Attention-Augmented LSTM Autoencoders for Improved Anomaly Detection in Time-Series Data

MSc Research Project

Master of Science in Data Analytics

Karthik Shankar Naik

Student ID: X23242701

School of Computing
National College of Ireland

Supervisor: Hamilton V. Niculescu

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Karthik Shankar Naik		
Student ID:	X23242701		
Programme:	Master of Science in Data Analytics	**Year:**	2024-2025
Module:	MSc (Research)Practicum Part 2		
Supervisor:	Hamilton V. Niculescu		
Submission Due Date:	11th August 2025		
Project Title:	Attention-Augmented LSTM Autoencoders for Improved Anomaly Detection in Time-Series Data		
Word Count:	7757	**Page Count**	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Karthik Shankar Naik

Date: 11th August 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Attention-Augmented LSTM Autoencoders for Improved Anomaly Detection in Time-Series Data

Karthik Shankar Naik

X23242701

Abstract

Detecting anomalies in time-series data is critical in many domains, including transportation, healthcare, and cybersecurity. Traditional autoencoders often rely solely on the final hidden state to reconstruct input sequences, limiting their ability to focus on subtle, localized deviations. This study leverages an attention-enhanced LSTM autoencoder that incorporates a temporal attention mechanism to dynamically weigh the importance of each timestep during reconstruction, thereby improving the model's sensitivity to anomalies, using a time series dataset and compared the leveraged model against baseline deep learning architectures including RNN, GRU, and vanilla LSTM autoencoders. Experimental results show that our model significantly outperforms these baselines, achieving a macro-averaged F1-score of 0.82 and an anomaly-specific F1-score of 0.66. These improvements highlight the benefit of integrating attention mechanisms into sequence modelling for more accurate anomaly detection. The findings suggest strong potential for the model's application in real-time monitoring systems. It also discusses limitations such as computational overhead and dataset specificity, and proposes future directions, including expansion to multivariate datasets and lightweight deployment on edge devices.

Keywords: Time-Series Anomaly Detection, LSTM Autoencoder, Attention Mechanism, Deep Learning, Temporal Modelling

1 Introduction

1.1 Problem Background

Anomaly detection on time-series is a vital task in many important fields, such as industrial equipment monitoring, medical diagnosis, fraud detection in financial services, and cybersecurity. In such areas, anomalous behaviors that occur in the real time can be detected and thus catastrophic failures will be avoided, operational risks will be minimized, and decision-making will be improved. The time-series anomalies are normally scattered, localized, and subtle, and thus identifying them is a difficult task. Anomaly detection, in contrast to typical classification tasks, is frequently unsupervised and the data to be learned is imbalanced and non-stationary, which makes learning more difficult (Belay et al.,2023).

Statistical thresholds, autoregressive models (such as ARIMA) and distance-based outlier detection methods have been used to address this challenge using traditional methods. Although these methods are easy to understand and interpret, they are very noise-sensitive, are restricted in their ability to learn complex temporal dependencies and usually necessitate stationarity assumptions that are not met in practice by real-world time-series data. These classical models often fail or do not perform well in new applications, where the data may be high-dimensional, non-linear, and with multi-scale dependencies.

1.2 Drawbacks and Limitations of Existing Works

To overcome the shortcomings of traditional approaches researchers have increasingly resorted to machine learning and deep learning-based anomaly detection in time-series data. Statistical methods are not as robust as classical machine learning models, such as Isolation Forests, Support Vector Machines, and ensemble methods. Nevertheless, these models remain dependent on engineered features and are not necessarily effective to model long-term dependencies by themselves-which is still lacking when it comes to anomalies that take place over time or on variable length contexts (Nguyen et al., 2021; Belay et al., 2023).
Deep learning techniques, especially autoencoders based on recurrent neural networks (RNNs), GRUs, and LSTMs, have been promising in learning compressed latent representations of regular temporal behavior. Among them, LSTM-based autoencoders have proven particularly effective due to their ability to capture long-term dependencies. Despite this, most LSTM autoencoder frameworks apply equal importance to every time step during encoding and decoding. This uniform attention disregards the contextual significance of certain time points, especially when anomalies manifest as subtle deviations embedded within long sequences. Furthermore, while Transformer-based models resolve the problem of uniform temporal weighting through self-attention mechanisms, they are often computationally expensive and require substantial labeled data and tuning, limiting their practicality in real-time, resource-constrained environments (Kim et al., 2024; Lin et al., 2025).

1.3 Research Question

Given the limitations of both classical and advanced deep learning-based approaches, this study aims to answer the following research question:

Can the integration of attention mechanisms into LSTM autoencoders improve the accuracy and efficiency of time-series anomaly detection?

This research question arises from the need to enhance the sensitivity of LSTM-based models to critical time steps while preserving the sequential modeling strengths inherent in their architecture. By embedding attention modules into LSTM autoencoders, we hypothesize that the model will be able to learn which time points contribute most to the reconstruction errors enabling more precise detection of anomalous patterns, especially in complex and evolving data streams.

This question also implicitly compares the proposed model not only with baseline LSTM, RNN, and GRU autoencoders but also with Transformer-based architectures. It seeks to determine whether the proposed hybrid approach can retain or exceed the detection performance of Transformers, while significantly reducing computational burden and training data requirements—two persistent challenges in deploying advanced models in real-world systems.

1.4 Justification

Incorporating attention mechanisms in LSTM autoencoders is a potential and unexplored avenue that can fill major gaps in existing models of anomaly detection. The attention feature allows the model to give time steps their due weight according to their contextual significance instead of treating them equally. It is especially handy when only a small number of time steps within a large series are relevant to the anomaly. Attention-augmented LSTMs allow dynamic attention to focus on important features, and therefore, are more sensitive to anomalies and have fewer false positives. Besides, the overhead incurred by attention modules is much less than the complexity of full Transformer architectures, and thus such modules are more appropriate in online or real-time scenarios.

Such a hybrid model is especially suitable in an environment where computational resources are scarce, e.g., embedded systems in industrial Internet of Things, portable health monitoring, or high-frequency financial systems, but where reliability and explainability are required. Attention-based LSTM autoencoders offer an alternative to Transformers, which can be computationally expensive to use due to the need to use large-scale hardware accelerators and fine-tune them extensively. This study also leverages the advantages of the known architectures and provides a solution to a trade-off that has existed since the beginning of time-series anomaly detection: between accuracy and efficiency.

2 Related Work

2.1 Machine Learning-Based Approaches

Classical machine learning has been used traditionally in time-series anomaly detection because of its simplicity, interpretability, and ease of implementation. Isolation Forest and Local Outlier Factor (LOF) are popular models that are applied to detect fraud in finance and, in IoT, diagnose faults. These models work well when there is relatively fixed data distribution. In practice, however, when time-series data are highly dynamic, non-stationary and usually noisy, their generalization capabilities become much worse. According to Belay et al. (2023), the models are especially ineffective in terms of subtle, sporadic, or time-dependent anomalies, and they are characterized by a large false positive rate or false negative rate.

In more domain-specific tasks such as industrial automation or environmental monitoring, Support Vector Machines (SVM), decision tree, and random forest machine learning models have been used to classify anomalies using handcrafted features. Although they provide sensible performance when the data are structured and well preprocessed, the need to rely on manual feature extraction hinders their flexibility. Nguyen et al. (2021) also highlighted that these models do not represent long-range temporal dependencies and changing data distributions, and thus they are not enough to handle complicated multivariate time-series where anomalies may not be in the form of sharp spikes but trends or gradual changes.

In response to the drawbacks of individual models, ensemble learning techniques have been studied to combine many base learners in an effort to increase robustness and generalization. These consist of stacking, boosting or bagging methods, which combine the merits of a variety of anomaly detectors. Iqbal et al. (2024) showed that ensemble methods, especially deep ensembles, improve the results on multivariate time-series anomaly detection. The drawback is, however, that this results in more computational overhead and complexity. In addition, ensemble models continue to share the inherent weakness of the inability to explicitly model time relationships in sequential data. This highlights a major shortcoming, in that there has been an increasing desire to develop methods that not only generalize over the various types of anomalies but also inherently capture temporal structure and dependencies, without undue resources overhead.

2.2 Deep Learning Autoencoder Models

With their ability to automatically find complex temporal and spatial patterns in data, deep learning methods have risen to prominence in time-series anomaly detection. One of these, autoencoder-based architectures, have been popularly used due to their unsupervised learning, which is especially useful when there is limited labeled anomalous data. Autoencoders are trained to represent the normal patterns in the data in a compressed form and detect anomalies by the reconstruction error. Recurrent designs like RNN and GRU autoencoders introduce time consciousness and are thus more appropriate than their non-temporal feedforward counterparts in analyzing time series. Yet, these models are also subject to serious problems. Standard RNNs as Provotar et al. (2019) mentioned are susceptible to the vanishing-gradient problem and can hardly memorize long sequences, a factor that compromises their anomaly-detection ability based on long-term temporal dependencies.

As a remedy to some of them, Gated Recurrent Units (GRUs) have been considered as an alternative. GRUs simplify the LSTMs gating, lessening the computational strain, but retaining some memory. Nguyen et al. (2021) demonstrated that GRU-based autoencoders can be adequate to use in such tasks as supply chain anomaly forecasting and offer a trade-off between speed and predictive power during training. Nevertheless, GRU themselves are inadequate when the patterns of anomalies are mostly dominated by long-term dependencies, and its more lightweight structure sometimes fails to learn more intricate non-linear temporal dependencies. By contrast, LSTM autoencoders have gained traction to be the most common choice to identify anomalies in time-series in a high-dimensional and sequential scenario. Their memory cell structure allows them to remember information over long sequences, which makes them many times more effective than the classic RNNs and GRUs in most applications, such as healthcare monitoring and power electronics. To cite one example, Liu et al. (2022) have applied LSTM autoencoders to ECG data to identify arrhythmia and managed to distinguish the slight irregularities in the cardiac activity. In a similar manner, Radaideh et al. (2022) proved the efficiency of LSTM-based models in identifying operational errors in power electronics systems by rebuilding normal signal patterns with high fidelity. Nevertheless, LSTM autoencoders traditionally treat all time steps in the same way during both encoding and decoding, regardless of the presence of a sequential dependency. Such undifferentiated attention to more informative or abnormal time steps may result in sensitivity loss in identifying subtle anomalies. Therefore, one promising way forward is to augment these models with some mechanism (including attention) that allows them to selectively focus on important components of the time series, which would allow them to be both more accurate and more interpretable in their anomaly detection.

2.3 Attention and Transformer-Based Architectures

Attention mechanisms have marked a paradigm shift in sequence modeling introducing the capacity to dynamically assign different weight to various time steps in the learning process. Unlike the traditional sequential models, which process the input in an even or fixed-length window of context, or attention-based models are able to selectively attend to temporally distant information that is related in context. The ability is especially effective when detecting anomalies in a time-series, where anomalies can occur in different time scales, and are not necessarily localized. Transformer models, which are based on self-attention, demonstrate this innovation as they can model both short-term and long-term dependencies, without recurrence. Among the most prominent uses of Transformer architectures, one can mention the Zhang and Luo (2025) study that proposed a decomposition-based multi-scale Transformer framework. Their model decomposes time-series data into the various time resolutions, which can enable its ability to isolate anomalies that are at various frequency bands. Such decomposition increases granularity and accuracy of the model in identifying sudden and gradual aberrations in complex multivariate data. In like manner, Kim et al. (2024) proved the strength of hybrid models, which combine Transformer encoders with time-series image representations with multi-channel Gramian Angular Difference Fields (GADF). Their system converts time-based data into spatial pictures, which capture complex spatial-temporal correlations, and the results are impressive in industrial anomaly detection activities like railway HVAC systems monitoring.

Transformer-based models have significant trade-offs in spite of their high accuracy and flexibility. Such models are generally characterized by millions of parameters and consume large amounts of computational resources both during training and inference. Similar to what has been noted by Lee et al. (2024) and Geiger et al. (2020), self-attention mechanisms require a significant amount of matrix operations, which translates to high memory and processing requirements, thus limiting real-time deployment in edge or IoT settings. In addition, Lin et al. (2025) stated that such models will probably require large quantities of training data and hyperparameter tuning to perform well, which is typically infeasible in practice unsupervised anomaly detection settings where labeled anomalies are minimal or nonexistent. Thus, despite the theoretical advantages of Transformers in modeling complex dependencies, they are rather complex in operation, which prevents their use in conditions where computational efficiency, interpretability, and minimal supervision are required.

2.4 Research Gaps and Proposed Solution

Despite the recent successes of deep learning-based anomaly detection models such as LSTM autoencoders, which have greatly improved the status quo of time-series anomaly detection, such models are still restricted by their fixed attention to every time step. This shortcoming can be addressed by directly introducing attention mechanisms into LSTM autoencoders, such that the model can focus on the most relevant time steps during reconstruction (Wei et al., 2023; Lee et al., 2024). Such selective focus increases the capability of the model to identify contextually significant anomalies that would otherwise be missed.

Attention-augmented LSTM autoencoders are a good trade-off between performance and computational efficiency compared to Transformer-based models. They share the temporal memory benefits of LSTMs but attain higher anomaly detection accuracy, guided attention, and do not require the resources characteristic of Transformers (Kim et al., 2022; Gjorgiev & Gievska, 2020). The proposed model has these features which make it particularly applicable in industrial, healthcare, and financial systems in real-time.

3 Research Methodology

The research takes a methodological research perspective to evaluate the success of the application of attention mechanisms to LSTM autoencoders in the identification of anomalies in time-series. It is organized into six major steps, which are data selection and preprocessing, development of the baseline model, attention-based LSTM autoencoder, training strategy and hyperparameter tuning, performance analysis through standard and robustness measures, and comparison with Transformer-based and classical models. The methodology perfectly corresponds to the previous research practices in the domain of temporal anomaly detection and adheres to reproducible deep learning principles.

3.1 Data Selection and Preprocessing:

To use the NYC Taxi dataset (https://github.com/numenta/NAB) to train a model, the values of the ride count column are initially normalized via the MinMaxScaler to scale all the inputs to the range [0,1]. It is done to make the neural network utilize input values in a similar way regardless of their scale and avoid biasness and achieve faster convergence during training. After normalization the data is divided into overlapping sequences via a sliding window method with each window being a fixed length segment of continuous hourly ride counts. The approach maintains time-dependency in the data and allows the model to learn contextual dependencies over time, which is critical in the detection of subtle or gradual anomalies.

They are then reorganized into a 3D tensor of shape (samples, time_steps,1) which is needed by the recurrent neural networks, e.g., RNNs, GRUs, LSTMs. Notably, the autoencoder is trained on sequences that do not include labeled anomalies (i.e., on sequences containing normal data only) according to the unsupervised learning paradigm. The strategy will enable the model to learn a parsimonious version of typical patterns, so that any deviation in the inference process expressed as excessive reconstruction error can be reported as an anomaly. The time-based indexing and the sequence-wise labeling guarantees temporal continuity of the model and learning with realistic, chronologically ordered data.

3.2 Baseline Model Development:

In order to form a solid basis against which the effectiveness of the proposed attention-augmented LSTM autoencoder could be tested, a number of baseline models are formulated based on the conventional recurrent architectures. They are LSTM autoencoder, GRU autoencoder and RNN autoencoder, which are commonly used in time-series anomaly detection because they can model temporal dependencies. All the baseline models have a symmetrical encoder-decoder architecture. The encoder maps sequential input data to a latent representation, and the goal of the decoder is to reproduce the original sequence using the latent representation. The main idea of anomaly detection in these models is that anomalous sequences will have a higher reconstruction error because the model has been trained on mostly normal patterns and cannot easily reproduce unseen/anomalous behaviour. This reconstruction error, usually determined by the Mean Squared Error (MSE), is then used to provide anomaly scores to each of the sequences or time step.

These baseline architectures can be meaningfully and controlled compared with the attention-based LSTM autoencoder. By ensuring uniformity in input processing, training procedures and evaluation criteria of all models, the research isolates the influence of incorporating attention mechanisms on the performance of anomaly detection. Additionally, the use of various recurrent cells (LSTM, GRU and vanilla RNN) makes it possible to investigate the extent to which some architectures are more sensitive or efficient to detect anomalies under different levels of sequence complexity. Whereas RNNs can have problems with the vanishing gradient in long sequences, LSTMs and GRUs are better at handling long-term dependencies, although with different complexity requirements. An analysis of these baselines gives answers on the strengths and weaknesses of each one in the reconstruction of time-series information and anomaly detection. Above all, such a comparative configuration will aid in the evaluation of whether the suggested attention mechanism can improve model accuracy and robustness without creating excessive overhead, especially when used in real-time deployment or when run on resource-limited systems.

3.3 Design of Attention-Enhanced LSTM Autoencoder:

The main idea of the innovation of this study is the architectural improvement of the conventional LSTM autoencoder with the addition of the soft attention mechanism. All steps in a sequence are treated equally during encoding and decoding by the traditional LSTM autoencoders and this can smooth out the sensitivity of the model to subtle yet vital anomalies that are localized in certain parts of the input sequence. To address this shortcoming, this paper introduces a soft attention block between the encoder and decoder parts so that the model can add learned importance weights to every hidden state generated by the encoder. These attention weights indicate the importance of each time step in general structure of the sequence and possible anomalies.

The encoder is designed architecturally as a stack of several LSTM layers that take input sequence and output a sequence of hidden states, one per time step. The attention mechanism is used on all of the hidden states of the encoder rather than being applied to one final hidden state and passed into the decoder. The soft attention mechanism computes a score of each time step in its relevance to the sequence context and then takes a weighted sum of these hidden states- giving a context vector that highlights the most informative portions of the sequence. This context vector weighted by attention is then fed to the decoder which tries to reconstruct the original sequence with better emphasis on important temporal patterns. In this way, the model increases its capacity to identify minor deviations or anomalies, especially in the sequences where only a limited number of time steps are anomalous. Such focused reconstruction does not only increase the performance of anomaly detection, but it also improves model interpretability since it makes the decision process more transparent due to visualizable attention weights.

3.4 Training Strategy and Hyperparameter Optimization:

The Adam optimizer is an adaptive learning rate optimization algorithm that has become very popular due to its effectiveness in training models with sparse gradients and non-stationary objectives and thus all the models in this study are trained using the Adam optimizer. Early stopping is also applied to the model to avoid overfitting and ensure generalization; validation loss is used to stop the training when the performance on new data starts to decline. The most common loss function is Mean Squared Error (MSE) that estimates the accuracy of reconstruction by calculating the difference between the original and reconstructed sequences. Important hyperparameters such as LSTM units, sliding window size, dimensionality of attention vectors, learning rate and batch size are optimized using a blend of grid search and Bayesian optimization within k-fold cross-validation scenarios to find stable combinations. Regularization measures like dropout layers and L2 weight decay are also introduced to reduce overfitting especially in deep models that are highly capable. All the experiments are performed on a GPU-based computing platform with TensorFlow and PyTorch that have the flexibility and computational performance needed to train complex models of deep learning on large-scale time-series data sets.

3.5 Comparative Analysis and Model Efficiency:

Lastly, in order to have a thorough evaluation of the performance of the proposed attention-based LSTM autoencoder, a benchmarking exercise is conducted against two major classes of models; the traditional recurrent models (such as LSTM, GRU, and RNN autoencoders). This comparative report is organized in terms of a multi-dimensional assessment framework. Instead of only looking at accuracy, the paper systematically examines the various aspects of performance such as model complexity (as parameter count), inference speed, training time, and memory consumption. Such measures are especially important to real-time, or resource-constrained applications, where detection performance may not be the only factor of importance when compared to computational efficiency. All models are trained and tested in the same experimental conditions by using the same datasets, preprocessing, and evaluation criteria. This makes the comparison fair and enables us to identify trade-offs between model performance and resource requirements. In this process, the study aims to find out the effectiveness of the proposed model in terms of balancing between the efficiency of computations and the ability to detect anomalies, but without making a conclusion on the results yet.

4 Design Specification

4.1 Overview of the Proposed Architecture

The Attention-Augmented LSTM Autoencoder (Attn-LSTM-AE) proposed is intended to improve time-series anomaly detection by incorporating a soft attention mechanism into a standard encoder-decoder framework. The model takes sequential input in the form of a sequence of modular stages: data preprocessing, LSTM-based encoding, attention-weighted decoding, anomaly scoring through reconstruction error, and threshold-based classification. The focus is on the trade-off between the accuracy of the detection, its interpretability, and computational cost, which makes the pipeline applicable both to offline and real-time use.

4.2 Key Techniques and Mechanisms

Fundamentally, the model takes a sequence-to-sequence learning approach. The attention layer is used as an intermediary that dynamically assigns weights to hidden states over time so that the decoder can attend to the most significant aspects of the sequence during reconstruction. The design with attention aims to overcome the weaknesses of the traditional models which considered all time steps equally and introduce greater sensitivity to short-term or subtle anomalies without sacrificing long-term temporal modeling due to LSTM layers.

4.3 Model Features and Tools

The model provides time sensitivity in that it uses LSTM to learn sequential relationships and increase interpretability with adaptive attention. It has an unsupervised reconstruction-based architecture that eliminates the requirement of labeled anomalies, which enables domain-agnostic generalization. To implement, TensorFlow and PyTorch are used to streamline deep learning processes; NumPy, Pandas and Scikit-learn to manage and test data; and Matplotlib/Seaborn to portray anomaly trends. Experiment and performance tracking tools such as Weights & Biases are useful.

4.4 System Workflow and Processing Pipeline

The operation pipeline starts with the ingestion and preprocessing of multivariate time-series data, normalization, imputation of missing values and windowing. Every sequence goes through the stages of encoding, attention weighting and decoding. The anomalies are identified by calculating the reconstruction errors via Mean Squared Error and a threshold is calculated with training or validation data. Such modular architecture enables the adaptation to a variety of areas, including energy, health, and industrial monitoring

5 Implementation

5.1 Problem Formulation and Model Justification

This paper will attempt to develop a practical system of identifying anomalies in time series data utilizing deep learning-based autoencoders. It suggests an **LSTM Autoencoder with an attention mechanism** that will enhance sensitivity to temporal anomalies. Conventional recurrent models like LSTM, GRU, or RNN autoencoders do not prioritize time steps, and thus the effect of rare anomalous events may become watered down. The model can dynamically concentrate on the most relevant time steps in a sequence by including attention, and this results in more precise reconstruction and anomaly detection. The rationale of adopting an LSTM-attention fusion is that it is stronger in sequence modelling than pure Transformers, but more interpretable and lightweight than complicated Transformer models.

5.2 Dataset Description

The dataset (https://github.com/numenta/NAB) that was utilized in the present study is the NYC Taxi passenger count dataset that contains hourly numbers of rides in NYC taxis as well as annotated anomalies that indicate unusual occurrences (e.g., holidays, extreme weather). The data set has a column with the value which represents the number of rides per hour and a binary anomaly flag. The dataset is chronologically split into training and testing sets, where the training set (80%) includes sequences to allow the model to learn standard patterns, while the test set (20%) contains both normal and anomalous sequences to evaluate model performance. Time-based indexing ensures the temporal continuity of sequences and preserves real-world temporal dynamics.

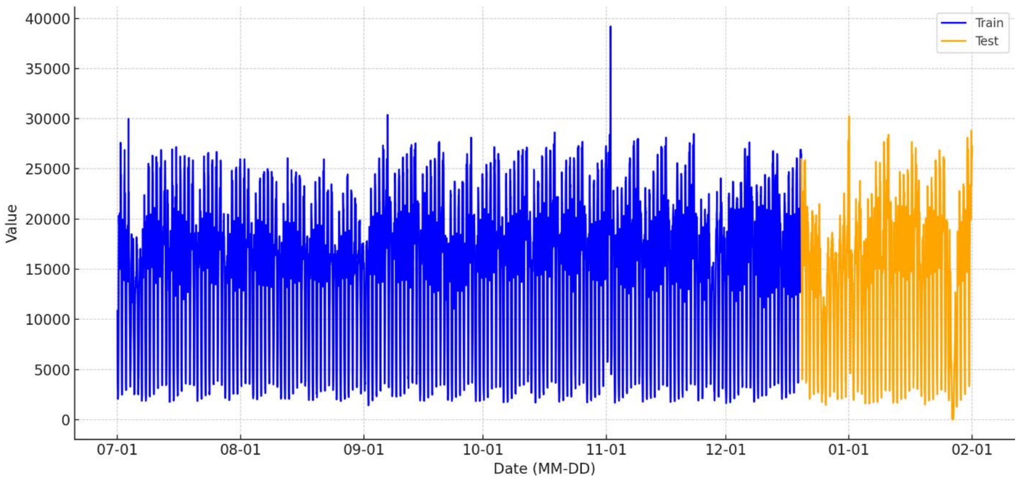

Figure 1. Train and Test Split of NYC Taxi Data

Figure 1 illustrates the temporal distribution of the train and test sets used in this study. The figure presents a time-series plot where the blue curve represents the training data, and the orange curve denotes the testing data. The x-axis is set to represent name of months, thus a clear indication on how the data stretches over a period. The sudden peaks in both parts are the results of abnormal fluctuations in the number of passengers, which are usually associated with large events, holidays (e.g., Thanksgiving Day, New Year), or severe weather conditions. As an example, the apparent peaks in early November and late December are probably the results of an aberration of the Thanksgiving travel and the holiday season. Such outliers are significant indicators that any anomaly detection model ought to identify and make sense of.

5.3 Model Design

The incorporation of attention solves one of the major drawbacks of the conventional LSTM autoencoders where they are unable to localize anomalies because time steps are equally weighted. Attention is a focus mechanism that can be practiced and thus increases precision and clarity. The LSTMs were selected instead of GRUs and vanilla RNNs because they have a better ability to capture long-term dependencies and do not have gradient vanishing problems (Ribeiro et al, 2020). PyTorch and TensorFlow were employed to exploit the flexibility of development and optimized deployment respectively, to ensure that the model is scalable and efficient.

5.4 Model Architecture

The suggested model is an encoder-attention-decoder model, which is founded on the models used in natural language processing. However, it is specifically optimized to the problem of time-series reconstruction and anomaly detection. The model is specifically built to learn more complex temporal dependencies and focus selectively on time steps of interest which are more informative or possibly anomalous.

5.4.1 Encoder

The encoder module is composed of a stack of Long Short-Term Memory (LSTM) layers. It primarily operates to analyze the serial input data i.e. time-series sensor readings, by extracting patterns and time dependencies. At every time step of the sequence, the LSTM layers calculate hidden states which capture the information of the context. These latent states which are a summation of all the temporal steps, are then transferred to the attention layer. This enables the model to learn rich temporal representations instead of just depending on the final state as in the case of traditional autoencoders.

5.4.2 Attention Mechanism

As opposed to traditional LSTM autoencoders, that usually only use the final hidden state of the encoder to create a context to decode with, the proposed model incorporates a soft attention to improve the representation of the sequence. Instead of considering the final time step to contain all important information, the attention mechanism is dynamic in the sense that it determines the significance of each hidden state in the input sequence. It learns to compute how much attention to pay to each time step, that is it learns which elements of the sequence are more informative or anomalous. The weights are deployed in creating a **context vector,** which is a weighted average of all the encoder outputs, consequently allowing the decoder to have access to more discriminative, richer features with which to reconstruct.

The mechanism is especially useful when anomalies are not very pronounced, not regular, or concentrated in parts of the time series. The model prioritizes the most important parts of the input by giving greater weight to more important or unusual time steps, as a result of which it concentrates its reconstruction effort on the most important parts of the input. This restricted attention makes sure that no abnormal trends are washed out or missed as is common when using uniform or fixed-length representations. Consequently, the attention-augmented autoencoder is more capable of detecting small or contextual anomalies which would otherwise be lost in the conventional designs.

5.4.3 Decoder

The decoder will be used to bring back the original input sequence based on the context vector produced by the attention mechanism. It is made up of one or more LSTM layers which successively unfold the condensed contextual data back into a sequence. Dense (fully connected) layer that gives the final reconstructed sequence follows. The decoder is learned to reduce the difference between the input and the reconstructed output. Training the model to correctly reconstruct only patterns observed during training renders the model very sensitive to deviations, and thus effective in anomaly detection.

The Adam optimizer is used to train all models, and early stopping is applied according to the validation loss to avoid overfitting. All the models will utilize the Mean Squared Error as the loss function. Hyperparameters are tuned through a grid search and Bayesian optimization, which includes LSTM units, attention dimensionality, learning rate, batch size, and window size. Model generalization is also achieved by dropout layers and L2 regularization. The training and assessment is done on a GPU-enabled environment, and is computationally efficient and scalable, using TensorFlow and PyTorch, to experiment with deep learning.

Anomaly detection in time-series data with sequences of 30-time steps using the Attention-Based LSTM Autoencoder was trained. The model structure entailed stacked LSTM layers and the soft attention mechanism on the outputs of the encoder that enabled the model to focus more on important time steps. The model was trained on 20 epochs with a batch of 32 and Adam optimizer and mean squared error (MSE) as loss. Training was done with a 10 percent validation split but not shuffled to maintain temporal order. The reconstruction error was used as a basis of anomaly detection and the value of 95 th percentile was selected as a threshold. This attention-augmented model increased the capacity to identify localized and subtle anomalies as opposed to baseline autoencoders.

The model performance is measured by such metrics as precision, recall, F1-score to determine the effectiveness of the model in the imbalanced tasks of anomaly detection. It also uses qualitative analysis by displaying attention weight and overlaying errors to increase interpretability. Thorough benchmarking pits the proposed attention-augmented LSTM autoencoder against baseline models in consistent conditions, in terms of both detection accuracy and computational efficiency on resource-constrained real-time applications.

6 Evaluation

In this section, the proposed attention-based LSTM autoencoder is evaluated in detail against three baseline models: RNN Autoencoder, GRU Autoencoder, Vanilla LSTM Autoencoder. Each of the models was trained and evaluated on the NYC Taxi passenger count dataset and performance was measured by standard classification measures (Macro Average Precision, Recall, and F1-Score) and by dedicated anomaly detection measures (Anomaly Precision, Recall, and F1-Score). These findings form a strong foundation to evaluate the capability of each model to reconstruct the temporal data and detect anomalous patterns, which is important in time-series anomaly detection tasks. The differences in the model performance are evaluated by statistical comparisons with the emphasis on the trade-off between the reconstruction accuracy and the anomaly sensitivity.

Model	**Macro Avg Precision**	**Macro Avg Recall**	**Macro Avg F1-Score**	**Anomaly Precision**	**Anomaly Recall**	**Anomaly F1-Score**
RNN Autoencoder	0.68	0.68	0.68	0.39	0.39	0.39
GRU Autoencoder	0.78	0.78	0.78	0.59	0.59	0.59
Vanilla LSTM	0.72	0.72	0.72	0.47	0.47	0.47
LSTM + Attention	**0.82**	**0.82**	**0.82**	**0.66**	**0.66**	**0.66**

Table 1: Classification Results of Different Autoencoder Models on Test Set (Metrics Higher the better)

Table 1 presents the classification accuracy of four distinct autoencoder-based deep learning models of time series anomaly detection, with attention to both macro-averaged performance measures and anomaly-specific precision, recall, and F1-score. The macro average metrics (precision, recall, and F1-score) are selected to provide an overall view of model performance across all classes, particularly helpful in imbalanced datasets like this one. However, since anomaly detection inherently prioritizes the accurate identification of rare events, and also report anomaly-specific metrics to evaluate how well each model isolates abnormal patterns. These metrics give insight into how many actual anomalies the model captures (recall), how many predicted anomalies are truly anomalous (precision), and their harmonic balance (F1-score). Such granularity enables a more targeted comparison when selecting the best model for real-world deployment, where false negatives (missed anomalies) or false positives can carry significant operational consequences.

6.1 RNN Autoencoder Performance

The RNN Autoencoder served as the simplest baseline in this study. Its macro-averaged scores were modest: a Precision, Recall, and F1-Score of **0.68** each. However, when evaluated specifically for anomaly detection, its performance significantly dropped, yielding **0.39** for Precision, Recall, and F1-Score. These results suggest that while the RNN Autoencoder can moderately reconstruct the general data pattern, it lacks sensitivity in highlighting anomalous sequences as show in figure 2. This limitation likely stems from the RNN's inability to effectively capture long-term dependencies and complex temporal structures, making it less suitable for subtle anomaly detection in noisy real-world time-series data.

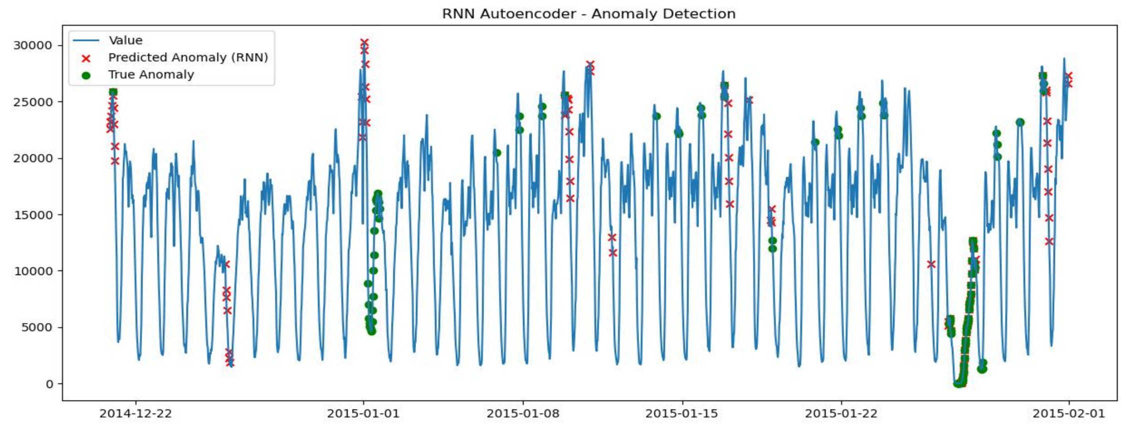

Figure 2. RNN Autoencoder Predictions test data

6.2 GRU Autoencoder Performance

The GRU Autoencoder demonstrated a noticeable performance improvement over the RNN model. It achieved macro-averaged Precision, Recall, and F1-Score values of **0.78**, indicating more reliable sequence reconstruction. For anomaly-specific metrics, it recorded **0.59** for all three evaluation parameters, reflecting enhanced anomaly sensitivity compared to the RNN. This improvement is attributed to GRU's gating mechanisms, which allow for better management of long-term temporal dependencies without the full complexity of LSTM units. The GRU's relative simplicity and strong performance make it a robust choice for time-series anomaly detection tasks where model interpretability and efficiency are also a concern in figure 3.

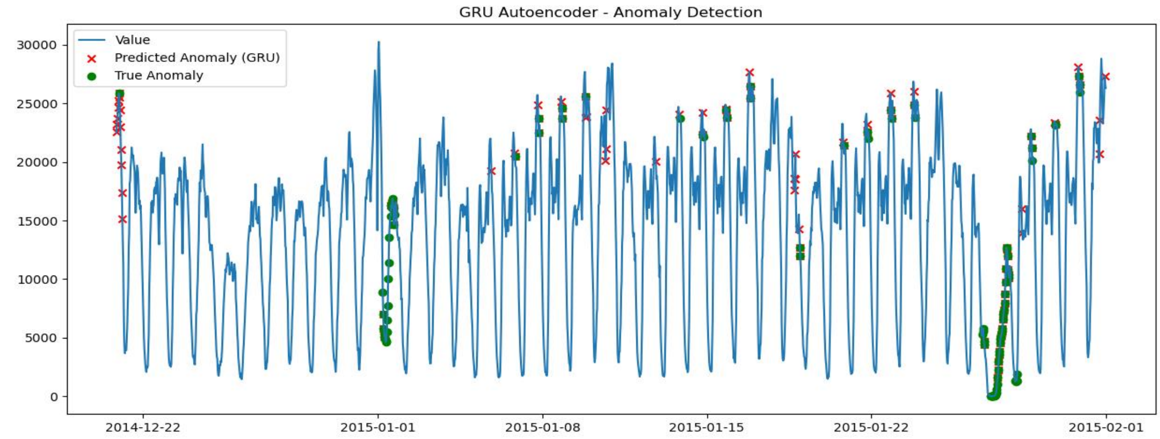

Figure 3. GRU Autoencoder prediction.

6.3 Vanilla LSTM Autoencoder Performance

The Vanilla LSTM model displayed consistent yet moderate results across both macro-averaged and anomaly-focused evaluations. It achieved a **0.72** score across Macro Precision, Recall, and F1-Score. For anomaly detection, it recorded **0.47** on all metrics. While the LSTM network better handles long-term temporal relationships than GRUs or RNNs, the lack of a dedicated attention mechanism likely limited its ability to assign importance to key time steps indicative of anomalies in figure 4. These results highlight the need for mechanisms that can dynamically prioritize informative temporal patterns, especially when the anomalous behavior is rare or subtle.

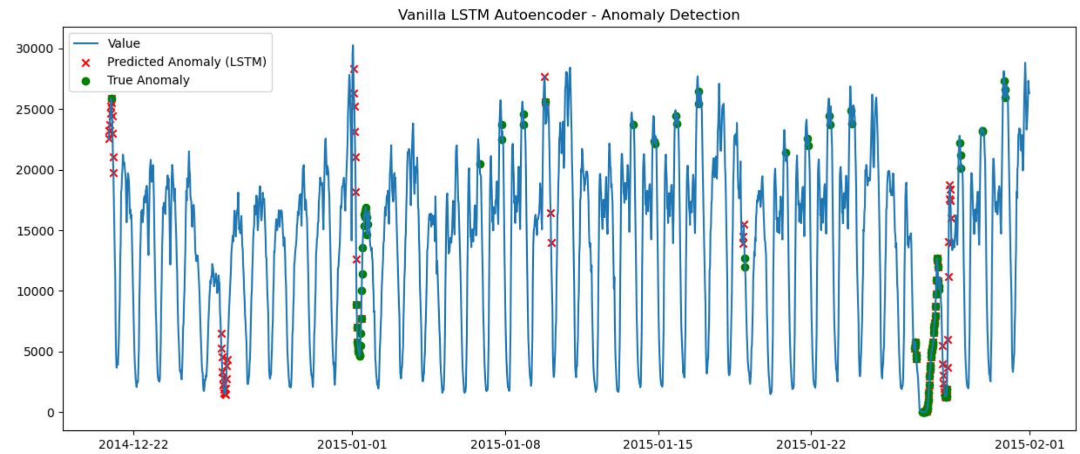

Figure 4. Vanila LSTM Autoencoder prediction.

6.4 Attention-Based LSTM Autoencoder Performance

The proposed model, which integrates an attention mechanism into the LSTM autoencoder architecture, outperformed all baseline models in both reconstruction accuracy and anomaly detection effectiveness. The highest being all the models, it got a Macro Precision, Recall, and F1-Score of 0.82. What is more important, as an anomaly-specific measure, it reported a

Precision, Recall, and F1-Score of 0.66, which has a considerable lead over the rest. This performance gain is critical on the attention mechanism which allows the model to dynamically weight time steps according to their contribution to the reconstruction task in figure 5. The design enables it to better show the deviations of normal behavior, thus is more resilient and precise in practice, such as predictive maintenance or real-time monitoring.

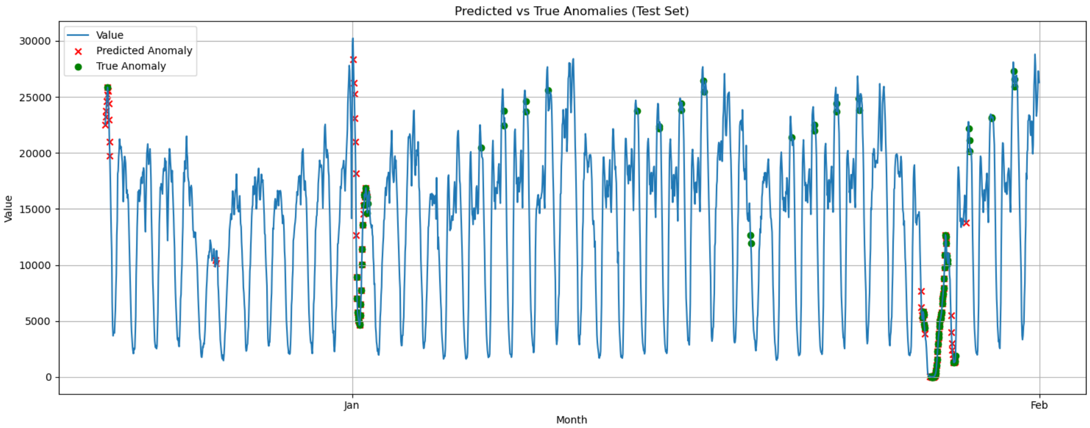

Figure 5. Attention LSTM Auto Encoder prediction.

6.5 Discussion

The findings of this paper offer sound evidence that the use of an attention mechanism in LSTM autoencoders can increase time-series anomaly detection significantly. Going back to the research question we started with in Section 1.3- *"Can the integration of attention mechanisms into LSTM autoencoders improve the accuracy and efficiency of time-series anomaly detection?"* - the findings decisively affirm this hypothesis. The attention-augmented LSTM autoencoder improved macro-averaged metrics over traditional RNN, GRU, and vanilla LSTM baselines and, additionally, the precision, recall, and F1-scores in anomaly-specific measures were much higher. This capability of the model to apply learned weights to each time step enhanced its concentration on the critical deviations, which can be translated into a concrete increase in detection sensitivity (Anomaly F1-score of 0.66). This aligns with the study's core objective of improving anomaly detection in resource-constrained environments while maintaining model efficiency.

Reflecting on the limitations highlighted in the literature (Section 2), conventional models like RNNs and LSTMs often treat all time steps equally, ignoring the temporal salience that anomalies often exhibit. This uniform attention limits the models' ability to localize subtle but impactful deviations in the sequence. As previously documented by Belay et al. (2023) and Nguyen et al. (2021), traditional autoencoders—while competent at reconstructing normal patterns—fall short in prioritizing informative time segments. The incorporation of an attention mechanism into the LSTM architecture addresses this critical gap by enabling the model to selectively emphasize the most informative portions of a sequence, making it more aligned with Transformer-like interpretability but without the associated computational complexity.

In terms of comparative evaluation (as detailed in Section 6), the GRU model provided a balance between simplicity and moderate performance, achieving an anomaly F1-score of 0.59—higher than RNN (0.39) and LSTM (0.47) autoencoders, but still below the proposed attention-based model. This further reinforces the added value brought by attention, especially in contexts where anomalies are not abrupt spikes but gradual or contextually masked deviations. Moreover, by achieving the highest overall macro F1-score (0.82) and anomaly-specific F1-score (0.66), the attention-enhanced LSTM model demonstrates its robustness across both general sequence reconstruction and targeted anomaly detection. The enhancements are especially useful in the context of real-time monitoring, where anomaly detection in the industrial Internet of Things, financial fraud detection, or critical infrastructure management can prevent operational and economic losses by providing timely and accurate anomaly detection.

Lastly, some limitations are also identified in this discussion that mitigates the generalizability of the results. Moreover, even though the computational complexity of the suggested model is still lower than Transformer-based models, the attention mechanism provides certain overhead to the baseline autoencoders. Despite that, such outcomes create a strong basis of future extensions, such as the testing on multivariate data or integration with real-time edge deployment frameworks. In conclusion, the discussion reaffirms the model's success in bridging the performance-efficiency gap, validating the research premise, and contributing a meaningful enhancement to the domain of time-series anomaly detection.

7 Conclusion and Future Work

This study explored the enhancement of time-series anomaly detection by integrating an attention mechanism into the LSTM autoencoder framework. The proposed architecture was evaluated against three baseline models—RNN, GRU, and vanilla LSTM autoencoders—using the NYC Taxi dataset, with performance assessed through both macro-averaged and anomaly-specific metrics. The attention-augmented model achieved superior results, recording the highest macro F1-score (0.82) and anomaly-specific F1-score (0.66). These outcomes indicate a tangible improvement in the model's ability to capture subtle, temporally localized anomalies while maintaining efficiency and interpretability.

The observed performance improvements directly address critical limitations discussed in existing literature, particularly the uniform weighting of time steps in traditional recurrent models and the computational burden of Transformer-based solutions. By enabling dynamic temporal focus through soft attention, the proposed model enhances both the accuracy and practicality of deep learning-based anomaly detection systems. This demonstrates that selective temporal modeling—achieved with relatively lightweight architectural modifications—can yield significant benefits without introducing prohibitive computational costs.

In research terms, the results have a positive answer to the key research question: *"Can the integration of attention mechanisms into LSTM autoencoders improve the accuracy and* *efficiency of time-series anomaly detection?"* The findings support the idea that not only is such integration possible, but it also enhances detection ability in a meaningful way, not compromising on the benefits of using LSTM-based models in deployment. Thus, the suggested solution adds a practical option to more complicated architectures in the situations that demand high sensitivity of detection together with limited resources or real-time constraints.

Further research will be devoted to the enhancement of the model generalization to various time-series areas, including multivariate data and the application of the domain of interest such as industrial monitoring, healthcare, and cybersecurity. Further focus will be on the optimization of the model to edge computing with the help of compression strategies and delving into the explainability framework to promote transparency in anomaly attribution. Architecture also has potential to be integrated in real world anomaly detection systems where modularity, interpretability and computational efficiency are key requirements in scalable and trustworthy deployment.

Bibliography

Liu, Pengfei, Xiaoming Sun, Yang Han, Zhishuai He, Weifeng Zhang, and Chenxu Wu. "Arrhythmia classification of LSTM autoencoder based on time series anomaly detection." *Biomedical Signal Processing and Control* 71 (2022): 103228.

Wei, Yuanyuan, Julian Jang-Jaccard, Wen Xu, Fariza Sabrina, Seyit Camtepe, and Mikael Boulic. "LSTM-autoencoder-based anomaly detection for indoor air quality time-series data." *IEEE Sensors Journal* 23, no. 4 (2023): 3787-3800.

Nguyen, Huu Du, Kim Phuc Tran, Sébastien Thomassey, and Moez Hamad. "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management." *International Journal of Information Management* 57 (2021): 102282.

Radaideh, M.I., Pappas, C., Walden, J., Lu, D., Vidyaratne, L., Britton, T., Rajput, K., Schram, M. and Cousineau, S., 2022. Time series anomaly detection in power electronics signals with recurrent and ConvLSTM autoencoders. *Digital Signal Processing*, *130*, p.103704.

Lee, B., Kim, S., Moon, J., & Rho, S. (2024). Advancing Autoencoder Architectures for Enhanced Anomaly Detection in Multivariate Industrial Time Series. *Computers, Materials & Continua*, *81*(1).

Iqbal, Amjad, Rashid Amin, Faisal S. Alsubaei, and Abdulrahman Alzahrani. "Anomaly detection in multivariate time series data using deep ensemble models." *Plos one* 19, no. 6 (2024): e0303890.

Lin, Y., Yu, Z., Yang, K. and Chen, C.P., 2025. Ensemble denoising autoencoders based on broad learning system for time-series anomaly detection. *IEEE transactions on neural networks and learning systems*.

Kim, B., Kang, J.W., Kim, C.S., Kwon, O.K. and Gwak, J., 2024. Hybrid Transformer for Anomaly Detection on Railway HVAC Systems Through Feature Ensemble of Spatial–Temporal with Multi-channel GADF Images. *Journal of Electrical Engineering & Technology*, *19*(4), pp.2803-2815.
Zhang W, Luo C. Decomposition-based multi-scale transformer framework for time series anomaly detection. Neural Networks. 2025 Jul 1;187:107399.

Lee, M.C., Lin, J.C. and Katsikas, S., 2024, August. Impact of recurrent neural networks and deep learning frameworks on real-time lightweight time series anomaly detection. In *International Conference on Information and Communications Security* (pp. 228-247). Singapore: Springer Nature Singapore.

Provotar, O.I., Linder, Y.M. and Veres, M.M., 2019, December. Unsupervised anomaly detection in time series using lstm-based autoencoders. In *2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT)* (pp. 513-517). IEEE.

Belay, M.A., Blakseth, S.S., Rasheed, A. and Salvo Rossi, P., 2023. Unsupervised anomaly detection for iot-based multivariate time series: Existing solutions, performance analysis and future directions. *Sensors*, *23*(5), p.2844.

Chen, Xuanhao, Liwei Deng, Feiteng Huang, Chengwei Zhang, Zongquan Zhang, Yan Zhao, and Kai Zheng. "Daemon: Unsupervised anomaly detection and interpretation for multivariate time series." In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pp. 2225-2230. IEEE, 2021.

Gjorgiev, Laze, and Sonja Gievska. "Time series anomaly detection with variational autoencoder using mahalanobis distance." In *ICT Innovations 2020. Machine Learning and Applications: 12th International Conference, ICT Innovations 2020, Skopje, North Macedonia, September 24–26, 2020, Proceedings 12*, pp. 42-55. Springer International Publishing, 2020.

Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A. and Veeramachaneni, K., 2020, December. Tadgan: Time series anomaly detection using generative adversarial networks. In *2020 ieee international conference on big data (big data)* (pp. 33-43). IEEE.

Kim, Siwon, Kukjin Choi, Hyun-Soo Choi, Byunghan Lee, and Sungroh Yoon. "Towards a rigorous evaluation of time-series anomaly detection." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, pp. 7194-7201. 2022.

Ribeiro, Antônio H., Koen Tiels, Luis A. Aguirre, and Thomas Schön. "Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness."

In *International conference on artificial intelligence and statistics*, pp. 2370-2380. PMLR, 2020.