# Neural Networks for Anomaly Detection in Time Series Data

Final Paper

Doga Deniz Ates

10.07.2025

**Abstract**

Neural networks have become a promising approach for anomaly detection in time series data, especially in industrial environments where complex dependencies and nonlinear behaviors are common. TimeGPT-1 utilizes contextual forecasting for anomaly detection over a long series of data, while the LSTM Autoencoder identifies local anomalies using the reconstruction error. Sensitivity detection was an issue with a dynamic thresholding approach on the LSTM model outputs. The findings demonstrate trade-offs between interpretability and sensitivity, where TimeGPT-1 allows for larger contextual understanding while the LSTM Autoencoder captures finer-grained shifts. Findings suggest that model choice should depend on the nature of the monitored system and the cost of false positives or undetected anomalies.

## 1 Introduction

Anomaly detection in time series data is critical in industrial control systems, as undetected anomalies may signify equipment failure, cyberattacks, or unsafe operating conditions. Given that industrial control systems are increasingly digitized and sensor-based, monitoring complex, multivariate time series processes in real time has become crucial.

Traditional anomaly detection methods, typically based on rules or statistical techniques, although interpretable and efficient, often have difficulty representing non-linear relationships or temporal dependencies of real-world sensor data. Additional deep learning informed, specifically sequence models such as LSTM (Long Short-Term Memory) and AutoEncoders provide advanced methods of learning normal behavior from raw time series data. These methods can also detect anomalies which may be subtle or delayed by contextualizing the temporal components of expected behavior and modelling the expected behavior of the signal analysis process.

More recently, transformer-based architectures originally developed for natural language processing have been adapted for time series tasks. Among these, TimeGPT-1[3] is a notable example, a general-purpose time series model that uses a pre-trained transformer to forecast and interpret sequential data across domains. TimeGPT-1 is not traditional in that it's trained per dataset, it's generalization induced across many time series types provides a flexible workflow for predictions and anomaly detection. Zero-shot and few-shot inference options make TimeGPT-1 very appealing for industrial-like applications, where labeled anomaly data is often scarce.

In this study, the application of neural network-based approaches for detecting anomalies in industrial control data is examined. Using the Intel Berkeley Research Lab Sensor Datase[6], a multivariate dataset collected from a physical testbed under both normal and attack conditions, LSTM AutoEncoder based anomaly detection with a transformer-based alternative using TimeGPT-1 is implemented and compared. The objective is to evaluate their practical effectiveness in identifying anomalies in a real-world, multivariable industrial process.

# 2 Related Work

Time series modeling has been an area of research in many different fields for a long time, from finance, healthcare, to industrial systems. Historically, the two widely used approaches were based on statistical models (e.g. ARIMA, Exponential Smoothing). Traditional statistical models are widely applicable to stationary and low noisy data but time series data is often much more complex, due to multivariate or non-linear patterns.With the growing volume and diversity of time series data, especially in sensor-rich industrial systems, recent research has increasingly focused on data-driven models based on neural networks and transformers.

Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) architectures and AutoEncoders, have been extensively used for time series anomaly detection. These models can exist with sequential dependencies and can represent expected behavior in time series data by reconstructing or predicting a time series. If there is high reconstruction or prediction error this suggests a potential anomaly. As anomalies typically occur in datasets that consist of multiple variables and their appropriate interactions in time, LSTM AutoEncoders have been effective in datasets such as Intel Berkeley Research Lab Sensor Dataset where multivariate sensor readings and cyber-physical interactions demand models that can generalize across time and variable interactions.

More recently, transformer-based models originally designed for natural language processing have been adapted for time series forecasting and anomaly detection. One such model, Wu et al. TimesNet[7], introduces a novel temporal 2D variation modeling approach to capture local and global temporal patterns across different timescales. TimesNet shows strong generalization performance across multiple complex domains and demonstrating improved forecasting accuracy and generalizability over standard attention mechanisms.

Liu et al.[4] perform inversion on the transformer architecture, such that instead of each time step being a token, the full temporal sequence for a variable is the token. Inversion improves the efficiency of attention, and permits the model to represent variable-level dependencies an important dimension of multivariate industrial systems like SWaT.

Zhou et al.[1] propose an even more boundary pushing approach, a prompt-based pretraining framework for time series. Initialized by the advancements of prompting with NLP, TEMPO pulls it concepts one-step further. More specifically, task specific prompts are available during inference, enabling generalization across multiple types and formats of time series. Subsequently, a single pre-trained model could be adapted for forecasting, imputation, or anomaly detection tasks without extensive fine-tuning. This characteristic is particularly useful within real world sensor environments, where labeled data is often sparse.

These recent pieces of literature demonstrate the initial shift away from hand-crafted or fixed modeling approaches towards flexible, general-purpose architectures for understanding time series. For industrial anomaly detection, transformer-based models can be quite valid when compared to traditional sequence models, specifically when dealing with high-dimensional, multivariate data. This study considered the use of transformer-based neural architectures, including TimeGPT-1, to detect anomalies in industrial control systems using the Intel Berkeley Research Lab Sensor dataset.

# 3 Methodology

This study compared two distinct deep learning models for time series anomaly detection: TimeGPT-1, a transformer-based foundation model (as illustrated in Figure 1) that implements attention based forecasting, and the LSTM Autoencoder. The LSTM autoencoder is pre-trained as a reconstruction-based model using LSTM units and temporal encoding. TimeGPT-1 is a large-scale transformer model designed for time series forecasting tasks. It is built on the Transformer encoder-decoder architecture, which utilizes self-attention, residual connections, and layer normalization. Unlike other transformers that are designed for natural language, TimeGPT-1 is capable of working with continuous timestamped multivariate data. Through proprietary positional encodings and variable embeddings, TimeGPT-1 effectively captures the temporal structure and feature-level relationships within the time series data. The authors refer to each window of previous observations as context, and then it creates probabilistic forecasts for future values. TimeGPT-1 was pre-trained on an extremely large and diverse time series dataset with over 100 billion data points across multiple domains. This version of TimeGPT-1 can generalize to unseen series without fine-tuning (zero-shot learning). During inference, TimeGPT-1 detects anomalies by comparing the model predictions to the observed values of the time series and examining the magnitude of the forecast errors. The forward residuals of the model with a certain defined threshold indicates potential anomaly detection. This capacity to detect anomalies, irrespective of the number of labeled anomalies available for supervised learning, and with retraining not always a feasible option, makes TimeGPT-1 suitable for some applications where there are limited labeled anomalies.

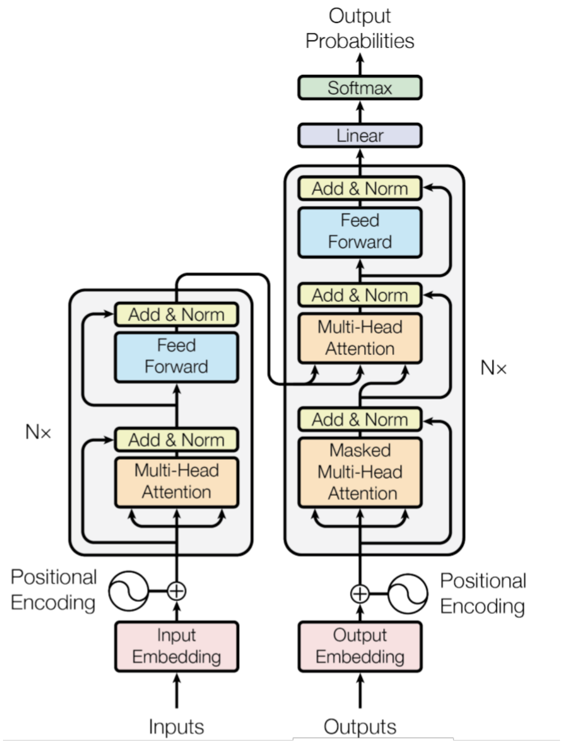


Figure 1: **Architecture of the TimeGPT-1 model adapted for time series forecasting.** Based on a Transformer encoder-decoder structure, the model utilizes multi-head attention, convolutional layers, and positional encoding to process continuous time series data. The model generates future values in an autoregressive manner while anomaly detection consists of comparing forecasted values to the observed values.

In contrast, the LSTM Autoencoder (Figure2), is both a reconstruction-based architecture that contains two sequential components, the encoder and the decoder, each consists of LSTM units. The first part of the LSTM Autoencoder is the encoder, which is responsible for capturing the temporal dependencies from the input time series and storing its output in a fixed-size latent representation in order to allow the second part of the LSTM Autoencoder, the decoder, to reconstruct the input. The training is based on a minimization reconstruction loss from normal (non-anomaly) data. During inference, the reconstruction error will be described as an anomaly score, based on the difference between the input and its reconstruction.Sequences with high reconstruction error are considered anomalous, under the assumption that the model can reconstruct normal patterns well but struggles with unfamiliar or rare patterns. In this manner, the LSTM Autoencoder has additional utility in the detection of finite and contained anomalies in otherwise regular or moderately complicated time series data [5].

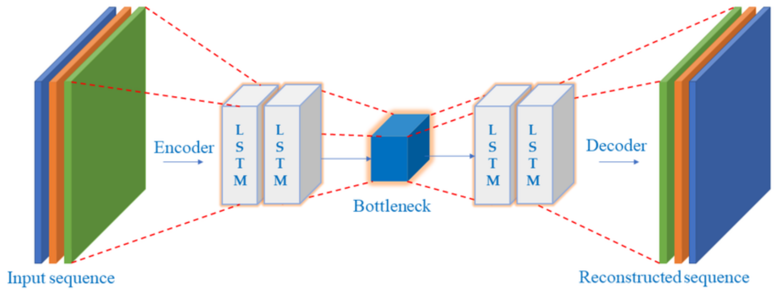


Figure 2: **Architecture of the LSTM Autoencoder model for anomaly detection.** The input time series is first passed through an encoder consisting of LSTM layers that generate a fixed-length latent representation (bottleneck). The decoder subsequently reconstructs the original input from this latent representation. The reconstruction error then serves as an anomaly score during inference[2].

By comparing these two architectures, the study aims to evaluate how well each model captures temporal dependencies and generalizes to complex real-world time series data under unsupervised settings.

Additionally, this study integrates two thresholding techniques to convert the reconstruction errors obtained from the LSTM Autoencoder into discrete anomaly labels. The first approach applies a static threshold, which is determined by calculating the mean and standard deviation of the reconstruction error on the validation data and setting a fixed threshold. While this method is simple and effective in detecting pronounced anomalies, it may overlook more nuanced deviations, especially in non-stationary data.

To improve anomaly detection, dynamic thresholding is applied to the reconstruction error of the LSTM Autoencoder. This technique generates a threshold that is time-variant using a rolling mean and standard deviation over a fixed window, enabling the model to adapt to local fluctuations and gradual changes in the error distribution. Unlike a static threshold, which assumes a stationary error profile, the dynamic threshold approach is more capable of detecting changes to a non-stationary time series while still being sensitive to context-specific deviations.

# 4 Results and Discussion

To compare the two methods, TimeGPT-1 and the LSTM Autoencoder, for the anomaly detection tasks, both models were run on both voltage and temperature sensor data, and the status of anomalies detected was reviewed.

The LSTM Autoencoder, as shown in Figure 3, relies on reconstruction error to identify anomalies. This method is particularly sensitive to sudden and local changes in patterns over time. Consequently, the model captures numerous small-scale fluctuations, resulting in a higher number of detected anomalies. In practical terms, sensitivity is useful in some situations; however, for these purposes, it makes it difficult to determine whether an observed anomaly is of actual importance or simply a minor fluctuation in the data.

On the contrary, TimeGPT-1, illustrated in Figure 4, exhibits stronger characteristics by focusing on broader contextual patterns. TimeGPT-1, with a transformer-based architecture trained on a large and diverse time series corpus, is able to model long-duration dependencies and trend dynamics more effectively. It discards short-term noise and detects deviations with more substance, leading to a resulting anomaly profile that is more conservative and interpretable. Rather than highly reacting to minor voltage drops, TimeGPT-1 is able to identify key points at which the time series behavior is meaningfully divergent from its predicted behavior.

To further refine the performance of the LSTM Autoencoder, a dynamic thresholding strategy is applied (Figure 5) based on a rolling average and standard deviation of the reconstruction error. This allows the threshold to change over time, making it more sensitive to the small changes and drifts in the data. The model is able to detect a lot more local fluctuations than would be possible with a fixed threshold.

In summary, this comparison demonstrates a trade-off between sensitivity and interpretability. The LSTM Autoencoder may be unsuited for scenarios where even slight deviations are critical, particularly when combined with dynamic thresholding, while TimeGPT1 provides a high-level abstraction that reduces false positives while still emphasizing major trend anomalies.

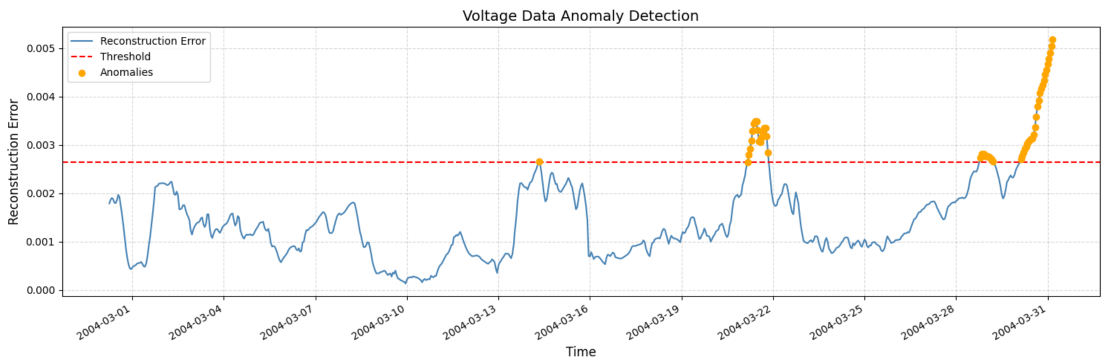


Figure 3: LSTM Autoencoder anomaly detection result in voltage sensor data

In the temperature anomaly detection results (Figure 6 for TimeGPT-1, Figure 7 for LSTM Autoencoder), both models successfully identify the major anomaly corresponding to the sudden and sharp increase in temperature. However, their detection behavior differs: while the LSTM Autoencoder marks the anomaly more locally and tightly around the spike through reconstruction error (Figure 6), TimeGPT-1 captures a broader segment of the shifted region as anomalous (Figure 7). This suggests that TimeGPT-1 is more sensitive to long-term distributional changes, whereas the LSTM Autoencoder reacts more to abrupt, short-term deviations. This distinction highlights TimeGPT-1's strength in contextual understanding over extended sequences.

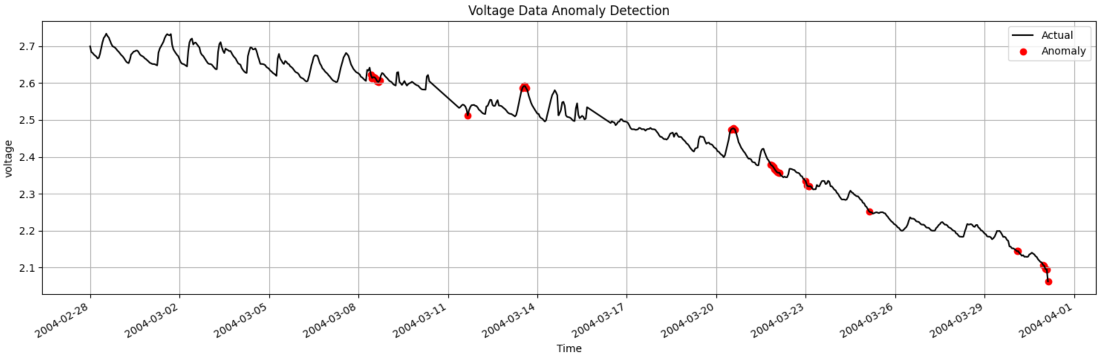


Figure 4: TimeGPT-1 anomaly detection result in voltage sensor data

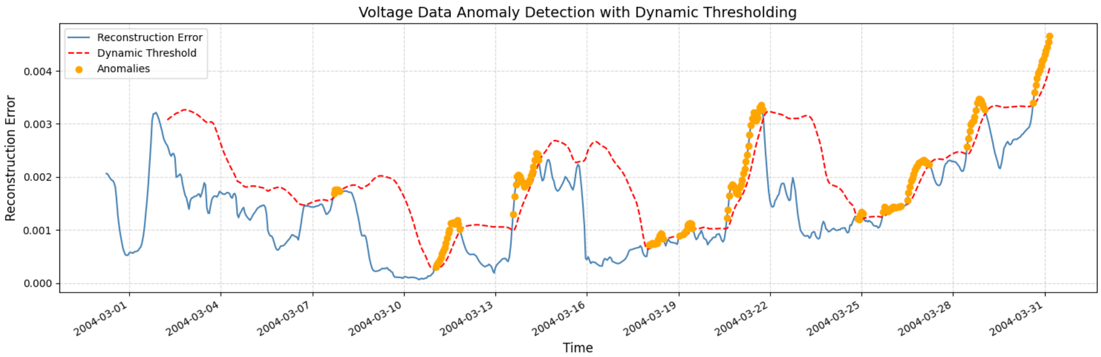


Figure 5: LTSM Autoencoder anomaly detection with dynamic thresholding result in voltage sensor data

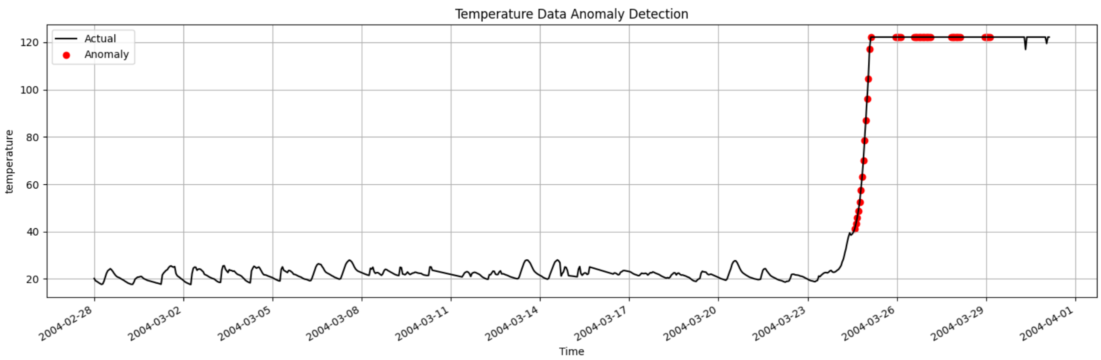


Figure 6: TimeGPT-1 anomaly detection result in temperature sensor data

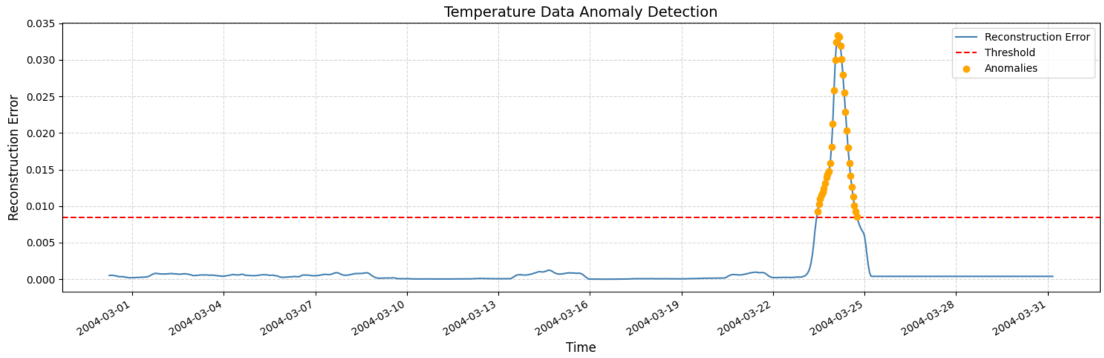


Figure 7: LSTM Autoencoder anomaly detection result in temperature sensor data

To further refine anomaly detection with the autoencoder, dynamic thresholding was additionally applied based on a rolling mean and standard deviation over the reconstruction error (Figure 8). This means it can use local variations of reconstruction error to adjust the threshold over time to identify not only the strong anomaly peak, but also nearby deviations in relation to the reconstruction error that may be lost if a static threshold were to be used. With dynamic thresholding, there is more sensitivity compared to the fixed MSE threshold with dynamic thresholding also being better suited under conditions of gradual drift and/or variable noise. Dynamic thresholding may result in a greater number of anomalies detected as well as more context-dependent interpretation compared to a static threshold that may miss important deviation effects.

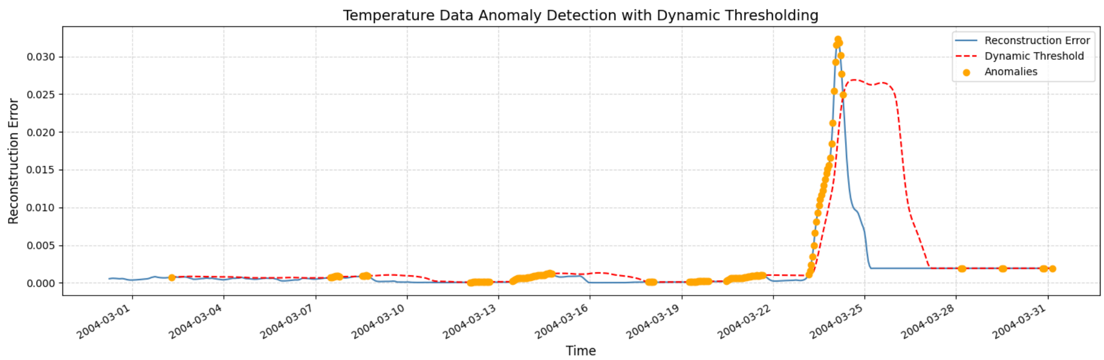


Figure 8: LTSM Autoencoder anomaly detection with dynamic thresholdig result in temperature sensor data

# 5 Conclusion

This study explored the effectiveness of deep learning approaches in detecting anomalies within multivariate industrial time series data. Both methods were applied and assessed via temperature and voltage sensor readings from the Intel

Berkeley Research Lab Sensor Dataset, which simulates a cyber-physical in the real world.

This results determined that both methods could demonstrate that both methods have both worked on identifying significant anomalies but each has unique characteristics associated with detection. The first method, the LSTM Autoencoder is much better at reconstructing errors to identify significant short-term changes in the anomaly which was more than likely a sudden change and is much more sensitive to local changes. Through dynamic thresholding, the LSTM Autoencoder can be emplaced to identify subtle or significant changes in anomalies when static thresholds would not have detected them. The ability to identify fine changes in anomalies also comes at the risk of more false positives, which is something to consider in reviewing actual practical applications.

On the other hand, TimeGPT-1 offers a more robust and generalized understanding of time series behavior. Long-term dependency modeling with time series helps us to filter out the transient noise of individual observations and detect more meaningful, context-driven deviations. Therefore, TimeGPT-1 is particularly useful in industrial contexts, where interpretability and limited false-alarm rates are vital.

Depending on the application's tolerance for noise and the cost and importance of missed anomalies, hybrid approach could be applied for anomaly detection. Future work could include refining these models with ensemble approaches or implementing them in a real-time deployment context to enhance robustness and usability.

# References


[1] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*, 2023.

[2] Jae Do, Akeem Bayo Kareem, and Jang-Wook Hur. Lstm-autoencoder for vibration anomaly detection in vertical carousel storage and retrieval system (vcsrs). *Sensors*, 23:1009, 01 2023.

[3] Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. Timegpt-1. *arXiv preprint arXiv:2310.03589*, 2023.

[4] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

[5] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*, 2016.

[6] MIT Computer Science and Artificial Intelligence Laboratory. Labsensor data, 2004. Accessed: 2025-07-10.

[7] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.