Research Statement

Muhammad Hasan Ferdous
University of Maryland, Baltimore County
Web: mhasanferdous.github.io · Email: h.ferdous@umbc.edu

Research Vision: Trustworthy AI for Complex and Dynamic Systems

Modern AI has achieved remarkable success in pattern recognition, powering applications from medical diagnosis to climate modeling. Yet a critical limitation persists: contemporary AI systems excel at finding correlations but struggle to understand causation. This distinction is vital in high-stakes domains, where decision-makers need systems that can explain why events occur rather than merely predicting what will happen next. The challenge intensifies with complex time-series data from real-world information systems, characterized by noise, non-stationarity, and irregular sampling that obscure true causal relationships.

My research in Causal AI develops robust, scalable, and interpretable methods for discovering causal relationships in complex time-series data, together with the open-source software through which those methods become accessible to other researchers, students, and practitioners. The work synthesizes three disciplines: the formal inference of **Statistics** (B.Sc., University of Dhaka), the systems-oriented perspective of **Information Systems** (M.S. and Ph.D., UMBC), and the cutting-edge methods of **Machine Learning and Artificial Intelligence** developed during my dissertation in the UMBC Causal AI Lab under the supervision of Dr. Md Osman Gani.

Completed Contributions

1. Constraint-Based Causal Discovery for Temporal Data: CDANs and eCDANs

Standard constraint-based discovery algorithms condition on the entire past, which is computationally expensive and statistically inefficient in autocorrelated settings. The **CDANs** framework [1] optimizes the conditioning sets used in constraint-based search by explicitly identifying lagged parents, reducing the search space and improving detection power. CDANs also introduces a mechanism for detecting changing modules: components whose causal structure varies over time due to distribution shifts, by considering both contemporaneous and lagged parents. **eCDANs** [2] refines the lagged-adjacency search strategy, preserving accuracy while reducing computational complexity, making the framework suitable for larger datasets and resource-constrained deployment such as bedside monitoring or edge computing. CDANs is publicly available as a Python package: `pip install cdans` (github.com/hferdous/CDANs).

2. Decomposition-Based Discovery for Multi-Seasonal Data: DCD

Real-world time series in domains like climate science and finance often exhibit multiple overlapping seasonal patterns and long-term trends that confound standard causal discovery. The **DCD** framework [3] separates each series into trend, seasonal, and residual components and applies appropriate inference machinery to each: stationarity tests for trends, kernel-based dependence measures for seasonal components, and constraint-based discovery for residuals. Component-level findings are integrated into a unified multi-scale causal graph. Empirical evaluations on synthetic benchmarks and real climate data demonstrate that DCD recovers ground-truth causal structure more accurately than state-of-the-art baselines under strong non-stationarity and autocorrelation.

3. Open-Source Benchmarks: TimeGraph and ClassyGlass

Progress in causal discovery has been constrained by inconsistent evaluation. Researchers test methods on different datasets under simplified conditions, making cross-method comparison difficult. **TimeGraph** [4] is a synthetic benchmark suite that simulates realistic temporal complexities, generating datasets with controlled causal structures, varying autocorrelation, non-stationarity, multiple noise types, and seasonal patterns. TimeGraph is released as open-source software to support reproducible evaluation and standardize comparison across methods. **ClassyGlass** [5] extends the benchmarking program to multimodal wearable-sensor data for activity and mobility analysis, providing high-quality annotated data for testing causal methods in human activity recognition and health monitoring.

4. Climate, Healthcare, and Interdisciplinary Applications

I have applied my methods through several interdisciplinary collaborations. With Dr. Aneesh Subramanian (University of Colorado Boulder), I co-developed a causal deep learning framework for Arctic sea ice prediction [6] and a causal time-series model of Greenland supraglacial lake evolution [7]. Both projects integrate causal discovery with deep learning to produce forecasts that maintain accuracy under distribution shift. In healthcare, I have applied causal methods to physiological time series and to clinical datasets including the Pima Indian Diabetes database, identifying physiological parameters that represent actionable intervention targets rather than mere correlational risk markers.

Five-Year Research Agenda

As a faculty member, I will establish a research program organized around four interconnected thrusts that advance foundational methods while maintaining strong connections to real-world applications.

Thrust 1: Foundational Causal AI for Streaming and Non-Stationary Systems

I will extend the CDANs and DCD frameworks to streaming and concept-drifting data, the regime in which most production information systems operate. Projects include online causal discovery with bounded memory, causal change-point detection, methods for handling missingness arising from sensor failure rather than missing-at-random assumptions, and the integration of causal reasoning with anomaly detection. These methods support trustworthy inference across cyber-physical systems, sensor networks, and operational monitoring.

Thrust 2: Synergistic Integration of Causal Inference and Deep Learning

Deep learning models excel at capturing complex patterns but lack transparency. Causal models provide explicit structure but face scalability challenges in high-dimensional settings. My second thrust develops hybrid neuro-causal frameworks that embed causal discovery into deep representation learning. The goal is to retain predictive performance while gaining the ability to reason about cause and effect under distribution shift. Specific directions include neuro-causal models for transfer learning and the development of interpretable explanations grounded in causal structure. This is a productive area for undergraduate involvement through course projects, summer research, and senior capstones.

Thrust 3: Open-Source Ecosystem for Reproducible Causal AI

My research program has already produced three distinct open-source contributions: the CDANs Python package, the TimeGraph benchmark suite, and the ClassyGlass multimodal benchmark (under review). I will expand this ecosystem with semi-synthetic benchmarks drawn from finance, climate, healthcare, and security data; libraries that implement state-of-the-art causal discovery baselines; and community engagement through workshop organization at KDD, AAAI, and adjacent venues focused on trustworthy machine learning. Software infrastructure provides departmental visibility disproportionate to its cost and is sustainable work with dedicated undergraduate research assistants.

Thrust 4: Interdisciplinary Collaborations and Societal Impact

Causal AI methods become valuable when they touch substantive problems. I will pursue collaborative projects with domain experts in three application areas:

Healthcare and digital health. Causal modeling of physiological time series, trustworthy clinical decision support, and identification of actionable intervention targets in chronic disease management.

Climate and environmental science. Extension of causal time-series methods to ice-sheet dynamics, sea ice forecasting, and climate teleconnections, building on existing collaborations.

Adversarial robustness and security analytics. Causal feature selection for machine learning models with greater resistance to input perturbation, causal discovery for intrusion detection on autocorrelated network traffic, and causal explanations for security incidents where analysts need root-cause reasoning beyond anomaly scores.

Funding Strategy

My initial funding plan is calibrated for early-career growth and combines federal awards with industry partnerships.

Source	Program	Focus	Years
NSF	CRII; later CAREER	Foundational causal AI methods; open-source infrastructure	1–3, then 3–5
NSF	SaTC; CICI	Causal and adversarially robust ML for security analytics	2–5
NIH	R03 / R21 / R-series	Trustworthy clinical decision support; physiological time series	2–5
DoE / NOAA	Data Science programs	Causal discovery for climate and environmental modeling	2–5
Industry	Sponsored research	Finance, enterprise IT, regional technology partners	Ongoing

Mentorship and Student Involvement

My research program is built for substantive student involvement. Students will be integral contributors and future leaders rather than mere assistants. Open-source benchmarks and ML pipelines are work in which well-supervised undergraduates produce real contributions, and published code, datasets, and reproducibility audits are concrete artifacts that strengthen graduate-school applications and industry resumes. My objective targets are at least one peer-reviewed publication per year with a student co-author by year three, and at least one student per year placed into industry internships or graduate programs.

Conclusion

My research philosophy holds that the most significant advances emerge from the synthesis of theoretical rigor and practical application. By integrating statistics, information systems, and AI, my work is positioned to contribute to the next generation of trustworthy, decision-driven information systems. I am eager to discuss how this research program would contribute to your department.

References

[1] M. H. Ferdous, U. Hasan, M. O. Gani. **CDANs: Temporal Causal Discovery from Autocorrelated and Non-Stationary Time Series Data.** *Machine Learning for Healthcare Conference (MLHC)*, pp. 186–207, 2023.

[2] M. H. Ferdous, U. Hasan, M. O. Gani. **eCDANs: Efficient Temporal Causal Discovery from Autocorrelated and Non-Stationary Data (Student Abstract).** *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37(13), p. 16208, 2023.

[3] M. H. Ferdous, M. O. Gani. **DCD: Decomposition-based Causal Discovery from Autocorrelated and Non-Stationary Temporal Data.** arXiv preprint arXiv:2602.01433, 2026. Under review at *Transactions on Machine Learning Research*.

[4] M. H. Ferdous, E. Hossain, M. O. Gani. **TimeGraph: Synthetic Benchmark Datasets for Robust Time-Series Causal Discovery.** *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '25)*, pp. 5425–5435, 2025.

[5] N. Mahmud, N. Emmert, M. H. Ferdous, et al. **ClassyGlass: A Benchmark Dataset for Activity and Mobility Analysis using Smart Eyewear.** Submitted to *ACM SIGKDD Dataset Track*, 2026.

[6] E. Hossain, M. H. Ferdous, J. Wang, A. Subramanian, M. O. Gani. **Correlation to Causation: A Causal Deep Learning Framework for Arctic Sea Ice Prediction.** *IEEE PerCom Workshops*, pp. 62–67, 2025.

[7] E. Hossain, M. H. Ferdous, D. Dunmire, A. Subramanian, M. O. Gani. **Causal Time Series Modeling of Supraglacial Lake Evolution in Greenland under Distribution Shift.** *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2025.