CHAPTER 52

Neurosymbolic AI for Autonomous Causal Discovery

Suraj Prajapati and Rajashree Khandekar***

ABSTRACT

The quest to derive causal insights from observational data represents a fundamental objective within artificial intelligence and data science. Contemporary techniques, including causal structure learning and neural network-based pattern detection, are capable of uncovering correlations and inferring potential causal directions. However, these methods predominantly depend on pre-defined variables and substantial human input for generating and validating hypotheses. This study introduces an innovative framework that merges the sub symbolic processing capabilities of deep neural networks with the explicit, rule-based reasoning of symbolic artificial intelligence to establish a self-sufficient causal discovery mechanism. The envisioned neurosymbolic system is engineered to explore high-dimensional, multi-modal datasets, produce viable causal hypotheses, formalize them within structured knowledge graphs, and independently verify these models through counterfactual analysis and symbolic logic. A prototype application, focused on complex medical data, seeks to reveal hidden causal pathways associated with diseases without requiring prior human annotation of relationships. Preliminary experiments suggest the capability to detect original, verifiable causal sequences, advancing from mere association to genuine explanatory power. This investigation underscores the revolutionary capacity of neurosymbolic AI to automate the process of scientific discovery, carrying significant consequences for areas including healthcare, economics, and social sciences.

Keywords: Causal discovery; Neurosymbolic AI; Observational data; Knowledge graphs; Counterfactual analysis.

1.0 Introduction

Disentangling causation from correlation is essential for scientific advancement. In the context of big data, machine learning algorithms demonstrate remarkable proficiency in detecting complex patterns and predictive relationships.

**Corresponding author; Postgraduate Student, Department of Computer Applications, Dr. Moonje Institute of Management and Computer Studies, Nashik, Maharashtra*
(E-mail: surajprajapati7509@gmail.com)
***Postgraduate Student, Department of Computer Applications, Dr. Moonje Institute of Management and Computer Studies, Nashik, Maharashtra*
(E-mail: rajashreekhandekar5@gmail.com)

DOI: 10.17492/JPI/ICISI2025/251252 *ISBN: 978-93-49790-69-8*

Nonetheless, these models typically function as opaque systems, offering limited transparency regarding the underlying reasons for identified associations. Causal discovery approaches, such as those utilizing Bayesian networks or structural equation modelling, supply a reasoning foundation but are frequently hampered by computational limitations and a need for expert-crafted variables and constraints. These restrictions curtail their scalability and operational independence. Neurosymbolic artificial intelligence emerges as a unifying methodology, blending the statistical learning advantages of neural networks – sub symbolic processing – with the clear, logical deduction inherent to symbolic systems, such as knowledge graphs and automated theorem provers. This project confronts the pivotal divide between pattern detection and causal interpretation by crafting a neurosymbolic framework equipped for independent causal discovery. By employing neural networks to distil features and potential relational structures from raw data, alongside symbolic AI to construct, refine, and authenticate causal hypotheses, this strategy intends to mechanize the preliminary phases of scientific investigation, thus hastening insight generation in data-intensive fields.

2.0 Literature Review

The domain of causal inference has been the subject of substantial academic inquiry. Seminal contributions concerning causal diagrams and the "do-calculus" (Pearl, 2009) established a mathematical basis for causal reasoning, albeit necessitating a pre-established causal graph. Algorithms dedicated to structure learning, including the PC and FCI algorithms, possess the ability to deduce graphs from datasets but encounter difficulties with high-dimensional information and unobserved confounders (Spirtes et al., 2000). Concurrently, deep learning architectures have attained exceptional performance in recognizing patterns (LeCun et al., 2015), but continue to exhibit deficiencies in symbolic processing and logical deduction.

The neurosymbolic AI discipline has risen as a response to this divide. Contemporary initiatives (for instance, Garcez & Lamb, 2020) have concentrated on merging these paradigms for applications like visual question answering, though their use in self-governing causal discovery from unprocessed data remains underdeveloped. Present-day AI platforms designed for scientific purposes, including those that automate experimental procedures, still depend on hypotheses formulated by humans. This research endeavour strives to amalgamate these diverse developments – causal inference, deep learning, and neurosymbolic unification – to produce a system that completes the cycle from data acquisition to causal theory formulation without necessitating ongoing human involvement.

DOI: 10.17492/JPI/ICISI2025/251252 *ISBN: 978-93-49790-69-8*

3.0 Research Objectives

The central aims of this investigation are:

- To conceive an original neurosymbolic structure for the self-generating production of causal models from observational data.
- To formulate algorithms that convert subsymbolic neural outputs into organized symbolic causal graphs.
- To create a symbolic reasoning component capable of postulating causal mechanisms, producing counterfactual scenarios, and eliminating unrealistic models.
- To appraise the system's effectiveness and operational efficiency in comparison to conventional causal discovery techniques using both artificially generated and authentic biomedical data sources.
- To examine the originality and practical significance of causal hypotheses produced from an intricate disease-related dataset.

4.0 Methodology

This research will employ a design science methodology, centring on the creation and assessment of a prototype system:

- *System Architecture:* The prototype will incorporate three fundamental modules: (a) a Neural Feature and Relationship Extractor employing transformers and graph neural networks to handle raw data; (b) a Symbolic Translator that converts neural signals into probabilistic logical statements; and (c) a Causal Reasoner that uses symbolic constraint resolution mechanisms and causal calculus to construct and evaluate directed acyclic graphs (DAGs).
- *Data:* The system will undergo training and testing utilizing both artificially created datasets (with predefined causal frameworks) and actual, anonymized electronic health record (EHR) information.
- *Evaluation:*
    - *Quantitative:* Success will be gauged by the structural precision (for example, via Structural Hamming Distance) of the derived graphs compared to established graphs on synthetic data. Measures of computational performance (duration and memory usage) will be documented.
    - *Qualitative:* A group of field specialists (such as medical investigators) will conduct blind assessments of the innovativeness and validity of causal hypotheses derived from the EHR data, contrasting them with outputs from standard methods.

DOI: 10.17492/JPI/ICISI2025/251252 *ISBN: 978-93-49790-69-8*

5.0 Proposed System / Model

The envisaged neurosymbolic causal discovery system is founded on a recursive structure:

1. *Perception Module (Sub symbolic):* A deep neural network (for instance, a Variational Autoencoder featuring a structured latent space) analyses unprocessed input data (like patient histories) to acquire condensed representations and preliminary dependency frameworks.
2. *Abstraction Module (Symbolic Translation):* This component assigns symbolic meaning to the neural network's results. It detects essential entities and ideas from the latent domain and depicts their interconnections as logical expressions inside a knowledge graph.
3. *Reasoning Module (Symbolic):* A symbolic AI mechanism, furnished with causal principles and limitations (such as temporal sequence and acyclicity), functions on this knowledge graph. It produces prospective causal DAGs, discards those breaching logical rules, and suggests interventions or counterfactuals to examine particular causal connections.
4. *Learning Loop:* Outcomes from the symbolic reasoning (like a filtered collection of feasible DAGs) are reintroduced to direct the neural network's focus and learning progression, establishing a (closed-loop) mechanism for repeated hypothesis enhancement.

6.0 Expected Outcomes

This study forecasts the following results:

- An operational prototype of an independent neurosymbolic causal discovery tool.
- A noticeable advancement in recognizing authentic causal frameworks from high-dimensional data relative to strictly statistical or purely symbolic reference methods.
- The production of a minimum of one new, clinically reasonable causal hypothesis for a multifactorial disease from EHR information.
- A standardized structure for merging subsymbolic and symbolic operations for causal assignments.
- Novel standards for assessing autonomous causal discovery platforms.

7.0 Discussion

The proposed framework signifies a transformational change from instruments that facilitate causal discovery to an autonomous entity that propels it. By utilizing neural

DOI: 10.17492/JPI/ICISI2025/251252 *ISBN: 978-93-49790-69-8*

networks to manage the intricacy of raw data and symbolic AI to enforce logical strictness, it holds potential to surmount the drawbacks of existing techniques. The possible applications are extensive, ranging from hastening medical investigations and pharmaceutical development to shaping economic strategies. Still, considerable obstacles persist. The "symbol grounding problem" – the accurate correspondence of neural activations to semantic notions – is highly complex. Guaranteeing that the formulated causal models are both precise and comprehensible to human specialists is vital for credibility and utilization. Moreover, the ethical considerations of an AI producing causal assertions, especially in delicate sectors like healthcare, demand strong verification protocols and human supervisory involvement for ultimate endorsement. This research intends not to substitute human researchers but to enhance their competencies by automating the initial, frequently monotonous, steps of hypothesis formation.

8.0 Conclusion

This research puts forward a pioneering neurosymbolic structure for independent causal discovery. It tackles a crucial shortcoming in modern AI – the separation between correlation and causation – through the synergistic merger of the pattern detection prowess of neural networks with the deductive reasoning capacity of symbolic AI. The projected system strives to spontaneously create and examine causal models from intricate observational data, presenting a potent instrument for scientific investigation. Although obstacles in integration, assessment, and ethics continue, this effort establishes the groundwork for a new category of AI systems that can prognosticate, elucidate, and rationalize phenomena, eventually quickening the rate of scientific advancement and insight.

References

1. Garcez, A. d., & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd Wave. *Artificial Intelligence Review*.
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.
3. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
4. Sprites, P., Glymour, C. N., & Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press.

DOI: 10.17492/JPI/ICISI2025/251252 *ISBN: 978-93-49790-69-8*