horizontes de 6 a 12 horas, enquanto o MLP ficou em segundo lugar em 10 horas. Em relação à métrica RMSE, o Flashformer obteve o melhor desempenho nos horizontes de 6 e 10 horas, enquanto o Flowformer teve o melhor desempenho em 12 horas.

Figura 24 - Visualização das métricas de avaliação para diferentes horizontes de previsão da potência eólica, para cada modelo avaliado neste estudo. MAE na parte superior, RMSE na parte inferior (Cenário A)



Fonte: Autoria própria (2025).

Figura 25 - Visualização das métricas de avaliação para diferentes horizontes de previsão da potência eólica, para cada modelo avaliado neste estudo. MAE na parte superior, RMSE na parte inferior (Cenário B)



Fonte: Autoria própria (2025).

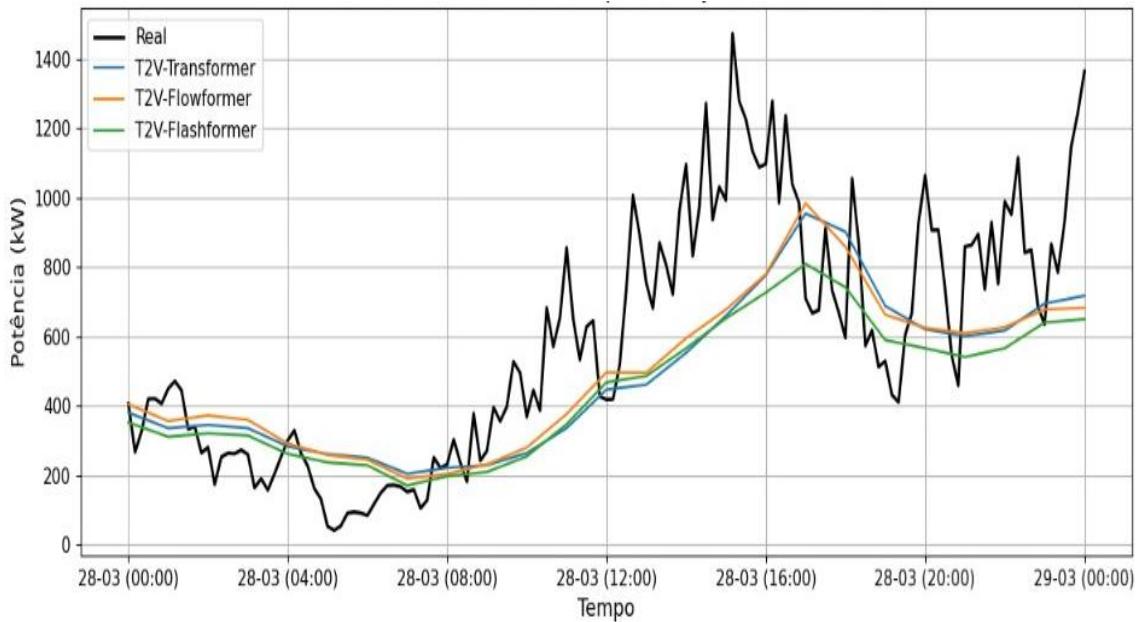
5.1.1 Discussão Física dos Erros de Previsão

Além da análise quantitativa baseada nas métricas MAE, RMSE, IoR-MAE e IoR-RMSE, é fundamental interpretar os erros à luz da física do processo de geração de energia eólica. A potência gerada por turbinas eólicas é fortemente influenciada por variações rápidas da velocidade do vento, especialmente durante eventos de rampa de subida e descida de potência, os quais representam um dos maiores desafios para modelos de previsão de curto prazo. As Figuras 26 e 27 apresentam os três modelos propostos neste trabalho em comparação com a potência eólica observada no Estudo de Caso 1, considerando, respectivamente, os Cenários A e B. Ambas as figuras apresentam uma janela representativa de 24 horas.

De acordo com a Figura 26, observa-se que os maiores desvios entre as séries ocorrem predominantemente durante eventos de rampa de descidas e subidas bruscas de potência, especialmente no período entre aproximadamente 12:00 às 16:00 horas. Nessas situações, a potência real apresenta variações abruptas e de alta frequência, associadas a mudanças rápidas nas condições do vento e ao aumento da turbulência atmosférica, fenômenos intrinsecamente difíceis de antecipar em horizontes de curto prazo. Os modelos tendem a subestimar a potência, evidenciando uma limitação na captura da inclinação negativa acentuada das rampas. Por outro lado, os modelos foram capazes de capturar o comportamento de tendência da série temporal ao longo da janela de tempo analisada, ainda que com pequenas defasagens temporais. Assim, os resultados evidenciam que os erros de previsão estão mais fortemente associados aos eventos de descidas e subidas abruptas de potência, os quais representam o principal fator limitante para a redução dos erros em aplicações operacionais de curto prazo.

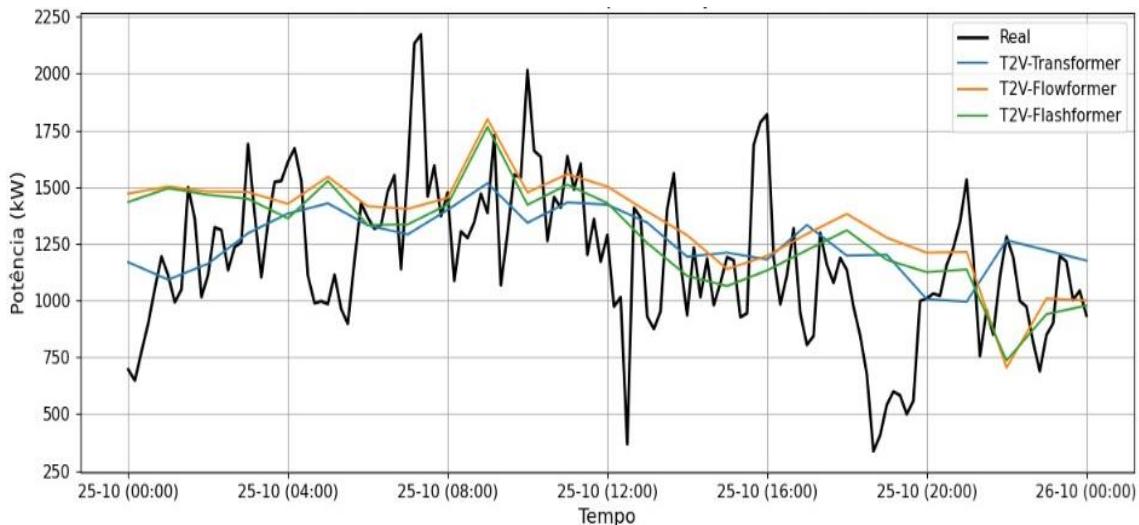
De acordo com a Figura 27, observa-se que os maiores desvios entre as séries ocorrem predominantemente durante os eventos de descidas e subidas abruptas de potência, comportamento análogo ao observado no Cenário A. Embora os modelos consigam capturar adequadamente a tendência geral da série temporal, verifica-se a ocorrência de maiores erros nos instantes associados a variações rápidas da geração eólica, especialmente durante quedas acentuadas de produção. Esse comportamento está relacionado a mudanças abruptas nas condições do vento, como o aumento da turbulência e a perda de persistência temporal, fenômenos intrinsecamente difíceis de antecipar em horizontes de curto prazo.

Figura 26 - Potência eólica observada e prevista em uma janela de 24 horas do Estudo de Caso 1 (Cenário A), entre 3 e 4 de março de 2020



Fonte: Autoria própria (2025).

Figura 27 - Potência eólica observada e prevista em uma janela de 24 horas do Estudo de Caso 1 (Cenário B), entre 25 e 26 de outubro de 2020



Fonte: Autoria própria (2025).

5.1.2 Impacto da integração do Time2Vec no desempenho dos modelos (Estudo de Caso 1)

Em geral, a adição do Time2Vec melhorou o desempenho dos modelos, conforme ilustrado na Tabela 13 para os Cenários A e B. Os valores são expressos em porcentagens, com números negativos indicando que a adição de Time2Vec não melhorou os modelos. Para o Cenário A, melhorias foram observadas em quase todos os horizontes e métricas dos X-formers. Notavelmente, o Transformador T2V demonstrou ganhos significativos nos horizontes de 6, 10 e 12 horas, alcançando uma melhoria de 4,56%, 4,47% e 4,80% em relação ao Transformador em MAE e 3,34%, 3,69% e 4,40% em RMSE, respectivamente. O T2V-Flowformer superou o Flowformer, com uma melhoria de 1,92%, 2,30% e 0,88% no MAE nos horizontes de 6, 10 e 12 horas. O T2V-Flashformer apresentou melhorias consistentes em todos os horizontes e métricas, com uma melhoria de 2,87% no RMSE no horizonte de 12 horas em comparação com o Flashformer. Em comparação com os modelos de referência, os ganhos foram menos pronunciados. No entanto, algumas melhorias foram observadas em horizontes e métricas específicos. Por exemplo, o T2V-MLP superou o MLP nos horizontes de 10 e 12 horas na métrica MAE, com melhorias de 1,75% e 1,70%, respectivamente. Por outro lado, para a métrica RMSE, o MLP superou consistentemente o T2V-MLP em todos os horizontes. O T2V-LSTM apresentou desempenho superior ao LSTM em todos os horizontes da métrica RMSE. Da mesma forma, o T2V-DLinear superou o DLinear no horizonte de 12 horas, tanto na métrica MAE quanto na RMSE.

De acordo com a Tabela 13, para o Cenário B, a adição do Time2Vec melhorou todos os formadores de X. O T2V-Flashformer apresentou melhorias consistentes em relação ao Flashformer em todas as métricas e horizontes, particularmente na métrica MAE, com ganhos de 3,79% e 4,47% nos horizontes de 10 e 12 horas, respectivamente. Para a métrica RMSE, as melhorias foram de 2,32%, 4,23% e 4,67% nos horizontes de 6, 10 e 12 horas, respectivamente. Da mesma forma, o T2V-Flowformer superou consistentemente o Flowformer de referência em todas as métricas de avaliação e horizontes de previsão. Nos horizontes de 6, 10 e 12 horas, obteve melhorias notáveis, com reduções no MAE de 4,27%, 4,27% e 5,05%, e no RMSE de 3,97%, 4,86% e 4,96%, respectivamente. Para os modelos de referência, as melhorias com a inclusão do Time2Vec foram menos pronunciadas, mas ainda presentes. O T2V-MLP superou

consistentemente o MLP padrão em termos de MAE em todos os horizontes de previsão, e também em RMSE, exceto no horizonte de 6 horas. O T2V-LSTM apresentou uma melhora notável em relação ao LSTM no horizonte de 12 horas para a métrica MAE (2,25%). Em contraste, a integração do Time2Vec ao modelo DLinear não resultou em ganhos significativos.

A incorporação do Time2Vec resultou em ganhos significativos de desempenho, especialmente para arquiteturas X-former, evidenciando uma sinergia estrutural entre *embeddings* temporais e modelos baseados em atenção. Arquiteturas X-former dependem inherentemente de codificações temporais ou posicionais para modelar dependências sequenciais por meio de mecanismos de atenção. O Time2Vec aprimora esse processo ao fornecer uma representação contínua, diferenciável e periódica do tempo, o que aumenta a capacidade do modelo de capturar dinâmicas temporais além das codificações posicionais fixas. Essa incorporação temporal enriquecida permite que o mecanismo de atenção alinhe as relações temporais de forma mais eficaz, aprimorando assim a modelagem de dependência de curto e longo prazo. Em contraste, modelos como MLP, LSTM e DLinear aprendem padrões temporais principalmente por meio de recorrência interna, convolução ou projeção linear, sem depender de incorporações temporais explícitas. Consequentemente, eles são menos capazes de explorar a riqueza representacional introduzida pelo Time2Vec, o que provavelmente explica as menores melhorias de desempenho observadas em comparação com os modelos X-formers.

Os resultados apresentados nesta seção demonstram que modelos baseados em Transformer — particularmente aqueles aprimorados com Time2Vec — são altamente eficazes para a previsão de energia eólica, superando consistentemente modelos estabelecidos na literatura em múltiplos horizontes de previsão. No Cenário A, os modelos T2V-Transformer e T2V-Flashformer superaram todos os modelos de referência (MLP, LSTM, DLinear, T2V-MLP, T2V-LSTM e T2V-DLinear) em praticamente todas as métricas e horizontes avaliados. No Cenário B, os modelos T2V-Flowformer e T2V-Flashformer também superaram os modelos de referência, confirmado sua robustez e precisão preditiva, conforme discutido ao longo desta tese.

Tabela 13 - Melhorias com a adição do Time2Vec

Cenário A						
Modelo	MAE			RMSE		
	6h	10h	12h	6h	10h	12h
MLP	222,349	245,996	254,032	294,207	318,141	326,291
T2V-MLP	222,947	241,690	249,697	297,990	324,162	330,686
Melhoria (%)	-0,27	1,75	1,70	-1,29	-1,89	-1,35
LSTM	225,592	248,002	250,645	299,486	325,241	331,262
T2V-LSTM	218,730	249,981	252,356	293,626	320,027	326,267
Melhoria (%)	3,04	-0,80	-0,68	1,96	1,60	1,51
DLinear	218,137	239,555	250,375	297,971	321,293	328,022
T2V-DLinear	213,386	257,894	246,195	298,052	327,494	324,737
Melhoria (%)	2,18	-7,66	1,67	-0,03	-1,93	1,00
Transformer	224,325	244,365	254,115	297,550	322,429	330,681
T2V-Transformer	214,096	233,430	241,911	287,613	310,545	316,146
Melhoria (%)	4,56	4,47	4,80	3,34	3,69	4,40
Flowformer	223,979	244,266	249,242	298,385	322,098	326,549
T2V-Flowformer	219,680	238,644	247,056	296,478	320,958	327,351
Melhoria (%)	1,92	2,30	0,88	0,64	0,35	-0,25
Flashformer	215,573	235,273	246,213	295,377	321,323	329,660
T2V-Flashformer	214,368	235,402	242,598	291,201	315,353	320,213
Melhoria (%)	0,56	-0,05	1,47	1,41	1,86	2,87
Cenário B						
Modelo	MAE			RMSE		
	6h	10h	12h	6h	10h	12h
MLP	401,633	422,769	428,720	492,236	514,570	519,293
T2V-MLP	399,556	421,157	425,582	496,024	513,812	514,389
Melhoria (%)	0,52	0,38	0,73	-0,77	0,15	0,94
LSTM	404,243	430,225	436,876	488,402	514,152	519,833
T2V-LSTM	406,040	430,061	427,028	497,665	521,738	516,666
Melhoria (%)	-0,44	0,04	2,25	-1,90	-1,48	0,61
DLinear	400,527	423,949	427,896	496,257	516,697	520,411
T2V-DLinear	404,430	427,453	427,903	496,896	514,842	517,046
Melhoria (%)	-0,97	-0,83	-0,01	-0,13	0,36	0,65
Transformer	402,317	427,499	433,263	492,053	515,418	519,182
T2V-Transformer	397,342	425,233	431,457	484,898	512,338	519,001
Melhoria (%)	1,24	0,53	0,42	1,45	0,60	0,03
Flowformer	405,141	425,564	429,519	492,491	512,717	511,864
T2V-Flowformer	387,832	407,396	407,848	472,943	487,814	488,135
Melhoria (%)	4,27	4,27	5,05	3,97	4,86	4,64
Flashformer	392,619	422,642	427,476	481,754	511,446	514,838
T2V-Flashformer	386,651	406,608	408,364	470,586	489,812	490,786
Melhoria (%)	1,52	3,79	4,47	2,32	4,23	4,67

Fonte: Autoria própria (2025).

5.2 Estudo de Caso 2

De acordo com a Tabela 14, é possível verificar o desempenho de cada modelo no Estudo de Caso 2. Para as métricas MAE e IoR-MAE, o modelo que apresentou o melhor desempenho no horizonte de 6 horas foi o Transformer, seguido pelos modelos T2V-Flowformer e T2V-Transformer, com valores de IoR-MAE de 57,55%, 56,42% e 56,06%, respectivamente. Nos horizontes de 10 e 12 horas, os melhores resultados foram obtidos pelos modelos T2V-Transformer, T2V-Flashformer e T2V-Flowformer, cujos IoR-MAE foram de aproximadamente 65,82%, 65,42% e 65,25% para 10 horas, e 66,80%, 66,62% e 66,39% para 12 horas.

Considerando as métricas RMSE e IoR-RMSE, o modelo Transformer também apresentou o melhor desempenho no horizonte de 6 horas, seguido pelos modelos T2V-Flashformer e T2V-Flowformer, com valores de IoR-RMSE de 54,25%, 53,86% e 53,42%, respectivamente. Para os horizontes de 10 e 12 horas, os modelos com melhor desempenho foram o T2V-Flashformer e o T2V-Flowformer, com IoR-RMSE de 60,97% e 60,79% para 10 horas, e 61,76% e 61,14% para 12 horas, respectivamente.

De acordo com a Tabela 14, observa-se que os modelos propostos superaram praticamente todos os modelos de benchmark, com exceção do Transformer no horizonte de 6 horas. Para os horizontes de 10 e 12 horas, os modelos propostos apresentaram os melhores resultados. Assim, de forma geral, demonstraram maior consistência em todos os horizontes analisados, sendo, portanto, mais indicados para a previsão de potência eólica no Estudo de Caso 2.

A Figura 28 apresenta uma visão detalhada do desempenho de todos os modelos ao longo dos diferentes horizontes, corroborando a conclusão de que os modelos propostos são os mais adequados, por apresentarem resultados superiores e consistentes no Estudo de Caso 2.

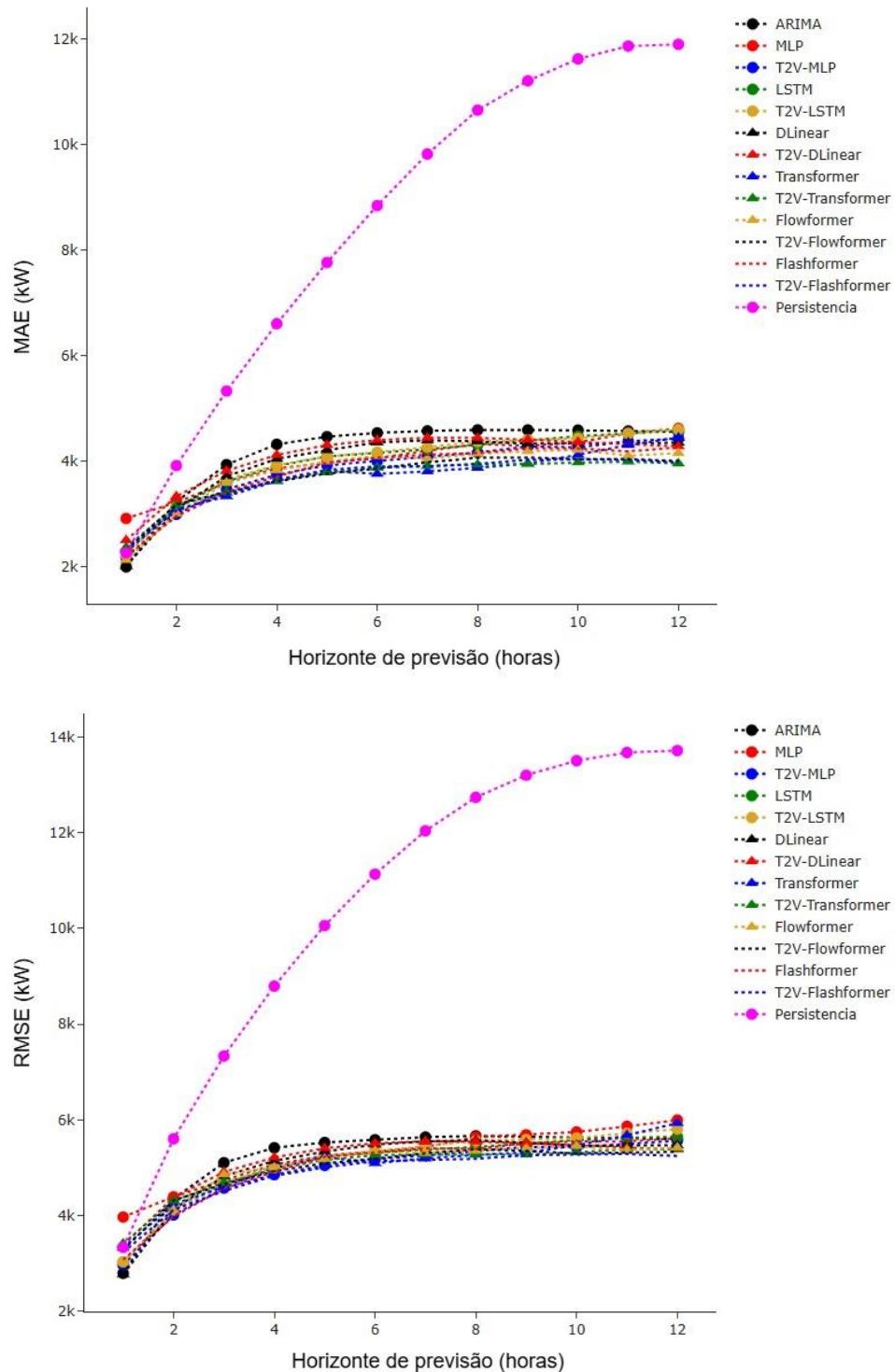
Tabela 14 - Comparação do desempenho dos modelos para diferentes horizontes de previsão
(Estudo de Caso 2)

Modelo	MAE / IoR – MAE (%)					
	6 h	10 h	12 h			
Persistência	8.840	0	11.620	0	11.897	0
ARIMA	4.530	48,75	4.582	60,56	4.546	61,78
MLP	4.051	54,17	4.349	62,56	4.614	61,21
T2V-MLP	3.992	54,84	4.255	63,37	4.422	62,82
LSTM	4.146	53,09	4.467	61,55	4.573	61,55
T2V-LSTM	4.162	52,91	4.428	61,89	4.592	61,39
DLinear	4.349	50,80	4.337	62,67	4.342	63,49
T2V-DLinear	4.394	50,28	4.397	62,49	4.274	64,07
Transformer	3.752	57,55	4.115	64,58	4.439	62,68
T2V-Transformer	3.870	56,06	3.971	65,82	3.948	66,80
Flowformer	4.090	53,73	4.207	63,79	4.146	65,14
T2V-Flowformer	3.852	56,42	4.037	65,25	3.998	66,39
Flashformer	4.061	54,06	4.235	63,55	4.255	64,23
T2V-Flashformer	3.896	55,92	4.017	65,42	3.971	66,62

Modelo	RMSE / IoR – RMSE (%)					
	6 h	10 h	12 h			
Persistência	11.135	0	13.509	0	13.720	0
ARIMA	5.583	49,85	5.628	58,33	5.596	59,21
MLP	5.334	52,09	5.746	57,46	5.995	56,30
T2V-MLP	5.178	53,50	5.427	59,82	5.565	59,43
LSTM	5.318	52,24	5.575	58,72	5.646	58,84
T2V-LSTM	5.343	52,01	5.621	58,39	5.806	57,68
DLinear	5.482	50,77	5.462	59,56	5.468	60,14
T2V-DLinear	5.509	50,52	5.462	59,56	5.381	60,77
Transformer	5.094	54,25	5.467	59,52	5.933	56,75
T2V-Transformer	5.261	52,75	5.321	60,61	5.396	60,67
Flowformer	5.348	51,96	5.452	59,63	5.422	60,48
T2V-Flowformer	5.186	53,42	5.296	60,79	5.331	61,14
Flashformer	5.318	52,24	5.549	58,92	5.616	59,06
T2V-Flashformer	5.136	53,86	5.273	60,97	5.245	61,76

Fonte: Autoria própria (2025).

Figura 28 - Visualização das métricas de avaliação para diferentes horizontes de previsão de potência eólica, para cada modelo avaliado neste estudo. MAE na parte superior, RMSE na parte inferior (Estudo de Caso 2)



Fonte: Autoria própria (2025).

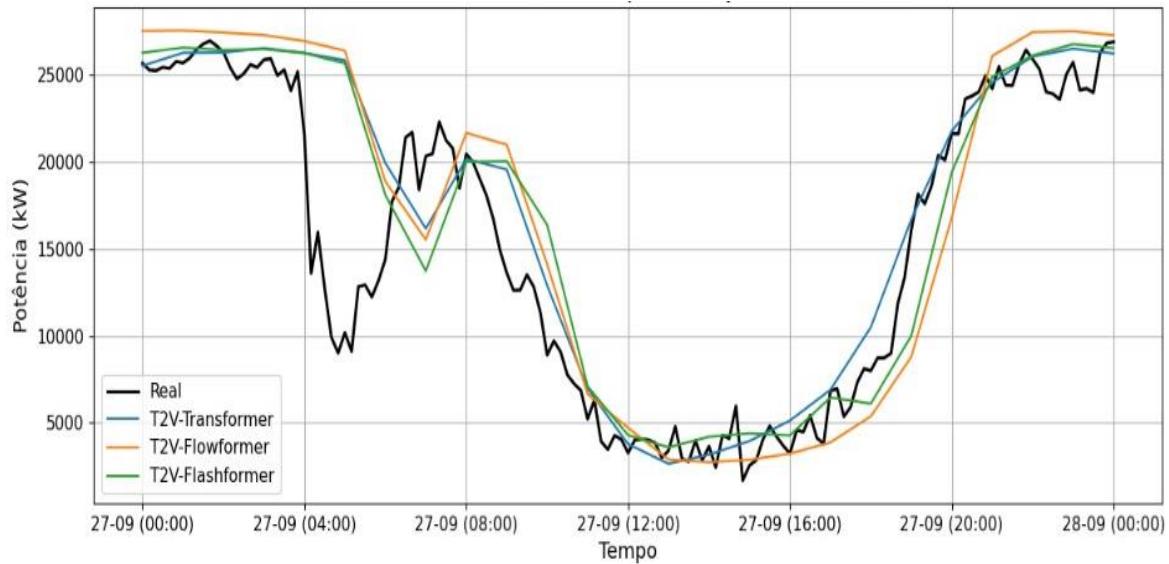
5.2.1 Discussão Física dos Erros de Previsão

Para a discussão física dos erros de previsão do Estudo de Caso 2, a Figura 29 apresenta os três modelos propostos neste trabalho em comparação com a potência eólica real. Por meio dela observa-se que os maiores desvios entre as previsões e a série observada ocorrem predominantemente durante os eventos de rampa de descida de potência. Nessas situações, a potência eólica real apresenta reduções abruptas associadas a mudanças rápidas nas condições do vento, como a passagem de frentes atmosféricas ou a intensificação da turbulência, fenômenos que são intrinsecamente difíceis de antecipar com precisão em horizontes de curto prazo. Os modelos propostos tendem a apresentar uma superestimação temporária da potência durante os períodos de queda acentuada, como observado após as 04:00 horas e no intervalo entre 08:00 e 12:00 horas, conforme ilustrado na Figura 29.

Por outro lado, durante as rampas de subida de potência, observa-se que os modelos conseguem capturar de forma mais consistente a tendência de crescimento da geração, ainda que com pequenas defasagens temporais, como observado após as 16:00 horas. Esse comportamento indica que a informação histórica disponível é suficiente para antecipar aumentos graduais da velocidade do vento, enquanto as quedas abruptas impõem maiores limitações preditivas devido à natureza estocástica e altamente não linear do escoamento atmosférico.

Como conclusão, observa-se que a maior concentração de erro permanece associada aos eventos de descida rápida de potência, evidenciando que tais eventos representam o principal fator limitante para a redução dos erros de previsão em curto prazo. Esses resultados corroboram a interpretação física do problema e reforçam que o desempenho dos modelos não deve ser avaliado apenas por métricas globais, mas também pela sua capacidade de acompanhar dinamicamente as rampas de potência, aspecto crucial para aplicações operacionais e para o gerenciamento de sistemas elétricos com elevada penetração de energia eólica.

Figura 29 - Potência eólica observada e prevista em uma janela de 24 horas do Estudo de Caso 2, entre 27 e 28 de setembro de 2020



Fonte: Autoria própria (2025).

5.2.2 Impacto da integração do Time2Vec no desempenho dos modelos (Estudo de Caso 2)

De forma análoga ao Estudo de Caso 1, a adição do Time2Vec melhorou o desempenho dos modelos de forma geral, conforme ilustrado na Tabela 15. Os valores são expressos em porcentagens, com números negativos indicando que a adição de Time2Vec não melhorou os modelos.

Para os X-formers, a incorporação do Time2Vec resultou em ganhos mais expressivos no Flashformer, com melhorias de 5,15% e 6,67% nos horizontes de 10 e 12 horas para a métrica MAE, e de 4,97% e 6,61% para a métrica RMSE nos respectivos horizontes. No caso do Flowformer, as maiores melhorias ocorreram nos horizontes de 6 e 10 horas, alcançando 5,82% e 4,04% para MAE, e 3,03% e 2,86% para RMSE. Já o Transformer apresentou o maior ganho entre todos os modelos no horizonte de 12 horas, com melhorias de 11,06% e 9,05% para as métricas MAE e RMSE, respectivamente.

Para os modelos de *benchmark*, a melhoria foi mais acentuada no modelo MLP, com valores de 4,16% e 7,17% no horizonte de 12 horas para as métricas MAE e RMSE, respectivamente. Para o modelo LSTM, a adição do Time2Vec no geral não foi positiva.

Já para o DLinear, houve uma pequena melhoria para os horizontes finais, mas nada tão acentuado em comparação com os X-formers.

Assim como no Estudo de Caso 1, a incorporação do Time2Vec resultou em ganhos significativos de desempenho, especialmente nas arquiteturas X-former, evidenciando uma sinergia estrutural entre as codificações temporais e os mecanismos de atenção. No Estudo de Caso 2, esse comportamento reforça as conclusões apresentadas na Seção 5.1.2. Conforme discutido anteriormente, nas arquiteturas X-former, a representação temporal é fundamental para capturar relações de dependência entre instantes da série. O Time2Vec aprimora esse processo ao introduzir uma codificação contínua e aprendível do tempo, permitindo que o modelo identifique padrões temporais com maior precisão. Dessa forma, sua utilização mostra-se altamente recomendável em arquiteturas baseadas em atenção.

Tabela 15 - Melhorias com a adição do Time2Vec (Estudo de Caso 2)

Modelo	MAE			RMSE		
	6h	10h	12h	6h	10h	12h
MLP	4.051	4.349	4.614	5.334	5.746	5.995
T2V-MLP	3.992	4.255	4.422	5.178	5.427	5.565
Melhoria (%)	1,46	2,16	4,16	2,92	5,55	7,17
LSTM	4.146	4.467	4.573	5.318	5.575	5.646
T2V-LSTM	4.162	4.428	4.592	5.343	5.621	5.806
Melhoria (%)	-0,39	0,87	-0,42	-0,47	-0,83	-2,83
DLinear	4.349	4.337	4.342	5.482	5.462	5.468
T2V-DLinear	4.394	4.397	4.274	5.509	5.462	5.381
Melhoria (%)	-1,03	-1,38	1,57	-0,49	0,00	1,59
Transformer	3.752	4.115	4.439	5.094	5.467	5.933
T2V-Transformer	3.870	3.971	3.948	5.261	5.321	5.396
Melhoria (%)	-3,14	3,50	11,06	-3,28	2,67	9,05
Flowformer	4.090	4.207	4.146	5.348	5.452	5.422
T2V-Flowformer	3.852	4.037	3.998	5.186	5.296	5.331
Melhoria (%)	5,82	4,04	3,57	3,03	2,86	1,68
Flashformer	4.061	4.235	4.255	5.318	5.549	5.616
T2V-Flashformer	3.896	4.017	3.971	5.136	5.273	5.245
Melhoria (%)	4,06	5,15	6,67	3,42	4,97	6,61

Fonte: Autoria própria (2025).

5.3 Análise comparativa entre o Estudos de Caso 1 e 2

Como demonstrado nas Seções 5.1 (Tabelas 11 e 12) e 5.2 (Tabela 14), observa-se uma diferença significativa entre as melhorias obtidas em relação ao modelo de Persistência nos dois estudos de caso. Considerando o IoR-RMSE como referência, no Estudo de Caso 1 os ganhos situaram-se na faixa de 15% a aproximadamente 17% para o Cenário A, e de 23% a aproximadamente 28% para o Cenário B. No Estudo de Caso 2, as melhorias variaram entre 56% e aproximadamente 62%. Essa discrepância indica que os modelos propostos apresentaram um desempenho substancialmente superior no Estudo de Caso 2. Essa diferença pode estar associada a fatores intrínsecos de cada conjunto de dados, como as características do regime de ventos, o nível de variabilidade temporal e a qualidade das medições SCADA, que influenciam diretamente a previsibilidade da série temporal.

Além disso, o Estudo de Caso 2, por se referir a uma usina eólica composta por várias turbinas, tende a apresentar menor variabilidade relativa e maior estabilidade na curva de potência agregada, o que favorece a aprendizagem dos padrões pelos modelos baseados em *deep learning*. Esse comportamento decorre do chamado efeito de agregação, no qual as flutuações individuais de potência entre as turbinas se compensam parcialmente, resultando em uma série temporal mais suave e previsível.

Em síntese, enquanto o Estudo de Caso 1 evidencia um ganho moderado em relação à persistência, o Estudo de Caso 2 demonstra uma melhoria expressiva e consistente, reforçando a robustez e a capacidade de generalização dos modelos propostos em cenários com maior estabilidade operacional e qualidade de dados.

Para uma comparação mais equitativa entre os Estudos de Caso 1 e 2 existem as métricas nMAE e nRMSE, que permitem normalizar os erros em relação à potência nominal, tornando a avaliação independente da escala absoluta de cada sistema eólico. Dessa forma, as Equações (19) e (20) apresentam as expressões utilizadas para o cálculo dessas métricas.

$$nMAE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{P_{nom}} \times 100 \% \quad (19)$$

$$nRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{P_{nom}} \right)^2} \times 100\% \quad (20)$$

Onde P_{nom} para o Estudo de Caso 1 vale 2.300 kW e 29,7 MW para o Estudo de Caso 2. A Tabela 16 apresenta os resultados para as métricas nMAE e nRMSE dos modelos deste estudo, para ambos os Estudos de Caso. De forma análoga ao MAE e ao RMSE, as métricas nMAE e nRMSE apresentam melhor desempenho quando assumem valores menores. Para fins de simplificação, considerou-se o horizonte de 12 horas como referência.

A Tabela 16 apresenta uma comparação mais equitativa entre os Estudos de Caso 1 e 2. Observa-se que os valores das métricas estão mais próximos entre si, uma vez que a normalização padroniza os resultados, eliminando o efeito das diferentes potências nominais de cada caso.

Tabela 16 - Valores de nMAE e nRMSE para os modelos avaliados no Estudo de Caso 1 (Cenário A e B) e Estudo de Caso 2, considerando o horizonte de previsão de 12 horas

Modelo	nMAE (%)			nRMSE (%)		
	Estudo de Caso 1		Estudo de Caso 2	Estudo de Caso 1		Estudo de Caso 2
	Cenário A	Cenário B		Cenário A	Cenário B	
Persistência	12,16	23,47	40,05	16,70	29,41	46,20
ARIMA	11,41	18,50	15,30	14,47	22,42	18,84
MLP	11,04	18,64	15,54	14,18	22,57	20,18
T2V-MLP	10,85	18,50	14,89	14,37	22,36	18,74
LSTM	10,89	18,99	15,40	14,40	22,60	19,01
T2V-LSTM	10,97	18,56	15,46	14,18	22,46	19,55
DLinear	10,88	18,60	14,62	14,26	22,62	18,41
T2V-DLinear	10,70	18,60	14,39	14,12	22,48	18,12
Transformer	11,04	18,83	14,94	14,37	22,57	19,98
T2V-Transformer	10,51	18,76	13,29	13,74	22,56	18,17
Flowformer	10,83	18,67	13,96	14,19	22,25	18,26
T2V-Flowformer	10,74	17,73	13,46	14,23	21,22	17,95
Flashformer	10,70	18,58	14,33	14,33	22,38	18,91
T2V-Flashformer	10,54	17,75	13,37	13,92	21,33	17,66

Fonte: Autoria própria (2025).

Para a métrica nMAE, os modelos T2V-Transformer, T2V-Flowformer e T2V-Flashformer apresentaram respectivamente 10,51%, 10,74% e 10,54% para o Cenário A; 18,76%, 17,73% e 17,75% para o Cenário B; e 13,29%, 13,46% e 13,37% para o Estudo

de Caso 2. Já para a métrica nRMSE, os modelos apresentaram respectivamente 13,74%, 14,23%, 13,92% para o Cenário A, 22,56%, 21,22% e 21,33% para o Cenário B; e 18,17%, 17,95% e 17,66% para o Estudo de Caso 2.

Em todos os cenários avaliados, esses valores não apenas superam de forma consistente os demais modelos, mas o fazem com estabilidade e regularidade ao longo de praticamente todo o horizonte de previsão. Embora as diferenças em relação a modelos de referência como ARIMA, MLP, LSTM e DLinear variem entre 1% e 3%, esse intervalo é altamente expressivo quando traduzido para a escala operacional dos sistemas estudados.

No Estudo de Caso 2, por exemplo, uma diferença de apenas 1% em nRMSE corresponde a aproximadamente 297 kW de erro evitado em uma usina com potência nominal de 29,7 MW. Já uma diferença de 3% representa cerca de 891 kW — quase 1 MW de erro a menos em cada horizonte de 12 horas. Como essas previsões são realizadas continuamente, essa redução de erro tende a se acumular ao longo do tempo, resultando em um ganho substancial de precisão operacional, com impacto direto na gestão energética e no planejamento da produção. Em ambientes operacionais onde decisões de despacho, compra e venda de energia e planejamento de manutenção dependem de estimativas confiáveis, essa melhoria se traduz em economia direta, melhor alocação de recursos e maior previsibilidade da geração.

Com o objetivo de avaliar a robustez estatística dos resultados obtidos nos Estudos de Caso 1 e 2, cada modelo proposto foi treinado e avaliado em 10 execuções independentes, utilizando diferentes *seeds* aleatórias. A Tabela 17 apresenta a média e o desvio padrão da métrica IoR-RMSE obtidos a partir dessas múltiplas execuções para os modelos avaliados no Estudo de Caso 1 (Cenários A e B) e no Estudo de Caso 2, considerando um horizonte de previsão de 12 h. Observa-se que o maior desvio padrão ocorreu para o modelo T2V-Flowformer no Estudo de Caso 1 (Cenário B), com valor de $\pm 2,29$, enquanto o menor foi observado para o T2V-Transformer no mesmo cenário, com $\pm 0,47$. De modo geral, os valores reduzidos de desvio padrão indicam estabilidade dos modelos frente à variação das inicializações aleatórias, evidenciando a robustez estatística das abordagens propostas.

Tabela 17 - Média ± desvio padrão do IoR-RMSE obtidos a partir de 10 execuções independentes para os modelos avaliados nos Estudos de Caso 1 (Cenários A e B) e Estudo de Caso 2, considerando horizonte de 12 h

Modelo	Estudo de Caso 1 (Cenário A)	Estudo de Caso 1 (Cenário B)	Estudo de Caso 2
T2V-Transformer	$15,88 \pm 1,71$	23.53 ± 0.47	$60,17 \pm 0,81$
T2V-Flowformer	$13,58 \pm 1,22$	25.63 ± 2.29	$59,94 \pm 1,40$
T2V-Flashformer	$16,68 \pm 1,01$	24.91 ± 1.98	$60,06 \pm 1,15$

Fonte: Autoria própria (2025).

Para verificar se as melhorias observadas em relação ao modelo de Persistência são estatisticamente significativas, aplicaram-se o *t*-test pareado e o teste não paramétrico de Wilcoxon sobre os valores da métrica RMSE, para todas as 10 seeds aleatórias obtidas nos experimentos realizados, considerando os três contextos avaliados — Estudo de Caso 1 (Cenário A e Cenário B) e Estudo de Caso 2.

Para verificar se as melhorias observadas em relação ao modelo de Persistência são estatisticamente significativas, aplicaram-se o *t*-test pareado e o teste não paramétrico de Wilcoxon sobre os valores agregados da métrica RMSE (um valor por execução), obtidos a partir das 10 *seeds* aleatórias realizadas nos experimentos, considerando os três contextos avaliados — Estudo de Caso 1 (Cenário A e Cenário B) e Estudo de Caso 2. O *t*-test pareado foi utilizado para comparar as médias das diferenças entre os resultados de cada modelo e o modelo de Persistência. Já o teste de Wilcoxon, de natureza não paramétrica, foi aplicado de forma complementar, dispensando o pressuposto de normalidade e avaliando a significância das diferenças com base nos postos ordenados das observações. A aplicação combinada de ambos os testes aumenta a robustez da análise, garantindo consistência mesmo diante de possíveis desvios de normalidade nos dados. Os testes estatísticos foram implementados com a biblioteca SciPy do Python (`scipy.stats.ttest_rel` e `scipy.stats.wilcoxon`), considerando um nível de significância de 5% ($p < 0,05$).

De acordo com a Tabela 18, observa-se que, no Estudo de Caso 2, os três modelos propostos apresentaram valores de p muito próximos tanto no *t*-test pareado quanto no teste de Wilcoxon. Esse comportamento decorre do grande tamanho do efeito observado em relação ao modelo de Persistência, aliado à baixa variabilidade entre as execuções, o

que conduz à saturação dos valores de p abaixo ou próximo do limiar de 0,001 no teste paramétrico e a valores muito próximos no teste não paramétrico. Assim, a coincidência dos resultados não indica equivalência de desempenho entre os modelos, mas sim que todos apresentam melhorias estatisticamente significativas e robustas em relação ao modelo de Persistência. Além disso, a consistência entre os resultados dos dois testes reforça a confiabilidade das conclusões, comprovando que as melhorias observadas refletem ganhos reais de desempenho. Resultados semelhantes entre o t -test e o teste de Wilcoxon também são relatados em cenários onde as diferenças entre pares são homogêneas e de mesma direção (FAY; PROSCHAN, 2010; MEEK; OZGUR; DUNNING, 2007).

Conforme evidenciado na Tabela 16, os modelos propostos (T2V-Transformer, T2V-Flowformer e T2V-Flashformer) mantêm o melhor desempenho, corroborando as análises e conclusões discutidas nas seções anteriores.

Tabela 18 - Valores de p obtidos nos testes estatísticos t -test e Wilcoxon aplicados à métrica RMSE para os modelos avaliados nos Estudos de Caso 1 (Cenários A e B) e Estudo de Caso 2, considerando o horizonte de previsão de 12 horas

Modelo	Estudo de Caso 1				Estudo de Caso 2	
	Cenário A		Cenário B		p (t -test)	p (Wilcoxon)
	p (t -test)	p (Wilcoxon)	p (t -test)	p (Wilcoxon)		
T2V-Transformer	<0.001	0.007	<0.001	0.002	<0.001	0.002
T2V-Flowformer	<0.001	0.007	0.026	0.026	<0.001	0.002
T2V-Flashformer	<0.001	0.001	0.021	0.021	<0.001	0.002

Fonte: Autoria própria (2025).

5.4 Análise comparativa de desempenho dos modelos, custo computacional e escalabilidade

A Tabela 19 apresenta uma avaliação comparativa dos modelos de predição utilizados neste estudo. Os modelos variam em sua sensibilidade a padrões temporais, com os X-formers geralmente exibindo uma alta capacidade de capturar tais dependências. A adição do Time2Vec aumenta ainda mais essa sensibilidade, pois codifica explicitamente informações temporais. Neste estudo, os dados de séries temporais são alimentados diretamente em cada modelo. Embora o MLP não possua memória temporal inerente, o LSTM e o Transformer capturam dependências por meio

de suas arquiteturas sequenciais — o LSTM por meio de sua memória interna e o Transformer por meio de mecanismos de auto atenção, que se concentram dinamicamente em passos de tempo relevantes. Os X-formers alavancam seus respectivos mecanismos de atenção para o aprendizado de representação temporal, enquanto o DLinear emprega uma técnica de decomposição que auxilia na modelagem de séries temporais.

Em relação ao desempenho computacional, o tempo total de experimento para cada modelo — compreendendo treinamento de mais de 100 tentativas e inferência (geração de predição) — foi medido usando as GPUs empregadas neste estudo, conforme discutido na Seção 4.6.

Para o Estudo de Caso 1, a GPU utilizada foi a Nvidia RTX A4000, com 16 GB de VRAM. Conforme mostrado na Tabela 19, o T2V-Transformer teve o maior tempo de experimento, aproximadamente 4 horas e 36 minutos para o Cenário A e 4 horas e 56 minutos para o Cenário B. Comparando os mecanismos de atenção, FlowAttention e FlashAttention apresentam custos computacionais menores que FullAttention, demonstrando vantagens significativas tanto no Cenário A quanto no Cenário B. O ARIMA teve o menor tempo de experimento, cerca de 24 minutos para o Cenário A e 28 minutos para o Cenário B. Pode-se observar que o Cenário B exigiu um pouco mais de tempo para todos os modelos. Isso pode ser atribuído às diferenças no comportamento temporal da série e às configurações de hiperparâmetros usadas. Modelos e dados com padrões temporais mais complexos normalmente requerem mais tempo de processamento e treinamento.

Para o Estudo de Caso 2, a GPU utilizada foi a Nvidia RTX 4060 Ti, com 8 GB de VRAM. De acordo com a Tabela 19, o T2V-Transformer apresentou o maior tempo de experimento, com aproximadamente 3 horas e 3 minutos; seguido pelo Transformer com aproximadamente 2 horas e 55 minutos. Como discutido no parágrafo anterior, os mecanismos FlowAttention e FlashAttention demonstraram mais vantajosos neste quesito específico, com menor tempo de processamento em comparação com o FullAttention. Os modelos com menor tempo de processamento foram ARIMA, MLP e LSTM, respectivamente com 15, 38 e 40 minutos.

Além disso, a inclusão do Time2Vec na arquitetura dos modelos aumentou a duração total dos experimentos, devido à computação adicional necessária para capturar padrões temporais específicos. O custo computacional de cada modelo foi avaliado com base na duração total do experimento. Modelos com tempos de execução abaixo de 2 horas foram classificados como tendo baixo custo computacional, aqueles acima de 2

horas e abaixo de 3 foram classificados como tendo custo moderado. E acima de 3 horas como tendo alto custo. Finalmente, em relação à escalabilidade para grandes conjuntos de dados, o MLP tem baixa escalabilidade devido à sua incapacidade de capturar dependências temporais de forma eficaz (TAUD; MAS, 2018). O LSTM tem escalabilidade moderada, pois processa sequências longas sequencialmente, o que pode se tornar um gargalo para grandes conjuntos de dados (YU, Y. *et al.*, 2019). O DLinear, beneficiando-se de sua abordagem de decomposição linear, atinge alta escalabilidade (ZENG *et al.*, 2023). O Transformer tem escalabilidade moderada, pois sua complexidade quadrática pode limitar sua eficiência para sequências muito longas (VASWANI *et al.*, 2017). Em contraste, o Flowformer e o Flashformer apresentam escalabilidade muito alta, pois seus mecanismos de atenção especializados são otimizados para sequências de séries temporais longas, melhorando significativamente a eficiência computacional (DAO *et al.*, 2022; WU, H. *et al.*, 2022).

Apesar das diferenças no custo computacional e na escalabilidade, todos os modelos avaliados são viáveis para previsões operacionais de curto prazo de energia eólica com um horizonte de previsão de 12 horas. Os tempos de experimento relatados nesta subseção correspondem a 100 ensaios; em contraste, o tempo unitário de inferência (latência média por predição), métrica determinante para aplicações em tempo real, é substancialmente menor e é explicitamente apresentado nas Tabelas 8, 9 e 10, confirmando a aplicabilidade prática dos modelos propostos."

Vale ressaltar que o tempo de processamento de cada modelo, considerando 100 ensaios, foi obtido com base na capacidade computacional dos equipamentos utilizados neste estudo, conforme descrito detalhadamente na Seção 4.6.

Tabela 19 - Critérios de avaliação comparativa dos modelos considerando 100 ensaios. O tempo unitário de inferência da melhor configuração é apresentado nas Tabelas 8, 9 e 10

Modelos	Sensibilidade a padrões temporais	Estratégia de aprendizado	Escalabilidade para grandes conjuntos de dados	Tempo total de experimento (100 ensaios)				Custo Computacional		
				Estudo de Caso 1		Estudo de Caso 2	Caso 1		Estudo de Caso 2	
				Cenário A	Cenário B		Cenário A	Cenário B		
ARIMA	Baixo	Modelagem estatística	Baixo	24 min	28 min	15 min	Baixo	Baixo	Baixo	
MLP	Baixo	Engenharia de recursos	Baixo	58 min	1 h 1 min	38 min	Baixo	Baixo	Baixo	
T2V-MLP	Moderado	Engenharia de recursos	Baixo	1h 8min	1 h 10 min	56 min	Baixo	Baixo	Baixo	
LSTM	Moderado	Aprendizagem implícita de padrões	Moderado	53 min	58 min	40 min	Baixo	Baixo	Baixo	
T2V-LSTM	Alto	Aprendizagem implícita de padrões	Moderado	1h 5min	1 h 10 min	53 min	Baixo	Baixo	Baixo	
DLinear	Moderado	Modelagem linear com decomposição	Alto	1 h 43 min	1 h 48 min	1 h 36 min	Baixo	Baixo	Baixo	
T2V-DLinear	Alto	Modelagem linear com decomposição	Alto	2 h 1 min	2 h 6 min	1 h 52 min	Moderado	Moderado	Baixo	
Transformer	Alto	Aprendizagem baseada em FullAttention	Moderado	4 h 10 min	4 h 26 min	2 h 55 min	Alto	Alto	Moderado	
T2V-Transformer	Muito Alto	Aprendizagem baseada em FullAttention	Moderado	4 h 36 min 40	4 h 56 min 40	3 h 3 min	Alto	Alto	Alto	
Flowformer	Alto	Aprendizagem baseada em FlowAttention	Muito Alto	3 h 43 min 20	4 h 3 min 20	2 h 33 min	Alto	Alto	Moderado	
T2V-Flowformer	Muito Alto	Aprendizagem baseada em FlowAttention	Muito Alto	4 h 5 min	4 h 26 min	2 h 47 min	Alto	Alto	Moderado	
Flashformer	Alto	Aprendizagem baseada em FlashAttention	Muito Alto	2 h 43 min	3 h 6 min	2 h 8 min	Moderado	Alto	Moderado	
T2V-Flashformer	Muito Alto	Aprendizagem baseada em FlashAttention	Muito Alto	3 h 18 min	3 h 36 min	2 h 25 min	Alto	Alto	Moderado	

Fonte: Autoria própria (2025).

6 CONCLUSÃO

Neste estudo, foram propostos três novos modelos para a previsão de energia eólica de curto prazo, aplicados a turbinas eólicas em operação localizadas na região Nordeste do Brasil. Para garantir uma análise de previsão robusta e avaliar o desempenho dos modelos sob diferentes condições temporais, dois casos foram considerados. No Estudo de Caso 1, realizou-se a previsão para uma única turbina eólica situada em um parque eólico no estado do Rio Grande do Norte, com dois cenários distintos: Cenário A, abrangendo o período de verão e outono, e Cenário B, correspondente à transição do inverno para a primavera. No Estudo de Caso 2, a análise de previsão foi realizada para uma usina eólica completa, localizada na Chapada do Araripe — região que se estende entre os estados do Ceará, Pernambuco e Piauí —, considerando apenas um cenário.

Os modelos propostos integram a camada Time2Vec para aprimorar a representação de padrões temporais nos dados. Uma análise de sensibilidade foi realizada com três arranjos, identificando a configuração que otimizou o desempenho do modelo. Os melhores resultados foram obtidos quando o Time2Vec foi aplicado apenas na entrada do codificador (Arranjo I), preservando a capacidade do decodificador de gerar saídas a partir das representações codificadas.

Além disso, este estudo explorou mecanismos alternativos de atenção, substituindo o mecanismo FullAttention pelos mecanismos FlowAttention e FlashAttention dos modelos Flowformer e Flashformer, com o objetivo de mitigar a complexidade quadrática da atenção tradicional resultante do mecanismo FullAttention. Este estudo trata-se da primeira aplicação do mecanismo FlashAttention à previsão de energia eólica, bem como da primeira integração do Time2Vec com múltiplos mecanismos de atenção nesse contexto.

Os modelos propostos foram comparados com MLP, LSTM e DLinear — cada um também testado com integração Time2Vec, além do ARIMA. No geral, os resultados demonstram que a abordagem proposta melhora significativamente a precisão da previsão e a eficiência computacional, confirmando sua eficácia para a previsão de energia eólica de curto prazo.

Com base na metodologia proposta e nos resultados apresentados, podemos destacar as principais conclusões deste trabalho da seguinte forma:

1. O processo metodológico utilizado neste estudo provou ser altamente eficaz,

incorporando pré-processamento, tratamento de dados e o uso do Optuna para otimização eficiente de hiperparâmetros. Essa abordagem ajudou a prevenir o *overfitting* e identificou as melhores configurações possíveis de cada modelo.

2. A metodologia proposta demonstrou sua eficácia na previsão da potência de turbinas eólicas, com os modelos apresentando melhorias substanciais em relação ao modelo de Persistência. Os resultados alcançados neste estudo contribuem para o avanço da área de previsão de energia eólica, oferecendo referências valiosos para a otimização de modelos preditivos em aplicações de energia renovável.
3. A análise de sensibilidade da integração do Time2Vec na arquitetura do Transformer facilitou a identificação da configuração ideal para esta aplicação. Essa adição foi particularmente vantajosa para os X-formers, com os modelos Flowformer e Flashformer apresentando melhorias em praticamente todos os cenários.
4. Para o Estudo de Caso 1, no Cenário A, os modelos com melhor desempenho foram o T2V-Transformer e o T2V-Flashformer, demonstrando maior consistência em todos os horizontes e métricas. Para a tarefa de previsão de 12 horas, esses modelos alcançaram valores de IoR-MAE de 13,55% e 13,30%, respectivamente, superando o ARIMA (6,20 %), MLP (9,22%), LSTM (10,43%) e DLinear (10,52%). Da mesma forma, na métrica IoR-RMSE, o T2V-Transformer e o T2V-Flashformer atingiram 17,73% e 16,67%, enquanto o ARIMA, MLP, LSTM e DLinear obtiveram 13,36%, 15,08%, 13,79% e 14,64%, respectivamente. No Cenário B, os modelos com melhor desempenho foram o T2V-Flowformer e o T2V-Flashformer. Para o horizonte de 12 horas, eles alcançaram valores de IoR-MAE de 24,47% e 24,37%, superando ARIMA (21,20 %), MLP (20,60%), LSTM (19,09%) e DLinear (20,75%). Na métrica IoR-RMSE, o T2V-Flowformer e o T2V-Flashformer atingiram 27,84% e 27,45%, enquanto ARIMA, MLP, LSTM e DLinear obtiveram 23,75%, 23,23%, 23,15% e 23,06%, respectivamente.
5. No Estudo de Caso 2, considerando o horizonte de 12 horas e a métrica IoR-MAE, os modelos T2V-Transformer, T2V-Flashformer e T2V-Flowformer apresentaram os melhores desempenhos, com valores de 66,80%, 66,62% e 66,39%, respectivamente, superando os modelos ARIMA, MLP, LSTM e DLinear, que obtiveram 61,78%, 61,21%, 61,55% e 63,49%. Para a métrica IoR-RMSE, destacaram-se os modelos T2V-Flashformer e T2V-Flowformer, com

valores de 61,76% e 61,14%, respectivamente, superando ARIMA, MLP, LSTM e DLinear, que apresentaram 59,21%, 56,30%, 58,84% e 60,14%.

6. Os mecanismos FlowAttention e FlashAttention demonstraram custos computacionais menores em comparação com o FullAttention, conforme evidenciado por tempos de teste mais curtos nos Estudos de Caso 1 e 2. Em relação ao desempenho preditivo, para o Estudo de Caso 1, o T2V-Transformer apresentou resultados superiores no Cenário A. No Cenário B, no entanto, o T2V-Flowformer e o T2V-Flashformer superaram o T2V-Transformer, sugerindo que esses modelos são mais adequados para este contexto específico. Para o Estudo de Caso 2, os três modelos apresentaram desempenho semelhante, com o T2V-Transformer performando melhor na métrica MAE; enquanto na métrica RMSE os modelos T2V-Flowformer e T2V-Flashformer no geral performaram melhor.
7. Os modelos propostos são os mais adequados para este estudo, apresentando consistentemente os melhores resultados em praticamente todas as métricas e horizontes temporais. Ao superar os *benchmarks* em quase todos os cenários, eles representam um avanço significativo no estado da arte. Sua precisão preditiva aprimorada aumenta a eficiência operacional dos parques eólicos, otimizando as estratégias de manutenção e a confiabilidade geral. Além disso, contribuem para um uso mais eficaz de recursos renováveis, como a energia eólica.

Esta pesquisa apresenta uma abordagem robusta, com tempo de execução e viabilidade aceitáveis para os modelos propostos, fornecendo previsões de energia com 12 horas a frente no horizonte temporal. Com base na metodologia adotada, nos modelos desenvolvidos e nos resultados alcançados, este trabalho pode contribuir para maximizar a eficiência produtiva de parques eólicos. Além disso, ao integrar codificações temporais interpretáveis, como Time2Vec, em arquiteturas baseadas em atenção, este estudo ajuda a reduzir a natureza de "caixa-preta" frequentemente associada ao modelo Transformer, promovendo maior confiança e transparência em suas aplicações práticas.

7 PERSPECTIVAS FUTURAS

Como perspectivas futuras, espera-se que os modelos desenvolvidos neste estudo sejam aplicados à previsão de energia eólica em horizontes de médio e longo prazo, possibilitando uma avaliação mais abrangente de seu desempenho em diferentes escalas temporais. Outra direção promissora consiste na aplicação desses modelos a parques eólicos situados em distintas regiões geográficas, de modo a verificar sua capacidade de generalização frente a diferentes regimes de vento e condições climáticas.

Além disso, os modelos propostos podem ser explorados em outras tarefas relacionadas à energia eólica, como a previsão da velocidade do vento e a detecção de anomalias em componentes críticos dos aerogeradores. Outra vertente relevante envolve a integração de técnicas de decomposição às arquiteturas desenvolvidas, uma vez que tais abordagens podem potencializar a precisão e a robustez das previsões.

As arquiteturas propostas também apresentam flexibilidade para adaptação a novos mecanismos de atenção — tanto os já existentes quanto aqueles que venham a ser desenvolvidos —, mantendo-se atualizadas frente à rápida evolução dos modelos de aprendizado profundo. Por fim, ressalta-se que esses modelos possuem ampla aplicabilidade na previsão de séries temporais em diversos domínios, como finanças, economia, saúde e ciências climáticas.

REFERÊNCIAS BIBLIOGRÁFICAS

- AASIM, M.; SINGH, S. N.; MOHAPATRA, A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. **Renewable Energy**, v. 136, p. 758–768, jun. 2019.
- ABEEÓLICA. Boletim Anual 2024. São Paulo: ABEEólica, 2024. Disponível em: https://abeeolica.org.br/wp-content/uploads/2025/05/424_ABEEOLICA_BOLETIM-ANUAL-DIGITAL-2025_PT_FINAL.pdf. Acesso em: 22 out. 2025.
- AKIBA, T. *et al.* Optuna. New York, NY, USA: **ACM**, 2019. p. 2623–2631.
- BAN, G. *et al.* The univariate model for long-term wind speed forecasting based on wavelet soft threshold denoising and improved Autoformer. **Energy**, v. 290, p. 130225, mar. 2024.
- BANGALORE, P.; PATRIKSSON, M. Analysis of SCADA data for early fault detection, with application to the maintenance management of wind turbines. **Renewable Energy**, v. 115, p. 521–532, jan. 2018.
- BENTSEN, L. Ø. *et al.* Spatio-temporal wind speed forecasting using graph networks and novel Transformer architectures. **Applied Energy**, v. 333, p. 120565, mar. 2023a.
- BENTSEN, L. Ø. *et al.* A Unified Graph Formulation for Spatio-Temporal Wind Forecasting. **Energies**, v. 16, n. 20, p. 7179, 20 out. 2023b.
- BISPO JUNIOR, D. A.; LEITE, G. de N. P.; VILLA, A. A. O.; BRENNAND, L. J. de P.; COSTA, A. C. A. da; SOUZA, O. V. C. de; CAVALCANTI, M. T. Wavelet-Flowformer: integração da decomposição wavelet a um mecanismo de atenção avançado para previsão de potência eólica a curto prazo. **Observatorio de La Economía Latinoamericana**, v. 23, n. 11, p. e12327, 2025. DOI: 10.55905/oelv23n11-138.
- BOMMIDI, B. S.; TEEPARTHI, K.; DULLA MALLESHAM, V. K. ICEEMDAN-Informer-GWO: a hybrid model for accurate wind speed prediction. **Environmental Science and Pollution Research**, 1 maio 2024, v. 31, n. 23, p. 34056–34081.
- BOMMIDI, B. S.; TEEPARTHI, K.; KOSANA, V. Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function. **Energy**, v. 265, p. 126383, fev. 2023.
- BOTTOU, L. Large-Scale Machine Learning with Stochastic Gradient Descent. In: **COMPSTAT'2010**. Heidelberg: Physica-Verlag HD, 2010. p. 177–186.
- CAO, Q.; EWING, B. T.; THOMPSON, M. A. Forecasting wind speed with recurrent neural networks. **European Journal of Operational Research**, v. 221, n. 1, p. 148–154, ago. 2012.

CHANG, W.-Y. A Literature Review of Wind Forecasting Methods. **Journal of Power and Energy Engineering**, v. 2, n. 4, p. 161–168, 2014.

CHAWLA, I. *et al.* Assessment of the Weather Research and Forecasting (WRF) model for simulation of extreme rainfall events in the upper Ganga Basin. **Hydrology and Earth System Sciences**, v. 22, n. 2, p. 1095–1117, 8 fev. 2018.

CHEN, Y. *et al.* WindFix: Harnessing the power of self-supervised learning for versatile imputation of offshore wind speed time series. **Energy**, v. 287, p. 128995, jan. 2024.

CLEVELAND, R. B. *et al.* STL: A seasonal-trend decomposition. **Journal of Official Statistics**, v. 6, n. 1, p. 3–73, 1990.

COOLEY, J. W.; LEWIS, P. A. W.; WELCH, P. D. The Fast Fourier Transform and Its Applications. **IEEE Transactions on Education**, v. 12, n. 1, p. 27–34, 1969.

COSTA, R. *et al.* Vector Representation and Machine Learning for Short-Term Photovoltaic Power Prediction. [S.l.]: IEEE, 2023. p. 1241–1246.

CRABTREE, C. J.; ZAPPALÁ, D.; TAVNER, P. J. Survey of commercially available condition monitoring systems for wind turbines. **Durham University**, v. 24, 2010.

DAI, K. *et al.* Environmental issues associated with wind energy – A review. **Renewable Energy**, v. 75, p. 911–921, mar. 2015.

DAO, T. *et al.* FLASHATTENTION: Fast and Memory-Efficient Exact Attention with IO-Awareness. [S.l.: s.n.], 2022. Disponível em: <https://github.com/HazyResearch/flash-attention>.

DONG, Z. *et al.* Wind-Mambaformer: Ultra-Short-Term Wind Turbine Power Forecasting Based on Advanced Transformer and Mamba Models. **Energies**, v. 18, n. 5, p. 1155, 26 fev. 2025.

DONTHU, N. *et al.* How to conduct a bibliometric analysis: An overview and guidelines. **Journal of Business Research**, v. 133, p. 285–296, set. 2021.

DRAGOMIRETSKIY, K.; ZOSSO, D. Variational Mode Decomposition. **IEEE Transactions on Signal Processing**, v. 62, n. 3, p. 531–544, fev. 2014.

DUTTA, S. *et al.* Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique. **Energy Procedia**, v. 143, p. 617–622, dez. 2017.

ELLABBAN, O.; ABU-RUB, H.; BLAABJERG, F. Renewable energy resources: Current status, future prospects and their enabling technology. **Renewable and Sustainable Energy Reviews**, v. 39, p. 748–764, nov. 2014.

ERDEM, E.; SHI, J. ARMA based approaches for forecasting the tuple of wind speed and direction. **Applied Energy**, v. 88, n. 4, p. 1405–1414, abr. 2011.

FAY, M. P.; PROSCHAN, M. A. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. **Statistics Surveys**, v. 4, p. 1–39, jan. 2010.

FU, T.; WANG, D.; JIN, J. Forecasting the daily evaporation by coupling the ensemble deep learning models with meta-heuristic algorithms and data pre-processing in dryland. **Scientific Reports**, v. 15, n. 1, p. 30502, 2025. GALLEGÓ CASTILLO, C. J. Statistical models for short-term wind power ramp forecasting. [S.l.]: **Universidad Politécnica de Madrid**, 2013.

GAO, Y. *et al.* Short-Term Wind Speed Prediction for Bridge Site Area Based on Wavelet Denoising OOA-Transformer. **Mathematics**, v. 12, n. 12, p. 1910, 20 jun. 2024.

GARCÍA MÁRQUEZ, F. P. *et al.* Condition monitoring of wind turbines: Techniques and methods. **Renewable Energy**, v. 46, p. 169–178, out. 2012.

GENG, D.; WANG, B.; GAO, Q. A hybrid photovoltaic/wind power prediction model based on Time2Vec, WDCNN and BiLSTM. **Energy Conversion and Management**, v. 291, p. 117342, set. 2023.

GIELEN, D. *et al.* The role of renewable energy in the global energy transformation. **Energy Strategy Reviews**, v. 24, p. 38–50, abr. 2019.

GONG, M. *et al.* Short-term wind power forecasting model based on temporal convolutional network and Informer. **Energy**, v. 283, p. 129171, 2023.

GU, A.; DAO, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. 31 maio 2024. Disponível em: <http://arxiv.org/abs/2312.00752>.

GWEC. GLOBAL WIND ENERGY COUNCIL. **Global Wind Report 2025**. Lisboa: GWEC, 2025. Disponível em: <https://26973329.fs1.hubspotusercontent-eu1.net/hubfs/26973329/2.%20Reports/Global%20Wind%20Report/GWEC%20Global%20Wind%20Report%202025.pdf>. Acesso em: 22 dez. 2025.

HARBOLA, S.; COORS, V. One dimensional convolutional neural network architectures for wind prediction. **Energy Conversion and Management**, v. 195, p. 70–75, set. 2019.

HASAGER, C. B. *et al.* Remote Sensing Observation Used in Offshore Wind Energy. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 1, n. 1, p. 67–79, mar. 2008.

HE, Y. *et al.* Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression. **Neurocomputing**, v. 430, p. 121–137, mar. 2021.

HE, Y.; LI, H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. **Energy Conversion and Management**, v. 164, p. 374–384, 2018.

HE, Y. *et al.* An Integrated CEEMDAN to Optimize Deep Long Short-Term Memory Model for Wind Speed Forecasting. **Energies**, v. 17, n. 18, p. 4615, 14 set. 2024.

HEIER, S. Grid Integration of Wind Energy. [S.l.]: Wiley, 2014.

HINTON, G.; SRIVASTAVA, N.; SWERSKY, K. Neural Networks for Machine Learning Lecture 6a Overview of mini-batch gradient descent. [S.l.]: [s.n.], 2012.

HONG, J.-T. *et al.* Dual-path frequency Mamba-Transformer model for wind power forecasting. **Energy**, v. 332, p. 137225, set. 2025.

HORVÁTH, L.; KOKOSZKA, P.; RICE, G. Testing stationarity of functional time series. **Journal of Econometrics**, v. 179, n. 1, p. 66–82, mar. 2014.

HU, J. *et al.* A novel transformer ordinal regression network with label diversity for wind power ramp events forecasting. **Energy**, v. 280, p. 128075, out. 2023.

HU, Z. *et al.* An efficient R-Transformer network with dual encoders for brain glioma segmentation in MR images. **Biomedical Signal Processing and Control**, v. 79, p. 104034, jan. 2023.

HUANG, S.; YAN, C.; QU, Y. Deep learning model-transformer based wind power forecasting approach. **Frontiers in Energy Research**, v. 10, 16 jan. 2023a.

HUANG, X.; JIANG, A. Wind Power Generation Forecast Based on Multi-Step Informer Network. **Energies**, v. 15, n. 18, p. 6642, 11 set. 2022.

ITALLO RIBEIRO DIAS DA SILVA, A. *et al.* Aerodynamic interference caused by wake effects of repowered wind farms on the annual energy production in neighboring wind farms. **Sustainable Energy Technologies and Assessments**, v. 64, p. 103704, abr. 2024.

JIN, Z. *et al.* Informer learning framework based on secondary decomposition for multi-step forecast of ultra-short term wind speed. **Engineering Applications of Artificial Intelligence**, v. 139, p. 109702, jan. 2025.

KATHAROPOULOS, Angelos *et al.* Transformers are rnns: Fast autoregressive transformers with linear attention. In: **International conference on machine learning**. PMLR, 2020. p. 5156-5165.

KAVASSERI, R. G.; SEETHARAMAN, K. Day-ahead wind speed forecasting using f-ARIMA models. **Renewable Energy**, v. 34, n. 5, p. 1388–1393, maio 2009.

KAZEMI, S. M. *et al.* Time2Vec: Learning a Vector Representation of Time. 11 jul. 2019. Disponível em: <http://arxiv.org/abs/1907.05321>

KHAN, S. *et al.* Transformers in Vision: A Survey. **ACM Computing Surveys**, v. 54, n. 10s, p. 1–41, 31 jan. 2022.

KHUDRI JOHARI, M.; AZIM A JALIL, M.; FAIZAL MOHD SHARIFF, M. Comparison of horizontal axis wind turbine (HAWT) and vertical axis wind turbine (VAWT). **International Journal of Engineering & Technology**, v. 7, n. 4.13, p. 74–80, 9 out. 2018.

KINGMA, D. P.; BA, J. Adam: A Method for Stochastic Optimization. 22 dez. 2014. Disponível em: <http://arxiv.org/abs/1412.6980>.

LEITE, G. DE N. P.; ARAÚJO, A. M.; ROSAS, P. A. C. Prognostic techniques applied to maintenance of wind turbines: a concise and specific review. **Renewable and Sustainable Energy Reviews**, v. 81, p. 1917–1925, jan. 2018.

LENG, Z. *et al.* Short-term wind speed forecasting based on a novel KANInformer model and improved dual decomposition. **Energy**, v. 322, p. 135551, maio 2025.

LI, N. *et al.* A novel EMD and causal convolutional network integrated with Transformer for ultra short-term wind power forecasting. **International Journal of Electrical Power & Energy Systems**, v. 154, p. 109470, dez. 2023.

LI, S. *et al.* Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. 29 jun. 2019. Disponível em: <http://arxiv.org/abs/1907.00235>.

LIN, S. *et al.* GAOformer: An adaptive spatiotemporal feature fusion transformer utilizing GAT and optimizable graph matrixes for offshore wind speed prediction. **Energy**, v. 292, p. 130404, abr. 2024.

LIN, T. *et al.* A survey of transformers. **AI Open**, v. 3, p. 111–132, 2022.

LIU, S. *et al.* PYRAFORMER: Low-Complexity Pyramidal Attention for Long-Range Time Series Modeling and Forecasting. [S.l.]: [s.n.], 2022. Disponível em: <https://github.com/alipay/Pyraformer>.

LYDIA, M. *et al.* A comprehensive review on wind turbine power curve modeling techniques. **Renewable and Sustainable Energy Reviews**, v. 30, p. 452–460, fev. 2014.

MARVUGLIA, A.; MESSINEO, A. Monitoring of wind farms' power curves using machine learning techniques. **Applied Energy**, v. 98, p. 574–583, out. 2012.

MCKENNA, R.; OSTMAN V.D. LEYE, P.; FICHTNER, W. Key challenges and prospects for large wind turbines. **Renewable and Sustainable Energy Reviews**, v. 53, p. 1212–1221, jan. 2016.

MEEK, G. E.; OZGUR, C.; DUNNING, K. Comparison of the t vs. Wilcoxon Signed-Rank Test for Likert Scale Data and Small Samples. **Journal of Modern Applied Statistical Methods**, v. 6, n. 1, p. 91–106, 1 maio 2007.

MO, S. *et al.* Powerformer: A temporal-based transformer model for wind power forecasting. **Energy Reports**, v. 11, p. 736–744, jun. 2024.

MÖHRLEN, C. *et al.* EirGrid's met mast and alternatives study. **IET Renewable Power Generation**, v. 16, n. 9, p. 1941–1954, 27 jul. 2022.

MURTHY, K. S. R.; RAHI, O. P. A comprehensive review of wind resource assessment. **Renewable and Sustainable Energy Reviews**, v. 72, p. 1320–1342, maio 2017.

NASCIMENTO, E. G. S.; MELO, T. A. C. DE; MOREIRA, D. M. A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy. **Energy**, v. 278, p. 127678, set. 2023.

NASH, R.; NOURI, R.; VASEL-BE-HAGH, A. Wind turbine wake control strategies: A review and concept proposal. **Energy Conversion and Management**, v. 245, p. 114581, out. 2021.

PAGE, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, p. n71, 29 mar. 2021.

PAN, X. *et al.* Short-term wind speed forecasting based on spatial-temporal graph transformer networks. **Energy**, v. 253, p. 124095, ago. 2022.

PARRI, S.; TEEPARTHI, K. SVMD-TF-QS: An efficient and novel hybrid methodology for the wind speed prediction. **Expert Systems with Applications**, v. 249, p. 123516, set. 2024.

PENG, T. *et al.* An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. **Energy**, v. 221, p. 119887, abr. 2021.

QIAO, W.; LU, D. A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis—Part II: Signals and Signal Processing Methods. **IEEE Transactions on Industrial Electronics**, v. 62, n. 10, p. 6546–6557, out. 2015a.

QU, K. *et al.* Short-term forecasting for multiple wind farms based on transformer model. **Energy Reports**, v. 8, p. 483–490, ago. 2022a.

RHIF, M. *et al.* Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. **Applied Sciences**, v. 9, n. 7, p. 1345, 30 mar. 2019.

SHENG, S.; O'CONNOR, R. Reliability of wind turbines. In: **Wind Energy Engineering**. [S.l.]: Elsevier, 2023. p. 195–211.

SHI, Z. *et al.* WGformer: A Weibull-Gaussian Informer based model for wind speed prediction. **Engineering Applications of Artificial Intelligence**, v. 131, p. 107891, maio 2024.

SOARES, A. *et al.* *Metodologia da Pesquisa Científica*. [S.l.]: [s.n.], 2018.

SONG, J.; WANG, J.; LU, H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. **Applied Energy**, v. 215, p. 643–658, abr. 2018.

TANAKA, T.; MANDIC, D. P. Complex Empirical Mode Decomposition. **IEEE Signal Processing Letters**, v. 14, n. 2, p. 101–104, fev. 2007.

TANG, W.; LIU, C.; ZHANG, B. Gated Fusion Based Transformer Model for Crack Detection on Wind Turbine Blade. **Energy Engineering**, v. 120, n. 11, p. 2667–2681, 2023.

TAO, H. *et al.* An Online Digital Imaging Excitation Sensor for Wind Turbine Gearbox Wear Condition Monitoring Based on Adaptive Deep Learning Method. **Sensors**, v. 24, n. 8, p. 2481, 12 abr. 2024.

TAUD, H.; MAS, J. F. Multilayer Perceptron (MLP). [S.l.]: [s.n.], 2018. p. 451–455.

TAUTZ-WEINERT, J.; WATSON, S. J. Using SCADA data for wind turbine condition monitoring – a review. **IET Renewable Power Generation**, v. 11, n. 4, p. 382–394, mar. 2017.

TAY, Y. *et al.* Efficient Transformers: A Survey. **ACM Computing Surveys**, v. 55, n. 6, p. 1–28, 30 jun. 2023.

TIAN, Y. *et al.* An Adaptive Hybrid Model for Wind Power Prediction Based on the IVMD-FE-Ad-Informer. **Entropy**, v. 25, n. 4, p. 647, 12 abr. 2023.

TIWARI, R.; BABU, N. R. Recent developments of control strategies for wind energy conversion system. **Renewable and Sustainable Energy Reviews**, v. 66, p. 268–285, dez. 2016.

VAJIRE, S. L. *et al.* A Comparative Analysis of Wind Turbine Power Generation Forecasting: Recurrent Neural Network vs. Multi-Head Self-Attention Transformer Approaches. [S.l.]: [s.n.], [s.d.].

VASWANI, A. *et al.* Attention Is All You Need. [S.l.]: [s.n.], 2017.

VEERS, P. *et al.* Grand challenges in the science of wind energy. **Science**, 25 out. 2019, v. 366, n. 6464.

VOYANT, C. *et al.* Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. **Energy**, mar. 2012, v. 39, n. 1, p. 341–355.

WAN, A. *et al.* Multistep forecasting method for offshore wind turbine power based on multi-timescale input and improved Transformer. **Journal of Marine Science and Engineering**, 31 maio 2024, v. 12, n. 6, p. 925.

WANG, Hai-Kun *et al.* GCNInformer: A combined deep learning model based on GCN and Informer for wind power forecasting. **Energy Science & Engineering**, 21 out. 2023, v. 11, n. 10, p. 3836–3854.

WANG, H.-K.; SONG, K.; CHENG, Y. A hybrid forecasting model based on CNN and Informer for short-term wind power. **Frontiers in Energy Research**, 24 jan. 2022, v. 9.

WANG, Lei; HE, Yigang; LI, Lie; *et al.* A novel approach to ultra-short-term multi-step wind power predictions based on encoder–decoder architecture in natural language processing. **Journal of Cleaner Production**, jun. 2022, v. 354, p. 131723.

WANG, Lei *et al.* M2TNet: Multi-modal multi-task Transformer network for ultra-short-term wind power multi-step forecasting. **Energy Reports**, 2022, v. 8, p. 7628–7642.

WANG, Yun *et al.* A convolutional Transformer-based truncated Gaussian density network with data denoising for wind speed forecasting. **Applied Energy**, mar. 2023, v. 333, p. 120601.

WANG, Zhongrui *et al.* Short-term offshore wind power multi-location multi-modal multi-step prediction model based on Informer (M3STIN). **Energy**, maio 2025, v. 322, p. 135616.

WEI, H.; WANG, W.; KAO, X. A novel approach to ultra-short-term wind power prediction based on feature engineering and informer. **Energy Reports**, dez. 2023, v. 9, p. 1236–1250.

WU, H. *et al.* Autoformer: Decomposition Transformers with Auto-Correlation for long-term series forecasting. [S.l.]: [s.n.], 2021. Disponível em: <https://github.com/thuml/Autoformer>.

WU, Haixu *et al.* Flowformer: Linearizing transformers with conservation flows. **arXiv preprint**, arXiv:2202.06258, 2022.

XIAO, H.; HE, X.; LI, C. Probability density forecasting of wind power based on transformer network with expectile regression and kernel density estimation. **Electronics**, 1 mar. 2023, v. 12, n. 5, p. 1187.

XINXIN, W. *et al.* Short-term wind speed forecasting based on a hybrid model of ICEEMDAN, MFE, LSTM and Informer. **PLOS ONE**, 8 set. 2023, v. 18, n. 9, p. e0289161.

YANG, W.; COURT, R.; JIANG, J. Wind turbine condition monitoring by the approach of SCADA data analysis. **Renewable Energy**, maio 2013, v. 53, p. 365–376.

YU, Chengqing *et al.* Attention mechanism is useful in spatio-temporal wind speed prediction: Evidence from China. **Applied Soft Computing**, nov. 2023, v. 148, p. 110864.

YU, Y. *et al.* A review of recurrent neural networks: LSTM cells and network architectures. **Neural Computation**, jul. 2019, v. 31, n. 7, p. 1235–1270.

YU, Z. *et al.* Multi-attribute adaptive aggregation transformer for vehicle re-identification. **Information Processing & Management**, mar. 2022, v. 59, n. 2, p. 102868.

ZENG, A. *et al.* Are transformers effective for time series forecasting? **Proceedings of the AAAI Conference on Artificial Intelligence**, 26 jun. 2023, v. 37, n. 9, p. 11121–11128.

ZERVEAS, George *et al.* A transformer-based framework for multivariate time series representation learning. In: **Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining**. New York: ACM, 2021. p. 2114–2124.

ZHA, W. *et al.* A wind speed vector–wind power curve modeling method based on data denoising algorithm and the improved Transformer. **Electric Power Systems Research**, jan. 2023, v. 214, p. 108838.

ZHANG, K.; LI, X.; SU, J. Variable support segment-based short-term wind speed forecasting. **Energies**, 1 jun. 2022, v. 15, n. 11, p. 4067.

ZHANG, Yunhao; YAN, Junchi. CROSSFORMER: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. [S.I.]: [s.n.], 2023.

ZHANG, Zhendong *et al.* Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression. **Applied Energy**, ago. 2019, v. 247, p. 270–284.

ZHAO, J. *et al.* A self-organizing forecast of day-ahead wind speed: Selective ensemble strategy based on numerical weather predictions. **Energy**, mar. 2021, v. 218, p. 119509.

ZHENG, H. *et al.* VMD-CAT: A hybrid model for short-term wind power prediction. **Energy Reports**, jun. 2023, v. 9, p. 199–211.

ZHENG, Y. *et al.* Hierarchical spatial–temporal autocorrelation graph neural network for online wind turbine pitch system fault detection. **Neurocomputing**, jun. 2024, v. 586, p. 127574.

ZHOU, Haoxuan *et al.* Convolution enabled transformer via random contrastive regularization for rotating machinery diagnosis under time-varying working conditions. **Mechanical Systems and Signal Processing**, jul. 2022, v. 173, p. 109050.

ZHOU, Haoyi *et al.* Informer: Beyond efficient transformer for long sequence time-series forecasting. **Proceedings of the AAAI Conference on Artificial Intelligence**, 18 maio 2021, v. 35, n. 12, p. 11106–11115.

ZHOU, T. *et al.* FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. [S.I.]: [s.n.], 2022. Disponível em: <https://github.com/MAZiqing/FEDformer>

APÊNDICE A

As *strings* de busca utilizados na base Scopus para o estudo de revisão sistemática descrito na Seção 3.5 encontram-se apresentados nos Quadros A1 e A2.

Quadro A1 - *String* de busca utilizada na Etapa 1 da revisão sistemática (SCOPUS)

```
TITLE-ABS-KEY("transformers networks" OR "transformers model" OR "transformers architecture"
OR "crossformer" OR "autoformer" OR "pyraformer" OR "informer" OR "reformer" OR "fedformer"
OR "itransformer" OR "Non-stationary Transformers" OR "LogSparse Transformer" OR "PatchTST"
OR "flowformer" OR "FlashAttention" OR "iTransformer") AND NOT TITLE-ABS-KEY("electrical
networks" OR "power transformers" OR "power transformer") AND ORIG-LOAD-DATE > 20241125
AND ( LIMIT-TO ( PUBSTAGE,"final" ) ) AND ( LIMIT-TO ( DOCTYPE,"ar" ) ) AND ( LIMIT-TO
( PUBYEAR,2017) OR LIMIT-TO ( PUBYEAR,2018) OR LIMIT-TO ( PUBYEAR,2019) OR LIMIT-
TO ( PUBYEAR,2020) OR LIMIT-TO ( PUBYEAR,2021) OR LIMIT-TO ( PUBYEAR,2022) OR
LIMIT-TO ( PUBYEAR,2023) OR LIMIT-TO ( PUBYEAR,2024) OR LIMIT-TO ( PUBYEAR,2025)
) AND ( LIMIT-TO ( LANGUAGE,"English" ) )
```

Fonte: Autoria própria (2025).

Quadro A2 - *String* de busca utilizada na Etapa 2 da revisão sistemática (SCOPUS)

```
TITLE-ABS-KEY("transformers networks" OR "transformers model" OR "transformers architecture"
OR "crossformer" OR "autoformer" OR "pyraformer" OR "informer" OR "reformer" OR "fedformer"
OR "itransformer" OR "Non-stationary Transformers" OR "LogSparse Transformer" OR "PatchTST" OR
"flashformer" OR "flashattention" OR "flowformer") AND TITLE-ABS-KEY("wind turbine" OR
"wind energy" OR "power forecast" OR "forecast wind" OR "wind power" OR "wind speed" OR "wind
turbine") AND NOT TITLE-ABS-KEY("electrical networks" OR "power transformers" OR "solar"
OR "diesel reformers") AND ORIG-LOAD-DATE > 20240903 AND ( LIMIT-TO ( PUBSTAGE,"final" )
) AND ( LIMIT-TO ( DOCTYPE,"ar" ) ) AND ( LIMIT-TO ( PUBYEAR,2017)
OR LIMIT-TO ( PUBYEAR,2018) OR LIMIT-TO ( PUBYEAR,2019) OR LIMIT-TO ( PUBYEAR,2020)
OR LIMIT-TO ( PUBYEAR,2021) OR LIMIT-TO ( PUBYEAR,2022) OR LIMIT-
TO ( PUBYEAR,2023) OR LIMIT-TO ( PUBYEAR,2024) OR LIMIT-TO ( PUBYEAR,2025) ) AND
( LIMIT-TO ( LANGUAGE,"English" ) )
```

Fonte: Autoria própria (2025).

APÊNDICE B

A análise de sensibilidade da incorporação do Time2Vec na arquitetura Transformer considerou os Arranjos I, II e III, conforme detalhado neste estudo. A otimização de hiperparâmetros usando o Optuna produziu as configurações mostradas nas Tabelas B1 e B2 para os Cenários A e B, respectivamente. Cada análise incluiu a avaliação de diferentes mecanismos de atenção — FullAttention, FlowAttention e FlashAttention — dentro dos respectivos Arranjos.

De acordo com a Tabela B1, o T2V-Transformer no Arranjo I utiliza apenas 1 camada de codificador e 1 de decodificador, enquanto os Arranjos II e III empregam 3 camadas cada. O Arranjo III apresenta a menor dimensão do modelo ($d_{model} = 32$), seguido pelo Arranjo I ($d_{model} = 64$) e pelo Arranjo II ($d_{model} = 128$). O comprimento da sequência (*seq len*) foi maior no Arranjo I, sugerindo uma capacidade aprimorada de capturar dependências temporais de longo prazo. Para o modelo T2V-Flowformer, o Arranjo I empregou valores menores tanto para $d_{model} = 32$ quanto para $d_{ff} = 96$, enquanto os Arranjos II e III utilizaram $d_{model} = 64$ e $d_{ff} = 384$. Esses valores mais altos sugerem um custo computacional maior, pois levam a mais operações tanto no mecanismo de atenção quanto nas camadas de *feed-forward*, resultando, consequentemente, em maiores demandas de inferência e treinamento. Para o modelo T2V-Flashformer, o Arranjo I apresentou os maiores valores para $d_{model} = 256$ e $d_{ff} = 512$. Em contraste, os Arranjos II e III utilizaram valores de d_{model} valores de 64 e 32, e d_{ff} valores de 256 e 96, respectivamente. Essas configurações indicam que o Arranjo I incorre no maior custo computacional devido ao aumento do número de operações tanto nos componentes de atenção quanto nos de propagação direta.

De acordo com a Tabela B2, para o T2V-Transformer, o Arranjo I foi configurado com apenas 1 camada de codificador e 1 camada de decodificador, enquanto que o Arranjo II com 1 camada de codificador e 2 camadas de decodificador. Já o Arranjo III apresentou 3 camadas para ambos. Em termos de *seq len*, o Arranjo I apresentou o valor de 44, já o Arranjo II apresentou o maior valor, com 92. O valor de d_{model} foi mais alto no Arranjo II (128), já para d_{ff} o Arranjo I apresentou o maior valor (384). Para o T2V-Flowformer, o Arranjo I apresentou o maior *seq len* (120). O número de camadas de codificador e decodificador foi de 2 para ambos. O Arranjo II utilizou 2 camadas para o codificador e 1 camada para o decodificador, enquanto o Arranjo III possui 1 camada para

ambos. O Arranjo I apresentou o maior valor para d_{model} e d_{ff} (128 e 512, respectivamente); enquanto o Arranjo II utilizou 64 e 256; e o Arranjo III utilizou 64 e 384, respectivamente, para estes parâmetros. Para o T2V-Flashformer, o Arranjo I demonstrou maior capacidade de capturar padrões temporais de longo prazo devido ao seu maior *seq len* (105), enquanto os Arranjos II e III apresentaram 43 e 23, respectivamente. Além disso, o Arranjo I apresentou valores mais altos para d_{model} , d_{ff} e número de cabeças (256, 1536 e 8, respectivamente), indicando um custo computacional maior em comparação com os Arranjos II e III.

Em relação à sensibilidade aos hiperparâmetros, as variações do *seq len*, d_{model} , d_{ff} , número de cabeças de atenção e demais parâmetros demonstraram que o desempenho do modelo não escala linearmente com o tamanho da arquitetura. Dimensões de incorporação e larguras de *feed-forward* maiores aumentaram o custo computacional, mas não garantiram maior precisão. De acordo com os valores dos hiperparâmetros apresentados nas Tabelas B1 e B2, conclui-se que o projeto do arranjo exerce influência mais significativa que a magnitude dos parâmetros. De modo geral, os Arranjos I apresentaram melhor desempenho nos Cenários A e B, sendo adotados neste trabalho para a arquitetura proposta.

Todos os modelos empregaram a função de ativação ReLU, o que garantiu convergência estável entre os arranjos, reforçando que a interação entre a codificação periódica do Time2Vec e os mecanismos de atenção desempenha um papel mais decisivo do que a escolha da função de ativação. Para o Cenário A, o Arranjo I de todos os modelos utilizou o RMSprop. Já para o Cenário B, o Arranjo I de todos os modelos utilizaram o Adam. Para o Arranjo II, os modelos apresentaram um uso equilibrado de RMSprop e Adam para ambos os Cenários. Já o Arranjo III para ambos os Cenários, todos os modelos utilizam o Adam, com exceção do T2V-Flashformer no Cenário B; nenhum utilizou SGD. Uma taxa de *dropout* de 0,1 foi aplicada consistentemente em todas as configurações. Essas descobertas destacam que os benefícios da escalabilidade dependem fortemente do mecanismo de atenção: o FlowAttention permanece eficiente com dimensionalidade moderada, enquanto o FlashAttention torna-se mais sensível ao crescimento de parâmetros devido às suas operações densas.

Em relação ao componente periódico do Time2Vec (seno versus cosseno), para o T2V-Transformer no Cenário A, os Arranjos I e II usam seno, enquanto o Arranjo III usa cosseno; já para o Cenário B, os Arranjos I e III usam cosseno, enquanto o Arranjo II usa

seno. Para o T2V-Flowformer no Cenário A, os Arranjos I e II usam cosseno e o Arranjo III usa seno; no Cenário B, o Arranjo I usa o seno, enquanto os Arranjos II e III usam o cosseno. Para o T2V-Flashformer, os Arranjos I e II usam seno, enquanto o Arranjo III usa cosseno (em ambos os cenários).

Os resultados da Análise de Sensibilidade são apresentados nas Figuras B1 e B2. A Figura B1 corresponde ao Cenário A e ilustra os resultados para os modelos T2V-Transformer, T2V-Flowformer e T2V-Flashformer. A Figura B2 apresenta os resultados correspondentes para o Cenário B. Em todos os casos, o Arranjo I produziu consistentemente os menores erros de previsão em todos os horizontes, indicando que essa configuração foi a mais adequada para as arquiteturas dos modelos. Portanto, o Arranjo I foi adotado para as previsões realizadas neste estudo.

Tabela B1- Análise de sensibilidade do Cenário A

Parâmetro	Cenário A								
	T2V-Transformer			T2V-Flowformer			T2V-Flashformer		
	Arranjo I	Arranjo II	Arranjo III	Arranjo I	Arranjo II	Arranjo III	Arranjo I	Arranjo II	Arranjo III
<i>seq len</i>	32	23	15	49	22	27	63	18	13
Camadas do codificador	1	3	3	2	2	3	2	2	3
Camadas do decodificador	1	3	3	3	2	2	1	2	2
Épocas	10	10	10	10	10	10	10	10	10
Otimizador	RMSprop	Adam	Adam	RMSprop	RMSprop	Adam	RMSprop	RMSprop	Adam
Função de ativação	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU
<i>Dropout</i>	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
<i>d_{model}</i>	64	128	32	32	64	64	256	64	32
Nº de cabeças	4	2	8	4	4	2	6	6	6
<i>d_{ff}</i>	128	512	128	96	384	384	512	256	96
Função	sen	sen	cos	cos	cos	sen	sen	sen	cos

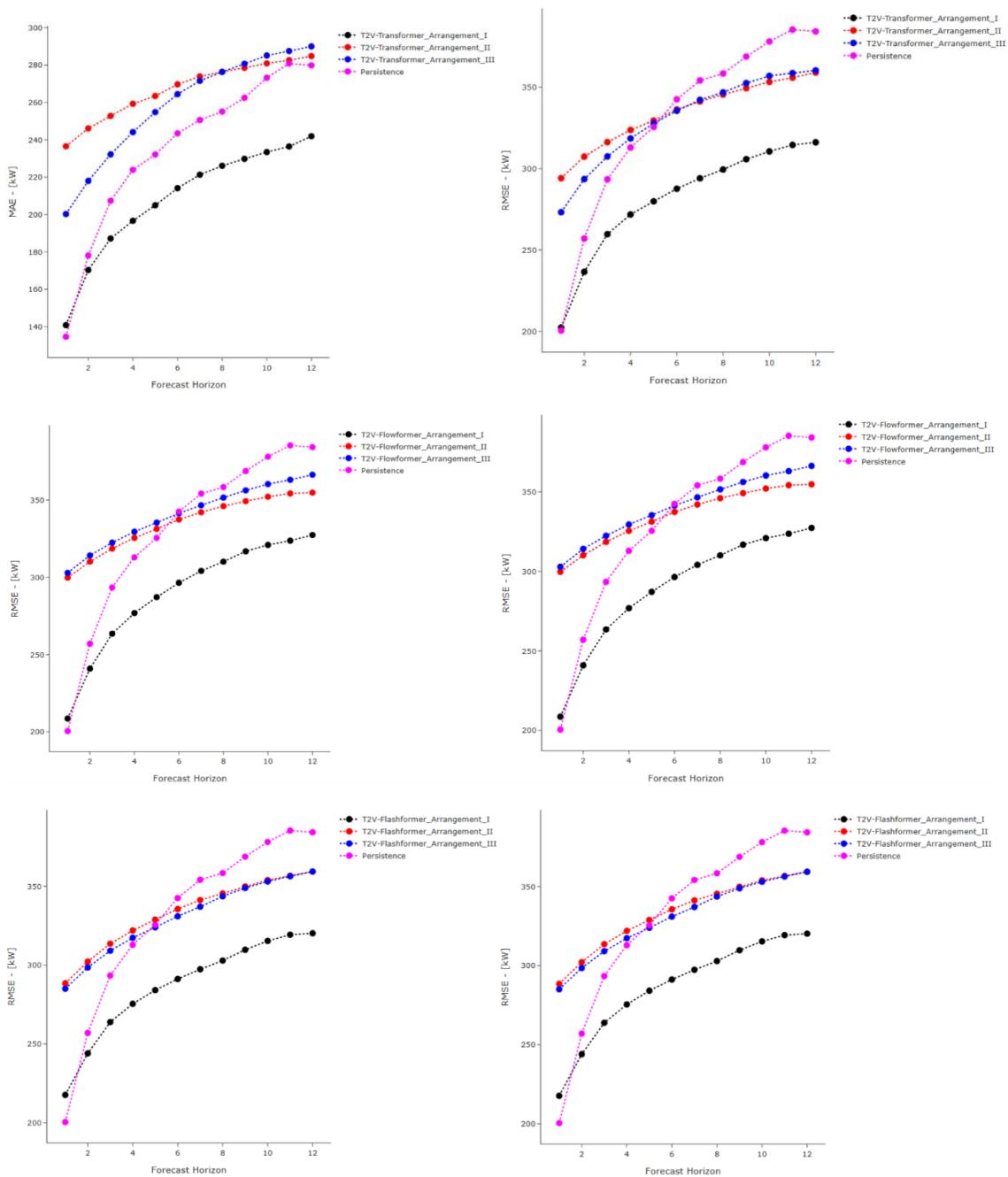
Fonte: Autoria própria (2025).

Tabela B2- Análise de sensibilidade do Cenário B

Parâmetro	Cenário B								
	T2V-Transformer			T2V-Flowformer			T2V-Flashformer		
	Arranjo I	Arranjo II	Arranjo III	Arranjo I	Arranjo II	Arranjo III	Arranjo I	Arranjo II	Arranjo III
<i>seq len</i>	44	92	25	120	44	22	105	43	23
Camadas do codificador	1	1	3	2	2	1	3	3	3
Camadas do decodificador	1	2	3	2	1	1	1	1	2
Épocas	10	10	10	10	10	10	10	10	10
Otimizador	Adam	Adam	Adam	Adam	RMSprop	Adam	Adam	RMSprop	RMSprop
Função de ativação	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU
<i>Dropout</i>	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
d_{model}	64	128	32	128	64	64	256	64	128
Nº de cabeças	6	6	6	2	2	6	8	6	6
d_{ff}	384	256	128	512	256	384	1536	256	640
Função	cos	sen	cos	sen	cos	cos	sen	sen	cos

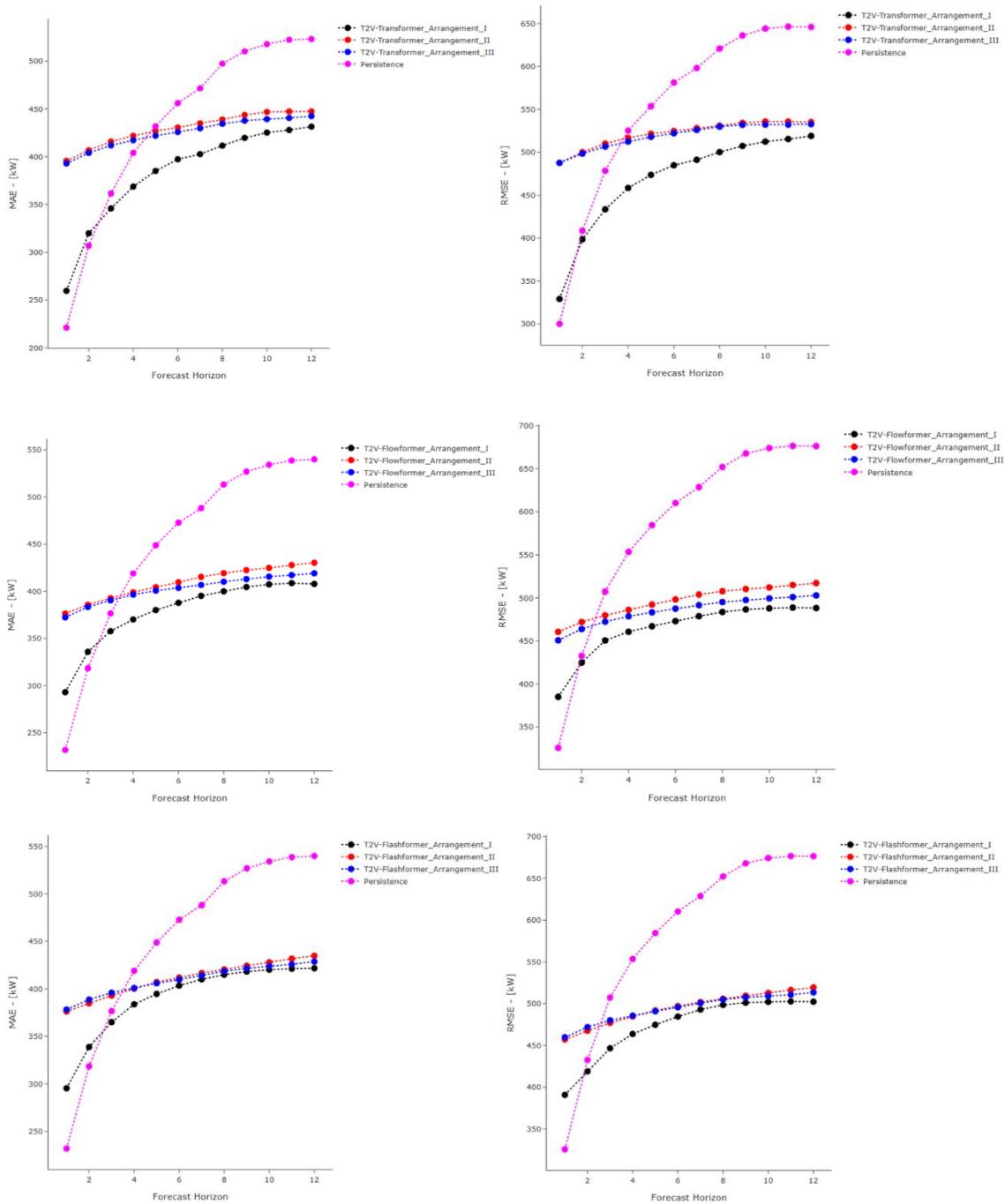
Fonte: Autoria própria (2025).

Figura B1 - Análise de sensibilidade para os modelos T2V-Transformer, T2V-Flowformer e T2V-Flashformer (Cenário A)



Fonte: Autoria própria (2025).

Figura B2 - Análise de sensibilidade para os modelos T2V-Transformer, T2V-Flowformer e T2V-Flashformer (Cenário B).



Fonte: Autoria própria (2025).