Exploring Causality and Explainability in Time Series Models

Masters Thesis Presentation

by

Ankan Kar, MCS202303

Thesis Advisor: Prof. K.V. Subrahmanyam

at

Chennai Mathematical Institute

on

May 18, 2025

Overview

Introduction

Causal inference is a key concept in the field of Statistics, Mathematics and Computer Science, using which we are able to determine the cause-and-effect relationships from observational data.

- For time-independent variables, such as X and Y, causality is inferred by analyzing asymmetry in evidence for models in the directions $X \to Y$ and $Y \to X$.
- In time-dependent data, causality is determined by combining asymmetry with temporal precedence and accounting for lagged dependencies, often using time-embedded causal graphs and dynamic models.

Causal inference in time series analysis plays a vital role in understanding the dynamic relationships between variables over time. Unlike simple correlations, this approach aims to uncover genuine cause-and-effect relationships, revealing the mechanisms driving observed phenomena.

However, the complexity of interrelated variables, external influences, and temporal dependencies makes establishing causality a challenging task. Advanced methodologies are essential to address these intricacies and guide informed decision-making.

Challenges in Time Series Causality:

- Complexity arises from interdependencies among variables and external factors.
- Traditional methods may fail to address these intricacies, leading to misinterpretations.
- Requires sophisticated approaches to disentangle causal relationships.

Methodologies for Addressing Challenges:

- Graphical models and recursive structures help capture dependencies.
- Directed acyclic graphs (DAGs) represent causal relationships effectively.
- Facilitate understanding of causal orders and dependencies.

Importance of Temporal Precedence:

- Ensures that causes precede effects in causal analysis.
- Combines temporal information with patterns of association for accuracy.
- Embeds time series data within DAGs to visualize and analyze causality.

Applications and Implications:

- Enhances empirical research by uncovering causal mechanisms in time series data.
- Provides a better understanding of complex systems.
- Guides informed decision-making across various domains.

Focus of Presentation

Focus of Presentation:

- Principles of causal inference in time series analysis.
- Discussion of methodologies for identifying causal relationships in time dependent variables.
- Find out Causal relation as well as the Causal strength.

Causal Models

The theory of inferred causation uses directed acyclic graphs (DAGs) to model causal relationships between variables. Arrows in the graph indicate causality, and by analyzing conditional independencies, we can infer the graph's structure and the direction of causation based on observed data.

Key concepts in this theory involve defining causal structures and models as follows:

Definition (Causal Structure in Pearl (2000) p.44)

A causal structure of a set of variables V is represented as a directed acyclic graph (DAG) where each node corresponds to a distinct variable in V, and each link indicates a direct functional relationship among the corresponding variables.

Definition (Causal Model in Pearl (2000) p.44)

A causal model is defined as a pair $M = \langle D, \Theta \rangle$, consisting of a causal structure D and a set of parameters Θ_D that are compatible with D. The parameters Θ_D assign a function $x_i = f_i(pa_i, u_i)$ to each variable $X_i \in V$ and a probability measure $P(u_i)$ to each random disturbance u_i, where P_A denotes the parents of X_i in D and each U_i is independently distributed according to $P(u_i)$.

Observational Equivalence

Proposition (VermaPearl1990)

Two directed acyclic graphs (DAGs) (models) are observationally equivalent if and only if they have the same skeletons and the same sets of v-structures, that is, two converging arrows whose tails are not connected by an arrow.

- Statistical methods are limited in inferring causal directions in DAGs due to observationally equivalent models.

- Only v-structures or causal directions creating new v-structures or cycles are inferrable.

- Some arrow directions in a DAG cannot be uniquely determined from data.

PC Algorithm

Here is an algorithm that uses the Causal DAG Model to infer causal relations:

Algorithm 1 PC Algorithm

Input: Observations of a set of variables X generated from a DAG model.
Output: A pattern (DAG) compatible with the data generating DAG.
Start with a full undirected graph.
for each pair of variables $(X_i, X_j) \in X$ **do**
  Search a subset $S_{ij} \subseteq X \setminus \{X_i, X_j\}$ such that $X_i \perp X_j | S_{ij}$ holds.
  Delete the edge between X_i and X_j.
end for
for each pair of non-adjacent variables X_i and X_j with a common neighbor X_k **do**
  if $X_k \in S_{ij}$ **then**
    Continue.
  else
    Add arrowheads pointing as $X_k : (X_i \rightarrow X_k \leftarrow X_j)$.
  end if
end for
In the partially directed graph that results, orient as many of the undirected edges as possible subject to: (i) The orientation should not create a new v-structure, (ii) The orientation should not create a directed cycle.

PC Algorithm at Work

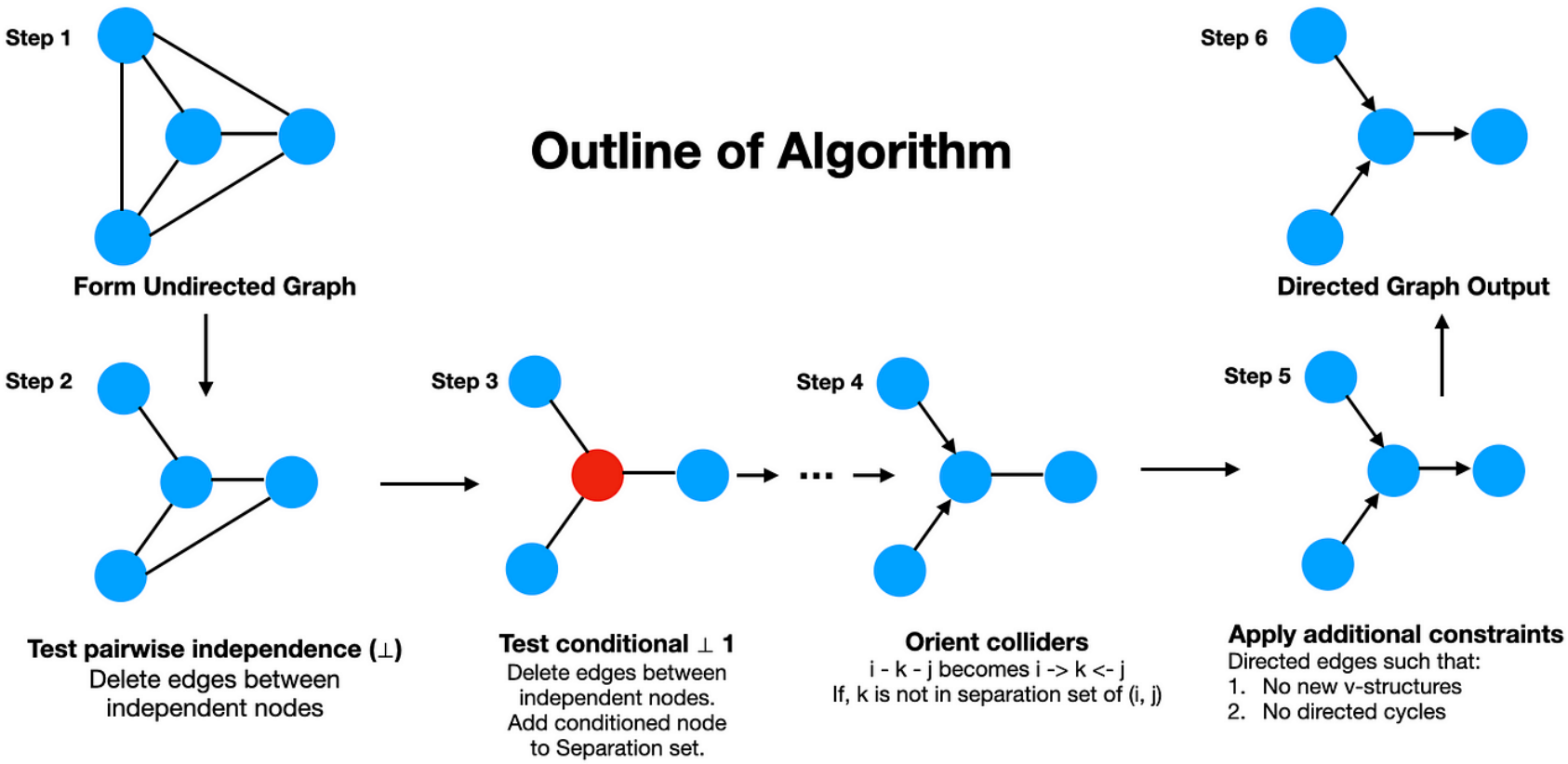

Figure 1: [1]Outline of PC Algorithm

[1]Image taken from *Causal Discovery Learning causation from data using Python*, https://towardsdatascience.com/causal-discovery-6858f9af6dcb

The PC algorithm's tests are designed to be consistent, meaning that as the number of observations increases and the significance level approaches zero, the likelihood of correctly identifying edges in the graph becomes nearly certain.

Proposition

Under the assumption of faithfulness, the PC-algorithm can consistently identify the inferrable causal directions, i.e. for $T \to \infty$ the probability of recovering the inferrable causal structure of the data generating causal model converges to one.

Greedy Search Algorithm

Here is an another algorithm that is used to infer causal relation using causal DAG model:

Algorithm 2 Greedy Search Algorithm

Input: Observations of a set of variables X generated from a DAG model.
Output: A pattern (DAG) compatible with the data generating DAG.
Step 1: Start with a DAG A_0.
Step 2: Calculate the score of the DAG according to BIC/AIC/likelihood criterion.
Step 3: Generate the local neighbor DAGs by either adding, removing, or reversing an edge of the network A_0.
Step 4: Calculate the scores for the local neighbor DAGs. Choose the one with the highest score as A_n.
if the highest score is larger than that of A_0 **then**
    Update A_0 with A_n and go to Step 2.
else
    Stop and output A_0.
end if

It's important to recognize that a causal model functions as a statistical model. When the score used in the greedy search algorithm is based on a consistent model selection criterion, such as the Bayesian Information Criterion (BIC), the algorithm will reliably recover the inferable causal directions, assuming that the search space encompasses the true directed acyclic graph (DAG). BIC is a method that helps to identify the best-fitting model while penalizing for the number of parameters to avoid overfitting.

Time Series Causal Model

DAGs and Recursive Structural Models

If an n-dimensional variable X is jointly normally distributed, a linear causal model for X is equivalent to a linear recursive structural equation model (SEM). In this framework, each variable is defined as a function of its parent variables, represented by the equation:

$$x_j = \sum_{k=1}^{j-1} a_{jk} x_k + u_j \quad \text{for } j = 1, 2, \ldots, n$$

Here, x_j is the j-th variable, a_{jk} are coefficients showing the causal influence of parent variables x_k, and u_j is a normally distributed error term capturing other influences.

We summarize this equivalence in the following proposition.

Proposition

If a set of variables X are jointly normal $X \sim N(0; \Sigma)$, a linear causal model for X can be equivalently formulated as a linear recursive structural equation model (SEM) that is represented by a lower triangular coefficient matrix A with ones on the principal diagonal. Any nonzero element in this coefficient matrix, say α_{jk}, corresponds to a directed edge from variable k to variable j.

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & 1 \end{pmatrix}$$

where A is the triangular decomposition matrix of Σ with $A\Sigma A' = \Lambda$ and Λ is a diagonal matrix.

TSCM

- The linear causal model is designed for independent data, but economic time series data are inherently dependent.
- Treat N time series, each with T observations, as realizations of NT random variables.
- These random variables can be integrated into a larger recursive structural equation model.
- Assuming the elements of the multivariate time series X_{it} (where $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$) are jointly normally distributed.
- Apply Proposition 2.2 to state that a causal model for the multivariate time series is a linear recursive structural model encompassing all NT components.

In this context, temporal information dictates a natural causal order, so the recursive structural model must respect this temporal sequence. We can express this recursive system as:

$$\begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_{T1} & A_{T2} & \cdots & A_{TT} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

- Here, $\epsilon_t \sim N(0, D)$ represents a vector of independent residuals, where D is a diagonal matrix.
- Additionally, the residuals ϵ_t and $\epsilon_{t-\tau}$ are independent.
- The random vector $X_t = (X_{1t}, X_{2t}, \ldots, X_{Nt})'$ captures the values of the time series at time t.

This structure allows us to model the dependencies among time series while respecting the temporal order of the data.

- Due to the limitation of having only one observation at each time point, the recursive system has too many parameters for effective statistical analysis.
- To make the system statistically manageable, reasonable constraints must be imposed on its parameters.
- Three key assumptions:
  - **Temporal Causal Constraint:** Causal relationships between variables remain consistent over time.
  - **Time-Invariant Causal Structure Constraint:** The causal structure at different time points is the same.
  - **Time-Finite Causal Influence Constraint:** A variable X_t can only influence $X_{t+\tau}$ if $\tau \leq p$, where p is a finite integer.

By applying these constraints, we can rewrite the linear recursive system for $p = 2$ as below.

$$\begin{pmatrix} A_0 & 0 & \cdots & & \cdots & 0 \\ A_1 & A_0 & 0 & \cdots & & 0 \\ A_2 & A_1 & A_0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & A_2 & A_1 & A_0 & 0 \\ 0 & \cdots & 0 & A_2 & A_1 & A_0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{T-1} \\ X_T \end{pmatrix} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{T-1} \\ \epsilon_T \end{pmatrix}$$

- The parameter matrices $A_1, A_2, \ldots, A_p$ in the t-th row indicate the causal influence of $X_{t-1}, \ldots, X_{t-p}$ on X_t.
- A_0 represents the contemporaneous causal influences among the elements of X_t.
- The time-finite constraint ensures that all parameter sub-matrices to the left of A_p are zero in each row.
- The causal model represented in this way is called a **Time Series Causal Model (TSCM)**.

Examples

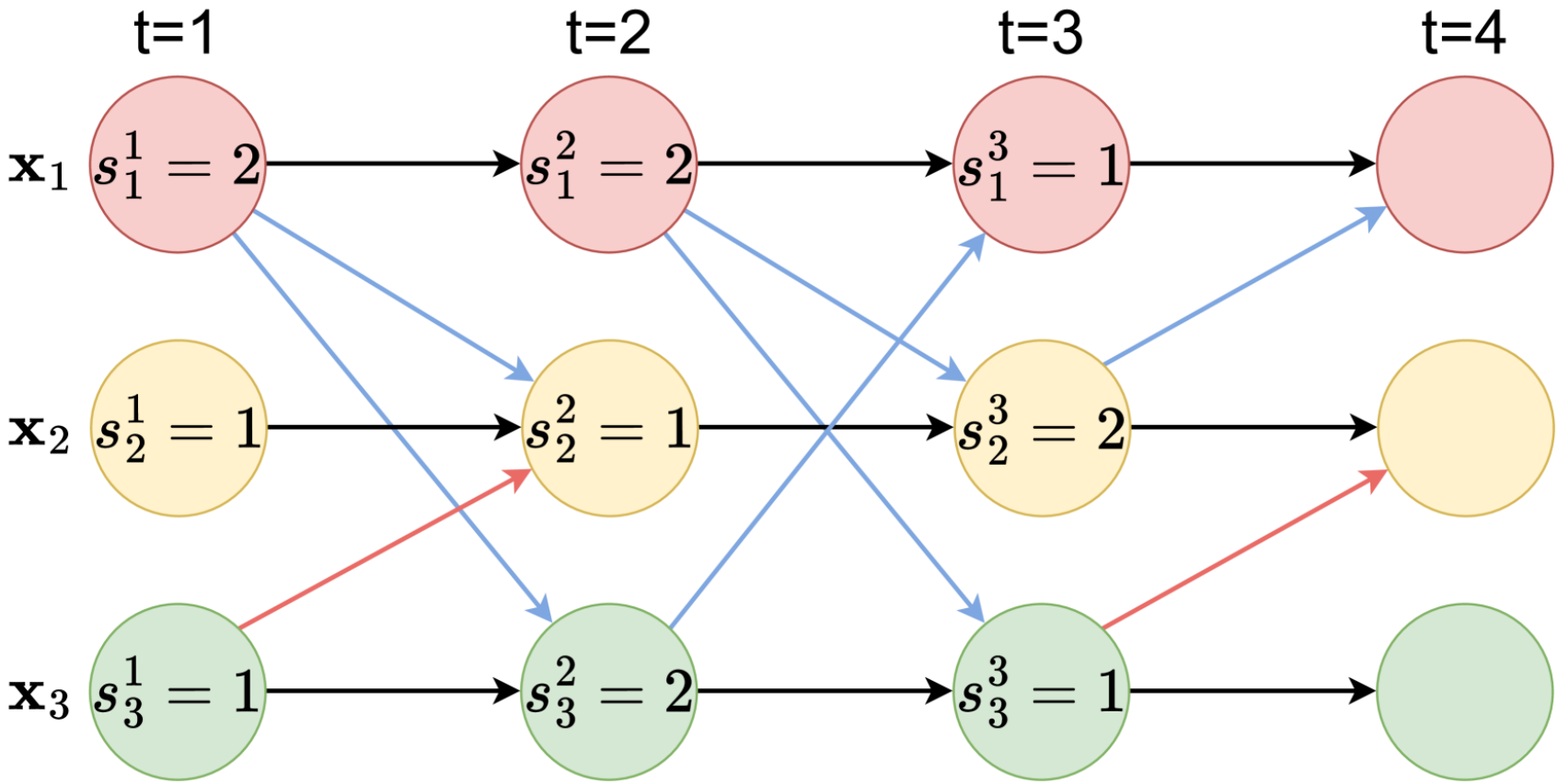

Figure 2: [2]Time Series Causal Model

[2]Image taken from *Causal Discovery from Conditionally Stationary Time Series, arXiv: https://arxiv.org/abs/2110.06257*

Observational Equivalence in TSCMs

Proposition

A partial Directed Acyclic Graph (DAG) has an observationally equivalent model if there are some arrows between elements of X_t that satisfy the following two conditions:

- The lagged parents of the connected elements of X_t are the same
- The change of the arrow directions will not lead to a new v-structure or a cycle in the partial DAG

Corollary: If in a partial DAG all the elements of X_t have different lagged parents, the partial DAG does not have an observationally equivalent model.

Vector Autoregression (VAR) Model

A **Vector Autoregression (VAR)** model is a statistical method used to capture the interdependencies between multiple time series. It extends the univariate autoregressive (AR) model by allowing for multivariate time series, making it useful for modeling systems where several variables evolve together over time, such as in economics.

Next we will explain some key concepts of VAR Model.

1. Endogenous Variables: In a VAR model, each of the variables (say, $y_{1,t}, y_{2,t}, \ldots, y_{k,t}$) depends on its own past values (lags) and the past values of all other variables in the system.

2. Model Structure: For a VAR model of order p (denoted as VAR(p)), the relationship is given by:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t \tag{1}$$

where:

- y_t is a k-dimensional vector of the variables at time t.
- c is a vector of constants (intercepts).
- $A_1, A_2, \ldots, A_p$ are time-invariant matrices (of size $k \times k$) representing the coefficients of the lagged values of the variables.
- e_t is the error term or residuals at time t, which captures the effect of shocks or noise.

3. Error Terms: The error terms e_t have three properties:

1. The mean of e_t is zero, $\mathbb{E}(e_t) = 0$.
2. The error terms are contemporaneously correlated, i.e., the covariance matrix $\mathbb{E}(e_t e_t') = \Omega$ is positive semi-definite.
3. The error terms are uncorrelated across time, meaning no serial correlation exists: $\mathbb{E}(e_t e_{t-k}') = 0$ for $k \neq 0$.

The order p in VAR(p) indicates how many lagged values of the variables are included in the model. Choosing the right number of lags is critical for the model's accuracy.

- I**(0) (Stationary)**: If all variables are stationary, the model is applied directly in levels.
- I**(d) (Non-stationary)**: If variables are non-stationary, they need to be differenced or cointegration techniques may be used, resulting in a Vector Error Correction Model (VECM).

In time series analysis, the notation I refers to the *integrated* nature of a time series, which indicates the number of differences needed to achieve stationarity.

Definitions

- $I(0)$: A time series that is stationary, meaning its statistical properties (such as mean and variance) do not change over time. This allows for direct application of models without transformations.
- $I(d)$: A time series that is non-stationary and requires differencing d times to become stationary. For example, $I(1)$ indicates that the series needs to be differenced once.

- Stationary series ($I(0)$) can be analyzed directly, providing reliable estimates and predictions.
- Non-stationary series ($I(d)$) often show trends or seasonality, requiring differencing or cointegration techniques.
- Stationarity is crucial for models like the Vector Error Correction Model (VECM) to capture both short-term dynamics and long-term relationships.

Relation of TSCM and VAR Model

Now we will discuss about the relation between TSCMs and the VAR models in time series econometrics.

Proposition

A TSCM is a restricted structural VAR model identified by the inferred causal relations among $\{X_t\}_{t=1}^{T}$, and hence it corresponds to a restricted VAR model.

Proposition

An unconstrained VAR model corresponds to a full partial DAG such that the TSCM does not contain any inferable causal relations except the temporal causal orders.

Granger Causality

The **Granger causality test** is a statistical method used to determine whether one time series (X) can predict another time series (Y). This test, introduced in 1969, checks if past values of X provide useful information in forecasting future values of Y beyond what Y's past alone can predict.

Granger causality exists when including the past values of X improves predictions of Y compared to using only Y's past. The test typically uses t-tests or F-tests on the lagged values of both X and Y to determine if X contains significant information about Y's future.
Granger's concept of causality is based on two key principles:

1. The cause must occur before the effect.
2. The cause provides unique information about the future values of the effect.

Mathematically, Granger causality is tested by comparing the probabilities of Y's future values with and without the information from X. If excluding X significantly changes this probability, then X is said to "Granger-cause" Y.

$$\mathbb{P}[Y(t+1) \in A \mid \mathcal{I}(t)] \neq \mathbb{P}[Y(t+1) \in A \mid \mathcal{I}_{-X}(t)] \tag{2}$$

where $\mathbb{P}$ denotes probability, A is an arbitrary non-empty set, $\mathcal{I}(t)$ represents all information available at time t, and $\mathcal{I}_{-X}(t)$ is the same information excluding X. If this condition holds, then X Granger-causes Y.

Granger causality is used to test if one time series (x) can predict another time series (y). The test involves running two regressions: one using only the lagged values of y and another that includes both lagged values of y and x. If the second regression provides more explanatory power than the first, it suggests that x "Granger-causes" y.

Testing Process

1. First, run a regression of y on its own lagged values:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_m y_{t-m} + \text{error}_t$$

2. Then, augment this regression by adding lagged values of x:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_m y_{t-m} + b_p x_{t-p} + \cdots + b_q x_{t-q} + \text{error}_t$$

The lags for x that are individually significant and collectively improve the model's fit (through an F-test) are retained. If no significant lags of x are retained, we fail to reject the null hypothesis that x does not Granger-cause y.

Multivariate Granger Causality

In multivariate cases, Granger causality is tested by fitting a vector autoregressive (VAR) model to a multivariate time series $X(t)$:

$$X(t) = \sum_{\tau=1}^{L} A_\tau X(t-\tau) + \epsilon(t)$$

where $\epsilon(t)$ is a white Gaussian noise vector, and A_τ are coefficient matrices. Time series X_i Granger-causes X_j if any of the elements $A_\tau(j,i)$ for $\tau = 1, \ldots, L$ is significantly different from zero.

Relation of Granger Causality and TSCM

Granger causality is an important concept in time series analysis that helps us understand whether one time series can predict another.

It measures the predictive power of one variable over another based on their past values. In the context of a Time Series Causal Model (TSCM), we can investigate how these relationships manifest among various time series variables.

Key Differences

It's essential to distinguish between Granger causality and graphical causal models:

- **Granger Causality**: This focuses on the ability of one time series to predict the future values of another time series. For example, if changes in variable X can be used to forecast changes in variable Y, we say that X Granger-causes Y.
- **TSCM**: This type of model emphasizes the causal relationships among time series variables at specific points in time. It provides a framework to explore how these variables influence one another directly.

Proposition

Let $X_{i,t}$ and $X_{j,t}$ be two time series variables in a TSCM. $X_{j,t}$ is a Granger cause of $X_{i,t}$ given other variables in the TSCM if and only if there is a directed path from some $X_{j,t-s}$ to $X_{i,t}$ for $s > 0$ in the partial Directed Acyclic Graph (DAG) of the TSCM.

Learning TSCM

In a Time Series Causal Model (TSCM), we only need to learn a partial DAG with $(p+1)N$ nodes, instead of the full DAG with TN nodes.

Lemma

Given the assumption of a causal model, an information set (joint distribution) containing a node and its parent variables is sufficient for the PC algorithm to connect the node to its parents and exclude non-descendants from connecting to it.

Proposition

To learn the partial DAG with arrows into X_t, the information set including $X_t, X_{t-1}, \ldots, X_{t-p}$ is sufficient.

PC Algorithm for TSCM

Here is an algorithm for discovering the causal relation between variables in a Time Series Model.

Algorithm 3 PC Algorithm for a Partial DAG in TSCM

Input: Observations of a set of time series variables X generated from a TSCM.
Output: A partial DAG compatible with the data-generating DAG.
Step 1: Choose a reasonable $\hat{p}$.
Step 2: Calculate the correlation matrix $\Sigma = \text{corr}(X_t, X_{t-1}, \ldots, X_{t-\hat{p}})$.
Step 3: Use Σ as input to obtain a DAG for $(X_t, X_{t-1}, \ldots, X_{t-\hat{p}})$.
Step 4: Delete all arrows and edges that do not connect at least one element of X_t.
Step 5: Orient all edges between X_{t-i} and X_t with arrowheads at X_t.
Step 6: Orient all edges between elements of X_t using the rules in the PC algorithm.

Greedy Search Algorithm

Now we can talk about the greedy search algorithm for the time series causal model.

- Evaluating graph scores is an alternative to uncovering the data-generating DAG model.
- For a partial DAG, the score can be based on the likelihood of the SVAR model.
- Since unconstrained models have higher likelihoods, a proper score includes a penalty term.

- The BIC criterion for a partial DAG of X_t is defined as:

$$\text{BIC} = \sum_{t=1}^{T} \log L(A_0, A_1, \ldots, A_p; X_t | X_{t-1}, \ldots, X_{t-p}) - (|E| + |V|) \log(T)$$

- Where $|E|$ is the number of arrows heading at X_t and $|V|$ is the number of elements in X_t.
- $(|E| + |V|)$ represents the number of free varying parameters in the TSCM.
- The BIC criterion is a sum of the log-likelihood function and a penalty factor.
- As $T \to \infty$, the BIC criterion becomes consistent for model selection.

Now we can summarize this details on greedy search into the following proposition:

Proposition

Under the assumption of TSCM, the BIC criterion is a consistent score, such that the probability of identifying the true model converges to 1 as $T \to \infty$, assuming the search space covers the true model.

Our Work & Results

Our work consists mainly of four parts as follows:

1. Random Forest Model with Feature Selection Using Causality
2. Causal DAG creation
3. Convolutional Neural Network (CNN) Model for ECG Image Classification
4. Causal Aware CNN Model for ECG Image Classification

The results will be discussed in the following slides.

Random Forest Model with Feature Selection Using Causality

The model structure is as follows:

- The top 25 features are selected using Granger Causality.
- Those features are used for training the Random Forest Model.
- To further improve performance, we conducted a comprehensive hyperparameter tuning process using grid search with 5-fold cross-validation. The grid included variations over the number of trees (`n_estimators`), tree depth (`max_depth`), minimum samples required to split and at leaves, and feature subset selection strategies.

AUROC of our Causal RF model

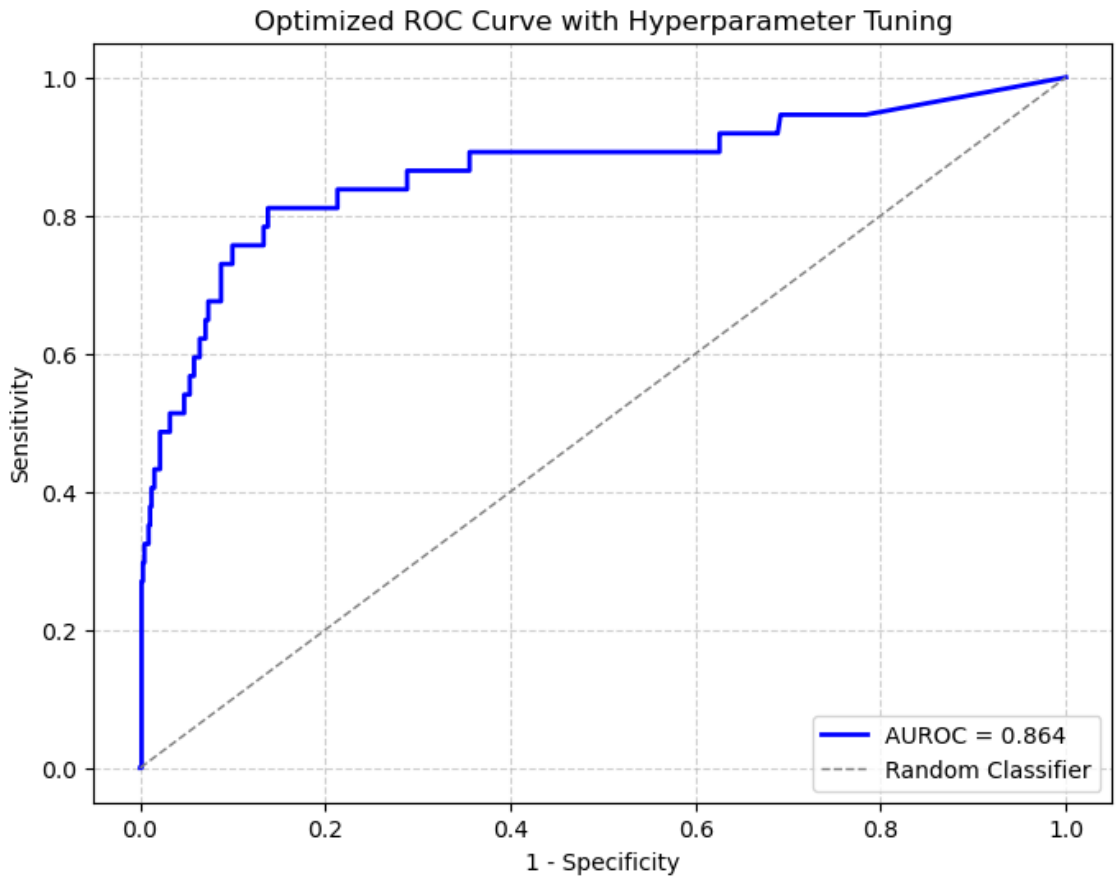

Figure 3: Optimized ROC Curve of Random Forest Classifier (AUROC = 0.864)

AUROC of the Benchmark RF model

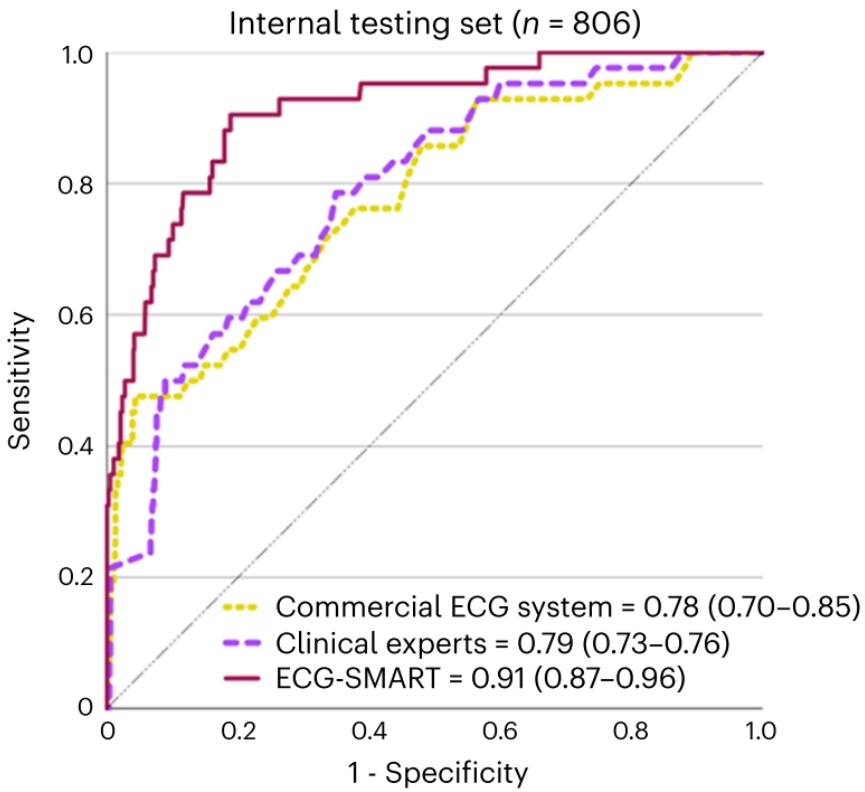

Figure 4: ROC Performance Comparison with ECG-SMART, Clinical Experts, and Commercial ECG Systems

Causal DAG Creation

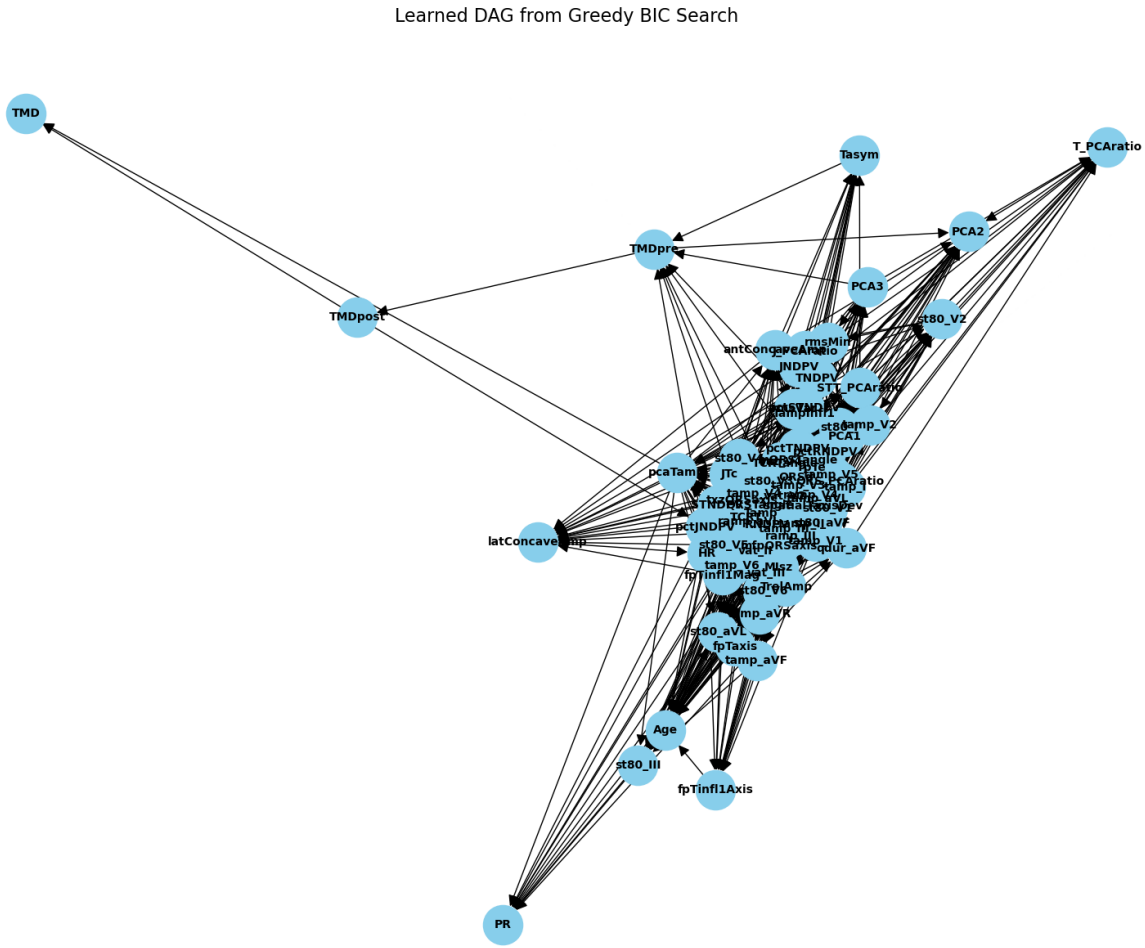

Figure 5: Causal DAG using BIC criterion based on all 74 features.

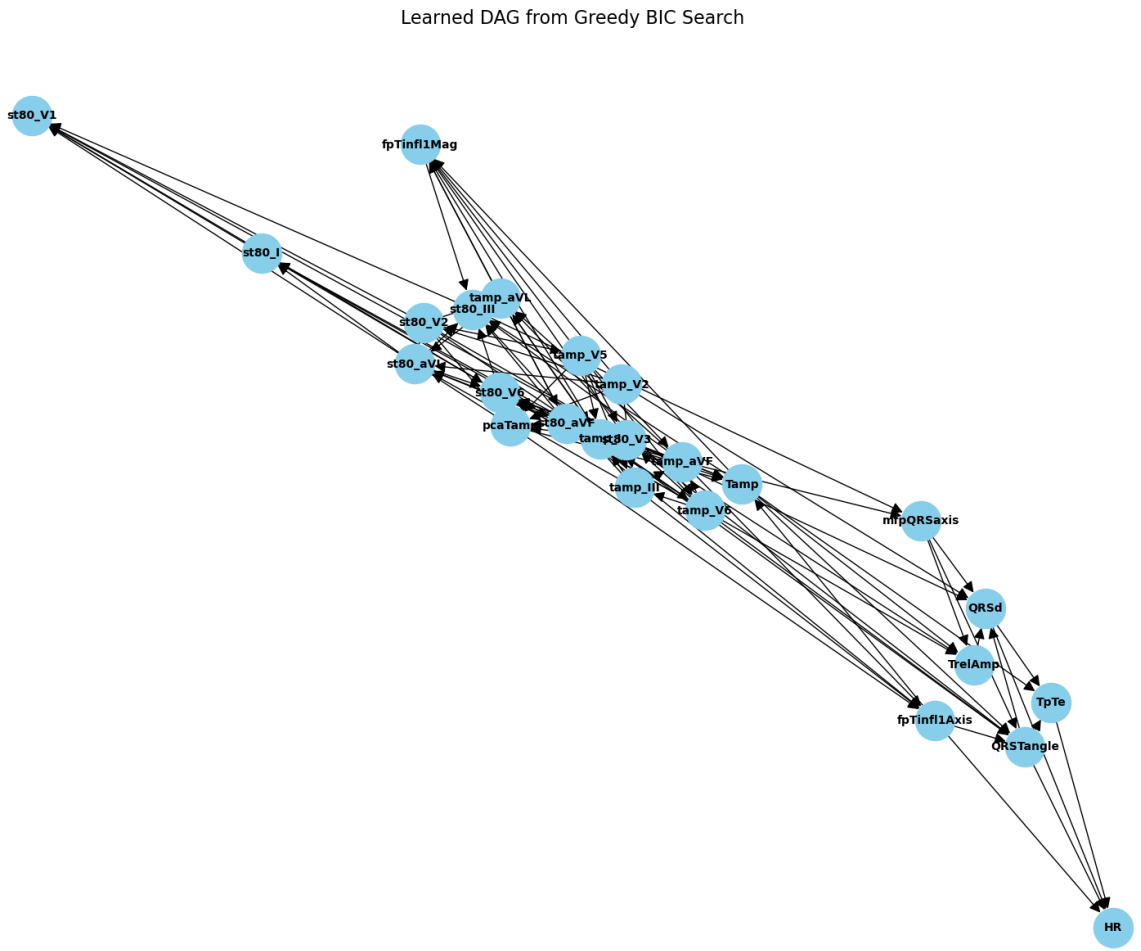

Figure 6: Causal DAG using BIC criterion based on the selected 25 features.

CNN Model for ECG Classification

We implemented a normal CNN model as described in Figure 7.

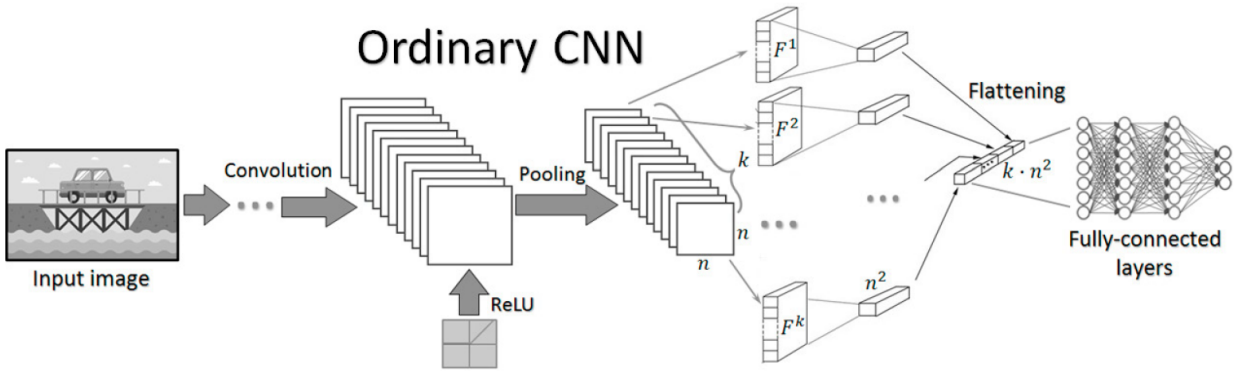

Figure 7: Ordinary CNN Model [3]

This model got an accuracy of 43.5%.

[3] Image is taken from 6

Causal Aware CNN Model for ECG Classification

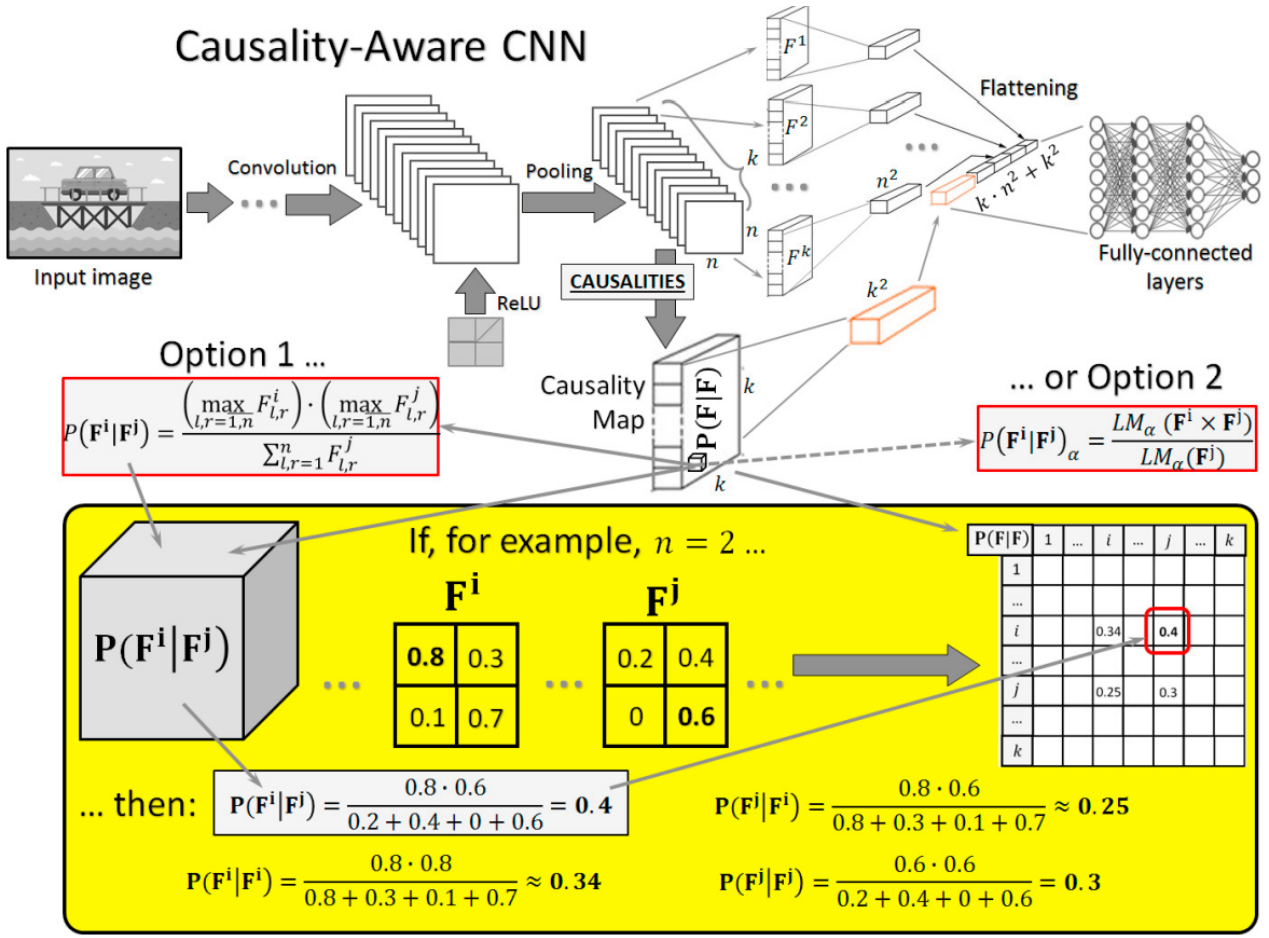

Figure 8: Causal Aware CNN Model [4]

[4] Image is taken from 6

We checked the performance of CA-CNN model using two methods explained below:

1. Input to the Fully Connected Layer is both the Causal Feature Map and the data samples. Then the CA CNN model achieved a test accuracy of approximately 77.48%, indicating way better performance than the ordinary CNN.
2. Input to the Fully Connected Layer is only the Causal Feature Map and not the data samples. Then the CA CNN model achieved a test accuracy of approximately 54.31%, still indicating better performance than the ordinary CNN.

Conclusion

This work demonstrates that combining causal inference with machine learning, whether through feature selection or architectural design, can lead to models that are both more interpretable and more effective. Further directions may include exploring richer causal mechanisms, incorporating temporal dependencies, and applying these techniques to other domains of medical imaging for classification and future prediction as well.

References

1. Chen, Pu, 2010; "A time series causal model,"; MPRA Paper 24841, University Library of Munich, Germany. https://ideas.repec.org/p/pra/mprapa/24841.html
2. Pearl, J. (2000). Causality. Cambridge University Press, 1st edition
3. Pearl, J. and Verma, T. (1991). A theory of inferred causation. In J.A. Allen, R. Fikes, and E. Sandewall(Eds.), Principles of Knowledge Representation and Rea- soning: Procedings of the 2nd International Conference, San Mateo, CA: Morgan Kaufmann, pages 441–452.
4. V. Del Tatto, G. Fortunato, D. Bueti, A. Laio, Robust inference of causality in high-dimensional dynamical processes from the Information Imbalance of distance ranks, Proc. Natl. Acad. Sci. U.S.A. 121 (19) e2317256121, https://doi.org/10.1073/pnas.2317256121 (2024).

5. H. Ni et al., "Time Series Modeling for Heart Rate Prediction: From ARIMA to Transformers," 2024 6th International Conference on Electronic Engineering and Informatics (EEI), Chongqing, China, 2024, pp. 584-589, doi: 10.1109/EEI63073.2024.10695966.
6. Vagan Terziyan, Oleksandra Vitko. (2023). Causality-Aware Convolutional Neural Networks for Advanced Image Classification and Generation, *Procedia Computer Science*, Volume 217, pp. 495-506, ISSN 1877-0509, doi: https://doi.org/10.1016/j.procs.2022.12.245.
7. Bernhard Schölkopf and Julius von Kügelgen (2022), Bernhard Schölkopf and Julius von Kügelgen, arXiv: https://arxiv.org/abs/2204.00607.
8. Al-Zaiti, S.S., Martin-Gill, C., Zègre-Hemsey, J.K. et al. Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction. Nat Med 29, 1804–1813 (2023). doi: https://doi.org/10.1038/s41591-023-02396-3.