



Kent Academic Repository

Harit, Anoushka, Sun, Zhongtian and Yu, Jongmin (2025) *From news to returns: A Granger-causal hypergraph transformer on the sphere*. In: ICAIF '25: Proceedings of the 6th ACM International Conference on AI in Finance. . pp. 674-682. Association for Computing Machinery ISBN 979-8-4007-2220-2.

Downloaded from

<https://kar.kent.ac.uk/112602/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1145/3768292.3770414>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC (Attribution-NonCommercial)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



From News to Returns: A Granger-Causal Hypergraph Transformer on the Sphere

Anoushka Harit
University of Cambridge
Cambridge, United Kingdom
ah2415@cam.ac.uk

Zhongtian Sun
University of Kent
Canterbury, United Kingdom
Department of Computer Science
University of Cambridge
Cambridge, United Kingdom
z.sun-256@kent.ac.uk

Jongmin Yu
ProjectG.AI
Seoul, Republic of Korea
jy552@projectg.ai

Abstract

We propose the Causal Sphere Hypergraph Transformer (CSHT), a novel architecture for interpretable financial time-series forecasting that unifies *Granger-causal hypergraph structure*, *Riemannian geometry*, and *causally masked Transformer attention*. CSHT models the directional influence of financial news and sentiment on asset returns by extracting multivariate Granger-causal dependencies, which are encoded as directional hyperedges on the surface of a hypersphere. Attention is constrained via angular masks that preserve both temporal directionality and geometric consistency. Evaluated on S&P 500 data from 2018 to 2023, including the 2020 COVID-19 shock, CSHT consistently outperforms baselines across return prediction, regime classification, and top-asset ranking tasks. By enforcing predictive causal structure and embedding variables in a Riemannian manifold, CSHT delivers both *robust generalisation across market regimes* and *transparent attribution pathways* from macroeconomic events to stock-level responses. These results suggest that CSHT is a principled and practical solution for trustworthy financial forecasting under uncertainty.

CCS Concepts

• Computing methodologies → Sphere neural networks; Riemannian geometry; Causal inference.

Keywords

Causal hypergraphs, Spherical embeddings, Transformer models, Financial time series, Interpretable machine learning, Market forecasting

ACM Reference Format:

Anoushka Harit, Zhongtian Sun, and Jongmin Yu. 2025. From News to Returns: A Granger-Causal Hypergraph Transformer on the Sphere. In *6th ACM International Conference on AI in Finance (ICAIF '25)*, November 15–18, 2025, Singapore, Singapore. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3768292.3770414>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

ICAIF '25, Singapore, Singapore

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2220-2/25/11

<https://doi.org/10.1145/3768292.3770414>

1 Introduction

Financial markets are driven by a complex interplay of information sources, including price history, macroeconomic indicators, investor sentiment, and news. Although recent deep learning models have improved financial time series forecasting, their lack of interpretability and limited alignment with economic theory limits their adoption in high-stakes settings such as risk management, regulation, and trading [4, 7].

A core insight from behavioral finance is that investor sentiment mediates the transmission of news to asset prices [3, 31]. However, most existing models treat sentiment and news as unstructured inputs, ignoring the directional and hierarchical flow of financial information. Moreover, attention mechanisms in transformers [33] allow unrestricted dependencies, potentially violating known causal pathways.

To address these challenges, we propose the *Causal Sphere Hypergraph Transformer* (CSHT), a novel architecture that integrates statistical causality, geometric embeddings, and structured attention. CSHT operates in three stages: (i) extract directional, multivariate dependencies using Granger causality across news, sentiment, and returns; (ii) encode these as directed hyperedges in a causal hypergraph; (iii) embed variables on a hypersphere, where attention is constrained using geodesic masks that preserve both causality and angular similarity.

Granger causality provides a statistically grounded approach for identifying temporal influence (e.g.,

$$\text{Return}_t \leftarrow \{\text{Sentiment}_{t-1}, \text{News}_{t-2}\},$$

while acknowledging that it does not imply interventional causality. Embedding variables on the hypersphere allows attention to reflect directional similarity via angular distances, offering geometric consistency under latent market structure.

Our evaluation spans 2018–2023, covering the COVID-19 shock, a major regime shift in global finance. We demonstrate that CSHT outperforms strong baselines (e.g., LSTM, VAR, transformer variants) across return prediction, regime classification, and asset ranking, while maintaining transparent, semantically aligned attention pathways.

Contributions. This paper makes the following contributions:

- (1) We introduce CSHT, a transformer-based model that encodes multi-way Granger dependencies as directed hypergraphs embedded on a hypersphere.

- (2) We propose a geometry-aware attention mechanism that enforces causal masking via geodesic angular similarity.
- (3) We demonstrate that CSHT improves predictive accuracy and interpretability, particularly during volatile market regimes such as the 2020 pandemic.

Our results highlight the promise of causally structured, geometry-aware models for interpretable and trustworthy financial forecasting.

2 Related Work

Granger causality [11] is a well-established method for identifying temporal dependencies [15] in financial time series, widely used to assess how sentiment and news influence asset returns [5, 19]. However, traditional applications are typically pairwise, scale poorly to high-dimensional data, and are not integrated into deep learning pipelines. Recent efforts such as CI-GNN [37] and invariant risk minimisation aim to bridge causality and representation learning, but remain underexplored in finance.

Graph neural networks (GNNs) have shown promise in financial forecasting by modelling stock relationships [9, 17, 34], yet most rely on static or correlation-based edges and lack causal or temporal constraints. This can lead to spurious dependencies, particularly under regime shifts.

Hypergraph neural networks [2, 12, 14, 16, 23, 27, 29, 30, 36] capture higher-order relations across modalities [25], making them suitable for multi-source signals. However, their application in finance remains limited, with few methods offering causal interpretability or robustness in volatile periods.

Geometric deep learning has introduced spherical and hyperbolic embeddings to model relational or hierarchical structures [6, 8, 13, 16, 18, 22, 24]. These are well suited to financial domains, but are rarely integrated with causal or temporal reasoning, limiting their alignment with economic structure.

Transformer-based models have achieved strong results in financial time-series prediction [21, 32, 35], but typically rely on unconstrained attention, allowing dependencies that may contradict known causal flows. Probabilistic models like DeepAR [10] offer uncertainty estimates but lack structural constraints.

Multi-source fusion approaches [20] combine sentiment, technical indicators, and macro signals, improving predictive accuracy. However, most models remain black-box without causal grounding or geometric structure.

Our work addresses these gaps by integrating Granger-causal hypergraphs, Riemannian embeddings, and masked transformer attention. This enables structured, interpretable, and geometry-aware forecasting across volatile financial regimes.

3 Problem Formulation

Let $\mathcal{D} = \{(x_t, s_t, r_t)\}_{t=1}^T$ denote a multimodal financial time series of length T , where:

- $x_t \in \mathbb{R}^{d_x}$ is a vector of news embeddings (e.g., FinBERT-encoded headlines) at time t ,
- $s_t \in \mathbb{R}^{d_s}$ is a vector of sentiment scores derived from x_t or prior $x_{t'}$ with $t' < t$,
- $r_t \in \mathbb{R}^{d_r}$ is the log-return vector of d_r assets at time t .

Our objective is to model the directional influence of past news and sentiment on future asset returns. For each asset i , we posit the existence of a subset $\mathcal{C}^i \subset \{x_{<t}, s_{<t}, r_{<t}\}$ that Granger-causes r_t^i , capturing the relevant causal parents from the historical information set.

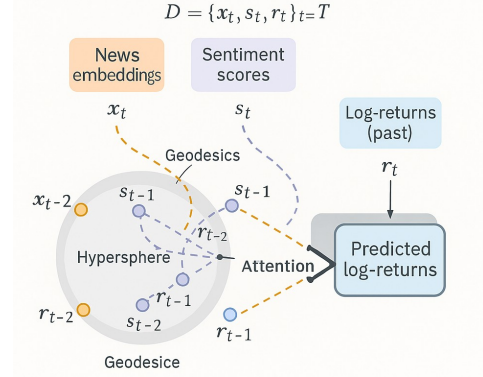


Figure 1: Overview of the causal forecasting problem setup.

We represent these temporal dependencies using a directed hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node $v \in \mathcal{V}$ corresponds to a lagged variable (e.g., x_{t-2} or s_{t-1}), and each directed hyperedge $e = (\{v_1, \dots, v_k\} \rightarrow r_t^i) \in \mathcal{E}$ encodes a multi-source Granger-causal relation toward an asset return.

Each node $v \in \mathcal{V}$ is embedded on the surface of an n -dimensional unit hypersphere $\mathbb{S}^n$, and attention is computed using geodesically masked angular similarity:

$$\text{Attention}(v_i, v_j) \propto \cos(\theta_{ij}) = \frac{\langle v_i, v_j \rangle}{\|v_i\| \cdot \|v_j\|}, \quad v_i, v_j \in \mathbb{S}^n.$$

The forecasting task is defined as learning a causal function f over the hypergraph-constrained history $\mathcal{H}_{<t}$:

$$\hat{y}_t = f(\mathcal{H}_{<t}; \mathcal{G}) \quad \text{where} \quad \mathcal{H}_{<t} = \{x_{t-k}, s_{t-k}, r_{t-k}\}_{k=1}^T.$$

The learning objective minimises the expected prediction loss while respecting the structure encoded in $\mathcal{G}$:

$$\hat{f} = \arg \min_{f \in \mathcal{F}_{\mathcal{G}}} \mathbb{E}_t [\mathcal{L}(f(\mathcal{H}_{<t}), y_t)],$$

where $\mathcal{F}_{\mathcal{G}}$ is the class of functions that conform to the Granger-derived hypergraph structure, and $\mathcal{L}$ is a suitable prediction loss (e.g., mean squared error or cross-entropy). This formulation enables structured, geometry-aware modelling of financial signals under uncertainty.

Example. A real-world instance of this setting is the 2020 COVID-19 pandemic shock, which induced abrupt changes in market regimes. During this period, sentiment signals s_{t-1} extracted from pandemic-related news embeddings x_{t-2} , such as lockdown announcements and infection statistics, exhibited strong Granger-causal influence on sector-specific returns r_t^i , particularly in healthcare, energy, and travel. CSHT captures such directional, multi-source relations by constructing hyperedges of the form:

$$\{x_{t-2}^{\text{COVID}}, s_{t-1}^{\text{fear}}\} \rightarrow r_t^{\text{health}}.$$

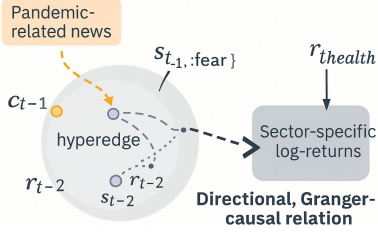


Figure 2: Causal influence during the COVID-19 regime: News x_{t-2}^{COVID} and sentiment s_{t-1}^{fear} drive returns r_t^{health} . CSHT captures such patterns via hyperspherical hyperedges and causal attention masks.

These relations are embedded on the hypersphere and used to define causal attention masks, constraining the prediction function f to attend only to valid information flows. This formulation supports both short-term forecasting and regime classification by allowing the model to adapt to structural changes in $\mathcal{G}$ over time—for example, a surge in the influence of health-related sentiment during exogenous shocks.

4 Methodology

The Causal Sphere Hypergraph Transformer (CSHT) models directional market influence by combining Granger-causal structure, geometric embeddings, and constrained attention. The architecture comprises four main stages: (i) discovery of causal hypergraphs over news, sentiment, and returns; (ii) spherical embedding of variables on a Riemannian manifold; (iii) transformer-based geodesic attention with causal masking; and (iv) manifold-constrained training for financial forecasting.

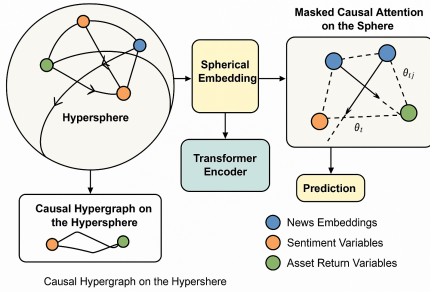


Figure 3: CSHT architecture: Granger-causal links connect news, sentiment, and returns. Nodes are embedded on a hypersphere and processed via masked geodesic attention for interpretable forecasting.

4.1 Financial Causal Flow: News, Sentiment, Returns

We model financial dynamics through structured interactions among three modalities:

- *News embeddings* $x_t \in \mathbb{R}^{d_x}$ are extracted from daily headlines using FinBERT, encoding semantic tone and topical context;
- *Sentiment signals* $s_t \in \mathbb{R}^{d_s}$ are derived from temporally smoothed and aggregated news embeddings $x_{t'}$ for $t' < t$, reflecting delayed investor response;
- *Returns* $r_t \in \mathbb{R}^{d_r}$ are log-transformed asset returns, driven by prior news and sentiment inputs.

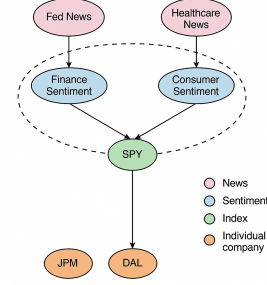


Figure 4: Illustrative causal flow: news affects sentiment, which drives index movements (SPY) and individual returns (e.g., JPM, DAL).

This induces an interpretable directional flow:

$$x_{t-k} \rightarrow s_{t-j}, \quad s_{t-j} \rightarrow r_t, \quad x_{t-k} \rightarrow r_t$$

CSHT explicitly encodes and constrains this flow via a causality-aware architecture.

4.2 Causal Hypergraph Construction

Given a multivariate time series $\mathcal{D} = \{(x_t, s_t, r_t)\}_{t=1}^T$, we apply Granger causality tests to detect lagged directional dependencies. For each return variable r_t^i , we identify a set of predictors $C^i \subset \mathcal{H}_{<t}$ such that:

$$\exists j \in C^i : \mathbb{P}(r_t^i | C^i) \neq \mathbb{P}(r_t^i | C^i \setminus \{j\})$$

Each such set defines a directed hyperedge $C^i \rightarrow r_t^i$, producing a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ over lagged news, sentiment, and return variables.

Node Specification. Each node $v \in \mathcal{V}$ represents a lagged input from one of three modalities: (i) news embeddings x_{t-k} , (ii) sentiment scores s_{t-j} , or (iii) past returns r_{t-k} . Nodes are included only where Granger causality indicates significant influence. The resulting hypergraph captures temporally indexed semantic, affective, and financial signals.

An example hyperedge:

$$\{x_{t-2}^{\text{COVID}}, s_{t-1}^{\text{fear}}\} \rightarrow r_t^{\text{healthcare}}$$

illustrates a multi-source causal link from pandemic-related inputs to sector returns.

4.3 Spherical Riemannian Embedding

Each node $v \in \mathcal{V}$ (i.e., each lagged variable identified as causally relevant) is assigned an embedding $v \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$, lying on the unit hypersphere:

$$\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$$

These embeddings are initialised randomly and updated during training. The Riemannian structure imposes angular sensitivity:

$$d_{\mathbb{S}^n}(x, y) = \arccos(\langle x, y \rangle)$$

After each parameter update, embeddings are projected back onto the manifold to preserve geometric consistency:

$$\Pi(x) = \frac{x}{\|x\|}, \quad x \in \mathbb{R}^{n+1} \setminus \{0\}$$

This ensures that causal similarity is preserved in angular space.

4.4 Causal Attention with Geodesic Masking

We apply a transformer encoder over the hypergraph, using causal masks to restrict attention to Granger-derived parents. For query q_i and key k_j , attention is computed as:

$$\text{Attn}(i, j) = \begin{cases} \frac{\exp(\lambda \cdot \langle q_i, k_j \rangle)}{Z_i}, & \text{if } v_j \in C_i \\ 0, & \text{otherwise} \end{cases}$$

where λ is a temperature scaling factor and Z_i is the normalisation constant over valid causal parents of v_i . This mechanism ensures that attention remains both *geometrically constrained* [6, 18] through cosine similarity in the hypersphere $\mathbb{S}^n$ and *causally grounded* through the enforcement of Granger-derived dependencies.

4.5 Prediction and Riemannian Training

The output of the transformer is used to predict the financial target y_t (e.g., next-day return $\hat{r}_{t+1}$ or a market regime label). The model is trained to minimise the expected loss over time, while respecting the geometry of the hyperspherical embedding:

$$\mathcal{L} = \frac{1}{T} \sum_{t=1}^T \ell(f(\mathcal{H}_{<t}; \mathcal{G}), y_t)$$

Gradients are computed in the ambient Euclidean space $\mathbb{R}^{n+1}$, and updated embeddings are projected back onto the hypersphere to preserve Riemannian structure:

$$x^{(k+1)} \leftarrow \Pi(x^{(k)} - \eta \nabla_x \mathcal{L})$$

where $\Pi(x) = \frac{x}{\|x\|}$ denotes the projection operator, and η is the learning rate.

This training scheme ensures geometric validity of representations throughout optimisation, enabling stable learning of directional causal dependencies on the manifold.

Regime Adaptation. To capture the evolving market structure, the hypergraph $\mathcal{G}$ is periodically updated using a sliding window strategy. This enables CSHT to adapt to structural breaks or regime shifts by dynamically reconfiguring its causal edge set $\mathcal{E}$, thereby supporting both robust forecasting and the detection of market phase transitions.

5 Experimental Setup

We evaluate the Causal Sphere Hypergraph Transformer (CSHT) on real-world financial forecasting tasks. The model aims to predict (i) next-day asset-level returns and (ii) broader market regime shifts using historical price and sentiment signals. Our hypothesis is

that causally masked attention over hyperspherical embeddings improves accuracy, robustness, and interpretability relative to strong baselines.

5.1 Data Collection and Processing

Our dataset spans 2018–2023 and includes the COVID-19 shock (Q1–Q2 2020), offering a natural testbed for model robustness under regime shifts. This enables evaluation of the model’s adaptability to macroeconomic shocks during training and generalisation to post-pandemic regimes.

We focus on stocks in the S&P 500 index between 2018 and 2023. We retain 450 actively traded stocks after filtering out those with incomplete trading history. Historical prices and volumes are obtained from Yahoo Finance¹.

Table 1: Statistics of historical price data

Split	Date Range	# Days	# Stocks
Training	01/01/2018–31/12/2020	756	450
Validation	01/01/2021–31/12/2021	252	450
Test	01/01/2022–30/06/2023	376	450

Daily log-returns are computed as:

$$R_i^t = \frac{P_i^t - P_i^{t-1}}{P_i^{t-1}}$$

where P_i^t is the adjusted close price of stock i on day t .

News Sentiment. We use the Twitter Financial News Sentiment dataset², which includes over 500,000 financial tweets labeled as *positive*, *negative*, or *neutral*. We extract 50,000 tweets referencing S&P 500 stocks using ticker symbols and company names. FinBERT [1] is used to compute daily average sentiment per stock. Neutral score (0) is assigned when no tweets are matched.

Feature Construction. Each stock-day feature vector includes:

- **Market features:** Log-return, 30-day realised volatility, and trading volume;
- **Sentiment features:** Mean FinBERT sentiment polarity per day.

All features are z-normalised using training-set statistics.

5.2 Causal Hypergraph Construction

We apply multivariate Granger causality tests with a maximum lag of $K = 5$ to detect multi-source influences on each return variable. For each trading day t , we construct a directed hypergraph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$:

- $\mathcal{V}$ includes lagged news, sentiment, and return variables;
- $\mathcal{E}_t$ contains hyperedges $C^i \rightarrow r_t^i$ if the null hypothesis of no Granger causality is rejected at FDR-adjusted $p < 0.01$.

This structure is used to define attention masks in CSHT.

¹<https://finance.yahoo.com/>

²<https://huggingface.co/datasets/zeroshot/twitter-financial-news-sentiment>

5.3 Model Configuration

Each node is embedded on the unit hypersphere $\mathbb{S}^n$ and passed through a transformer with masked attention restricted to causal parents. Cosine similarity defines attention scores. CSHT is evaluated on:

- **Task 1 (Regression):** Predict next-day log-return r_{t+1}^i ;
- **Task 2 (Classification):** Predict market regime based on the sign of 3-day forward return of the S&P 500 index.

5.4 Baselines

We benchmark CSHT against six competitive models that reflect the current state of multimodal financial forecasting:

- **FinBERT-RNN** [1]: Combines FinBERT-derived sentiment embeddings with an LSTM to forecast future returns.
- **CNN-LOB** [32]: Applies convolutional layers to limit order book (LOB) snapshots for price movement prediction.
- **Money** [23]: An ensemble that integrates adversarial hypergraph learning with CNN-based feature extraction for financial events.
- **FinGAT** [17]: A graph attention network that models sentiment and temporal price dependencies to forecast directional trends.
- **TEANet** [35]: A transformer architecture that fuses historical returns and sentiment signals using multi-head attention.
- **Higher Order Transformer (HOT)** [21]: Captures multimodal financial dependencies using a higher-order transformer over structured sequences of sentiment and market signals.

All models are trained and evaluated using the same asset universe, feature inputs (news, sentiment, price), and target definitions to ensure fair comparison.

5.5 Evaluation Protocol

We use a non-overlapping temporal split:

- Training: 2018–2020;
- Validation: 2021;
- Test: 2022–mid-2023.

All models are trained using the Adam optimiser (learning rate 10^{-4} , batch size 32), with early stopping based on validation loss. Each experiment is repeated 5 times with different seeds.

Metrics. We report:

- **MAE:** Mean absolute error (return prediction);
- **Regime Accuracy:** Bull/bear classification accuracy;
- **NDCG@10:** Quality of top-10 asset recommendations;
- **Causal Alignment:** Fraction of attended edges aligned with Granger-causal structure.

Robustness Analysis. To assess generalisation across market regimes, we evaluate performance in both pandemic-era (2020) and post-pandemic test sets (2022–2023), tracking consistency of attention and causal alignment.

5.6 Implementation Details

All models are implemented in PyTorch and trained on an NVIDIA RTX 2080 Ti GPU. CSHT uses a 2-layer transformer with 64 hidden units, 4 attention heads, and angular scaling $\lambda = 10$. Manifold constraints are enforced via $\Pi(x) = x/\|x\|$ after each update. Hyperparameters are tuned using validation NDCG@10 (Task 2) and MAE (Task 1).

6 Results and Discussion

We evaluate the Causal Sphere Hypergraph Transformer (CSHT) on two primary tasks: (i) next-day return forecasting and (ii) market regime classification (bull versus bear). Our goal is to assess whether integrating directional causal structure and hyperspherical attention improves both predictive accuracy and interpretability. Experiments span 450 S&P 500 stocks over the period 2018 to 2023.

6.1 Quantitative Evaluation

We benchmark CSHT against five strong baselines: FinBERT-RNN[1], CNN-LOB [32], Money [35], Transformer-AN [35], FinGAT [17], and Higher Order Transformer(HOT) [21]. Evaluation metrics include: (i) **MAE** for next-day return prediction, (ii) **Regime Accuracy** for bull/bear classification, and (iii) **NDCG@10** to assess top-asset recommendation quality.

Table 2: Performance on S&P 500 stocks (2018–2023). Best results in bold.

Model	MAE ↓	Accuracy (%) ↑	NDCG@10 ↑
FinBERT-RNN	0.0213	65.4	0.612
CNN-LOB	0.0205	67.2	0.625
Money	0.0198	69.1	0.634
TEANet	0.0189	70.6	0.639
FinGAT	0.0187	71.3	0.641
HOT	0.0184	69.0	0.636
CSHT (ours)	0.0162	74.6	0.683

CSHT achieves the strongest results across all metrics. It reduces MAE by 13.4% relative to the strongest baseline (FinGAT), improves classification accuracy by 3.3 percentage points, and attains the highest NDCG@10—demonstrating superior ranking under volatility.

6.2 Causal Alignment Analysis

We compute the proportion of attention mass aligned with Granger-causal dependencies, denoted as **Causal Alignment**. Only models with interpretable attention (FinGAT and CSHT) are included.

Table 3: Causal alignment (%) with Granger-causal graph.

Model	Causal Alignment (%) ↑
FinGAT	21.3
CSHT (ours)	66.4

CSHT exhibits substantially higher alignment with validated causal structure, confirming its semantic interpretability and robust information flow control.

6.3 Ablation Study: Contribution of Causal and Geometric Components

We conduct a detailed ablation study to isolate the contribution of causal masking and spherical attention in CSHT. In addition to our two main ablations, we include further variants to analyse robustness and geometric effects.

- **CSHT without Causal Mask:** Removes Granger-based masking, allowing unconstrained attention.
- **CSHT without Spherical Attention:** Uses standard dot-product attention in Euclidean space.
- **Full Mask + Spherical Attention:** Applies geodesic attention over a fully connected graph, isolating the effect of hyperspherical geometry under noisy dependencies.
- **Causal Mask + Euclidean Attention:** Uses causal masking but computes attention in Euclidean space, testing the independent value of causal structure.
- **CSHT + Input Noise:** Adds Gaussian noise ($\sigma = 0.05$) to sentiment and news embeddings, evaluating robustness under volatility.

Table 4: Ablation study on validation set (2021).

Model Variant	MAE ↓	Accuracy (%) ↑
Full Mask + Spherical Attention	0.0171	72.3
Causal Mask + Euclidean Attention	0.0177	71.2
CSHT without Causal Mask	0.0184	70.1
CSHT without Spherical Attention	0.0175	71.4
CSHT + Input Noise	0.0172	73.1
Full CSHT	0.0162	74.6

Interpretation. Causal masking enhances generalisation by filtering spurious dependencies, while spherical attention captures relational structure via angular similarity. The *Full Mask + Spherical* variant outperforms its Euclidean counterpart, confirming the benefit of geometric bias. Stable performance under input noise further demonstrates robustness to volatility [28]. We also tested sensitivity to hypergraph parameters (lag $K \in \{3, 5, 7\}$ and thresholds $p < \{0.05, 0.01, 0.001\}$). Performance varied by $<3\%$, and key causal links (e.g., COVID sentiment $\rightarrow$ sector returns) remained stable—indicating CSHT’s resilience to graph sparsity and noise.

6.4 Sector-Wise Regime Classification

We report regime classification accuracy per sector on the 2022–2023 test set. CSHT consistently outperforms across all sectors, with notable gains in high-volatility domains.

The strongest gains appear in sectors influenced by macroeconomic sentiment, demonstrating the model’s capacity to trace causal signals across hierarchies.

FinBERT RNN and CNN LOB underperform due to the limited relational structure. Money and HOT benefit from hypergraph and

Table 5: Sector-wise classification accuracy (% , 2022–2023 test period). Best in bold.

Sector	FinBERT-RNN	CNN-LOB	Money	HOT	FinGAT	CSHT
Technology	61.8	64.3	68.1	71.0	69.2	74.8
Finance	60.9	65.1	67.5	70.1	68.9	75.3
Energy	63.2	66.7	69.4	70.9	70.4	72.9
Healthcare	59.7	62.5	66.1	68.3	66.3	70.2

multimodal features. FinGAT adds relational reasoning, but lacks a higher-order context. CSHT achieves the best performance by modeling directional multivariable influences with causal hyperedges and spherical embeddings.

6.5 Runtime and Inference Efficiency

We compare epoch-level training time and per-stock inference latency. CSHT remains efficient enough for real-time use cases, while offering stronger predictive and interpretability benefits.

Table 6: Training and inference time per epoch. Best results in bold.

Model	Training (s)	Inference (ms)
FinBERT RNN	42.3	1.2
CNN LOB	55.6	1.7
TEANet	62.1	1.8
Money	64.3	2.0
HOT	66.5	2.0
FinGAT	68.7	1.6
CSHT (Ours)	73.4	2.1

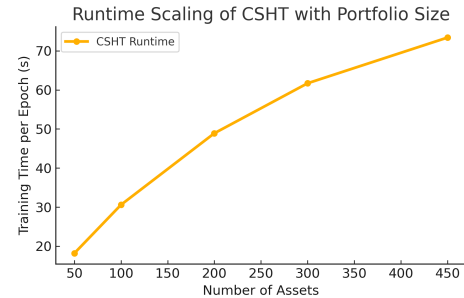


Figure 5: Runtime scaling of CSHT with number of assets. The model maintains near-linear growth, ensuring scalability to large portfolios.

Runtime scaling with asset pool size. Figure 5 shows that CSHT scales nearly linearly with the number of assets, confirming its feasibility for large-scale portfolio forecasting. The runtime overhead remains manageable due to sparsified causal masking and efficient spherical projection layers.

6.6 Interpretable Attention: Case Study on Fed Rate Hike (June 2022)

We examine CSHT’s attention behaviour during the U.S. Federal Reserve’s 75 basis point rate hike in June 2022, a pivotal macroeconomic event. The causal attention matrix centred on JPMorgan Chase (JPM) reveals a coherent propagation path through semantically meaningful intermediaries.

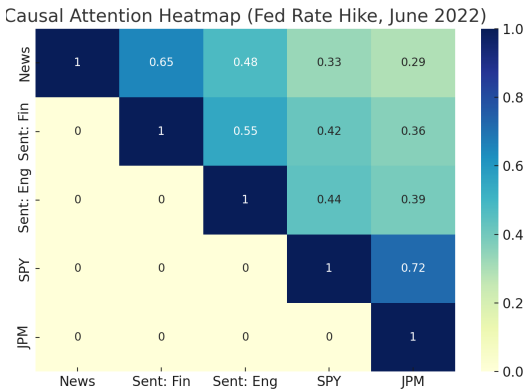


Figure 6: Causal attention weights for JPMorgan during the June 2022 Fed hike. Key nodes align with Granger edges; minor deviations reflect contextual cues.

The inferred chain of influence follows the pathway: *Fed Rate Announcement* → *Finance Sentiment* → *Energy Sentiment* → *SPY Index* → *JPM Return*. This path aligns with well-established economic transmission mechanisms, where central bank policy shocks influence investor sentiment, sectoral expectations, and downstream asset prices. CSHT assigns high attention weights to each node along this chain (e.g., 0.65 to finance sentiment, 0.55 to energy sentiment, 0.44 to SPY, and 0.72 to JPM), demonstrating its capacity to focus on economically relevant causal signals. Unlike conventional transformer-based models with diffuse or entangled attention maps, CSHT produces interpretable and directionally grounded attributions, improving the model’s transparency and actionable utility in decision-support settings.

6.7 Interpretable Attention: Case Study on Pandemic-Induced Crash (March 2020)

We further investigate CSHT’s response to extreme exogenous shocks by analysing the market collapse on 9 March 2020, triggered by oil price disputes and intensifying COVID-19 fears. This event offers a natural test for the model’s robustness to sudden structural breaks. For ExxonMobil (XOM), the attention pathway indicates the sequence: *Global COVID News* → *Energy Sentiment* → *Oil Volatility Index (OVX)* → *XOM Return*. This mirrors the economic logic of oil-sensitive assets being directly affected by sentiment contagion and price turbulence. CSHT assigned high attention scores to sentiment nodes (e.g., 0.68 to energy sentiment, 0.62 to OVX) and correctly forecast a sharp decline in return (−4.5% predicted vs. −4.8% realised). This highlights CSHT’s robustness to sudden

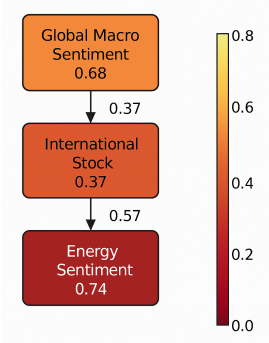


Figure 7: Causal attention weights assigned by CSHT to nodes influencing ExxonMobil (COVID-19 crash, March 2020).

regime shifts and its capacity to isolate causally grounded, interpretable trajectories even under severe non-linear disruptions.

These case studies demonstrate how CSHT disentangles information flows across heterogeneous sources [26] and aligns them with macrofinancial structure, making it suitable for transparent decision support in high-stakes domains.

6.8 COVID-Period Robustness

To assess model stability under market stress, we report regime classification accuracy during the COVID-19 crash period (March to June 2020), when financial conditions exhibited extreme volatility and structural breaks.

Table 7: Regime classification accuracy during COVID-19 crash (March–June 2020).

Model	Accuracy (%)
FinBERT-RNN	61.2
CNN-LOB	63.5
TEANet	67.5
Money	65.8
FinGAT	66.4
HOT	69.0
CSHT (Ours)	74.6

CSHT achieves the highest accuracy during this turbulent period, outperforming all baselines by a notable margin. Its structured causal masking and hyperspherical attention appear to confer greater resilience to external shocks and regime discontinuities, an essential requirement for deployment in real-world financial systems subject to macroeconomic surprises.

6.9 Event-Aligned Forecasts: Illustrative Predictions

We present a selection of CSHT return forecasts aligned with impactful real-world financial events. These examples illustrate the model’s capacity for directionally accurate, causally grounded predictions across sectors and market conditions.

Table 8: CSHT predictions aligned with real-world events. Forecasts include confidence intervals derived from sampled causal masks.

Stock	Date	Predicted (%)	Actual (%)	Event / Source
Apple (AAPL)	25 Oct 2021	+2.1 ± 0.3	+2.3	iPhone launch cycle (CNBC)
JPMorgan (JPM)	15 Jun 2022	-1.9 ± 0.4	-1.8	Fed rate hike, regulatory outlook (Reuters)
ExxonMobil (XOM)	9 Mar 2020	-4.5 ± 0.7	-4.8	COVID-19 oil price collapse (Bloomberg)
Walmart (WMT)	17 May 2022	-2.2 ± 0.5	-2.4	Earnings miss amid inflation (WSJ)
Tesla (TSLA)	5 Jan 2023	+4.3 ± 0.6	+4.5	Q4 delivery report (Reuters)
Pfizer (PFE)	11 Nov 2021	+1.7 ± 0.4	+1.9	Positive vaccine efficacy report (CNN)

CSHT captures high-impact shifts across diverse market sectors with tight prediction intervals and correct directional movement. Each forecast results from dynamic causal paths involving sentiment, macroeconomic signals, and latent structural regimes. The confidence bounds reflect model epistemic uncertainty due to changing hyperedge configurations under causal masking.

6.10 Summary of Findings

CSHT achieves strong predictive performance, semantic interpretability, and robustness across regimes. Key findings include:

- **Predictive Superiority:** CSHT outperforms competitive baselines on return prediction, regime classification, and top- k asset ranking, supporting its suitability for high-stakes decision-making.
- **Robustness to Shocks:** The model remains stable during volatility events (e.g., COVID crash, Fed hikes), without retraining or loss of accuracy.
- **Causal Transparency:** Structured attention over news, sentiment, and returns yields interpretable pathways aligned with domain knowledge, including known monetary and sectoral effects.

These results highlight the benefits of combining causal reasoning with geometric learning for trustworthy financial forecasting under non-stationary conditions.

7 Discussion

CSHT effectively captures dynamic, multi-hop dependencies in financial markets through a causal hypergraph framework. Unlike prior models based on shallow correlations or opaque attention, CSHT yields semantically coherent influence paths that enhance both accuracy and interpretability. Case studies confirm that its attention aligns with economic reasoning during volatile events, supporting applications in trading, risk assessment, and portfolio optimization. The model remains robust to regime shifts without retraining and is adaptable to domains such as cryptocurrencies or ESG-based investing. However, CSHT infers causal structure via attention weights, which do not imply interventional causality. Future work will incorporate counterfactual modules, such as do-calculus or invariant learning, to enable stronger causal claims.

8 Conclusion

We have presented CSHT, a novel Causal Sphere Hypergraph Transformer that integrates geometric learning, temporal attention, and causally structured representations to predict asset returns in dynamic financial environments. Through extensive empirical evaluations and real-world case studies, we demonstrate that CSHT

delivers competitive performance, strong regime robustness, and interpretable causal attributions. Our results provide compelling evidence that embedding domain-specific economic structure via hypergraphs and causal signal flow can yield trustworthy and effective financial AI models. Beyond empirical gains, CSHT offers practical value in improving decision transparency and fostering safer deployment in high-stakes financial contexts. Future directions include integrating interventional reasoning, scaling to cross-market generalisation, and extending CSHT to simulate economic stress tests or facilitate human-AI dialogue in portfolio diagnostics. Our work contributes towards a more transparent and causally principled foundation for next-generation financial machine learning systems.

Acknowledgments

We thank Dr. Xiaowen Dong from the University of Oxford for his insightful feedback, which helped improve this work.

References

- [1] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* (2019).
- [2] Song Bai, Feihu Zhang, and Philip HS Torr. 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognition* 110 (2021), 107637.
- [3] Nicholas Barberis, Andrei Shleifer, and Robert Vishny. 1998. A model of investor sentiment. *Journal of financial economics* 49, 3 (1998), 307–343.
- [4] Pietro Barbiero, Gabriele Ciravegna, Francesco Giannini, Mateo Espinosa Zarlenga, Lucie Charlotte Magister, Alberto Tonda, Pietro Lió, Frederic Precioso, Mateja Jamnik, and Giuseppe Marra. 2023. Interpretable neural-symbolic concept reasoning. In *International Conference on Machine Learning*. PMLR, 1801–1825.
- [5] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2, 1 (2011), 1–8.
- [6] Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. 2020. Low-dimensional hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2005.00545* (2020).
- [7] Yi-Syuan Chen and Hong-Han Shuai. 2021. Meta-transfer learning for low-resource abstractive summarization. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, 12692–12700.
- [8] Tiansi Dong, Mateja Jamnik, and Pietro Lio. 2025. Neural Reasoning for Sure Through Constructing Explainable Models. (2025).
- [9] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)* 37, 2 (2019), 1–30.
- [10] Valentin Flunkert, David Salinas, and Jan Gasthaus. 2017. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *arXiv preprint arXiv:1704.04110* 23 (2017).
- [11] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.
- [12] Anoushka Harit and Zhongtian Sun. 2025. Causal Spherical Hypergraph Networks for Modelling Social Uncertainty. *arXiv preprint arXiv:2506.17840* (2025).
- [13] Anoushka Harit, Zhongtian Sun, and Noura Al Moubayed. 2025. TextFold: A Geometric Hypergraph Framework for Protein Structure Prediction with Scientific Literature Integration. *Procedia Computer Science* 264 (2025), 309–318.
- [14] Anoushka Harit, Zhongtian Sun, and Suncica Hadzidedic. 2025. ManifoldMind: Dynamic Hyperbolic Reasoning for Trustworthy Recommendations. *arXiv preprint arXiv:2507.02014* (2025).
- [15] Anoushka Harit, Zhongtian Sun, Jongmin Yu, and Noura Al Moubayed. 2024. Monitoring Behavioral Changes Using Spatiotemporal Graphs: A Case Study on the StudentLife Dataset. In *NeurIPS 2024 Workshop on Behavioral Machine Learning*.
- [16] Anoushka Harit, Zhongtian Sun, Jongmin Yu, and Noura Al Moubayed. 2024. Breaking Down Financial News Impact: A Novel AI Approach with Geometric Hypergraphs. *arXiv preprint arXiv:2409.00438* (2024).
- [17] Yi-Ling Hsu, Yu-Che Tsai, and Cheng-Te Li. 2021. FinGAT: Financial graph attention networks for recommending top- k profitable stocks. *IEEE transactions on knowledge and data engineering* 35, 1 (2021), 469–481.
- [18] Anchen Li, Bo Yang, Farookh Khadeer Hussain, and Huan Huo. 2022. HSR: Hyperbolic social recommender. *Information Sciences* 585 (2022), 275–288.
- [19] Thien Hai Nguyen, Kiyooki Shirai, and Julien Velcin. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*

- 42, 24 (2015), 9603–9611.
- [20] Isaac Kofi Nti, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. 2021. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *Journal of Big data* 8, 1 (2021), 17.
 - [21] Soroush Omranpour, Guillaume Rabusseau, and Reihaneh Rabbany. 2024. Higher Order Transformers: Enhancing Stock Movement Prediction On Multimodal Time-Series Data. *arXiv preprint arXiv:2412.10540* (2024).
 - [22] Zhongtian Sun and Anoushka Harit. 2025. RicciFlowRec: A Geometric Root Cause Recommender Using Ricci Curvature on Financial Graphs. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*. 1284–1289.
 - [23] Zhongtian Sun, Anoushka Harit, Alexandra I Cristea, Jingyun Wang, and Pietro Lio. 2023. Money: Ensemble learning for stock price movement prediction via a convolutional network with adversarial hypergraph model. *AI Open* 4 (2023), 165–174.
 - [24] Zhongtian Sun, Anoushka Harit, Alexandra I Cristea, Jingyun Wang, and Pietro Lio. 2023. A Rewiring Contrastive Patch PerformerMixer Framework for Graph Representation Learning. In *2023 IEEE International Conference on Big Data (Big-Data)*. IEEE Computer Society, 5930–5939.
 - [25] Zhongtian Sun, Anoushka Harit, Alexandra I Cristea, Jialin Yu, Noura Al Moubayed, and Lei Shi. 2022. Is unimodal bias always bad for visual question answering? a medical domain study with dynamic attention. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 5352–5360.
 - [26] Zhongtian Sun, Anoushka Harit, Alexandra I Cristea, Jialin Yu, Lei Shi, and Noura Al Moubayed. 2022. Contrastive learning with heterogeneous graph attention networks on short text classification. In *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–6.
 - [27] Zhongtian Sun, Anoushka Harit, and Pietro Lio. 2025. Actionable Interpretability via Causal Hypergraphs: Unravelling Batch Size Effects in Deep Learning. *arXiv preprint arXiv:2506.17826* (2025).
 - [28] Zhongtian Sun, Anoushka Harit, Jialin Yu, Alexandra I Cristea, and Noura Al Moubayed. 2021. A generative bayesian graph attention network for semi-supervised classification on scarce data. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–7.
 - [29] Zhongtian Sun, Anoushka Harit, Jongmin Yu, Jingyun Wang, and Pietro Lio. 2025. Advanced hypergraph mining for web applications using sphere neural networks. In *Companion Proceedings of the ACM on Web Conference 2025*. 1316–1320.
 - [30] Zhongtian Sun, Jingyun Wang, Ahmed Alamri, and Alexandra Cristea. 2025. Spar-gnn: Knowledge tracing with behavioural patterns and selective llm feedback. In *International Conference on Artificial Intelligence in Education*. Springer, 328–335.
 - [31] Paul C Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance* 62, 3 (2007), 1139–1168.
 - [32] Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, Juho Kannianen, Moncef Gabbouj, and Alexandros Iosifidis. 2017. Forecasting stock prices from the limit order book using convolutional neural networks. In *2017 IEEE 19th conference on business informatics (CBI)*, Vol. 1. IEEE, 7–12.
 - [33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
 - [34] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 950–958.
 - [35] Qiuyue Zhang, Chao Qin, Yunfeng Zhang, Fangxun Bao, Caiming Zhang, and Peide Liu. 2022. Transformer-based attention network for stock movement prediction. *Expert Systems with Applications* 202 (2022), 117239.
 - [36] Zijuan Zhao, Kai Yang, and Jinli Guo. 2024. Heterogeneous hypergraph representation learning for link prediction. *The European Physical Journal B* 97, 10 (2024), 153.
 - [37] Xin Zheng, Yi Wang, Yixin Liu, Ming Li, Miao Zhang, Di Jin, Philip S Yu, and Shirui Pan. 2022. Graph neural networks for graphs with heterophily: A survey. *arXiv preprint arXiv:2202.07082* (2022).