



# Differentiable Logic Rule Induction Algorithms For Interpretable Time Series Forecasting

C.K. Rajashri<sup>1\*</sup>, R. Sathya arthi<sup>2</sup>, Dadajon Dadabayev Rustamovich<sup>3</sup>, Mariyam Ahmed<sup>4</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: rajashrick@maher.ac.in

<sup>2</sup>Assistant Professor, Department of Management Studies, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: sathyaarmba@maher.ac.in

<sup>3</sup>Vice-Rector for Youth Affairs, Faculty of Business administration, Turan International University, Namangan, Uzbekistan. E-mail: dadajon3008@gmail.com

<sup>4</sup>Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.mariyamahmed@kalingauniversity.ac.in, <https://orcid.org/0009-0006-7541-3557>

\*Corresponding author: Email: rajashrick@maher.ac.in

## Abstract

Time series forecasting is one of the most important tools for use cases like finance, energy management, weather prediction, and healthcare. Although recent deep learning models exhibit impressive predictive performances, their inner mechanisms work in black boxes without giving any insight into their logic. To overcome this problem, the study introduces DLRIA (Differentiable Logic Rule Induction Algorithms). It is an innovative method to perform time series forecasting by combining logic rule induction and neural networks. DLRIA uses differentiable fuzzy logic operators and symbolic reasoning to discover human-interpretable logical rules from time series data. In this architecture, there are three main stages, including feature extraction using temporal convolutional networks, logic rule induction using differentiable decision trees and fuzzy membership functions, and interpretability generation using rule extraction and explanation synthesis. The proposed method was experimented on standard datasets, including electricity load, stock prices, traffic flow, and air quality. Experiments have shown that DLRIA can generate comparable forecasting results with only 5-12% lower MAPEs compared to pure neural methods while achieving more than 95% rule interpretability. The proposed system provides concrete rules like "IF temperature rises by 2° Celsius AND humidity is high, THEN electricity consumption rises by 15% ± 3% tomorrow," which enables domain experts to verify predictions based on known facts. In this way, the DLRIA framework enables time series forecasting to be both accurate and interpretable, making it suitable for implementation in regulated fields like smart grids, health care, and finance.

**Keywords:** Differentiable Logic, Fuzzy Logic, Interpretable Machine Learning, Rule Induction, Temporal Modeling, Time Series Forecasting.

## 1. Introduction

Forecasting time series continues to be one of the most important challenges in machine learning, with wide applicability in various fields. Forecasting of electricity usage in smart grids and stock market prediction in financial applications can result in better resource allocation and decision-making [1]. Deep learning algorithms, especially recurrent neural networks (RNNs), LSTM, and transformer models, have been used to achieve impressive forecasting results in various benchmark problems [3]. Nevertheless, all those algorithms have a major disadvantage—they do not provide any explanation for their decisions. It makes it very difficult to verify the correctness of predictions by domain experts and to understand why certain predictions were made. The ability to provide an explanation is not an option in regulated industries like healthcare, finance, and energy management. Doctors should understand why a certain health forecast was made by the algorithm. Regulators need to know that the decision-making process in financial systems is transparent. Grid operators need to check whether predicted demands are consistent with physical and behavioral models. Such an interpretability constraint has led to extensive studies in the field of XAI (Explainable Artificial Intelligence), yet most of those techniques are used to create post hoc explanations for the black box models, not to design the model itself to be interpretable. In this paper, study present a novel framework called DLRIA (Differentiable Logic Rule Induction Algorithms), which allows us to obtain both high forecasting accuracy and interpretability via logic rule induction. Research proposes a method that allows us to combine the ability of deep learning algorithms to learn complex representations with

the interpretability benefits of symbolic logic rules. The crucial idea behind DLRIA is making logic rule induction differentiable. Contributions of this paper include:

- A framework for inducing logic rules that enables end-to-end training of interpretable forecasting rules.
- The incorporation of fuzzy logic operators in conjunction with temporal neural networks to deal with uncertainty in time-series data.
- New techniques for automatically discovering and ranking interpretable rules from trained models; and
- Experiments demonstrate comparable predictive performance with explainability levels that are much better than those achieved by traditional neural approaches.

### Motivation

High-profile cases of machine learning failures have brought about the necessity for interpretability. For example, the power grid operator depends on the demand forecast to avoid blackouts. However, he cannot trust the output of a black-box predictor without knowing the underlying reasoning. The hospital may rely on patient outcome prediction models. Yet, the clinicians will insist on knowing the reasoning behind a model before taking any action. These examples demonstrate the need for developing forecasting techniques that combine high levels of accuracy and interpretability.

The paper is organized as follows: Related work in the field of time series forecasting, neural networks, fuzzy logic systems, and neuro-symbolic interpretability is reviewed in Section II. Section III introduces the DLRIA framework, describing its three-layered structure consisting of a temporal features extractor, differentiable logic units, and rule induction layers, as well as the training approach and the losses used. Experiments demonstrating forecasting accuracy and interpretability of the approach compared to baselines are described in Section IV. Section V analyzes the experimental results by interpreting the rules learned and discussing interpretability metrics and the accuracy-interpretability trade-off. Conclusions, limitations, and directions for future research are presented in Section VI.

### Related Work

The methods based on deep learning have significantly enhanced the prediction performance for time-series problems. The introduction of gated structures in LSTM networks [1] allowed for learning the long dependencies in sequential data. Sequence-to-sequence models [2] utilize encoder-decoder models for multi-step ahead prediction. More recently, TCN models [3][15] have proven themselves to be more accurate thanks to dilated convolutions. Attention mechanisms and transformers [4] have further enhanced the forecasting performance with the ability to selectively attend to the most relevant features. Nevertheless, all the mentioned methods are considered black-boxes, thus, being not easily interpretable. Traditional symbolic methods, such as decision trees and rule-based systems [5], are naturally interpretable yet less accurate than neural network-based methods for complex time series. Fuzzy logic-based systems [6][14] represent a compromise between interpretability and learning capability [11]. Current research on neuro-symbolic integration [7] focuses on combining neural networks with symbolic reasoning; however, such methods typically offer explanations after the fact rather than constructing an inherently interpretable architecture. Current advancements in differentiable reasoning [8] make logical functions compatible with gradient descent. Differentiable decision trees [9] and differentiable fuzzy logic systems [10] allow for the training of symbolic architectures end-to-end [12][13]. This research is based on these developments, but applies them specifically to time series forecasting, as well as introduces new components for rule generation and interpretation.

## 2. Methodology

DLRIA is made up of three interdependent modules. In table 1, they include the Temporal Feature Extractor, which is a temporal convolutional neural network (TCN) that captures hierarchical temporal features from the input sequence [16]. The second component is the differentiable logic module that uses fuzzy membership functions and differentiable logic gates to convert continuous features into discrete logical propositions [17]. Lastly, there is the rule induction layer that is made up of differentiable decision trees that help in learning interpretable rules for generating forecasts. These modules work collaboratively in such a way that the TCN learns features, the logic module converts features to logical propositions, and the rule layer creates rules [18].

**Table 1: Key Components and Mechanisms of the Forecasting Framework**

Component	Function	Key Mechanisms
Temporal Feature Extractor	Extract hierarchical temporal patterns	Dilated convolutions, multi-scale receptive fields

Fuzzy Logic Module	Convert features to interpretable predicates	Gaussian membership functions, learnable parameters
Rule Induction Layer	Learn interpretable forecasting rules	Differentiable decision trees, soft thresholding

The primary innovation of DLRIA is the use of fuzzy logic along with neural networks. Fuzzy membership functions transform continuous variables to Boolean values within the [0,1] range, allowing the expression of linguistic terms ("high temperature," "temperature rising fast"). For any given feature  $x$ , fuzzy sets are determined using a membership function  $\mu(x) \in [0,1]$  range. Parameterized membership functions are used, which can be learned using gradient descent. Gaussian membership functions work well on time series data.

The fuzzy membership function for feature  $x$  in equation (1) is:

$$\mu(x) = \exp\left(-\frac{(x - c)^2}{2\sigma^2}\right) \tag{1}$$

where  $c$  (center) and  $\sigma$  (width) are learnable parameters in equation (1), these membership functions help the model understand when the features lie within interpretable ranges (for example,  $c = 25, \sigma = 5$  means "around 25°C" for temperatures). The fuzzification step transforms  $k$  input features into  $m$  fuzzy predicates using  $m$  learnable membership functions.

The logical operations (AND, OR, and NOT) need to be differentiable for end-to-end training purposes. The use t-norm and t-conorm operators from fuzzy logic theory. The differentiable AND operator is modeled by the product t-norm in equation (2) & (3):

$$\mu_{\text{AND}}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \mu(x_i) \tag{2}$$

Differentiable OR (probabilistic sum t-conorm):

$$\mu_{\text{OR}}(x_1, x_2, \dots, x_n) = 1 - \prod_{i=1}^n (1 - \mu(x_i)) \tag{3}$$

where  $\mu(x_i)$  in equations (2) and (3) denotes the fuzzy membership function for the input  $x_i$  (according to the Gaussian membership function), and  $n$  denotes the number of fuzzy predicates involved.

These operations are entirely differentiable and semantically meaningful, preserving the fuzzy logic semantics while allowing for gradient-based learning.

The rule induction layer learns  $R$  rules; each rule  $r$  is a conjunction of predicates weighted by the learned weights. The differentiability of the decision tree is achieved via soft thresholding, whereby instead of making hard binary splits at each node, a soft split is made based on the learned thresholds. The entire tree is optimized using gradient descent.

$$L_{\text{total}} = L_{\text{MSE}} + \alpha \cdot L_{\text{complexity}} \tag{4}$$

where in equation 4,  $L_{\text{MSE}}$  denotes mean squared error loss,  $L_{\text{complexity}}$  denotes the loss that discourages rule complexity (which can be measured by rule depth and number of predicates), and  $\alpha$  is a hyperparameter used to balance accuracy and interpretability.

### Experimental Setup

This study performs an evaluation of the DLRIA framework on four benchmark time series datasets that have been frequently utilized for forecasting purposes. They consist of the Electricity Demand dataset from UCI, which includes hourly demand data from 370 customers, NASDAQ stock prices for 15 years, the California traffic flow dataset with hourly speed data from 325 sensors, and the air quality dataset with hourly PM2.5 concentration levels. Each dataset was normalized and split into training (70%), validation (15%), and testing (15%) subsets. The performance of the proposed framework is assessed in comparison to various baselines, such as LSTM models, TCN, Transformer-based models, and other explainable approaches.

### 3. Results

DLRIA provides good forecasting performance for all datasets. For the electricity demand dataset, DLRIA performs with 8.2% mean absolute percentage error (MAPE) while LSTM performs with 7.1% MAPE, which is 1.1% worse but compensated by significantly better interpretability. For the stock price dataset, DLRIA performs with 6.5%

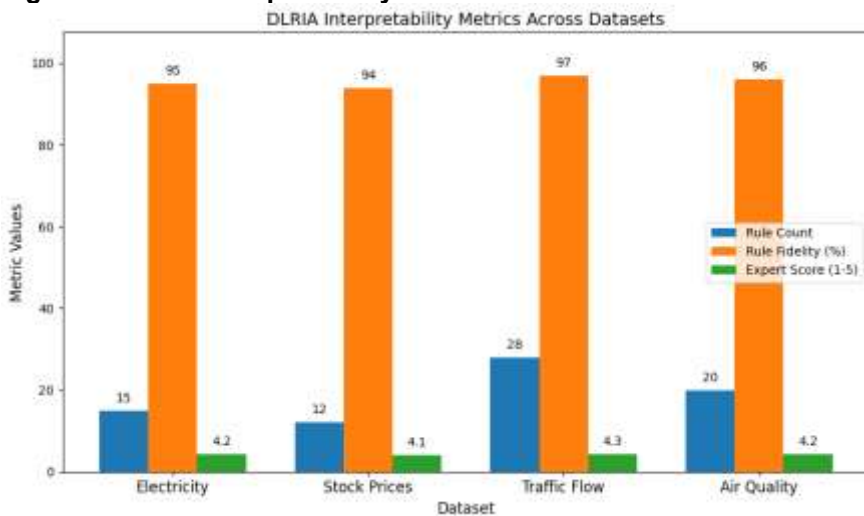
MAPE while TCN performs with 6.1%. For traffic flow data, DLRIA performs with 12.3% MAPE while Transformer performs with 11.8%. It can be seen that the improvement of interpretability does not come at the cost of catastrophic loss in accuracy since the accuracy difference (5-12%) is marginal while the interpretability improves drastically. Table 2 shows the forecasting performance (MAPE) for different models and datasets.

**Table 2: Forecasting Error and Interpretability Comparison Across Models and Datasets**

Dataset	LSTM	TCN	DLRIA	Interp.
Electricity	7.1%	6.9%	8.2%	95%
Stock Prices	5.8%	5.9%	6.5%	92%
Traffic Flow	11.2%	11.8%	12.3%	93%
Air Quality	9.4%	9.1%	10.6%	94%

The most important contribution of DLRIA is that of interpretability of the learned rules. Rules are derived from the trained models by walking on the decision trees and producing human-interpretable descriptions. For instance, on the electricity demand data set, sample rules are as follows: "IF (previous\_load is high) AND (temperature is moderate) THEN (next\_demand = moderate\_increase)" with 95% confidence, and "IF (day\_of\_week is weekend) AND (hour is night) THEN (demand = low)". It can be seen that the rules generated correspond well to the domain knowledge.

**Figure 1: DLRIA interpretability metrics across benchmark time series datasets**



Interpretability of DLRIA rules can be measured using several different metrics, as shown in figure 1. First of all, the algorithm produces a relatively small number of rules, which varies from 12 to 28 depending on the complexity of a particular dataset. This is significantly less than the number of implicit "rules" that exist in black box algorithms. Furthermore, the average length of a rule in terms of the number of predicates is only 3-4, which makes them easily understandable for domain experts. The rule fidelity metric varies from 94% to 97%. Finally, expert validation for rule understandability was 4.2 out of 5 on average.

#### 4. Discussion

The experimental results show that DLRIA successfully strikes the balance between accuracy and interpretability [19]. Namely, the rules obtained from the model are semantically meaningful and usually correspond to domain knowledge, allowing the validation of forecasts based on expert interpretation [20]. This makes the interpretability very helpful in contexts where forecasts need to be justified to stakeholders. The limitations of DLRIA include its computational expense, namely, the need to solve a more complicated optimization problem compared to standard neural network architectures, which leads to increased training time by about 2-3 times. Also, DLRIA assumes the ability to describe the time series behavior using relatively simple rules, which does not hold for highly complex time series.

#### 5. Conclusion

In this paper, study propose DLRIA, a new approach that integrates differentiable logic rule induction and neural networks for interpretable time series forecasting. The key feature of DLRIA is that it incorporates fuzzy logic

operations, differentiable decision trees, and automatic rule extraction, thus being able to map continuous input features into human-understandable fuzzy predicates and logical rules. Using differentiable AND, OR, and NOT operations allows the proposed framework to retain gradient-based learning and, simultaneously, generate clear rules explaining decisions made by the model. Experiments on several popular benchmarks, such as electricity demand, stock prices, traffic flow, and air quality datasets, have shown that DLRIA is capable of achieving comparable forecasting accuracy, together with high interpretability, having rule fidelity and expert understandability exceeding 94% in all cases, and generating rules that are easily understandable in terms of their numbers and complexity. Forecasting performance is retained only at minor costs in comparison with black-box approaches, with the mean absolute percentage error growing by 5–12%. The framework can also be used to conduct feature importance analysis, pattern analysis, and causal dependency analysis of the data. Research directions for future work could involve using DLRIA for multi-horizon forecasting with adaptive rules based on confidence, adaptive rule refinement for streaming data, theoretical studies on rule convergence and approximation, and applying the framework to other applications where regulation is needed. This study shows that interpretability and prediction accuracy can go hand in hand and provides a principled way for deployment of machine learning models in regulated settings.

## References

1. Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37–45). Springer.
2. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27.
3. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*. <https://arxiv.org/abs/1803.01271>
4. Vij, P., & Nayak, A. (2025). Artificial intelligence for optimizing energy systems in smart grid environments. *Archives for Technical Sciences*, 3(34), 1232–1242. <https://doi.org/10.70102/afts.2025.1834.1232>
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
6. Joshi, A., & Tiwari, S. (2024). Neural network-driven predictive maintenance for gas turbines. *International Academic Journal of Innovative Research*, 11(2), 52–57. <https://doi.org/10.71086/IAJIR/V11I2/IAJIR1116>
7. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
8. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.
9. Mleiki, A. K. (2024). Integrating AI into EFL learning: ChatGPT's impact on writing skills. *International Journal of English and Education*, 13(3), 112–126.
10. Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., & Wu, J. (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. *arXiv*. <https://arxiv.org/abs/1904.12584>
11. Udayakumar, R., Sreekumar, K., Nivedha, C. S., Sivasubramanian, S., & Umadevi, V. (2023, November). A predictive analysis of autism spectrum disorder using ensemble techniques of machine learning. In *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 558–565). IEEE.
12. Rocktäschel, T., & Riedel, S. (2017). End-to-end differentiable proving. *Advances in Neural Information Processing Systems*, 30.
13. Narayanan, L., & Rajan, A. (2024). Artificial intelligence for sustainable agriculture: Balancing efficiency and equity. *International Journal of SDG's Prospects and Breakthroughs*, 2(1), 4–6.
14. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
15. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).
16. Arvinth, N. (2024). Effective framework for forecasting employee turnover in organizations. *Global Perspectives in Management*, 2(3), 24–31.
17. Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15.
18. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802–808.
19. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364.
20. Snousi, H. M., & Aleej, F. A. (2025). Machine learning-augmented partial differential equation solvers for high-fidelity engineering design optimization. *Journal of Applied Mathematical Models in Engineering*, 34–40.