# VixLSTM: Visual Explainable LSTM for Multivariate Time Series

Tommy Dang
Tommy.Dang@ttu.edu
Texas Tech University
Lubbock, Texas, USA

Huyen N. Nguyen
Huyen.Nguyen@ttu.edu
Texas Tech University
Lubbock, Texas, USA

Ngan V.T. Nguyen
Ngan.V.T.Nguyen@ttu.edu
Texas Tech University
Lubbock, Texas, USA

## ABSTRACT

Neural networks are known for their predictive capability, leading to vast applications in various domains. However, the explainability of a neural network model remains enigmatic, especially when the model comes short in learning a particular pattern or features. This work introduces a visual explainable LSTM network framework focusing on temporal prediction. The hindrance to the training process is highlighted by the irregular instances throughout the entire architecture, from input to intermediate layers and output. The framework provides interactive features to support users in customizing and rearranging the structure to obtain different network representations and perform what-if analysis. To evaluate the usefulness of our approach, we demonstrate the application of VixLSTM on the various datasets generated from different domains.

## CCS CONCEPTS

• **Computing methodologies** → **Temporal reasoning**; **Machine learning**; • **Human-centered computing** → **Visualization application domains**.

## KEYWORDS

Neural Networks, Deep Learning Scatterplot, Time Series Visualization

**ACM Reference Format:**

Tommy Dang, Huyen N. Nguyen, and Ngan V.T. Nguyen. 2021. VixLSTM: Visual Explainable LSTM for Multivariate Time Series. In *The 12th International Conference on Advances in Information Technology (IAIT2021), June 29-July 1, 2021, Bangkok, Thailand.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3468784.3471603

## 1 INTRODUCTION

Time-series data comprises the values from observed variables collected at different points in time. Multivariate time series data is a subclass of this category regarding various data record attributes, leading to complexity in the analytics tasks. Regarding the machine learning model for time series data, the long short-term memory (LSTM) network is a popular architecture with the ability to learn temporal sequences.

At large, the neural network remains enigmatic as the details of the components may be found difficult for users to interpret. Furthermore, the interpretation becomes more complicated when there exist features that the network is unable to learn or predict correctly. An explainable model that can trace backward from the irregular output to the corresponding points in inputs would help alleviate such issues and pave the way for successive steps. To enable the ability to neural network explainability, this paper proposes a scalable visual prototype for representing the long short-term memory (LSTM) networks so that the expert users can understand the rationales behind the learned models. The main contributions of this work can be summarized as:

- We present a visual prototype, called VixLSTM, for the presentation of LSTMs on multivariate time series data. We present the visual arrangements to establish an intuition on the relationship between input and output, especially the irregularities among the instances using the SHAP (SHapley Additive exPlanations).
- We propose an interactive approach to exploring the underlying mechanism of LSTMs, with flexible customization for users to structure the visual components. This supports users in understanding the important variables/data instances for the LSTM outcomes.
- We demonstrate our framework on the High-Performance Computing Center health monitoring and US monthly employment datasets to evaluate the visual presentation's usefulness and effectiveness.

## 2 RELATED WORK

This section will discuss previous related literature regarding LSTMs models for time series data and explainable neural networks.

### 2.1 Long Short-Term Memory Networks for Time Series Data

Along with the development of machine learning models, the family of artificial neural networks marked its growth by various network architectures. The topology of the LSTM network is similar to standard Recurrent neural network (RNN) models [14], demonstrating its remarkable ability with sequential data [4], where time series is a typical instance. The LSTM network's ability helps to learn long and short patterns within the input dataset [6], also with long correlations [10]. LSTM structure eliminates the issue of vanishing gradient [15] in training RNN.

Regarding time series data, LSTM models present vast applications [10]. A typical structure of an ANN model with LSTM is depicted in Figure 1. Here, among hidden layers, we distinguish LSTM layers and dense – fully connected layers [7].

### 2.2 Explainable Neural Network

Machine learning models are seen as non-intuitive [16], especially neural networks are often considered "black boxes", hence the limitation of understanding such models. The field of explainable neural networks is getting more and more attention as the increasing attempts to open these black boxes [2, 5]. If we can present such

*IAIT2021, June 29-July 1, 2021, Bangkok, Thailand*

ACM ISBN 978-1-4503-9012-5/21/06…$15.00
https://doi.org/10.1145/3468784.3471603

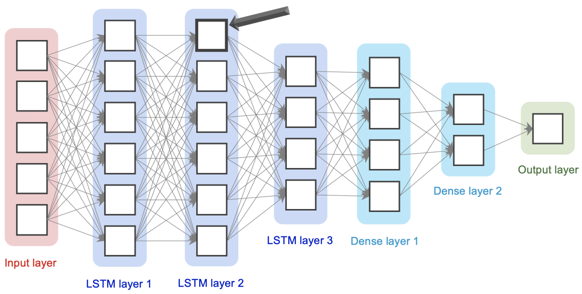


**Figure 1: Typical structure of an ANN model with LSTM. We distinguish between LSTM layers and dense layers among hidden layers.**

a complex machine learning model's internal structure, we can trace the reasoning process and inject human knowledge where necessary. DeepVix [3] propose a framework for explaining the neural network for the multivariate time-series predictions. Each node in the network is represented by a heatmap showing the input data's time series and the intermediate layers in the network. However, DeepVix relies on the viewer's visual capability to detect the patterns in the heatmap and how they evolve over time and through different layers in the network. In this paper, we adapt the game theoretically optimal Shapley Values [11] to support the explainability of our framework.

Furthermore, it is also crucial to explain and investigate the contributions of individual features or feature values to the predictions of the built models. In this research direction, SHAP (SHapley Additive exPlanations) is gaining favors. SHAP [9] is an approach to explain the output of any machine learning model or python function. SHAP provides a high level of interpretability for a model. The SHAP values offer two significant advantages: Global interpretability vs. Local interpretability. SHAP libraries also generate visualizations to support these interpretability advantages. In particular, the negative vs. positive contributions (for each feature or individual instance) are calculated by the mean absolute value of the SHAP values. In our work, we present each node as a time series visualization and highlight the relationship between the irregular outputs and their corresponding instances in the input and the intermediate layers.

## 3 METHODOLOGY

We use line graphs for representing time series because they are simple, standard, and familiar to different classes of users (from novice to expert). However, line graphs induce visual cluttering issues when visualizing many time series in a single plot. Heatmaps mitigate this issue by organizing individual series vertically and encoding cell colors by their values [2]. Consequently, heatmaps allow users to convey the overall temporal pattern of a neural [5]. We provide two options (line-graph and heatmap) for one node in the LSTM layer, as depicted in Figure 2.

Figure 3 shows a closeup view of the output node in the LSTM model. Notably, we highlight the top 5 percent of the highest absolute predicted values from MSE in orange, i.e., the ones that produce large errors among all instances, versus their actual values in green.

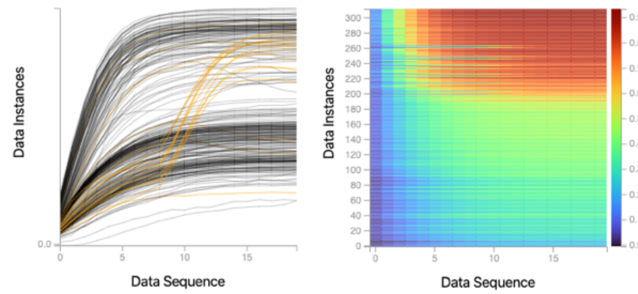


**Figure 2: Two visual representations of the same node in the LSTM layer: Line chart (left) and Heatmap (right).**

These top 5 percent of MSE errors are consistently highlighted in the LSTM node in the left panel of Figure 2. This effect is performed without interfering with the training process (highlighting orange data instances after the training is done).

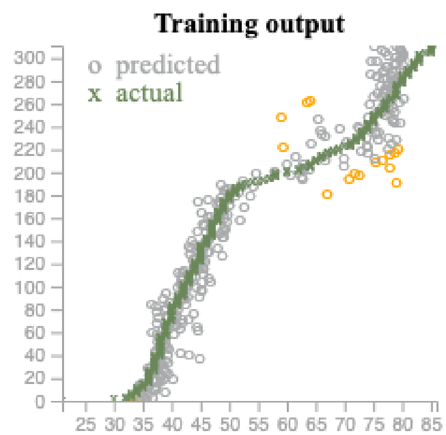


**Figure 3: The example scatterplot of training output: grey circles are predicted values, orange circles are irregular values, and green points are actual values.**

## 4 USE CASE

### 4.1 Predicting computer behaviors though the past usages

This multivariate time series data contains computer readings, including CPU temperature, Memory Usage, Fan Speed, and Power Consumption. In this use case, these computer health metrics are reported every 5 minutes for the entire interval of 2 hours [8]. After preprocessing, we have a dataset with 467 instances (associating to 467 computers in the HPC center), and each data instance has roughly 20 timesteps [1]. The input data is divided into two sets: the training set contains 300 computers, and the testing set contains 167 computers. The target variable is the *CPU Temperature*.

We set up a configuration with one input layer (includes ten nodes representing ten health readings of 467 computing nodes in the HPC center), two eight-node LSTM layers, two dense layers.

Figure 4 shows the LSTM neural network of a trained model using this architecture (we stopped the training after 30 epochs). VixLSTM orders the neural nodes in the same layers by their similarity (using cross-correlations) so that viewers can keep track of their pattern changes. In other words, neural nodes with similar trends should be listed closely in the vertical order.

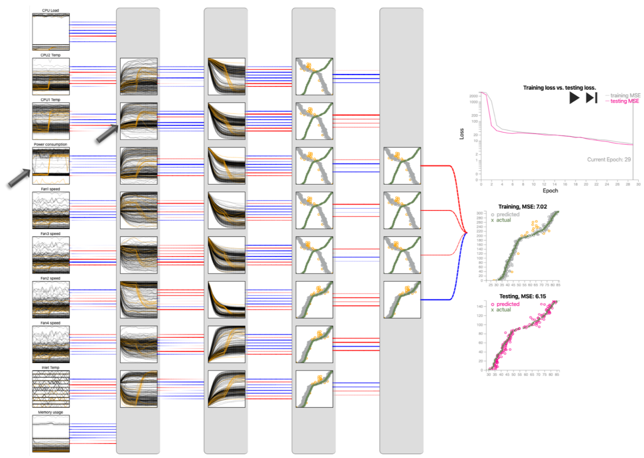

**Figure 4: The overview of our VixLSTM on HPC data with the following configuration from left to right: input layer (of ten health metrics), two LSTM layer (each has eight nodes), two dense layers.**

In this use case, we highlight outlying samples in the training data, colored in orange. We can quickly notice that the top 5 percent MSE errors exhibit very different temporal patterns than the other data instances.

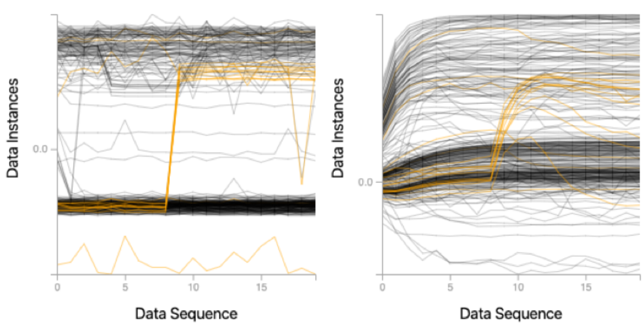


**Figure 5: The line chart of samples in a hidden node from the LSTM layer. Orange lines are computing exhibiting thermal throttling issues.**

Figure 5 shows closeup views of two sample neural nodes at the arrows in Figure 4: One from the input layer and one from the first LSTM layer. As depicted in these line graphs, the abnormal instances (in orange) represent sudden pumps in their time series at the time step 7 to the time step 13. This unusual behavior in the computer health metrics time series might indicate a new job is allocated on these computing nodes [12] or unusual events that cause the CPU thermal throttling [13]. This explainable view allows system administrators to filter and focus on a subset of health readings/computing hosts/timestamps of interest for system debugging. This example also suggests that the temporal profiles with high MSE errors in the output exhibit abnormal behaviors on the input time series and potentially can be highlighted as time series outliers [17].

## 4.2 US Employment data

This use case demonstrates the monthly unemployment rate in the US from 1999 to 2018. Each temporal profile is a state. Input variables include *Goods producing*, *Service providing*, *Manufacturing*, *Transportation*, and so on. In the Output layout, we wanted to predict the unemployment rate in December 2018. In this use case, we want to focus on the explainability of the SHAP [9] values.

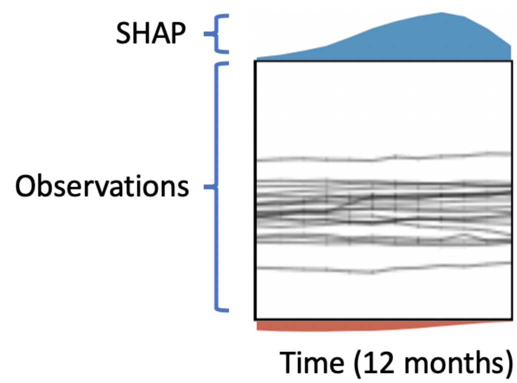


**Figure 6: Example of a neural node in our VixLSTM network: time series are presented in the box with the positive (blue) vs. negative (red) SHAP streams.**

Before getting into this use case, we discuss some design elements in our framework. Figure 6 shows an example of a node (in the input layer) in our network. As depicted, the time series (or observations) are presented as line graphs in the box. In this example, the time series is the monthly unemployment rate of a selected state (each year) on a given economic sector. To avoid occlusion, we use the space on the top and at the bottom to present the calculated SHAP values. In particular, the SHAP streams are the aggregated SHAP contributions of individual time series to the final prediction in the output layers. The blue stream shows a positive contribution, while the red stream shows a negative contribution. As we look at the monthly data, the blue shape indicates that the last few months (September, October, and November) have a significant positive contribution in predicting the unemployment rate of the selected state in December.

Figure 7 shows examples of the top three economy factor that contribute to Hawaii monthly unemployment rates (from left to right): *Wholesale trade*; *Transportation and Utilities*; and *Service Providing*. Hawaii has a large tourism industry relying on the season: Winter (high season) in January through March; Spring (low season) in April through June; Busy summer vacation in July and August; Autumn (off-season) in September to Mid-December. As depicted,

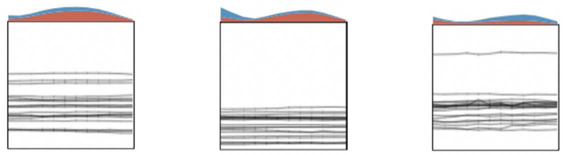

**Figure 7: The top three contributors to Hawaii unemployment rates (from left to right): Wholesale trade; Transportation and Utilities; and Service Providing.**

*Wholesale trade* has a high negative contribution in the Hawaii unemployment rates during the summer months while *Transportation and Utilities* vs. *Service Providing* have a similar contribution pattern in other months. Our SHAP value representation allows users o quickly capture and compare the trends.

**Figure 8: Our VixLSTM visual network focusing on Texas monthly employment over 20 years.**

Figure 8 shows the resulting model when a state is selected (Texas in this example). This LSTM model contains 3 LSTM layers (16-16-8) and 2 dense layers (8-4). From the SHAP value contributions, it is not surprising that farming-related sectors have substantial contributions to Texas's employment rate prediction. Notice that the neural nodes in intermediate layers have been ranked by correlations (similar nodes are put next to each other). By selecting different states, users can observe and compare various economic sectors' contributions to other states' employment rates in the US.

### 4.3 Reconstructing learning process with S&P500 stock data

For this data, we use stock attributes on the other days of the week to learn and predict Friday. The stock data has a dynamic nature; hence we focus on one-week data prior to making the predictions. Figure 9 depicts the structure and explainability of VixLSTM for the this data. Notice that we overlay the SHAP [9] values of different variables and time steps on the top (blue area) and at the bottom of each neural (red area). These streams become larger toward the beginning. This pattern indicates that the stock values on Monday are more important for predicting stock value on Friday.

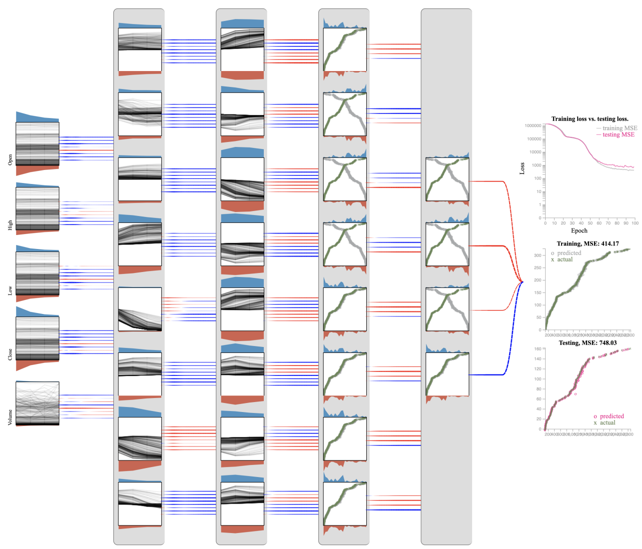

**Figure 9: VixLSTM visual network for the stock values: SHAP [9] values of different variables and time steps on the top (blue for positive contributions) and at the bottom (red for negative contributions) of each neural.**

In this use case, we want to show the evolution of weights in the model. During the learning process, the weights can change from positive (blue link segment) to negative (red link segment) or vice versa. We reconstruct the learning process by logging all epoch information.

Figure 10 shows how to simulate the learning process with the control buttons. These buttons allow users to go back and forth between different epoch. For example, when users move the time indicator (the vertical gray line) to the *epoch* with a significant drop, such as *epoch*=10. The LSTM network in Figure 9 (including the learned time series in each neural node and the weights of connecting lines between neural nodes) will be reverted back to show the current learning statuses at *epoch*=10. When users click on the play buttons, our VixLSTM network will stimulate the learning process in real-time. Notice that we plot the training loss vs. testing loss on a *log* scale, which allows us to see the changes on the high-end of the plot better.

r


**Figure 10: Play and skip button on training - testing losses chart: click on the play button to simulate the learning process, skip button to jump to the last epoch.**

## 5 CONCLUSION AND FUTURE WORK

This paper proposed an improved visual explainable approach for visualizing LSTM neural networks focusing on the multivariate time series. The proposed system allows the flexibility to customize the network architectures and settings and present how the new configuration impacts learning outcomes. VixLSTM also supports a full range of interactivities to rank, filter, and highlight various outstanding patterns in the dense network. The first use case depicts highlighting abnormal time series, which affects the overall accuracy of the learned model. The second use case presents how user users can filter and focus on temporal profiles. The last use case focuses on how SHAP values explain the temporal contributions vs. variable contributions in the final results. For future work, we will focus on using the SHAP explainability for the individual instance level. Moreover, we will investigate if the accuracy in the final results can highlight time-series outliers as these temporal profiles can significantly negatively impact the performance of the learned models.

## REFERENCES

[1] [n.d.]. HiperView: Visualizing and Monitoring Health Status of High Performance Computing Systems. https://idatavisualizationlab.github.io/HPCC/HiperView/demo.html.

[2] Roy Assaf and Anika Schumann. 2019. Explainable deep neural networks for multivariate time series predictions. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 6488–6490.

[3] Tommy Dang, Hao Van, Huyen Nguyen, Vung Pham, and Rattikorn Hewett. 2020. DeepVix: Explaining Long Short-Term Memory Network With High Dimensional Time Series Data. In *Proceedings of the 11th International Conference on Advances in Information Technology* (Bangkok, Thailand) *(IAIT2020)*. Association for Computing Machinery, New York, NY, USA, Article 38, 10 pages. https://doi.org/10.1145/3406601.3406643

[4] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. 2017. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems* 28, 10 (Oct 2017), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

[5] Tian Guo, Tao Lin, and Nino Antulov-Fantulin. 2019. Exploring Interpretable LSTM Neural Networks over Multi-Variable Data. *arXiv preprint arXiv:1905.12034* (2019).

[6] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[7] Andrej Karpathy et al. 2016. Cs231n convolutional neural networks for visual recognition. *Neural networks* 1 (2016).

[8] Jie Li, Ghazanfar Ali, Ngan Nguyen, Jon Hass, Alan Sill, Tommy Dang, and Yong Chen. 2020. MonSTer: An Out-of-the-Box Monitoring Tool for High Performance Computing Systems. In *Cluster'20: Proceedings of IEEE International Conference on Cluster Computing*. IEEE.

[9] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*. 4765–4774.

[10] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. 2015. Long short term memory networks for anomaly detection in time series. In *Proceedings*. Presses universitaires de Louvain, 89.

[11] Christoph Molnar. 2019. *Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/.

[12] Ngan Nguyen, Jon Hass, Yong Chen, Jie Li, Alan Sill, and Tommy Dang. 2020. RadarViewer : Visualizing the Dynamics of Multivariate Data. In *Practice and Experience in Advanced Research Computing* (Portland, OR, USA) *(PEARC '20)*. Association for Computing Machinery, New York, NY, USA, 555–556. https://doi.org/10.1145/3311790.3404538

[13] Ngan V. T. Nguyen, Bao D. Q. Nguyen, Tommy Dang, and Jon Hass. 2020. ScagnosticsViewer: Tracking Time Series Patterns via Scagnostics Meatures. In *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction* (Eindhoven, Netherlands) *(VINCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 30, 4 pages. https://doi.org/10.1145/3430036.3430072

[14] Chris Nicholson. 2019. A Beginner's Guide to LSTMs and Recurrent Neural Networks. *Skymind. Saatavissa: https://skymind. ai/wiki/lstm. Hakupäivä* 6 (2019), 2019.

[15] Daniel Soutner and Luděk Müller. 2013. Application of LSTM neural networks in language modelling. In *International Conference on Text, Speech and Dialogue*. Springer, 105–112.

[16] Joel Vaughan, Agus Sudjianto, Erind Brahimi, Jie Chen, and Vijayan N Nair. 2018. Explainable neural networks based on additive index models. *arXiv preprint arXiv:1806.01933* (2018).

[17] Leland Wilkinson. 2017. Visualizing Big Data Outliers through Distributed Aggregation. *IEEE transactions on visualization and computer graphics* (2017).