INVERTIBLE NEURAL NETWORK FOR TIME SERIES ANOMALY DETECTION

Malgorzata Schwab and Ashis Biswas

Department of Computer Science and Engineering, University of Colorado at Denver

ABSTRACT

In this paper we explore the applicability of Invertible Neural Network architecture for anomaly detection techniques on time series data and hypothesize that a reversible network designed with embedded convolutional transformations is an excellent fit for that task. We leverage previous findings on autoencoders as well as deep generative maximum-likelihood training focused primarily on processing images and apply them in the innovative way to the time series data exemplified by electrocardiograms or industrial sensor data. We recognize a challenge of common denominator patterns that occur across the entire sample domain, which might dominate the likelihoods and introduce intrinsic bias. We then mitigate it by applying wavelet transforms to decompose a time series into a set of subcomponents to eliminate low-level similarities between the healthy and abnormal samples. We conclude that the Invertible Neural Network designed to solve inverse problems learns data reconstructions extremely well, and thus provides a remarkable solution for anomaly detection that is applicable to medical diagnostics, as well as other use cases in the similar problem space, such as predictive maintenance or detecting out-of-distribution inputs to protect integrity of systems relying on machine learning components.

KEYWORDS

Invertible, Autoencoder, Anomaly.

1. INTRODUCTION

One of the keys ideas behind digital transformation is the use of technology to extract useful information from the overwhelming amount of omnipresent data and provide us with new optimized ways to get additional insight into complex phenomena, deepen our understanding of observed outcomes, and enhance educated decision making. We constantly engage in broadly understood evaluations and diagnostics of systems and processes, machines and living creatures to determine what works well and what does not, what is “normal” and what is “abnormal” in each context. Whereas desired behaviours can be profiled with standards or specifications, the undesirable ones have countless possibilities and thus are difficult to be systematically described, other than “not” the desirable ones.

Anomaly detection is a process of identifying items that stand out, objects that do not belong, measurements that do not fit into a pattern observed by examining a collection of data samples considered as standard, normal, or typical. The pattern learning mechanism in machine learning is scoped to what is considered a sample set rep-resenting normal behaviours, and everything that looks sufficiently different in that context is regarded to be an anomaly.

David C. Wyld et al. (Eds): SOEN, SIPP, PDTCA, ITE, CCSIT, NLPCL, DaKM, BIGML, AISC -2023
pp. 227-239, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131220

Outliers in the phenomena characterized by low dimensionality can be easily assessed visually, by an algorithmic approach based on acceptable value ranges, or simple clustering techniques. Conventional methods of anomaly detection can be sorted out in several categories, such as statistical-based methods, probability-based methods, similarity-based methods, and most recent prediction-based methods. They have been well summarized in Giannoni [1] and subsequently by Yin [2].

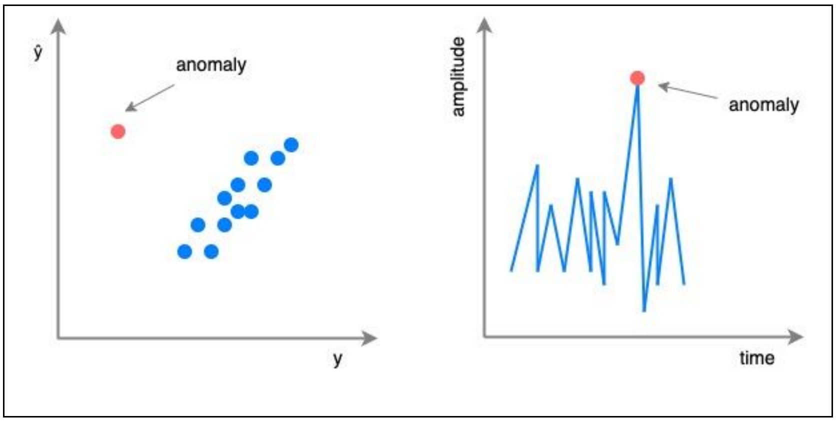


Figure 1. Anomaly detection techniques

For high dimensional use cases, multivariate statistics based on probability distributions and its counterpart machine learning techniques, capable of extracting complex patterns, come to the rescue.

In this work we leverage previous findings and principles regarding several types of autoencoders and reversible neural networks and propose an improved network architecture for better anomaly detection on time series data using an Invertible Neural Network (INN) [3]. We hypothesize that an INN trained as a convolutional autoencoder and used on 2D-transformed time series data is an effective alternative naturally suited to solve anomaly detection. Our proposed technique is applicable to various domains in this problem space, such as medical diagnostics or the Internet of Things' (IoT) discipline of predictive maintenance described in [4].

The remainder of this paper is organized as follows: Section 2 reviews related work about anomaly detection of time series data using deep learning techniques, focusing on the autoencoder network architecture [1]. It then elaborates on normalizing flows [5][6] and Invertible Neural Network architecture together with the previously established Framework for Easily Invertible Architectures (FrEIA), by Ardizone [3]. In Section 3 we learn how an INN can be trained as a probabilistic autoencoder [7] and introduce our proposed methodology to apply this architecture to time series type of data. Section 4 summaries our findings leading to a conclusion, which is offered in Section 5.

2. DEPENDENCIES AND LIMITATION

The anomaly detection solution presented in this work is based on the revolutionary Invertible Neural Network architecture, which was first introduced by Dinh [5] in 2016. The experiments leverage a concrete INN implementation described in [3] wrapped by the Framework for Easily Invertible Architectures, which offers an API mechanism to stack, infused with bijective functions, invertible network nodes to achieve reversible deep learning capability.

3. RELATED WORKS

3.1. Autoencoders

Autoencoders (AE) belong to the family of unsupervised dimensionality reduction deep learning neural network (DNN) models and have been described extensively in numerous works, such as [1], [2] and [8]. The idea around this type of neural network is to extract relevant features from input data and then learn how to reconstruct original data with these features.

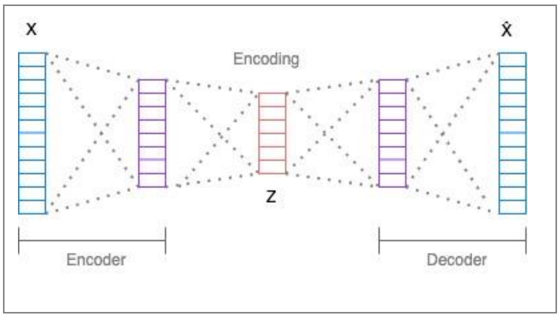


Figure 2. Classic Autoencoder

As shown in Fig. 2, an autoencoder consists of an encoder and a decoder, implemented as fully connected neural networks. It first encodes the network input x into a lower dimensional latent representation z, then decodes the latent representation back to the original input $\hat{x}$. The information preserved in hidden neurons is considered as the encoded features. The learning process is based on minimizing the reconstruction error, which is assessed by comparing a reconstruction result to the original input. Thus, an autoencoder learns to compress the data into a bottleneck to a point of minimal reconstruction error. The learned representation corresponds to the final hidden state of the encoder network and acts like a summary of the input sequence. Autoencoders can be applied for dimension reduction, as well as for further classification leading to prediction.

Convolutional autoencoders (CAEs), depicted in Fig. 3, use convolutional layers to fulfil the general autoencoder purpose of creating compressed representations [9][10].

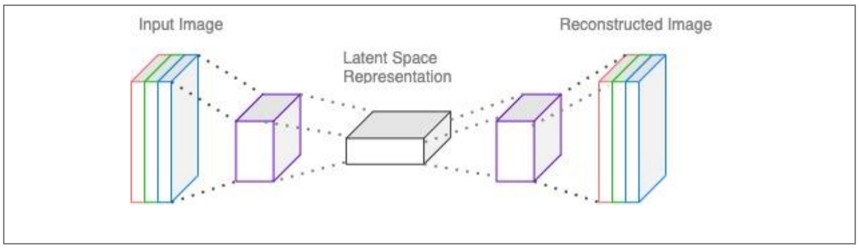


Figure 3. Convolutional Autoencoder

CAEs are particularly suited to learn internal representations of image data and are mainly utilized in image compression and noise reduction.

An important step in the autoencoder evolution was a model based on a Recurrent Neural Network (RNN) architecture proficient in encoding/decoding sequential data, proposed by Sutskever [11].

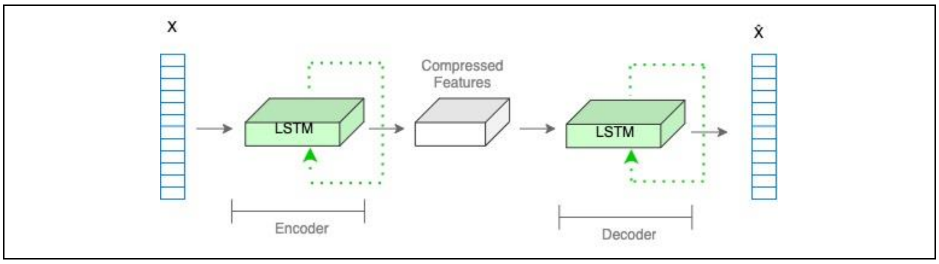


Figure 4. LSTM Autoencoder

An RNN-based encoder, which has been proven useful for time series reconstruction, receives the input sequence in subsequent time steps, extracts and memorizes the temporal features of the input data, which is preserved in hidden units as encoded information or embedding. The RNN-based autoencoder architecture has been improved with Long Short-Term Memory (LSTM), depicted in Fig. 4, which introduced memory cells to help remember sequences across intervals of time, capturing temporal dependencies within the data, together with gates to control the information flow across the network.

Variational Autoencoders are probabilistic generative models that learn a mapping from some latent random variable z to the probability density function on input x, so the regenerated input data can follow the exact same distribution while the reconstruction error is minimized.

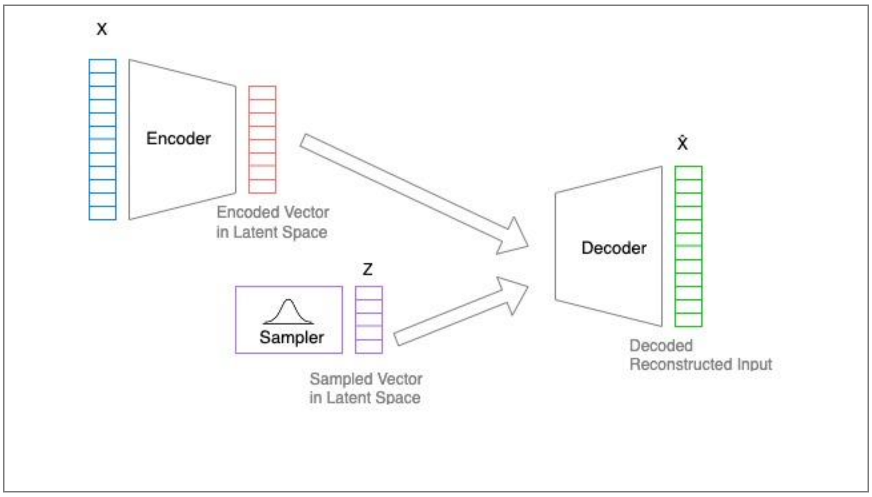


Figure 5. Variational Autoencoder

The decoding phase is augmented with random samples drawn from the variational distribution in the latent space so it can also learn effective decoding for the values it has not seen, and thus variational autoencoders are capable of generating new content matching the probability density of x. Both networks are trained together with the usage of a reparameterization trick to mitigate the necessity of derivatives on random variable z.

4. INVERTIBLE NEURAL NETWORKS

An Invertible Neural Network, described by Ardizone [3], is a class of networks suited to solve ambiguity that characterizes inverse problems, where multiple parameter sets can produce the same observed outcome. To express this ambiguity, the posterior probability of those parameters' distribution, given that outcome, must be learned so the most appropriate set can be selected. Such a model can perform log-density estimation of data points, leading to efficient inference and precise reconstruction of the inputs from the hierarchical features extracted by the model. This extraordinary capability to reconstruct the inputs corresponding to the encoder-decoder functionality makes INN a natural candidate to help solve the problem of anomaly detection.

An Invertible Neural Network guarantees reversibility by its construction and solves the ambiguous inverse relationships directly. INN is trained simultaneously in the forward and reverse directions. The forward learning process uses additional latent output variables to capture information otherwise lost, making the learning of the inverse process explicit.

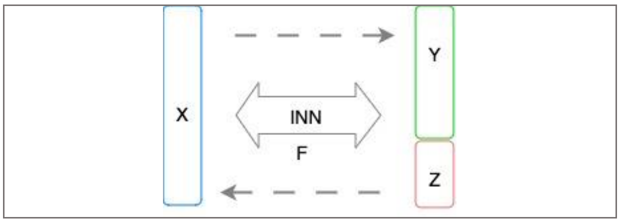


Figure 6. Invertible Neural Network Conceptual Diagram

To solve the general inverse problem, we augment the observation space Y with a latent variable Z which follows a normal distribution and look for a bijective function F that can map Z back to $\hat{X}$. Being Bayesian, INN learns the invertible mapping between the data distribution P_X and the latent distribution P_Z, typically Gaussian.

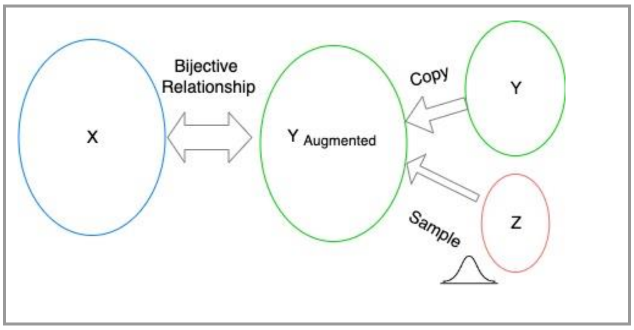


Figure 7. Reconstructing phenomenon X from observation Y

Invertibility of neural networks was first spearheaded by Dinh [5] as "real-valued non volume preserving transformations" (Real NVP) architecture, who introduced a stack of invertible affine coupling blocks (Fig 8.), arranged in hidden layers.

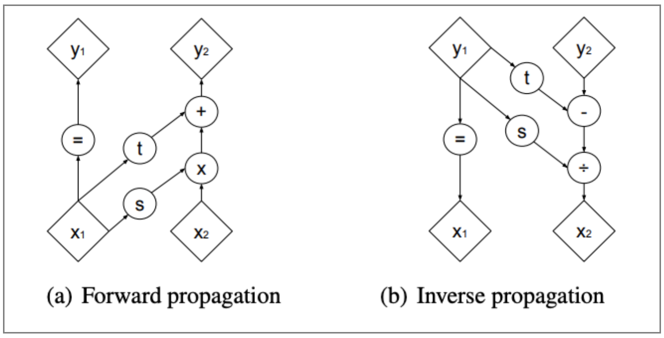


Figure 8. Real NVP Affine Coupling Block [5]

As described in [5], given a D-dimensional input x and d < D, the output y of an affine coupling layer follows the following equations:

$$y_{1:d} = x_{1:d} \tag{1}$$

$$y_{d+1:D} = x_{d+1:D} \odot \exp\left(s(x_{1:d}) + t(x_{1:d})\right) \tag{2}$$

where s and t are functions from $R^d \rightarrow R^{D-d}$, and $\odot$ is the Hadamard product or element-wise product.

Each block splits its input and output into two parts and applies transformations s (scale) and t (translation), which themselves do not have to be invertible – they can be quite complex and are often implemented as artificial neural networks, such as a CNNs. It has been proven [3] that a stack of such invertible blocks makes the end-to-end layout also invertible.

5. PROPOSED ANOMALY DETECTION SCHEME

5.1. Solution Architecture

Our proposed anomaly detection scheme builds on Autoencoders and Invertible Neural Networks and introduces a competitively performant INN Time Series Variational Autoencoder architecture with convolutional feature extraction and wavelet transformations for time series data. The concepts and models described in the previous section provide a foundational canvas for what is being architected and experimented in this work.

The concept of an INN entails bijective input-output mapping, so the dimensions of input x and output y augmented with z must be equal. We then construct an artificial bottleneck to achieve autoencoder-like behaviour [5], which is accomplished by zeroing the latent **z** to make sure that no extra information is retained by the network in the inverse process of representation learning. This is depicted in Fig. 9 below.

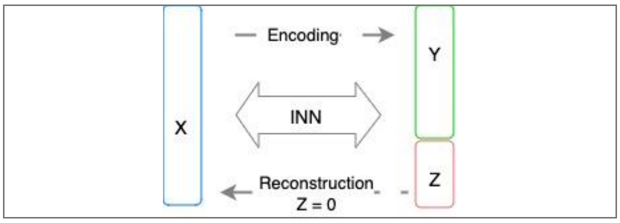


Figure 9. Invertible Neural Network as Autoencoder

To create a fully invertible neural network, the solution follows the architecture proposed by Dinh [4] and wrapped by the Framework for Easily Invertible Architecture [3] which encapsulates known patterns for constructing invertible affine coupling blocks and exposes an interface to apply custom configurations to the stackable network nodes.

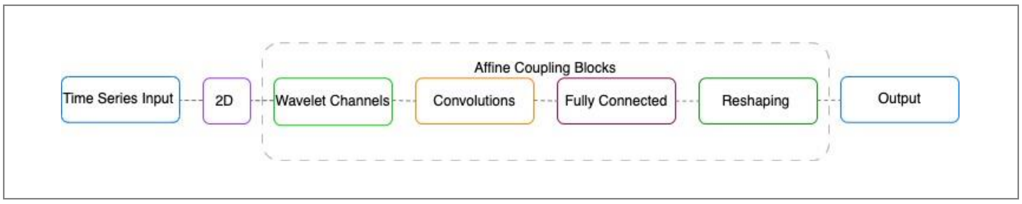


Figure 10. INN Autoencoder Network Layout

As depicted in Fig. 10, our INN Time Series Autoencoder configuration consists of three affine coupling layers leveraging convolutional coupling functions **s** and **t** with the size of the hidden channel equal to 100, a kernel size of 3 and leaky ReLU, followed by a coupling block with fully connected coupling function. A multiplexing Haar wavelets transformation layer was applied to split each channel into 4 channels, with half the width and height.

We acknowledge traditional constraints regarding an autoencoder bottleneck necessity but confirm the validity of a non-bottleneck approach [12], which naturally fits the INN autoencoder architecture. Indeed, an INN based autoencoder with large bottleneck sizes performs increasingly better, and as also hypothesized in [5], display no intrinsic information loss.

5.2. Experimental Setup

The goal of the experiments is to construct and assess an INN Time Series Autoencoder, as compared to the classical LSTM time series autoencoder. To understand the rationale of the necessity of a bottleneck [12] and the consequences of its size, we train the network with various bottleneck sizes.

The experiments were conducted on two different time series datasets: a 140-dimensional ECG heart diagnostics dataset with 5,000 electrocardiogram [13] and the 61,440-dimensional Airbus helicopter accelerometer dataset [14] from the IoT predictive maintenance domain (PdM), whose dimensionality was reduced to 240, where each of which corresponds to the means of 256 consecutive values. This allowed to apply the same depth and bottleneck sizes of an INN configuration in both cases. The ECG dataset was simplified from its 5-class classification format, where each example has been labeled “0” for an abnormal rhythm, or “1” for a normal rhythm, and only normal samples were utilized in the unsupervised fashion. We are interested in identifying the abnormal rhythms as not recognizable by the model.

The visualization of healthy and abnormal data samples is depicted in Fig. 11 below.

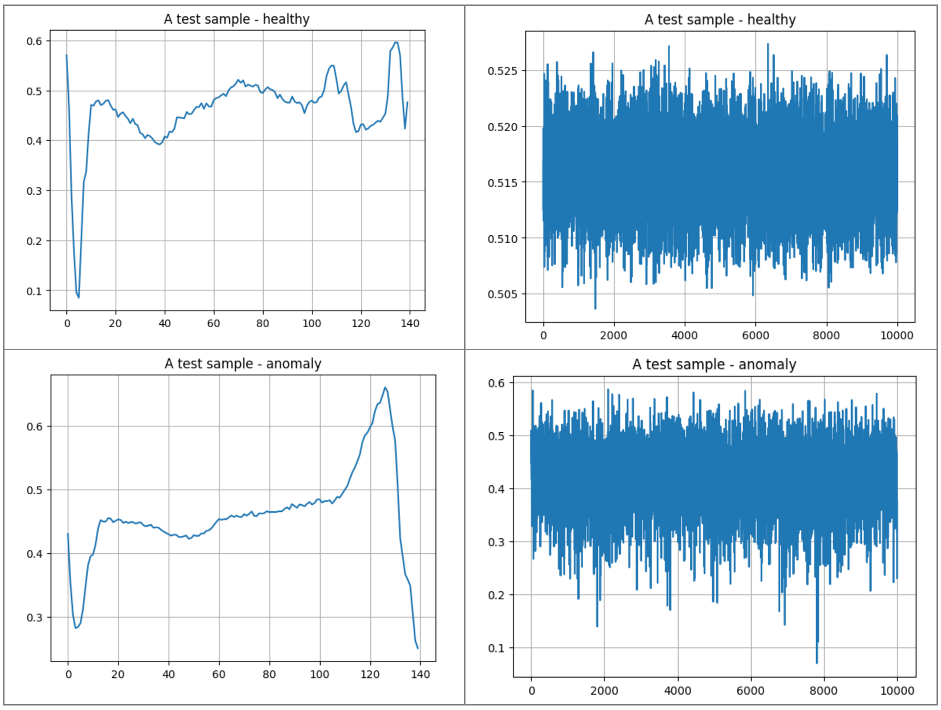


Figure 11. Healthy and Abnormal Samples – ECG (left) and PdM Vibration (right)

5.3. Methodology

The INN Time Series Autoencoder scheme unfolds as follows:

1) An INN is constructed with several nodes, described in Table 1, representing a wavelet transformation layer, CNN feature extraction layers, and one fully connected layer.
2) For high dimensional time series input, data is transformed to reduce the number of dimensions. This is described in section 3.2.
3) Time series samples X = (x_1, x_2, …, x_T) are then transformed into a 2-d format, where N x M = T
4) The INN is trained as an autoencoder on the healthy dataset only, while creating a benign artificial bottleneck by zero-padding the latent variable space [5]. The training was performed with a batch size of 64 and in 10 epochs.
5) The reconstruction error threshold TH for healthy samples is captured.
6) The model is applied to the validation set using a percent of the threshold to detect anomalies (example: 95%)
7) A truth table is drawn out to calculate the Recall number, since misdiagnosed abnormal measures (False Negative) are a concern.
8) The process may be repeated for various bottleneck sizes, to find the best solution.

5.4. Results

Both datasets were treated in a comparable manner, with the INN configuration described in the tables below, for the ECG and PdM datasets respectively. Tables 1 and 4 show the network structure, with wavelet transformation and three convolutions, followed by a fully connected layer; tables 2 and 5 visualise how well the network has learned to reconstruct the input data; tables 3 and 6 report on the inference reconstruction loss, much larger for anomalies as compared to the healthy training samples. The results are further reiterated in Figures 12 and 13, where the anomaly curves stand far apart from what the network was trained on, which is exactly what we were hoping to see.

5.4.1. ECG Dataset

Table 1. INN Nodes Input Dimensions

Node	Input Dimensions
Wavelet	(1, 10, 14)
Conv1, Conv2, Conv3	(4, 5, 7)
Fully Connected	(140)
Output	(1, 10, 14)

Table 2. Difference between original and reconstructed data

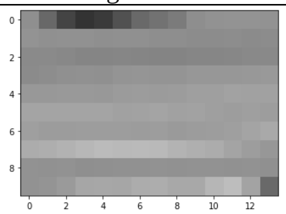
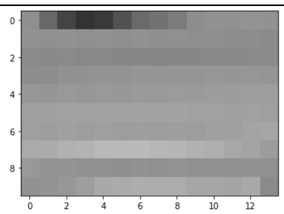
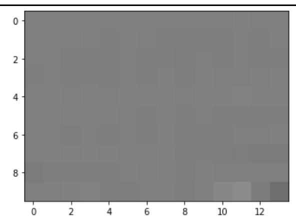
Original Data	Reconstructed Data	Difference
		

Table 3. Reconstruction loss for various bottleneck sizes

Latent Dimension	Healthy Reconstruction Loss	Abnormal Reconstruction Loss
8	0.019	0.110
16	0.017	0.096
32	0.013	0.094

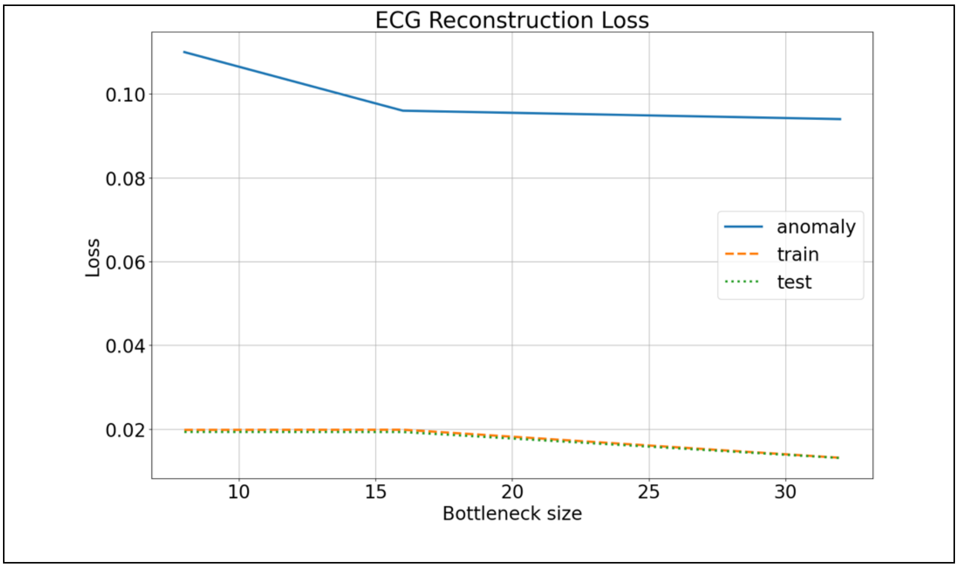


Figure 12. Reconstruction loss on the ECG dataset

5.4.2. Helicopter Accelerometer Vibration Dataset

Table 4. INN Nodes Input Dimensions

Node	Input Dimensions
Wavelet	(1, 12, 20)
Conv1, Conv2, Conv3	(4, 6, 10)
Fully Connected	(240)
Output	(1, 12, 20)

Table 5. Difference between original and reconstructed data (batches)

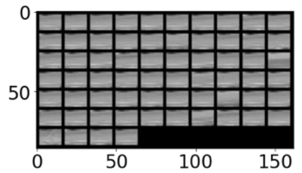
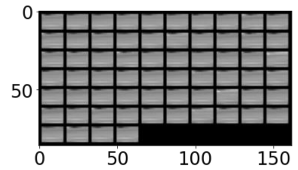
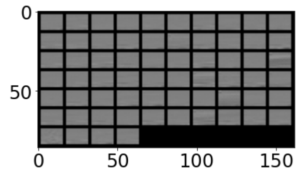
Original Data	Reconstructed Data	Difference
		

Table 6. Reconstruction loss for various bottleneck sizes

Latent Dimension	Healthy Reconstruction Loss	Abnormal Reconstruction Loss
8	0.027, 0.028	0.122
16	0.023, 0.027	0.120
32	0.021, 0.022	0.118

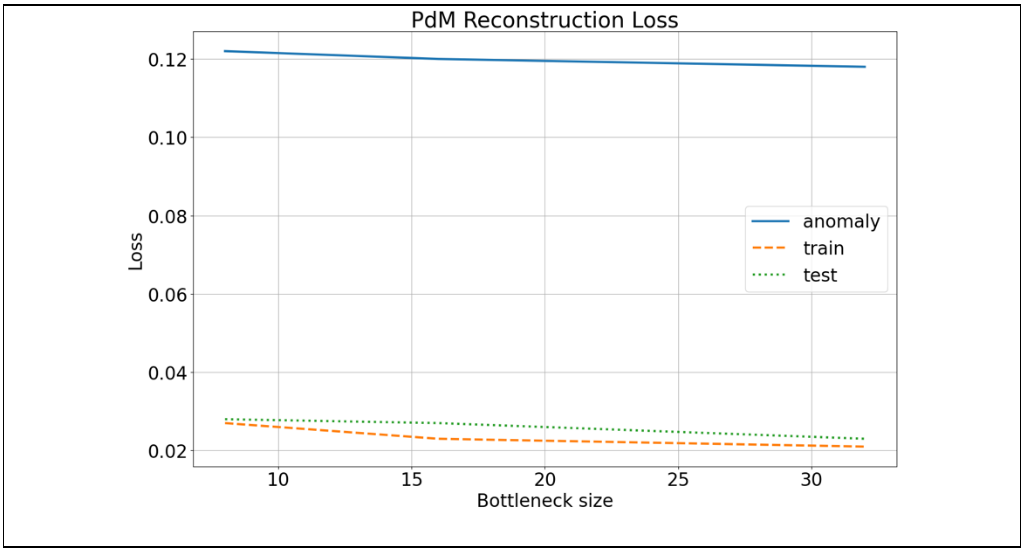


Figure 13. Reconstruction loss on the helicopter accelerometer PdM dataset

6. RESULTS DISCUSSION

We depicted the results in a plot gauging the ability of the network to reconstruct the inputs it was trained on against not seen data considered "out-of-distribution", or an anomaly. The reconstruction loss on the anomalous samples is significantly greater (an order of magnitude) as compared to the reconstruction error on the healthy validation data. There seems to be an improvement in reconstruction error in the network configuration with increasingly larger artificial bottlenecks, where zero-padding ensures that additional dimensions in the latent space are not used in representation learning.

Tables 8 and 9 below depict the results in terms of model evaluation metrics, defined in Table 7, where False Negative are underdiagnosed, abnormal samples presenting a risk of undetected anomaly. That risk in the performed experiments is lower with the Invertible Neural Network used as an autoencoder for time series, as compared to its classical or LSTM counterparts. Recall, which measures the proportion of actual positives that was identified correctly appears more than 10% better for the vibration dataset, and 2% better on the ECG dataset for the INN based solution.

Table 7. Truth Table Legend

True Positive (TP)	Anomaly predicted as Anomaly
False Negative (FN)	Anomaly predicted as Healthy (red flag)
False Positive (FP)	Healthy predicted as Anomaly
True Negative (TN)	Healthy predicted as Healthy

Table 8. PdM Truth Table

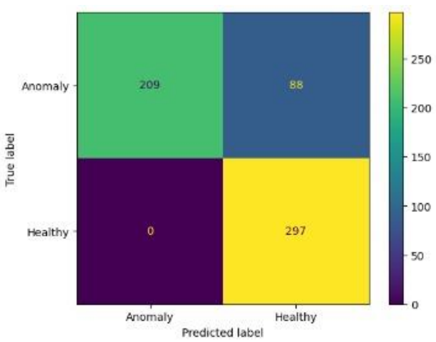
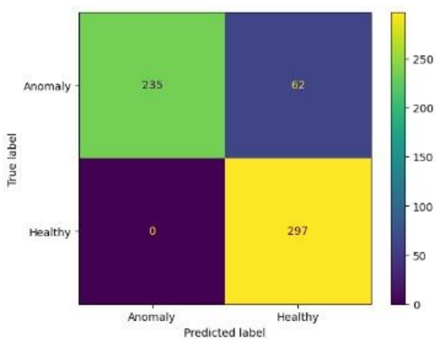
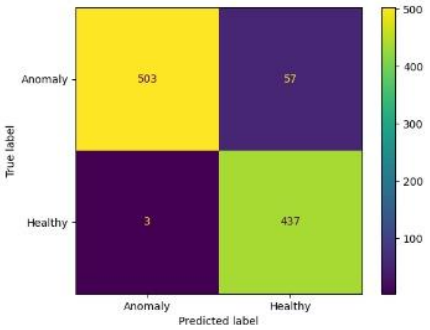
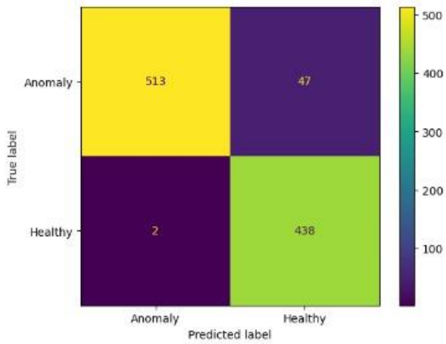
LSTM EA	INN EA
	
Precision: TP/(TP+FP) = 209/(209) = 100% Recall: TP/(TP+FN) = 209/(209+88) = 70%	Precision: TP/(TP+FP) = 235/(235) = 100% Recall: TP/(TP+FN) = 235/(235+62) = 79%

Table 9. ECG Truth Table

<table>
<tr><th>EA</th><th>INN AE</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Precision: TP/(TP+FP) = 503/(503+3) = 99%
Recall: TP/(TP+FN) = 503/(503+57) = 89%</td><td>Precision: TP/(TP+FP) = 513/(513+2) = 99%
Recall: TP/(TP+FN) = 513/(513+47) = 91%</td></tr>
</table>

7. Conclusion

In conclusion, the experiments indicate that the proposed model for time-series anomaly detection, which leverages an Invertible Neural Network as a convolutional autoencoder, proves effective and has better recall outcome from the classic stat-of-the-art autoencoder-based approaches. The proposed solution appears to be generalizable, as it was applied with similar success rate to the datasets from various domains and much different time-series characteristics. Effectiveness of the solution results from the natural ability of Invertible Neural Networks to perfectly reconstruct inputs from the learned latent space, and even though the model may seem conceptually complex, frameworks exist to abstract such intricacy away and easily configure invertible network architectures.

References

[1] Giannoni, F., Mancini, M., Marinelli F., “Anomaly detection models for IoT time series data.”, 2018. [Online]. Available: arXiv:1812.00890.
[2] Yin, C., Zhang, S., Wang, J., Xiong, N., “Anomaly Detection Based on Convolutional Recurrent Autoencoder for IoT Time Series.”, In: IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 1, pp. 112-122, Jan. 2022, doi: 10.1109/TSMC.2020.2968516.
[3] Ardizzone, L., Kruse, J., Wirkert, S.J., Rahner, D., Pellegrini, E., Klessen, R.S., Maier-Hein, L., Rother, C., Köthe, U., “Analyzing Inverse Problems with Invertible Neural Networks”, ArXiv, abs/1808.04730.
[4] Kamat, P., Sugandhi, R., “Anomaly Detection for Predictive Maintenance in Industry 4.0- A survey.”, E3S Web Conf. 170 02007 (2020), DOI: 10.1051/e3sconf/202017002007
[5] Dinh, L., Sohl-Dickstein, J. & Bengio, S.,“Density estimation using Real NVP”, 2016
[6] Rezende, D., Mohamed, R., ”Variational inference with normalizing flows.”, arXiv preprint arXiv:1505.05770, 2015.
[7] Nguyen, TG.L., Ardizzone, L., Köthe, U., “Training Invertible Neural Networks as Autoencoders.”, In: Fink, G., Frintrop, S., Jiang, X. (eds) Pattern Recognition. DAGM GCPR 2019. Lecture Notes in Computer Science(), vol 11824. Springer, Cham. https://doi.org/10.1007/978-3-030-33676-9_31
[8] Wen, L., Gao, L., Li, X., “A new deep transfer learning based on sparse auto-encoder for fault diagnosis.”, In: IEEE Trans. Syst., Man, Cybern., Syst., vol. 49, no. 1, pp. 136–144, Jan. 2019.
[9] Kingma D., Welling M., “Auto-Encoding Variational Bayes.”, In: arXiv:1312.6114

[10] Pintelas E., Livieris I.E., Barotsis N., Panayiotakis G., Pintelas P., "An Autoencoder Convolutional Neural Network Framework for Sarcopenia Detection Based on Multi-frame Ultrasound Image Slices.", In: Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations; Crete, Greece. 25–27 June 2021; pp. 209–219.
[11] Sutskever, I., Vinyals, O., Le, O., " Sequence to sequence learning with neural networks.", In: Proc. Adv. Neural Inf. Process. Syst., Montreal, QC, Canada, 2014, pp. 3104–3112.
[12] Yong, B., Brintrup, A.: "Do Autoencoders Need a Bottleneck for Anomaly Detection?", In: IEEE Access, vol. 10, pp. 78455-78471, 2022, doi: 10.1109/ACCESS.2022.3192134.
[13] ECG Dataset, http://www.timeseriesclassification.com/description.php?Dataset=ECG5000
[14] Predictive Maintenance Dataset, https://www.research-collection.ethz.ch/handle/20.500.11850/415151

AUTHORS

Malgorzata Schwab is an Enterprise Architect and IT executive leading digital transformation focused on innovative business strategies through cloud migrations and AI-augmented technologies. She is a PhD candidate at the Department of Computer Science and Engineering at the University of Colorado in Denver.