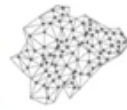




III MCSM

Modelagem Computacional no
Sertão Mineiro



PPGMCS
Programa de Pós-Graduação em
Modelagem Computacional e Sistemas

TIMESNET PARA PREVISÃO DE LONGO PRAZO DA IRRADIÂNCIA GLOBAL HORIZONTAL NOS ESTÁDIOS-SEDE DA COPA DO MUNDO FIFA 2026

Luis Eduardo Ramalho Gerino¹ - l117785@dac.unicamp.br

Gustavo Fraidenraich¹ - gfraiden@unicamp.br

Thales Wulfert Cabral¹ - t264377@dac.unicamp.br

¹Universidade Estadual de Campinas, UNICAMP - Campinas, SP, Brazil

Abstract. *Este trabalho propõe o emprego da arquitetura TimesNet para a previsão de longo prazo de Global Horizontal Irradiance (GHI), grandeza vital para o dimensionamento de plantas fotovoltaicas em grandes estabelecimentos. O dimensionamento de longo prazo exige séries históricas extensas, raramente disponíveis para o local de interesse; as séries preditas de GHI suprem essa lacuna ao reproduzir as propriedades estatísticas do clima local. Como estudo de caso, os modelos foram aplicados aos dezesseis estádios-sede da Copa do Mundo FIFA de 2026, distribuídos por EUA, México e Canadá, escolhidos por reunirem grande consumo, ampla cobertura e regimes climáticos diversos. A análise inédita do desempenho preditivo do TimesNet para GHI em estádios-sede da Copa do Mundo FIFA 2026 revelou que a proposta obteve o melhor desempenho médio (MAE: 12,76; RMSE: 16,44; R^2 : 0,925), superando modelos consolidados na literatura como eXtreme Gradient Boosting (XGBoost), Gated Recurrent Unit (GRU) e a Long Short-Term Memory (LSTM).*

Keywords: *Aprendizado de máquina, TimesNet, GHI, Energia fotovoltaica*

1. INTRODUÇÃO

A energia solar fotovoltaica se tornou uma das principais apostas para a expansão da geração elétrica mundial, respondendo, só em 2025, por cerca de três quartos de toda a nova capacidade renovável instalada no planeta (IRENA, 2025a). A queda acentuada no custo dos painéis (IRENA, 2025b), o amadurecimento da regulação da geração distribuída e a abundância do recurso explicam essa virada. Por ser renovável, modular e passível de produção no ponto de consumo, a fonte é especialmente atraente para grandes construções, como galpões, aeroportos e arenas, que têm amplas áreas de telhado expostas ao sol e sem uso. Para grandes construções, o desafio deixa de ser a adoção da geração fotovoltaica e passa a ser o dimensionamento da capacidade instalada que otimiza o compromisso entre investimento, geração esperada e risco. Para investigar essa pergunta é preciso olhar para o recurso solar disponível, medido pela

irradiância global horizontal (*Global Horizontal Irradiance*, GHI), em W/m^2 . Um único ano de dados mostra só uma fatia do clima do lugar, e por isso a prática de engenharia recomenda séries longas, da ordem de décadas, para estimar o risco de a geração ficar abaixo do esperado. Entretanto, dados históricos longos e confiáveis quase nunca estão disponíveis para o local exato de interesse. Neste caso, não se quer prever a irradiância das próximas horas ou dias, tarefa de curto prazo, mas sim prever séries longas e plausíveis, fiéis ao clima do lugar.

A predição das séries de irradiância, como as de GHI, construídas a partir do histórico real que mantêm suas características estatísticas, como a sazonalidade, a variação mês a mês e a forma como um mês influencia o seguinte, é vital para a indústria fotovoltaica. Com elas, é possível implementar um dimensionamento que leva a incerteza em conta, em vez de apostar em um único ano como sendo representativo. Avanços recentes no campo de aprendizado de máquina permitiram a implementação de modelos robustos na previsão de séries temporais (Hiremath et al., 2024; Li et al., 2026). O eXtreme Gradient Boosting (XGBoost) trata a previsão como uma regressão e aprende a relação entre o valor de uma amostra e as amostras anteriores usando árvores de decisão com *gradient boosting* (Chen & Guestrin, 2016; Cabral et al., 2024). A Gated Recurrent Unit (GRU) e a Long Short-Term Memory (LSTM) são redes neurais recorrentes, que carregam uma memória interna que leva informação de uma amostra para a próxima (Cho et al., 2014; Hochreiter & Schmidhuber, 1997). Recentemente, foi proposta a arquitetura TimesNet, que é um modelo de aprendizado de máquina que se destaca por descobrir sozinho os ciclos presentes na série, em especial o ciclo anual, em vez de recebê-los prontos (Wu et al., 2023). Diante das lacunas ainda existentes na previsão de longo prazo da irradiância global horizontal (GHI), especialmente em aplicações voltadas a grandes estabelecimentos, este trabalho propõe a utilização da arquitetura TimesNet para a previsão de longo prazo da GHI nos estádios-sede da Copa do Mundo FIFA 2026. Portanto, seguem as principais contribuições deste trabalho:

- Proposição da aplicação do TimesNet para a previsão de longo prazo de GHI;
- Análise inédita do desempenho preditivo do TimesNet para GHI em estádios-sede da Copa do Mundo FIFA 2026, estabelecendo uma comparação robusta entre o TimesNet e modelos *baseline* consolidados na literatura, como XGBoost, GRU e LSTM.

2. PROPOSTA PARA PREVISÃO DE GHI BASEADA NO MODELO TIMESNET

O diagrama ilustrado na Figura 1 resume o pipeline de ciência de dados estruturado em quatro etapas, voltado à previsão de GHI. O fluxo começa com a Aquisição dos Dados de GHI a partir das coordenadas de latitude e longitude dos estádios-sede da Copa do Mundo FIFA de 2026. Após a coleta, cada série passa por uma inspeção de qualidade, nomeada de Limpeza dos Dados, que incluiu a verificação da continuidade temporal, a checagem de valores ausentes e a inspeção de valores fisicamente implausíveis, como negativos. A série horária coletada foi então agregada em uma média mensal. Em seguida, na etapa de Processamento, os dados validados são submetidos à normalização Min-Max, que reescala as variáveis numéricas para o intervalo $[0,1]$, reduzindo discrepâncias de magnitude entre os atributos e proporcionando maior estabilidade e eficiência ao treinamento dos modelos de aprendizado de máquina. Por fim, a última fase é dedicada à Regressão, cujo objetivo é a Previsão de GHI.

Embora a proposta deste trabalho esteja centrada na aplicação do TimesNet, seu desempenho foi comparado com um *baseline* competitivo, incluindo XGBoost, GRU e LSTM.

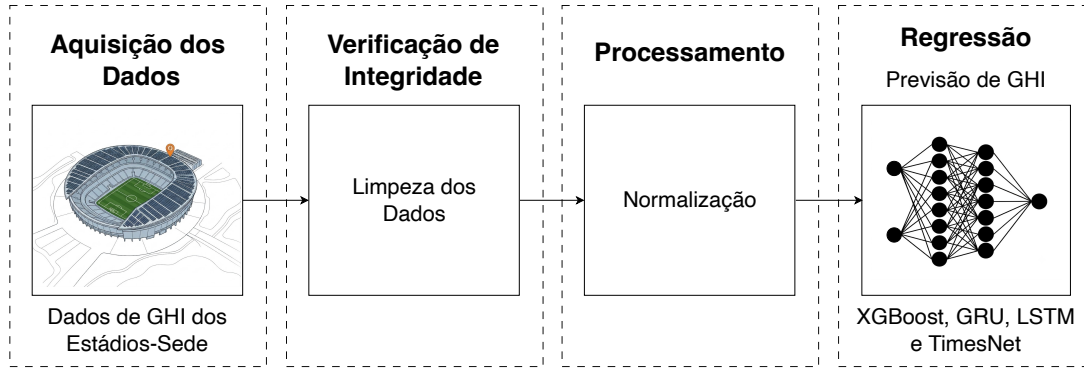


Figure 1: Visão geral do pipeline desenvolvido.

Aqui, o XGBoost, um método de *gradient boosting*, constrói árvores de decisão de forma sequencial, cada uma corrigindo o erro residual da anterior; a série mensal é transformada em uma tabela de atributos, com as defasagens, o mês em seno e cosseno e o indicador de ano. A GRU e a LSTM são redes neurais recorrentes que processam a janela de dados mantendo um estado interno atualizado por portões, lido na última janela. Entretanto, o TimesNet parte de uma ideia diferente: em vez de percorrer a série instante a instante, descobre os ciclos dominantes pela amplitude do espectro de Fourier, reorganiza a série em uma representação bidimensional segundo cada período e aplica convoluções no estilo Inception, capturando ao mesmo tempo as variações dentro de um período e entre períodos. A GRU e a LSTM usaram as camadas recorrentes prontas do PyTorch (Géron, 2022); o TimesNet foi implementado seguindo a arquitetura original. O Algoritmo 1 apresenta a sequência de instruções necessárias para a previsão via TimesNet, empregadas na previsão de GHI.

Algoritmo 1 — Previsão via TimesNet

1. Cada janela de dados é projetada para uma representação de maior dimensão por uma camada linear:

$$H \leftarrow \text{LayerNorm}(\text{Linear}(W))$$

2. Em cada bloco, o modelo descobre os ciclos dominantes pela amplitude do espectro de frequências, descartando a componente contínua, e obtém os períodos correspondentes:

$$A \leftarrow |\overline{\text{FFT}(H)}|, \quad A[0] \leftarrow 0, \quad p_i \leftarrow \lfloor T/f_i \rfloor$$

3. Para cada período, a representação é reorganizada em uma grade bidimensional, sobre a qual se aplicam convoluções no estilo Inception, que captam as vizinhanças nas duas direções:

$$Z_i \leftarrow \text{reshape}_{p_i}(H), \quad Z_i \leftarrow \text{Inception2D}(\text{GELU}(\text{Inception2D}(Z_i)))$$

4. As grades retornam ao formato de série e são recombinadas por uma média ponderada pelas amplitudes, com soma residual:

$$w \leftarrow \text{softmax}(A_{f_1}, \dots, A_{f_k}), \quad H \leftarrow \sum_i w_i Y_i + H$$

5. Após os blocos, a previsão da janela seguinte é lida da representação do último mês por uma camada linear, $\hat{y} \leftarrow \text{Linear}(H_{L_{seq}})$.

3. RESULTADOS E DISCUSSÃO

A robustez do método foi avaliada por meio de um conjunto complementar de métricas de desempenho: erro médio absoluto (MAE), raiz do erro quadrático médio (RMSE) e coeficiente de determinação (R^2), definidos em (1)

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}, \quad (1)$$

onde y_t é a irradiância global horizontal observada no período t ; $\hat{y}_t$ é a irradiância prevista pelo modelo no período t , e n é o número total de observações no conjunto de teste.

Para cada um dos dezesseis estádios-sede da Copa do Mundo FIFA 2026, os dados foram coletados no ponto correspondente às coordenadas do estádio, com resolução temporal horária, ao longo do período de 1998 a 2022, totalizando vinte e cinco anos de histórico por localidade. Desses vinte e cinco anos, os primeiros 85% de cada série foram destinados ao treino e os 15% finais ao teste, em divisão cronológica. Os dados provêm do *National Solar Radiation Database* (NSRDB), base de irradiância globalmente utilizada e acessada por meio da biblioteca pvlib. Aqui, a variável de interesse é o GHI, em W/m^2 .

Os resultados experimentais foram conduzidos por uma rotina unificada, aplicada de forma idêntica às dezesseis localidades: para cada uma, o pipeline prepara os dados, treina os quatro modelos com as mesmas configurações, gera as previsões sobre os 15% finais do período (reservados para teste em divisão cronológica) e calcula as métricas. Manter as configurações fixas permite avaliar como cada modelo se comporta em diferentes localidades.

A Tabela 1 traz a média de cada métrica entre as dezesseis localidades. O TimesNet obteve o melhor resultado em todas elas e foi o mais regular, com o menor desvio do R^2 entre localidades, seguido por LSTM, GRU e XGBoost. A diferença entre os três primeiros é pequena, mas há um salto de desempenho entre eles e o XGBoost.

Table 1: Comparação de desempenho médio dos modelos.

Modelo	MAE [W/m^2]	RMSE [W/m^2]	R^2	Desvio R^2
XGBoost	14,96	18,73	0,904	0,056
GRU	13,43	17,62	0,908	0,075
LSTM	13,00	16,73	0,919	0,060
TimesNet	12,76	16,44	0,925	0,046

A contagem de vitórias por localidade (Fig. 2, à esquerda) reforça a hierarquia: o TimesNet venceu em nove das dezesseis localidades, o LSTM em cinco, o GRU em duas e o XGBoost em nenhuma. O mapa de calor do R^2 (Fig. 2, à direita) detalha esse quadro. No Estádio Banorte/Azteca (R^2 0,816) e o Estádio BBVA, em Monterrey (0,838), a vantagem do TimesNet é evidente, e é onde o GRU mais sofre (0,687 no Azteca). Em Lumen Field (0,972), os quatro modelos apresentam desempenho similar.

A Figura 3 apresenta o contraste de desempenho, ao comparar as séries previstas e observadas de GHI para os dados de teste, em localidades que apresentaram o melhor e o pior desempenho. Na imagem esquerda da Fig. 3 (Lumen Field), as curvas previstas acompanham de perto o ciclo anual e os picos da série. Na imagem direita (Estádio Azteca), de altitude elevada e com um regime menos marcado pela sazonalidade anual, todos os modelos apresentam

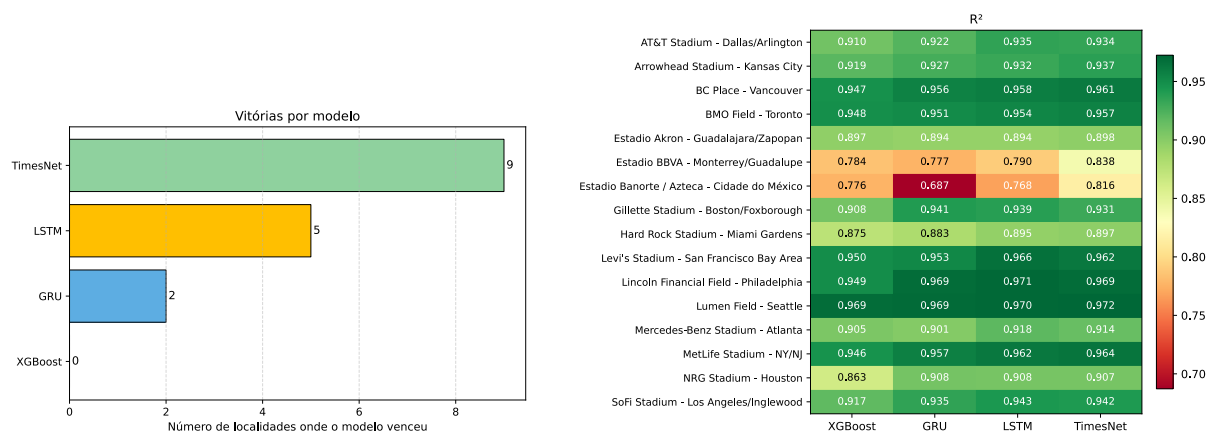


Figure 2: À esquerda, número de localidades em que cada modelo obteve o maior R^2 ; à direita, mapa de calor do R^2 por localidade.

maior dificuldade em capturar as oscilações, e é nesse caso que o TimesNet apresenta melhor desempenho sobre os demais.

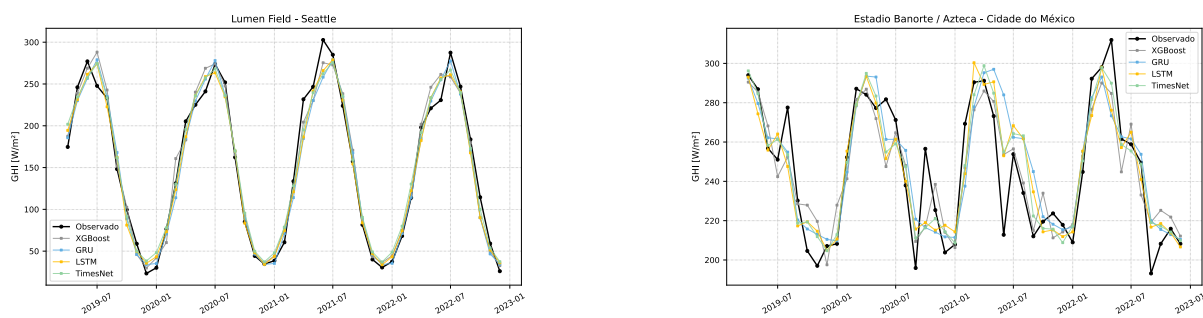


Figure 3: Séries temporais previstas frente às observadas no teste, nas localidades de melhor desempenho (Lumen Field, Seattle, à esquerda) e de pior desempenho (Estadio Banorte/Azteca, Cidade do México, à direita).

4. CONCLUSÕES

Este trabalho propôs a aplicação da arquitetura TimesNet para a previsão de longo prazo de GHI. Além disso, comparou modelos consolidados na literatura como XGBoost, GRU e LSTM, frente à TimesNet na previsão de longo prazo de GHI mensal. Esta análise inédita do desempenho preditivo do TimesNet visou avaliar se a arquitetura bidimensional do TimesNet, somada à busca automática de periodicidades, traria desempenho superior ao dos modelos mais consolidados, para a previsão de longo prazo de GHI. É importante destacar que a comparação foi conduzida nos dezesseis estádios-sede da Copa do Mundo FIFA de 2026, escolhidos por reunirem grande consumo, amplas coberturas e uma diversidade climática que serviu para testar a robustez dos modelos. Na análise estabelecida, o TimesNet obteve o melhor desempenho médio em todas as métricas e foi o modelo mais consistente entre as localidades, vencendo na maioria delas, seguido por LSTM, GRU e XGBoost, este último sem liderar em nenhuma.

A vantagem do TimesNet quando comparado com XGBoost mostrou-se relevante. Nota-se que a arquitetura do TimesNet é adequada para a previsão de longo prazo, mas o ganho sobre as recorrentes ainda é modesto. Como perspectiva de trabalhos futuros, pretende-se investigar a aplicação dos modelos a séries temporais de maior resolução temporal, como as diárias e horárias, nas quais coexistem múltiplas periodicidades. Esse cenário pode explorar de forma mais efetiva as características da arquitetura TimesNet, concebida para identificar simultaneamente diferentes padrões cíclicos.

Agradecimentos

Os autores agradecem à FAPEMIG (grant n OET-00243-26). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERENCES

- Cabral, T. W. et al. (2024), “Analysis of variance combined with optimized gradient boosting machines for enhanced load recognition in home energy management systems”, *Sensors*, Vol. 24, No. 15, p. 4965.
- Chen, T. e Guestrin, C. (2016), “XGBoost: a scalable tree boosting system”, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, 785-794.
- Cho, K. et al. (2014), “Learning phrase representations using RNN encoder-decoder for statistical machine translation”, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, ACL, Stroudsburg, 1724-1734.
- Géron, A. (2022), *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*, 3ª ed., O'Reilly Media, Sebastopol.
- Hiremath, S. V. et al. (2024), “GHI prediction based on weather data using machine learning”, *Proceedings of the 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*, IEEE, pp. 1–6.
- Hochreiter, S. e Schmidhuber, J. (1997), “Long short-term memory”, *Neural Computation*, vol. 9, nº 8, 1735-1780.
- International Renewable Energy Agency (IRENA) (2025a), *Renewable capacity statistics 2025*, IRENA, Abu Dhabi.
- International Renewable Energy Agency (IRENA) (2025b), *Renewable power generation costs in 2024*, IRENA, Abu Dhabi.
- Li, M. et al. (2026), “An enhanced monthly GHI prediction framework for China using a hybrid dynamical-machine learning approach”, *Theoretical and Applied Climatology*, Vol. 157, No. 2, p. 120.
- Wu, H. et al. (2023), “TimesNet: temporal 2D-variation modeling for general time series analysis”, *Proceedings of the 11th International Conference on Learning Representations (ICLR)*, OpenReview, Kigali.