

TimeMKG: Knowledge-Infused Causal Reasoning for Multivariate Time Series Modeling

Yifei Sun^{1,2}, Junming Liu², Yirong Chen², Xuefeng Yan^{1*}, Ding Wang^{2*}

¹Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai, 200237, P. R. China

²Shanghai Artificial Intelligence Laboratory
yifeisun@mail.ecust.edu.cn, xfyan@ecust.edu.cn, wangding@pjlab.org.cn

Abstract

Multivariate time series data typically comprises two distinct modalities: variable semantics and sampled numerical observations. Traditional time series models treat variables as anonymous statistical signals, overlooking the rich semantic information embedded in variable names and data descriptions. However, these textual descriptors often encode critical domain knowledge that is essential for robust and interpretable modeling. Here we present TimeMKG, a multimodal causal reasoning framework that elevates time series modeling from low-level signal processing to knowledge informed inference. TimeMKG employs large language models to interpret variable semantics and constructs structured Multivariate Knowledge Graphs that capture inter-variable relationships. A dual-modality encoder separately models the semantic prompts—generated from knowledge graph triplets—and the statistical patterns from historical time series. Cross-modality attention aligns and fuses these representations at the variable level, injecting causal priors into downstream tasks such as forecasting and classification—providing explicit and interpretable priors to guide model reasoning. The experiment in diverse datasets demonstrates that incorporating variable-level knowledge significantly improves both predictive performance and generalization.

Introduction

Multivariate time series data (MTSD) are critical in domains such as industrial automation (Wang, Zhu, and He 2024), finance (Sezer, Gudelek, and Ozbayoglu 2020), energy systems (Zhou et al. 2021), and healthcare (Seong, Lee, and Chae 2024). However, their complex, high-dimensional and often non-stationary nature poses persistent challenges to modeling, interpretation, and decision making (Huang et al. 2025; Zhao et al. 2025). Traditional methods - ranging from statistical methods to modern deep learning (Wang et al. 2024b) - have mainly focused on numerical temporal patterns, often overlooking the rich semantic information encoded in the names and descriptions of variables.

However, MTSDs are inherently multimodal: the *semantic information* conveyed by header names (textual modal-

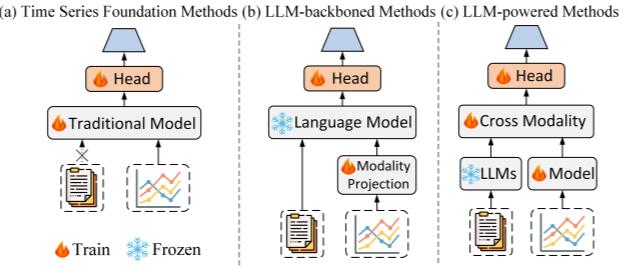


Figure 1: (a) Time Series Foundation & Traditional methods overlook header names and domain knowledge; (b) LLM-backboned methods tend to produce ambiguous semantic representations, as pretrained language models lack inherent temporal modeling capabilities; (c) LLM-powered methods integrate the strong semantic understanding of LLMs with the temporal modeling strengths of foundational models to enable more effective forecasting.

ity) and the temporal dynamics captured in *varying observations* (numerical modality). This dual structure naturally characterizes MTSD as a multimodal task. Textual descriptors often encode critical domain knowledge, such as causal relationships, physical meaning, and system roles, that can guide model reasoning (Wang et al. 2025). Crucially, identical numerical values can have vastly different meanings across variables. For instance, a reading of 100 could represent normal engine temperature, but a life-threatening fever in medical contexts. Ignoring this semantic modality leads to blind spots in both interpretability and robustness, especially in safety-critical or knowledge-intensive applications.

Traditional time series models face significant challenges in leveraging causal information (Mameche et al. 2025). On the one hand, statistical frameworks lack flexible mechanisms to encode prior domain knowledge (Li et al. 2023). On the other hand, most data-driven approaches (Zheng et al. 2020) often rely on task-specific architectures or hardcoded inductive biases to incorporate prior causal relationships among variables, making them difficult to generalize across tasks. As a result, modeling true causal interactions remains challenging, and models often depend on spurious correlations rather than grounded reasoning.

Large language models (LLMs) (Yang et al. 2025) of-

TimeMKG: Çok Değişkenli Zaman Serisi Modellemesi için Bilgiyle Zenginleştirilmiş Nedensel Akıl Yürütme

Yifei Sun^{1,2}, Junming Liu², Yirong Chen², Xuefeng Yan^{1*}, Ding Wang^{2*}

¹Enerji Kimyasal Süreçlerinde Akıllı Üretim Anahtar Laboratuvarı, Eğitim Bakanlığı, Doğu Çin Bilim ve Teknoloji Üniversitesi, Şanghay, 200237, Çin Halk Cumhuriyeti

²Şanghay Yapay Zeka Laboratuvarı yifeisun@mail.ecust.edu.cn, xfyan@ecust.edu.cn, wangding@pjlab.org.cn

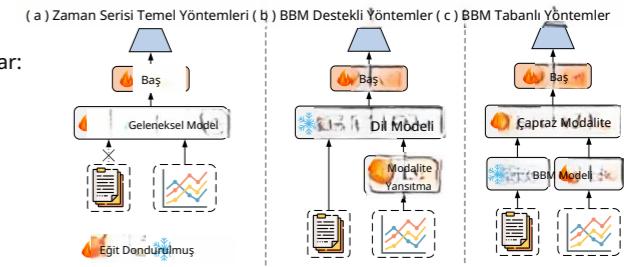
Özet

Cok değişkenli zaman serisi verileri genellikle iki farklı modaliteyi kapsar: değişken semantiği veya örneklenmiş sayısal gözlemler. GelenekSEL zaman serisi modelleri, değişkenleri anonim istatistiksel sinyaller olarak ele alır ve değişken adlarına ile veri açıklamalarına gömülü zengin semantik bilgi göz ardı eder. Ancak, bu metinsel tanımlayıcılar genellikle sağlam ve yorumlanabilir modelleme için kritik olan alan bilgisini kodlar. Burada, zaman serisi modellemesini düşük seviyeli sinyal işleme düzeyinden bilgi odaklı çıkarım düzeyine yükselten, çok modlu nedensel akıl yürütme çerçevesi TimeMKG'yi sunuyoruz. TimeMKG, değişken semantiğini yorumlamak için büyük dil modellerini kullanır ve değişkenler arası ilişkileri yakalayan yapılandırılmış Çok Değişkenli Bilgi Grafikleri (ÇBBK) oluşturur. Bir çift-modalite kodlayıcı, bilgi grafiği ütemelerinden üretilen semantik temsilleri ve geçmiş zaman serilerinden elde edilen istatistiksel örüntülerini ayrı ayrı modeller. Çapraz-modalite dikkat, bu temsilleri değişken düzeyinde hizalar ve birleştirir; tahmin ve sınıflandırma gibi aşağı yukarı görevlerine nedensel öncüler enjekte ederek model akıl yürütme yönlendirilecek açık ve yorumlanabilir öncüler sağlar. Çeşitli veri kümelerindeki deneyler, değişken düzeyinde bilginin dahil edilmesinin hem tahmin performansını hem de genelleme önemde iyileştirdiğini göstermektedir.

Giriş

Cok değişkenli zaman serisi verileri (MTSD); endüstriyel otomasyon (Wang, Zhu ve He 2024), finans (Sezer, Gudelek ve Ozbayoglu 2020), enerji sistemleri (Zhou vd. 2021) ve sağlık hizmetleri (Seong, Lee ve Chae 2024) gibi alanlarda kritik öneme sahiptir. Ancak, karmaşık, yüksek boyutlu ve genellikle durağan olmayan yapıları, modelleme, yorumlama ve karar alma süreçlerinde sürekli zorluklar ortaya koymaktadır (Huang ve diğerleri 2025; Zhao ve diğerleri 2025). İstatistiksel yöntemlerden modern derin öğrenmeye (Wang ve diğerleri 2024b) kadar değişen geleneksel yöntemler, esas olarak sayısal zamansal örüntülere odaklanmış, değişkenlerin adları ve açıklamalarında kodlanmış zengin semantik bilgiyle sıkılıkla göz ardı etmiştir. Ancak, ÇZVD'ler doğası gereği çok modludur: başlık adlarının iletişimi semantik bilgi (metinsel modalite)

* Bu yazarlar, eş-sorumlu yazarlar olarak bu çalışmaya eşit derecede katkıda bulunmuştur.
Telif Hakkı © 2026, Yapay Zekanın Geliştirilmesi Derneği (www.aaai.org). Tüm hakları saklıdır.



Şekil 1: (a) Zaman Serisi Temel ve Geleneksel yöntemler başlık adlarını ve alan bilgisini göz ardı etmemektedir; (b) BBM destekli yöntemler, önceden eğitilmiş dil modellerinin doğal zamansal modelleme yeteneklerinden yoksun olması nedeniyle belirsiz semantik temsiller üretme eğilimindedir; (c) BBM destekli yöntemler, daha etkili tahmin sağlamak amacıyla BBM'lerin güçlü anlamsal anlayışını temel modellerin zamansal modelleme yetenekleriyle bütünleştirir.

sallık) ve değişen gözlemlerde (sayısal modalite) yakalanan zamansal dinamikler. Bu ikili yapı, MTSD'yi doğal olarak çok modlu bir görev olarak karakterize eder. Metinsel tanımlayıcılar; nedenel iliskiler, fiziksels anlam ve sistem rolleri gibi kritik alan bilgisini sıkılıkla modellerin akıl yürütmesine rehberlik edebilecek şekilde kodlar (Wang vd. 2025). En önemli, aynı sayısal değerler değişkenler arasında büyük ölçüde farklı anımlara sahip olabilir. Örneğin, 100'lük bir okuma normal motor sıcaklığını temsil edebilirken, tıbbi bağışıklarda hayatı tehdit eden bir ateş'i ifade edebilir. Bu anlamsal modalitenin göz ardı edilmesi, özellikle güvenlik açısından kritik veya bilgi yoğun uygulamalarda hem yorumlanabilirlik hem de sağlamlık açısından kör noktalara yol açar.

Geleneksel zaman serisi modelleri, nedensel bilgiden yararlanmadı önemli zorluklarla karşılaşmaktadır (Mameche vd. 2025). Bir yandan, istatistiksel çerçeveler, önceden edinilmiş alan bilgisini kodlamak için esnek mekanizmalardan yoksundur (Li vd. 2023). Öte yandan, çoğu veri odaklı yaklaşım (Zheng vd. 2020), değişkenler arasındaki önceden edinilmiş nedensel iliskileri dahil etmek için genellikle görevde özgü mimarilere veya sabit kodlanmış tümevarımsal önyargılara dayanır, bu da onların görevler arasında genelleştirilmesini zorlaştırır. Sonuç olarak, gerçek nedensel etkileşimleri modellemek zorlu olmaya devam etmekte ve modeller genellikle temellendirilmiş akıl yürütmeden ziyade yaniltıcı korelasyonlara bağımlı olmaktadır.

Büyük dil modelleri (BBM'ler) (Yang vd. 2025) -

fer a powerful means of uncovering causal relationships between variables. Pretrained on large-scale text corpora, LLMs encode extensive common-sense and domain-specific knowledge, enabling them to interpret variable semantics and infer meaningful dependencies (Jin et al. 2023b). These capabilities make LLMs valuable tools for identifying latent relationships in MTSD. However, the knowledge they acquire remains implicit—embedded in opaque neural weights—lacking the transparency and structure needed for interpretability and refinement. This motivates the need for explicit, structured representations of causal knowledge.

In contrast, knowledge graphs (KGs) explicitly encode variable relationships as triplets, offering a clear and interpretable structure. KGs organize information through a clear entity–relation format, making them an ideal structure for storing causal relationships and highly suitable for injecting semantic priors into models. Susanti (Susanti and Färber 2024) showed that incorporating KGs as prompts allows LLMs to outperform fine-tuned models in few-shot causal discovery. Furthermore, Kim et al. (Kim et al. 2024) demonstrated that graph-based prompts enhance LLMs’ understanding of causal chains via *random walks*, significantly improving reasoning performance. KG is also conducive to expansion and dynamic updates: by simply adding or removing entities or edges, relevant knowledge can be instantly revised upon receiving new contexts.

In this work, we introduce **TimeMKG**, an LLMs-powered framework that leveraging KG for multivariate time series modeling. TimeMKG views MTSD as a dual-modality task, aligning: (1) The textual modality, where LLMs interpret variable names, descriptions, and expert knowledge to construct a causal knowledge graph; (2) The numerical modality, where statistical dependencies are extracted from temporal observations. By aligning these modalities through cross-attention, TimeMKG injects causal priors from the semantic space into temporal modeling, enabling more interpretable and robust reasoning. Our contributions are as follows:

- Variable-level semantic modeling:** We are the first to introduce variable names as an input modality and explicitly incorporate a textual branch for MTSD modeling. Using the semantic understanding capabilities of LLMs, we extend domain-level causal inference to the fine-grained variable level, enabling consistent interpretation across datasets.
- Explicit causal KG construction:** We automatically extract causal triplets between variables using LLMs to construct a KG that is human-auditable and incrementally updatable. This explicit structure significantly improves the performance of causal reasoning and prompt generation tasks.
- LLM-powered dual-modality framework:** TimeMKG adopts a dual-branch design, LLMs extract causal relationships from textual semantics, while time-series models capture statistical dependencies in numerical observations. This unified framework supports diverse time series tasks while explicitly capturing causal dependencies.
- SOTA performance on multiple tasks:** TimeMKG

achieves consistently superior performance across long-term and short-term forecasting and classification tasks in diverse domains, with notable gains in interpretability and accuracy.

Related Work

Deep learning models have been widely explored for MTSD modeling. **CNN-based** methods (Bai, Kolter, and Koltun 2018; Wu et al. 2022) extract intra and inter-variable features by applying sliding window convolutions. **RNN-based** methods (Qin et al. 2017; Salinas et al. 2020; Oreshkin et al. 2019) accumulate historical information through recurrent hidden states and alleviate the vanishing gradient problem using gated mechanisms. **Linear-based** models (Das et al. 2023; Zeng et al. 2023) improve predictive performance by separately modeling trend and seasonal components. Early **Transformer-based** methods (Zhou et al. 2021) capture nonlinear dependencies across time steps via self-attention mechanisms. However, multivariate inputs are typically projected into a unified dimension at the embedding stage, sacrificing the preservation of variable-specific statistical properties.

Due to the sequential nature of similarities between textual and time series modality, recent LLMs advances have been extended to time series analysis. As shown in Figure 1, these methods can be broadly categorized into three types:

(1) **Time Series Foundation Methods** are trained on large-scale time series data using strategies inspired by LLMs, aiming to learn generalizable temporal representations across diverse domains. Time-MoE (Shi et al. 2024) introduces a sparse mixture-of-experts mechanism to reduce computational and inference costs. Moirai (Woo et al. 2024) addresses cross-domain dimensional inconsistency by unfolding multivariate time series into a unified format during training. Although these models aim to solve zero-shot forecasting in the time series domain, their performance on private, domain-specific datasets remains limited.

(2) **LLM-backboned Methods**: To enable LLMs to process temporal information, numerical sequences must be discretized into univariate tokens and mapped into pseudo-word representations. For example, TimeLLM (Jin et al. 2023a) concatenates prompts with numerical tokens and feeds them into a frozen LLM, using the final hidden states as model outputs. However, since LLMs are pre-trained on natural language corpora rather than structured temporal data, their modeling effectiveness in this context has been questioned (Tan et al. 2024).

(3) **LLM-powered Methods**: These methods (Liu et al. 2024a) adopt a dual-modality paradigm, where LLMs focus on the textual modality—leveraging their strong language understanding and generation capabilities—while standard time-series models handle temporal dependencies in numerical data. This “division-of-responsibility” framework enables complementary strengths: LLMs contribute semantic and causal insights, while temporal models capture statistical dynamics. This facilitates an organic fusion of semantic knowledge and temporal modeling.

değişkenler arasındaki nedensel ilişkileri ortaya çıkarmak için güçlü bir araç sunar. Büyük ölçekli metin korpuslarında önceden eğitilmiş Büyük Dil Modelleri (BBM’ler), kapsamlı sağduyu ve alana özgü bilgiyi kodlayarak değişken semantığını yorumlamalarına ve anlamlı bağımlılıkları çıkarmalarına olanak tanır (Jin ve diğerleri 2023b).

Bu yetenekler, BBM’leri Çok Değişkenli Zaman Serisi Verilerindeki (ÇZSV) gizli ilişkileri tanımlamak için değerli araclar haline getirmektedir. Ancak, edindikleri bilgi örtük kalır—opak nöral ağırlıklara gömülü olarak—yorumlanabilirlik ve geliştirme için gerekli şeffaflık ve yapıdan yoksundur. Bu durum, nedensel bilginin açık ve yapılandırılmış temsillerine olan ihtiyacı ortaya koymaktadır.

Aksine, bilgi grafikleri (KGL’er), değişken ilişkilerini açıkça üçlüler olarak kodlayarak net ve yorumlanabilir bir yapı sunar. KGL’er, bilgiyi açık bir varlık-ilişki formatında düzenler; bu da onları nedensel ilişkileri depolamak için ideal bir yapı ve semantik ön bilgileri modellere enjekte etmek için son derece uygun hale getirir. Susanti (Susanti ve Farber 2024), KGL’eri ipuçları olarak dahil etmenin BBM’lerin az sayıda örnekle nedensel keşifte ince ayarlı modellerden daha iyi performans göstermesini sağladığını göstermiştir. Ayrıca, Kim ve diğerleri (Kim ve diğerleri 2024), grafik tabanlı ipuçlarının BBM’lerin nedensel zincirleri *random walks* aracılığıyla anlamalarını geliştirdiğini ve muhakeme performansını önemli ölçüde artırdığını ortaya koymıştır. KB ayrıca genişlemeye ve dinamik güncellemedere de elverişlidir: Yalnızca varlıklar veya kenarları ekleyerek veya kalıdirarak, yeni bağamlar alındığında ilgili bilgi alanında revize edilebilir.

Bu çalışmada, çok değişkenli zaman serisi modellemesi için KB’den yararlanan BBM destekli bir çerçeve olan **TimeMKG**’yi sunuyoruz. TimeMKG, MTSD’yi ikili modaliteli bir görev olarak ele almakta ve şunları hizalamaktadır: (1) BBM’lerin değişken adlarını, açıklamalarını ve uzman bilgisini nedensel bilgi grafiği oluşturmak üzere yorumladığı metinsel modalite; (2) İstatistiksel bağımlılıkların zamansal gözlemlerden çıkarıldığı sayısal modalite. Bu modaliteleri çapraz dikkat aracılığıyla hizalayarak TimeMKG, anlamsal uzaydan nedensel öncüleri zamansal modellemeye enjekte eder ve böylece daha yorumlanabilir ve sağlam bir akıl yürütme sağlar. Katkılarımız şunlardır:

- **Değişken düzeyinde anlamsal modelleme :** Değişken adlarını bir girdi modalitesi olarak tanıtan ve MTSD modellemesi için açıkça metinsel bir dali dâhil eden ilk bizi. BBM’lerin anlamsal anlaması yeteneklerini kullanarak, alan düzeyindeki nedensel karışımı ince taneli değişken düzeyine genişletiyor, böylece veri kümeleri arasında tutarlı yorumlama sağlıyoruz.
- **Açık nedensel KG oluşturma :** İnsan tarafından denetlenebilir ve aşamalı olarak güncellenebilir bir Bilgi Grafiği (KG) oluşturmak amacıyla, BBM’leri kullanarak değişkenler arasındaki nedensel ilişkileri otomatik olarak çıkarırız. Bu açık yapı, nedensel akıl yürütme ve istem oluşturma görevlerinin performansını önemli ölçüde iyileştirir.
- **BBM destekli çift modlu çerçeve :** TimeMKG, çift dallı bir tasarım benimser; BBM’ler metinsel anlamlardan nedensel ilişkileri çıkarırken, zaman serisi modelleri sayısal gözlemlerdeki istatistiksel bağımlılıkları yakalar. Bu birleşik çerçeve, nedensel bağımlılıkları açıkça yakalarken çeşitli zaman serisi görevlerini destekler.
- **Birden çok görevde SOTA performansı:** TimeMKG

çeşitli alanlarda uzun vadeli ve kısa vadeli tahmin ve sınıflandırma görevlerinde tutarlı bir şekilde üstün performans sergilemeyecek, yorumlanabilirlik ve doğrulukta dikkate değer kazanımlar elde etmektedir.

İlgili Çalışmalar

Derin öğrenme modelleri, çok değişkenli zaman serisi modellemesi (MTSD) için yaygın olarak araştırılmıştır. CNN tabanlı yöntemler (Bai, Kolter ve Koltun 2018; Wu vd. 2022), kayan pencere evrimleri uygulayarak değişken içi ve değişkenler arası öznitelikleri çıkarır. RNN tabanlı yöntemler (Qin vd. 2017; Salinas vd. 2020; Oreshkin vd. 2019), tekrarlayan gizli durumlar aracılığıyla geçmiş bilgiyi biriktirir ve geçitli mekanizmalar kullanarak kaybolan gradyan sorununu hafifletir. Doğrusal tabanlı modeller (Das vd. 2023; Zeng vd. 2023), trend ve mevsimsel bileşenleri ayrı ayrı modelleyerek tahmin performansını artırır. Erken Transformer tabanlı yöntemler (Zhou vd. 2021), öz-dikkat mekanizmaları aracılığıyla zaman adımları arasındaki doğrusal olmayan bağımlılıkları yakalar. Ancak, çok değişkenli girdiler genellikle genome aşamasında tek bir boyuta yansıtılır; bu durum, değişkene özgü istatistiksel özelliklerin korunmasını olumsuz et-

Metinsel ve zaman serisi modaliteleri arasındaki sıralı benzerlikler göz önüne alındığında, BBM’lerdeki son gelişmeler zaman serisi analizine de uyaranmıştır. Şekil 1’de görüldüğü üzere, bu yöntemler genel olarak üç ana kategoriye ayrılabilir:

(1) **Zaman Serisi Temel Yöntemleri :** BBM’lerden esinlenen stratejiler kullanılarak büyük ölçekli zaman serisi verileri üzerinde eğitilirler ve farklı alanlarda genellenebilir zaman-sal temsiller öğrenmeyi amaçları. Time-MoE (Shi vd. 2024), hesaplama ve çıkışım maliyetlerini azaltmak için seyrek bir uzmanlar karışımı mekanizması sunar. Moirai (Woo vd. 2024), çok değişkenli zaman serilerini eğitim sırasında birleşik bir formata dönüştürerek alanlar arası boyutsal tutarsızlığı giderir. Bu modeller zaman serisi alanında sıfır-ataş tahminiini çözmemi hedefleseler de, özel, alana özgü veri kümelerindeki performansları sınırlı kalmaktadır.

(2) **BBM Destekli Yöntemler :** BBM’lerin zamansal bilgiyi işleyebilmesi için, sayısal dizilerin tek değişkenli token’lara ayrıştırılması ve sözde kelime temsillerine dönüştürülmesi gerekmektedir. Örneğin, TimeLLM (Jin ve diğerleri 2023a), istemleri sayısal jetonlarla birleştirir ve bunları dondurulmuş bir BBM’ye besler, son gizli durumları model çıktıları olarak kullanır. Ancak, BBM’ler yapılandırılmış zamansal veriler yine doğal dil korpusları üzerinde önceden eğitildiğinden, bu bağlamdaki modelleme etkinlikleri sorgulanmıştır (Tan ve diğerleri 2024).

(3) **BBM Destekli Yöntemler :** Bu yöntemler (Liu ve diğerleri 2024a) çift modlu bir paradigmayı benimser; burada BBM’ler metinsel modaliteye odaklanarak güçlü dil anlaması ve üretme yeteneklerinden yararlanırken, standart zaman serisi modelleri sayısal verilerdeki zamansal bağımlılıkları ele alır. Bu “sorumluluk paylaşımı” çerçevesi tamamlayıcı güçleri mümkün kılar: BBM’ler anlamsal ve nedensel içgörüler sağlarken, zamansal modeller istatistiksel dinamikleri yakalar. Bu, anlamsal bilginin ve zamansal modellemenin organik birleşimini kolaylaştırır.

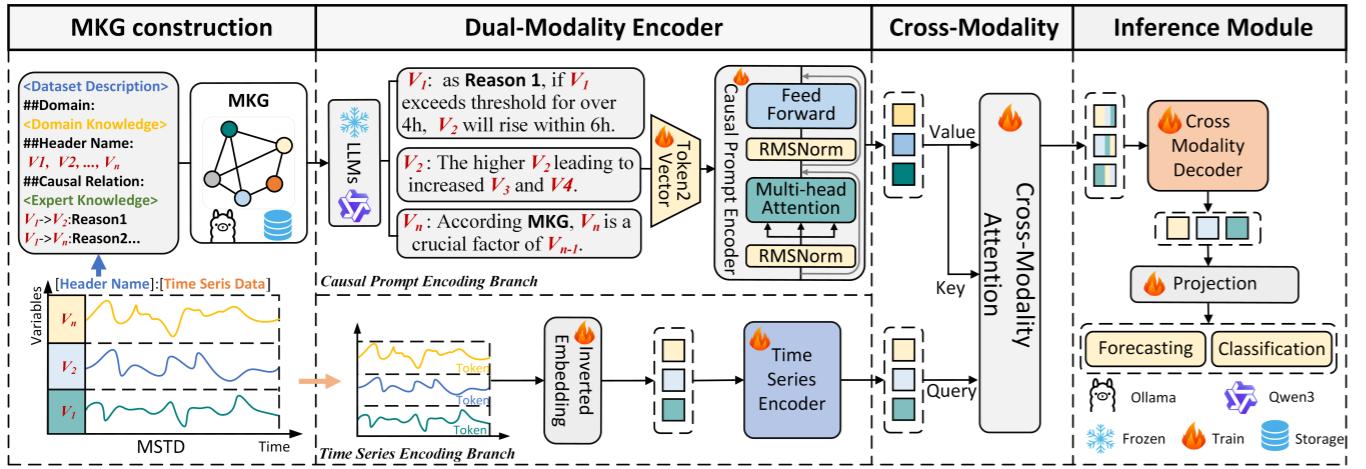


Figure 2: Overall structure of TimeMKG. First, LLMs are used to extract causal knowledge and construct a knowledge graph. Then, a dual-modality encoder is used to model the prior causal relationship and statistical correlation between variables respectively. Finally, cross-modality attention is used to summarize the two types of representations and apply to the target tasks.

Preliminaries

Multivariate Time Series Data. Numerical modality is denoted as $\mathbf{X}_{1:T} = \{x_1, x_2, \dots, x_T\} \in \mathbb{R}^{T \times N}$, with T time steps and N variables. Textual modality $\mathcal{V} = \{v_1, v_2, \dots, v_N\} \in \mathbb{R}^N$ contains N header variable names.

Knowledge Graph Definition. Define a Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ to express semantic and causal relationships between variables. Where \mathcal{V} denotes variable nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{R} \times \mathcal{V}$ represents directed edges between them, \mathcal{R} defines the relation types.

Problem Definition. Given a historical sequence $\mathbf{X}_{1:T}$, **Forecasting** task is to predict the future L time steps $\mathbf{Y}_{T:T+L} = \{x_{T+1}, \dots, x_{T+L}\} \in \mathbb{R}^{L \times N}$, $f_{\text{fore}} : \mathbf{X}_{1:T} \mapsto \mathbf{Y}_{T:T+L}$. **Classification** is to classify $\mathbf{X}_{1:T}$ into categories, $f_{\text{cls}} : \mathbf{X}_{1:T} \mapsto \hat{c} \in \mathbb{R}^C$.

Methodology

Overall Architecture

As shown in Figure 2, the overall structure of TimeMKG consists of four key modules: multivariate knowledge graph, dual-modality encoder, cross-modality attention, and inference module.

The multivariate knowledge graph is built by prompting a pretrained LLM with variable descriptions and causal knowledge to support domain reasoning. The Dual-Modality Encoder consists of causal prompt and time series branches, and has been demonstrated in TimeCMA (Liu et al. 2024a) to be effective for multimodal time series tasks. The causal branch uses a frozen LLM and a causal prompt encoder to refine inter-variable logic, while the time series branch employs inverted embeddings to capture meaningful inter-variable patterns from the numerical modality. Cross-Modality Attention aligns the two modalities into unified representations. The Inference Module decodes them via a cross-modality decoder and maps the result to task-specific outputs.

Multivariate Knowledge Graph

Recent advances (Huguet Cabot and Navigli 2021) in LLMs have shown remarkable capabilities in extracting entities, relationships, and structured patterns from unstructured data. Building on this strength, we propose the **Multivariate Knowledge Graph (MKG)**, a domain-specific graph representation tailored for MTSD. In MKG, nodes represent time-dependent variables, while directed edges capture causal or functional dependencies, serving as a prior graph to inform temporal modeling.

To build the proposed MKG, we first refine the generated variable-level textual descriptions $\hat{\mathcal{S}}$, optionally enriching them with external domain-specific knowledge \mathcal{T} , forming a comprehensive input for graph construction. The combined knowledge is fed into the LightRAG model (Guo et al. 2024; Liu et al. 2025), chosen for its efficiency in multi-hop reasoning and dynamic knowledge integration.

$$\mathcal{G}_M = \text{LightRAG}(\mathcal{I}), \quad \mathcal{I} = \hat{\mathcal{S}} \oplus \mathcal{T}, \quad (1)$$

Here, \oplus denotes optional concatenation with external textual knowledge \mathcal{T} . The combined input is then fed into LightRAG, which employs an LLM to infer graph structures that capture meaningful relationships among variables.

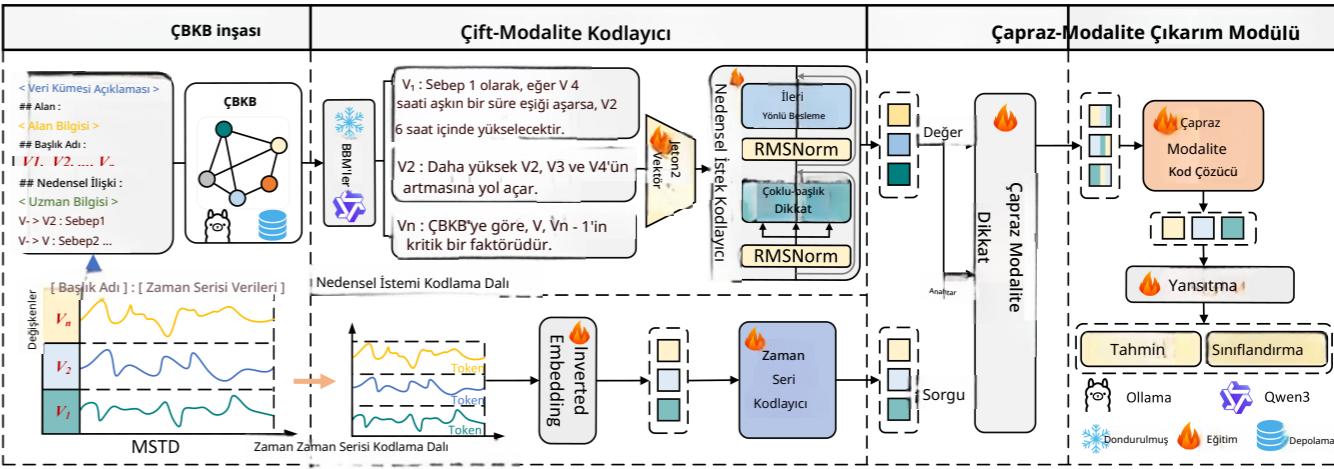
The obtained MKG, $\mathcal{G}_M = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, consists of a set of nodes \mathcal{V} representing variables, a set of directed edges \mathcal{E} , and a corresponding set of relations \mathcal{R} .

$$\mathcal{G}_M = \{(v_i, r, v_j) \mid v_i, v_j \in \mathcal{V}, r \in \mathcal{R}\}, \quad (2)$$

Each edge $e \in \mathcal{E}$ is represented as a triplet (v_i, r, v_j) , where $v_i, v_j \in \mathcal{V}$ and $r \in \mathcal{R}$, indicating that a directed causal relation linking v_i to v_j via r . MKG construction leverages LLMs to automatically extract meaningful inter-variable relationships without manual annotation.

Dual-Modality Encoder

Causal Prompt Encoding Branch. Given an input header name $v_k \in \mathcal{V}$, we construct a variable-specific query $\mathcal{Q}(v_k)$.



Şekil 2: TimeMKG'nin genel yapısı. İlk olarak, nedensel bilgi çıkarmak ve bir bilgi grafiği oluşturmak için BBM'ler kullanılır. Ardından, değişkenler arasındaki öncü nedensel ilişkisi ve istatistiksel korelasyonu sırasıyla çift-modalite kodlayıcı kullanılır. Son olarak, iki tür temsili özetlemek ve hedef görevlere uygulamak için çapraz-modalite dikkat mekanizması kullanılır.

Ön Bilgiler

Çok Değişkenli Zaman Serisi Verileri. Sayısal modalite, T zaman adımı ve N değişkenle birlikte $X_{1:T} = \{x_1, x_2, \dots, x_T\} \in \mathbb{R}^{T \times N}$ olarak belirtilir. Metinsel modalite $V = \{v_1, v_2, \dots, v_N\} \in \mathbb{R}^N$, N adet başlık değişken adı içerir.

Bilgi Grafiği Tanımı. Değişkenler arasındaki semantik ve nedensel ilişkileri ifade etmek için $G = (V, E, R)$ bir graf tanımlayın. Burada V , değişken düğümlerini; $E \subseteq V \times R \times V$, aralarındaki yönlendirilmiş kenarları; ve R , ilişki türlerini tanımlar.

Problem Tanımı. Verilen bir geçmiş dizisi $\mathbf{X}_{1:T}$, Tahmin görevi, gelecekteki L zaman adımlarını $\mathbf{Y}_{T:T+L} = \{x_{T+1}, \dots, x_{T+L}\} \in \mathbb{R}^{L \times N}$, $f_{\text{fore}} : \mathbf{X}_{1:T} \mapsto \mathbf{Y}_{T:T+L}$ olarak öngörmektir. Sınıflandırma, $X_{1:T}$ kategorilere ayırmaktır, $f_{\text{cls}} : \mathbf{X}_{1:T} \mapsto \hat{c} \in \mathbb{R}^C$.

Metodoloji

Genel Mimari

Şekil 2'de gösterildiği gibi, TimeMKG'nin genel yapısı dört temel modülden oluşmaktadır: çok değişkenli bilgi grafiği, çift-modalite kodlayıcı, çapraz-modalite dikkat ve çıkarım modülü.

Çok değişkenli bilgi grafiği, alan muhakemesini desteklemek amacıyla önceden eğitilmiş bir BBM'nin değişken açıklamaları ve nedensel bilgi ile yönlendirilmesiyle oluşturulur. Çift-Modalite Kodlayıcı, nedensel komut ve zaman serisi dallarından oluşur ve TimeCMA'da (Liu ve diğerleri 2024a) çok modlu zaman serisi görevleri için etkili olduğu gösterilmiştir. Nedensel dal, değişkenler arası mantığı iyileştirmek için dondurulmuş bir BBM ve nedensel bir komut kodlayıcı kullanırken, zaman serisi dalı sayısal modaliteden değişkenler arası örüntülerini yakalamak için ters çevrilmiş gömmeleri kullanır. Çapraz-Modalite Dikkat, iki modaliteyi birleşik göstergelere hızlar. Çıkım Modülü, bunları bir çapraz-modalite kod çözücü aracılığıyla çözer ve sonucu görevye özgü çıktılarına eşler.

Çok Değişkenli Bilgi Grafiği

BBM'lerdeki son gelişmeler (Huguet Cabot ve Navigli 2021), yapılandırılmamış verilerden varlıklarını, ilişkileri ve yapılandırılmış örtütüleri çalışma konusunda dikkat çekici yetenekler göstermiştir. Bu gücü temel olarak, MTSD'ye özel olarak tasarlanmış, alana özgü bir grafik gösterimi olan Çok Değişkenli Bilgi Grafiği (ÇBKB)'yı öneriyoruz. ÇBKB'de düğümler zamana bağlı değişkenleri temsil ederken, yönlü kenarlar nedensel veya işlevsel bağımlılıkları yakalar ve zamansal modellemeyi bilgilendirmek için önsel bir grafik görevi görür.

Önerilen ÇBKB'yi oluşturmak için, öncelikle oluşturulan değişken düzeyindeki metinsel açıklamaları \mathcal{S} rafine ediyor, isteğe bağlı olarak bunları harici alana özgü bilgi T ile zenginleştirerek grafik yapısı için kapsamlı bir girdi oluşturuyoruz. Birleştirilmiş bilgi, çok aşamalı akıl yürütme ve dinamik bilgi entegrasyonundaki verimliliği nedeniyle seçilen LightRAG modeline (Guo vd. 2024; Liu vd. 2025) aktarılır.

$$\mathcal{G}_M = \text{LightRAG}(\mathcal{I}), \quad \mathcal{I} = \hat{\mathcal{S}} \oplus \mathcal{T}, \quad (1)$$

Burada, \oplus harici metinsel bilgi T ile isteğe bağlı birleştirilmeyi ifade eder. Birleştirilmiş girdi daha sonra, değişkenler arasındaki anlamlı ilişkileri yakalayan grafik yapılarını çıkarmak için bir BBM kullanan LightRAG'e beslenir.

Elde edilen ÇBKB, $\mathcal{G}_M = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, değişkenleri temsil eden bir düğüm kümesi \mathcal{V} , bir yönlendirilmiş kenarlar kümesi \mathcal{E} ve buna karşılık gelen bir ilişkiler kümesi \mathcal{R} 'den oluşur.

$$\mathcal{G}_M = \{(v_i, r, v_j) \mid v_i, v_j \in \mathcal{V}, r \in \mathcal{R}\}, \quad (2)$$

Her kenar $e \in \mathcal{E}$, $v_i, v_j \in \mathcal{V}$ ve $r \in \mathcal{R}$ olmak üzere, v_i 'yi v_j 'ye bağlayan yönlendirilmiş bir nedensel ilişkisi gösteren bir üçlü (v_i, r, v_j) şeklinde temsil edilir. ÇBKB inası, değişkenler arası anlamlı ilişkileri manuel açıklamaya gerek kalmadan otomatik olarak çıkarmak için BBM'leri kullanır.

Çift-Modalite Kodlayıcı

Nedensel İstek Kodlama Dalı. Verilen bir girdi başlık adı $v_k \in \mathcal{V}$ olduğunda, değişkeni özgü bir soru $Q(v_k)$ oluştururuz.

Following LightRAG’s hybrid retrieval strategy, relevant triplets are extracted from \mathcal{G}_M by leveraging structural patterns during LLM inference.

$$\mathcal{G}_{M_{v_k}} = \text{Retrieve}_{\text{global+local}}(\mathcal{Q}(v_k), \mathcal{G}_M), \quad (3)$$

where $\text{Retrieve}_{\text{global}}$ captures broader semantic or functional groupings, while $\text{Retrieve}_{\text{local}}$ targets direct relational dependencies of v_k . The augmented prompt incorporates inter-variable causal evidence and domain knowledge of the dataset:

$$p_{v_k} = \mathcal{Q}(v_k) \parallel (\bigcup_{\mathcal{G}_{M_{v_k}}} [v_i] \rightarrow r \rightarrow [v_j]). \quad (4)$$

Note that both the \mathcal{G}_M and the task-specific causal prompts $\mathcal{P} = \{p_1, p_2, \dots, p_N\} \in \mathbb{R}^{N \times l_{seq}}$ for the N variables are precomputed and stored in a database. This design significantly reduces redundant computation during training and inference.

Token2Vector. To incorporate causal prompts into the TimeMKG, causal prompts \mathcal{P} are first tokenized and truncated to a predefined maximum length l_{\max} , resulting in $\hat{\mathcal{P}} \in \mathbb{R}^{N \times l_{\max} \times D}$, where D is the embedding dimensions of the LLM. However, this high-dimensional representation is incompatible with the TimeMKG encoder, which requires a fixed-size vector per variable. To resolve this, the Token2Vector module maps each tokenized prompt into a compact vector via a two-layer perceptron.

$$\hat{\mathcal{P}} = \mathbf{W}_{D2d} \cdot \sigma(\mathbf{W}_{\text{pool}} \mathcal{P} + \beta_{\text{pool}}) + \beta_{D2d} \quad (5)$$

Here, $\mathbf{W}_{\text{pool}} \in \mathbb{R}^{l_{\max} \times 1}$ serves as a learnable pooling operator that aggregates token-level embeddings into a single vector of size D , and $\mathbf{W}_{D2d} \in \mathbb{R}^{D \times d}$ projects this vector into a lower-dimensional space \mathbb{R}^d . The final output is $\hat{\mathcal{P}} \in \mathbb{R}^{N \times d}$, where d is the target embedding dimension of TimeMKG.

Causal Prompt Encoder. The processed prompts are then encoded by the causal prompt encoder $\text{CP Encoder}(\cdot)$. We adopt the Pre-LN Transformer architecture to ensure training stability and efficient gradient flow. At the i_{th} layer, the causal embedding $\hat{\mathcal{P}}^i$ is normalized via RMSNorm:

$$\tilde{\mathcal{P}}^i = RN(\hat{\mathcal{P}}^i) = \gamma \odot \frac{\hat{\mathcal{P}}^i}{\sqrt{\frac{1}{D} \sum_{k=1}^D (\hat{\mathcal{P}}^i)_k^2}} \quad (6)$$

where $\tilde{\mathcal{P}}^i$ is the normalized embedding, γ is a learnable scaling parameter, and \odot denotes element-wise multiplication.

Subsequently, $\tilde{\mathcal{P}}^i$ passes through the Multi-Head Self-Attention layer $MHSA(\cdot)$ and is combined with $\hat{\mathcal{P}}^i$ via a residual connection.

$$\bar{\mathcal{P}}^i = MHSA(\tilde{\mathcal{P}}^i) + \hat{\mathcal{P}}^i \quad (7)$$

$$MHSA(\tilde{\mathcal{P}}^i) = \rho_o(Attention(\rho_q \tilde{\mathcal{P}}^i, \rho_k \tilde{\mathcal{P}}^i, \rho_v \tilde{\mathcal{P}}^i)) \quad (8)$$

where $\bar{\mathcal{P}}^i$ represents the output after the residual connection, and ρ_o , ρ_q , ρ_k , and ρ_v are the linear projection weight matrices. $MHSA(\cdot)$ models the causal relationships among variables and enables effective information aggregation.

After normalization $\vec{\mathcal{P}}^i = RN(\bar{\mathcal{P}}^i)$, the attention embedding $\vec{\mathcal{P}}^i$ is fed into the feed-forward network $FFN(\cdot)$.

$$\vec{\mathcal{P}}^i = FFN(\vec{\mathcal{P}}^i) + \bar{\mathcal{P}}^i \quad (9)$$

$$FFN(\vec{\mathcal{P}}^i) = \rho_2 \cdot \sigma(\rho_1 \vec{\mathcal{P}}^i + \beta_1) + \beta_2 \quad (10)$$

where $\vec{\mathcal{P}}^i \in \mathbb{R}^{N \times d}$, σ denotes the activation function, ρ_1 and ρ_2 serve as the learnable weight matrices, β_1 and β_2 act as the bias vectors. Finally, $\vec{\mathcal{P}}$ denotes the output of the $\text{CP Encoder}(\cdot)$ after multiple layers of feature extraction.

Time Series Encoding Branch. In MTSD, the numerical modality input $\mathbf{X}_{1:T} \in \mathbb{R}^{T \times N}$ spans both temporal (T) and variable (N) dimensions. Traditional methods focus on T in $MHSA(\cdot)$, generating a self-attention score matrix $\mathbf{S} \in \mathbb{R}^{T \times T}$, while projecting the variable dimension N into a shared latent space. This, however, emphasizes temporal patterns but overlooks inter-variable dependencies. To address this limitation, we adopt an inverted embedding strategy (Liu et al. 2023) that treats variables as tokens to explicitly captures their statistical correlations.

Inverted Embedding. To treat variables as tokens, we first transpose the input to obtain a variable-oriented representation: $\hat{\mathbf{X}} = \mathbf{X}_{1:T}^\top \in \mathbb{R}^{N \times T}$. Each variable i corresponds to a temporal sequence of length T , denoted as $\hat{\mathbf{X}}_i \in \mathbb{R}^T$. We embed each temporal sequence into a unified latent space using the following transformation:

$$\tilde{\mathbf{X}}_i = \mathbf{W}_{\text{emb}} \hat{\mathbf{X}}_i + \beta_{\text{emb}} \quad (11)$$

where \mathbf{W}_{emb} and β_{emb} are learnable projection weights and bias terms, respectively. $\tilde{\mathbf{X}} \in \mathbb{R}^{N \times d}$ denotes the embedded representation of all variables in the shared latent space.

Time Series Encoder. Similar to the $\text{CP Encoder}(\cdot)$, the Time Series Encoder $\text{TSEncoder}(\cdot)$ adopts a Pre-LN Transformer to capture pairwise similarities among variables.

$$\bar{\mathbf{X}}^i = MHSA(RN(\tilde{\mathbf{X}}^i)) + \tilde{\mathbf{X}}^i \quad (12)$$

$$\dot{\mathbf{X}}^i = FFN(RN(\bar{\mathbf{X}}^i)) + \bar{\mathbf{X}}^i \quad (13)$$

where the attention weight matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ captures the relative importance among variables. The output $\bar{\mathbf{X}} \in \mathbb{R}^{N \times d}$, obtained after $MHSA(\cdot)$ and $FFN(\cdot)$ with residual connections, effectively encodes the statistical correlations within the numerical modality.

Cross-Modality Attention

To integrate the numerical modality of MTSD with the textual modality embedded from the MKG, we employ cross-modality attention $CMA(\cdot)$ to align modalities at the variable level. In TimeMKG, the numerical embedding $\hat{\mathbf{X}} \in \mathbb{R}^{N \times d}$ is used as the *Query*, while the textual embedding $\vec{\mathcal{P}} \in \mathbb{R}^{N \times d}$ serves as the *Key* and *Value*.

$$\mathbf{S}_N = \text{softmax} \left(\frac{\mathbf{W}_q \vec{\mathcal{P}} \otimes \mathbf{W}_k \hat{\mathbf{X}}^\top}{\sqrt{d}} \right) \quad (14)$$

LightRAG’ın hibrit erişim stratejisini takiben, BBM çıkarımı sırasında yapışal örüntülerden faydalananlarak ilgili üçüler $G M$ ’den çıkarılır.

$$\mathcal{G}_{M_{v_k}} = \text{Retrieve}_{\text{global+local}}(\mathcal{Q}(v_k), \mathcal{G}_M), \quad (3)$$

burada $\text{Retrieve}_{\text{global}}$ daha geniş semantik veya işlevsel gruplandırımları yaralarken, $\text{Retrieve}_{\text{local}}$ ise v_k ’nin doğrudan ilişkisel bağımlılıklarını hedefler. Zenginleştirilmiş istem, değişkenler arası nedensel kanıtları ve veri setinin etki alanı bilgisini içerir:

$$p_{v_k} = \mathcal{Q}(v_k) \parallel (\bigcup_{\mathcal{G}_{M_{v_k}}} [v_i] \rightarrow r \rightarrow [v_j]). \quad (4)$$

Hem $G M$ ’nin hem de N değişkenleri için görevi özgü nedensel istemlerin $P = \{p_1, p_2, \dots, p_N\} \in \mathbb{R}^{N \times 1}$ seq önceden hesaplandığı ve bir veritabanında saklandığını unutmayın. Bu tasarım, eğitim ve çıkışım süreçlerinde gereksiz hesaplamayı önemli ölçüde azaltmaktadır.

Token’dan Vektöre Dönüşüm. Nedensel istemleri TimeMKG’ye dahil etmek amacıyla, nedensel istemler \mathcal{P} öncelikle tokenlere ayrılır ve önceden tanımlanmış bir maksimum uzunluğa l_{\max} kadar kısaltılır; bu işlem sonucunda $\mathcal{P} \in \mathbb{R}^{N \times l_{\max} \times D}$ elde edilir, burada D BBM’nin gömme boyutlarını temsil eder. Ancak, bu yüksek boyutlu temsil, her değişken için sabit boyutlu bir vektör gerektiren TimeMKG kodlayıcısıyla uyumsuzdur. Bu durumu çözmek için, Token2Vector modülü her tokenleştirilmiş istemi, iki katmanlı bir algılayıcı aracılıyla kompakt bir vektöre dönüştürür

$$\hat{\mathcal{P}} = \mathbf{W}_{D2d} \cdot \sigma(\mathbf{W}_{\text{pool}} \mathcal{P} + \beta_{\text{pool}}) + \beta_{D2d} \quad (5)$$

Burada, $\mathbf{W}_{\text{pool}} \in \mathbb{R}^{l_{\max} \times 1}$, token düzeyindeki gömmeleri D büyüğünde tek bir vektöre toplayan öğrenilebilir bir havuzlama operatörü işlevi görür; ve $\mathbf{W}_{D2d} \in \mathbb{R}^{D \times d}$ ise bu vektöre daha düşük boyutlu bir uzay olan \mathbb{R}^d ’ye yansır. Nihai çıktı $\hat{\mathcal{P}} \in \mathbb{R}^{N \times d}$ şeklidendir; burada d , TimeMKG’nin hedef gömme boyutudur.

Nedensel İstek Kodlayıcı. İşlenmiş istekler daha sonra nedensel istek kodlayıcı $\text{CP Encoder}(\cdot)$ tarafından kodlanır. Eğitim kararlılığını ve verimli gradyan akışını sağlamak için Pre-LN Transformer mimarisini benimsiyor. i ’inci katmanda, nedensel gömme \mathbf{P}^i , RMSNorm aracılığıyla normalize edilir:

$$\tilde{\mathcal{P}}^i = RN(\hat{\mathcal{P}}^i) = \gamma \odot \frac{\hat{\mathcal{P}}^i}{\sqrt{\frac{1}{D} \sum_{k=1}^D (\hat{\mathcal{P}}^i)_k^2}} \quad (6)$$

Burada $\tilde{\mathcal{P}}^i$ normalize edilmiş gömmedir, γ öğrenilebilir bir ölçeklendirme parametresidir ve \odot öğe tabanlı çarpımı belirtir.

Daha sonra, $\tilde{\mathcal{P}}^i$ Çok Başlı Öz-Dikkat katmanı $MHSA(\cdot)$ aracılığıyla geçer ve bir kalıntı bağlantı ile $\vec{\mathcal{P}}^i$ ile birleştirilir.

$$\bar{\mathcal{P}}^i = MHSA(\tilde{\mathcal{P}}^i) + \vec{\mathcal{P}}^i \quad (7)$$

$$MHSA(\tilde{\mathcal{P}}^i) = \rho_o(Attention(\rho_q \tilde{\mathcal{P}}^i, \rho_k \tilde{\mathcal{P}}^i, \rho_v \tilde{\mathcal{P}}^i)) \quad (8)$$

burada $\tilde{\mathcal{P}}^i$ artık bağlantı sonrası çıktıya temsil ederken, ρ_o , ρ_q , ρ_k ve ρ_v doğrusal izdüşüm ağırlık matrisleridir. $MHSA(\cdot)$ değişkenler arasındaki nedensel ilişkileri modeller ve etkin bilgi toplamasını sağlar.

Normalizasyondan sonra $\mathcal{P}^i = RN(\vec{\mathcal{P}}^i)$, dikkat gömülümesi \mathcal{P}^i , ileri beslemeli ağ $FFN(\cdot)$ içine beslenir.

$$\dot{\mathcal{P}}^i = FFN(\mathcal{P}^i) + \bar{\mathcal{P}}^i \quad (9)$$

$FFN(\mathcal{P}^i) = \rho_2 \cdot \sigma(\rho_1 \mathcal{P}^i + \beta_1) + \beta_2$ (10) burada $\mathcal{P}^i \in \mathbb{R}^{N \times d}$ olup, σ etkinleştirme fonksiyonunu gösterir; ρ_1 ve ρ_2 öğrenilebilir ağırlık matrisleri, β_1 ve β_2 ise bias vektörleri olarak işlev görür. Son olarak, $\dot{\mathcal{P}}^i$ birden çok özellik çıkarma katmanından sonra $\text{CP Encoder}(\cdot)$ ’ın çıktılarını belirtir.

Zaman Serisi Kodlama Dali. Çok Değişkenli Zaman Serisi Kodlayıcıda (MTSD), sayısal modalite girdisi $\mathbf{X}_{1:T} \in \mathbb{R}^{T \times N}$ hem zamansal (T) hem de değişken (N) boyutlarını kapsar. Geleneksel yöntemler, $MHSA(\cdot)$ ’ındaki T boyutuna odaklanarak $S \in \mathbb{R}^{T \times T}$ boyuttunda bir öz-dikkat skor matrisi üretirken, değişken boyutu N ortak bir gizli uzaya yansır. Ancak bu durum, zamansal örüntülerin vurgulamakla birlikte, değişkenler arası bağımlılıkları göz ardı etmektedir. Bu sınırlamayı ele almak amacıyla, değişkenlerin istatistiksel korelasyonlarını açıkça yakalamak için onları jeton olarak ele alan tersine çevrilmiş bir gömme stratejisi (Liu ve diğerleri 2023) benimsenmiştir.

Tersine Gömme. Değişkenleri jeton olarak işlemek için, öncelikle değişken odaklı bir temsil elde etmek üzere girdiyi dönüştürüruz: $\tilde{\mathbf{X}} = \mathbf{X}_{1:T}^\top \in \mathbb{R}^{N \times T}$. Her bir i değişkeni, uzunluğu T olan bir zamansal dizisi karşılık gelir ve $\tilde{\mathbf{X}}_i \in \mathbb{R}^T$ olarak gösterilir. Her zamansal dizisi, aşağıdaki dönüşümü kullanarak birleşik gizli uzaya gömüyoruz:

$$\tilde{\mathbf{X}}_i = \mathbf{W}_{\text{emb}} \tilde{\mathbf{X}}_i + \beta_{\text{emb}} \quad (11)$$

burada \mathbf{W}_{emb} ve β_{emb} sırasıyla öğrenilebilir projeksiyon ağırlıkları ve yanılılık terimleridir. $\tilde{\mathbf{X}}_i \in \mathbb{R}^{N \times d}$, tüm değişkenlerin paylaşılan gizli uzaydaki gömülü temsilini ifade eder.

Zaman Serisi Kodlayıcı. $\text{CP Encoder}(\cdot)$ ’a benzer şekilde, Zaman Serisi Kodlayıcı $\text{TSEncoder}(\cdot)$, değişkenler arasındaki ikili benzerlikleri yakalamak için bir Pre-LN Transformer kullanır.

$$\bar{\mathbf{X}}^i = MHSA(RN(\tilde{\mathbf{X}}^i)) + \tilde{\mathbf{X}}^i \quad (12)$$

$$\dot{\mathbf{X}}^i = FFN(RN(\bar{\mathbf{X}}^i)) + \bar{\mathbf{X}}^i \quad (13)$$

burada dikkat ağırlık matrisi $\mathbf{S} \in \mathbb{R}^{N \times N}$ değişkenler arasındaki göreceli önemini yakalar. $MHSA(\cdot)$ ve $FFN(\cdot)$ fonksiyonlarından sonra artık bağlantılarla elde edilen çıkış $X^i \in \mathbb{R}^{N \times d}$, sayısal modalite içindeki istatistiksel korelasyonları etkili bir şekilde kodlar.

Çapraz-Modalite Dikkat

Çok Değişkenli Zaman Serisi Verilerinin (MTSD) sayısal modalitesini, Çok Boyutlu Bilgi Grafiği’nden (ÇBKB) türetilen metinsel modalite ile entegre etmek amacıyla, değişken düzeyinde modaliteleri hizalamak için çapraz-modalite dikkat mekanizmasını $CMA(\cdot)$ kullanmaktayız. TimeMKG’de, sayısal gömme $\tilde{\mathbf{X}}_i \in \mathbb{R}^{N \times d}$, Sorgu olan; metinsel gömme $\mathcal{P}^i \in \mathbb{R}^{N \times d}$ ise Anahtar ve Değer görevi üstlenen olarak kullanılır.

$$\mathbf{S}_N = \text{softmax} \left(\frac{\mathbf{W}_q \dot{\mathcal{P}} \otimes \mathbf{W}_k \tilde{\mathbf{X}}_i^\top}{\sqrt{d}} \right) \quad (14)$$

Method	TimeMKG		TimeCMA		TimeLLM		UniTime		Time-MoE		PatchTST		iTransformer		TimesNet		DLinear		
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh1	96	0.373 <u>0.387</u>	0.373 <u>0.391</u>	0.398	0.410	0.397	0.418	0.453	0.463	0.389	0.412	0.379 0.400	0.379 0.400	0.396	0.411				
	192	0.426 <u>0.425</u>	0.427 <u>0.421</u>	0.451	0.440	0.434	0.439	0.505	0.537	0.429	0.432	0.449	0.441	0.427 0.432	0.445	0.440			
	336	0.457 <u>0.441</u>	0.458 <u>0.448</u>	0.473	0.451	0.468	0.457	0.543	0.591	0.478	0.464	0.492	0.465	0.478	0.464	0.487	0.465		
	720	0.468 <u>0.454</u>	0.449 <u>0.460</u>	0.469	0.470	0.469	0.477	0.621	0.722	0.522	0.506	0.513	0.510						
	avg	0.431 <u>0.427</u>	0.427 <u>0.430</u>	0.448	0.443	0.442	0.448	0.531	0.578	0.455	0.454	0.460	0.453	0.452	0.451	0.460	0.457		
ETTh2	96	0.290 <u>0.336</u>	0.286 <u>0.336</u>	0.295	0.345	0.508	0.593	0.321	0.364	0.364	0.395	0.329	0.370	0.387	0.434				
	192	0.362 <u>0.386</u>	0.363 <u>0.387</u>	0.386	0.399	0.374	0.394	0.685	1.031	0.406	0.413	0.450	0.442	0.393	0.409	0.516	0.504		
	336	0.386 <u>0.417</u>	0.406 <u>0.421</u>	0.419	0.429	0.415	0.427	0.882	1.670	0.444	0.446	0.442	0.445	0.471	0.468	0.636	0.568		
	720	0.418 <u>0.438</u>	0.417 <u>0.438</u>	0.425	0.442	0.425	0.444	1.144	2.673	0.447	0.457	0.445	0.457	0.510	0.493	0.885	0.684		
	avg	0.364 <u>0.394</u>	0.368 <u>0.396</u>	0.381	0.404	0.378	0.403	0.805	1.492	0.404	0.420	0.425	0.435	0.426	0.435	0.606	0.548		
ETTm1	96	0.303 <u>0.347</u>	0.312 <u>0.351</u>	0.359	0.381	0.322	0.363	0.564	0.730	0.361	0.377	0.347	0.376	0.359	0.392	0.364	0.385		
	192	0.362 <u>0.377</u>	0.361 <u>0.378</u>	0.383	0.393	0.366	0.387	0.652	0.863	0.399	0.395	0.382	0.397	0.413	0.419	0.396	0.402		
	336	0.391 <u>0.392</u>	0.392 <u>0.401</u>	0.416	0.414	0.398	0.407	0.698	0.929	0.431	0.416	0.442	0.429	0.438	0.435	0.426	0.423		
	720	0.462 <u>0.436</u>	0.453 <u>0.438</u>	0.483	0.449	0.454	0.440	0.795	1.095	0.488	0.448	0.486	0.456	0.505	0.466	0.483	0.458		
	avg	0.379 <u>0.388</u>	0.380 <u>0.392</u>	0.410	0.409	0.385	0.399	0.677	0.904	0.420	0.409	0.414	0.415	0.429	0.428	0.417	0.417		
ETTm2	96	0.175 <u>0.257</u>	0.173 <u>0.258</u>	0.193	0.280	0.183	0.266	0.468	0.480	0.203	0.286	0.185	0.267	0.208	0.309				
	192	0.235 <u>0.301</u>	0.238 <u>0.301</u>	0.257	0.318	0.251	0.310	0.655	0.913	0.265	0.323	0.250	0.306	0.300	0.377				
	336	0.296 <u>0.336</u>	0.297 <u>0.338</u>	0.317	0.353	0.319	0.351	0.886	1.680	0.312	0.348	0.326	0.360	0.318	0.348	0.398	0.440		
	720	0.386 <u>0.392</u>	0.393 <u>0.394</u>	0.419	0.411	0.420	0.410	1.194	2.894	0.411	0.403	0.424	0.412	0.425	0.409	0.565	0.531		
	avg	0.273 <u>0.322</u>	0.275 <u>0.323</u>	0.297	0.341	0.293	0.334	0.800	1.492	0.298	0.340	0.304	0.345	0.294	0.333	0.368	0.414		
Weather	96	0.162 <u>0.208</u>	0.167 <u>0.211</u>	0.198	0.235	0.171	0.214	0.281	0.243	0.200	0.238	0.211	0.252	0.200	0.263				
	192	0.210 <u>0.244</u>	0.212 <u>0.253</u>	0.240	0.269	0.217	0.254	0.361	0.338	0.245	0.275	0.255	0.285	0.251	0.270	0.239	0.300		
	336	0.260 <u>0.284</u>	0.270 <u>0.292</u>	0.295	0.308	0.274	0.293	0.436	0.441	0.296	0.310	0.306	0.320	0.293	0.309	0.284	0.335		
	720	0.327 <u>0.346</u>	0.350	0.348	0.368	0.353	0.351	0.343	0.541	0.610	0.369	0.355	0.377	0.365	0.367	0.356	0.347	0.384	
	avg	0.240 <u>0.271</u>	0.250 <u>0.276</u>	0.275	0.291	0.253	0.276	0.405	0.408	0.277	0.294	0.287	0.306	0.274	0.293	0.268	0.321		
ILI	24	1.985 <u>0.965</u>	1.996 0.998	2.383	1.004	2.460	0.954	3.650	1.245	2.344	0.926	2.284	0.926	2.569	0.957	4.541	1.605		
	36	1.902 <u>0.898</u>	1.906 0.915	2.390	0.993	1.998	0.912	4.320	1.369	2.410	0.954	2.062	0.932	2.046	0.899	4.221	1.508		
	48	1.860 <u>0.859</u>	1.867 <u>0.868</u>	2.394	1.003	1.979	0.912	4.737	1.461	2.250	0.928	2.229	0.965	2.473	1.461	2.250	0.928		
	60	1.916 <u>0.910</u>	1.920 <u>0.904</u>	2.562	1.049	2.109	0.938	5.197	1.547	2.175	0.940	2.149	0.952	2.124	0.919	4.169	1.460		
	avg	1.916 <u>0.908</u>	1.922 <u>0.921</u>	2.432	1.012	2.137	0.929	4.476	1.405	2.295	0.937	2.181	0.944	2.247	0.924	4.240	1.506		

1st/2nd | **38** 10 | **13** 32 | **0** 2 | **1** 3 | **0** 0 | **1** 0 | **1** 1 | **0** 3 | **0** 1

Linear layers \mathbf{W}_q , \mathbf{W}_k , and \mathbf{W}_v are applied along the variable dimension, followed by dot-product similarity to compute the cross-modality score $\mathbf{S}_N \in \mathbb{R}^{N \times N}$, which is scaled by $\frac{1}{\sqrt{d}}$ to stabilize the values.

Inference Module

We adopt a Transformer-based Cross Modality Decoder $CMD(\cdot)$ to decode the fused cross-modality causal embeddings learned by the dual-modality encoder. This decoder captures both inter-variable and intra-variable dependencies, as well as modality interactions, crucial for downstream in-

ference. A task-specific projection head maps the decoder output to the target space.

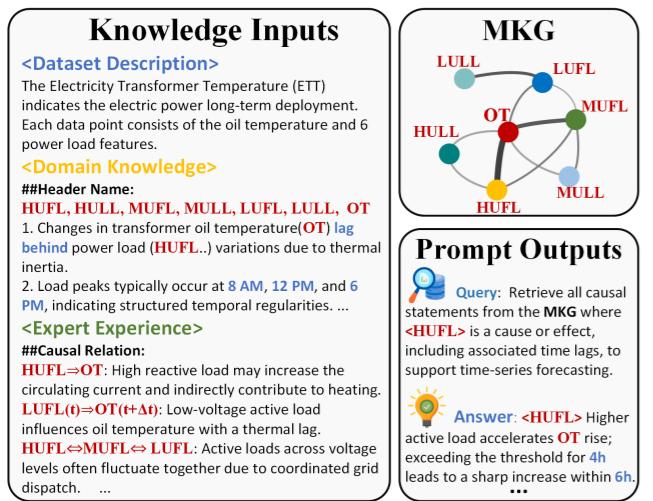


Figure 3: An Example of MKG Construction for the ETT Datasets. <> represents a variable-specific identifier, unique to each variable.

tion: Time-MoE (Shi et al. 2024); (4) **Transformer-based**: PatchTST (Nie et al. 2022), iTransformer (Liu et al. 2023); (5) **TCN-based**: TimesNet (Wu et al. 2022); (6) **Linear-based**: DLlinear (Zeng et al. 2023). We further compare SOTA models for specific tasks, such as TimeXer (Wang et al. 2024a) and FEDformer (Zhou et al. 2022) for short-term forecasting, XGBoost (Chen and Guestrin 2016) and LSTNet (Lai et al. 2018) for classification, and so on (listed in Figure 4) (Berndt and Clifford 1994; Hochreiter and Schmidhuber 1997; Zhang et al. 2022; Bai, Kolter, and Koltun 2018). In total, we compare TimeMKG against more than 20 baseline models across different categories.

Evaluation. We follow the evaluation protocol established in TimesNet (Wu et al. 2022). Long-term forecasting: MSE and MAE; Short-term forecasting: SMAPE, MASE and OWA; Classification: Accuracy. All tested methods use the same test batch size to ensure fairness. Each experiment was repeated at least three times using different seeds on a single NVIDIA A100 GPU.

MKG Example. Recent studies suggest that textual LLMs are relatively insensitive to numerical patterns in time series, rendering sample-specific prompt designs (e.g., TimeLLM (Jin et al. 2023a)) less effective for modeling real-time dynamics. To overcome this limitation, TimeMKG focuses on inferring causal and semantic relationships among variables and stores knowledge in an explicit MKG. Figure 3 illustrates the construction of MKG and prompt design for the ETT dataset. Specifically, *Dataset Description* provides contextual information about the application domain, *Domain Knowledge* defines the graph nodes (i.e., variable headers) along with relevant public knowledge retrieved by LLMs, and *Expert Experience*, optional and incorporated when available, encodes private domain-specific insights. Given a variable query, LLMs extract a subgraph from the MKG to generate prompt outputs. This structure enables the

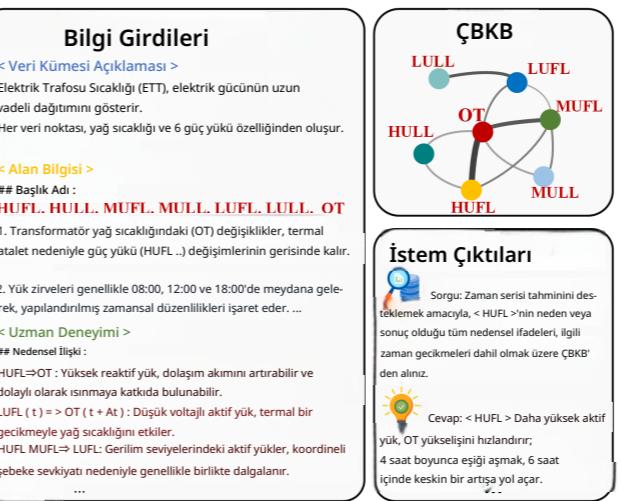
Method	TMKG	TXer	Patch*	iTrans*	FED*	TsNet	DLinear	
ICL	SMAPE	38.79	43.77	42.59	44.04	71.26	47.81	72.24
	MASE	0.353	0.422	0.391	0.411	0.725	0.461	0.790
	OWA	0.456	0.527	0.502	0.505	0.880	0.577	0.921
IoTFlow	SMAPE	84.78	85.15	84.91	84.82	90.92	84.85	88.22
	MASE	1.053	1.056	1.056	1.055	1.181	1.058	1.316
	OWA	1.542	1.548	1.545	1.543	1.675	1.545	1.695
Nasdq	SMAPE	22.51	23.55	22.99	23.06	26.44	25.32	27.86
	MASE	3.136	3.349	3.214	3.221	3.986	3.763	4.503
	OWA	1.116	1.184	1.143	1.146	1.380	1.307	1.509
Internet	SMAPE	62.24	62.88	63.25	62.37	64.04	62.78	66.21
	MASE	0.838	0.843	0.855	0.854	0.860	0.845	0.898
	OWA	1.017	1.027	1.036	1.028	1.046	1.027	1.086
Battery	SMAPE	3.898	4.712	3.907	4.611	12.15	4.156	19.44
	MASE	7.186	8.922	7.232	8.739	22.14	7.752	32.77
	OWA	0.958	1.197	0.959	1.167	3.167	1.039	5.075

Table 2: Short-term forecasting task. The historical sequence length T is set as 24. All the results are averaged from 4 different prediction lengths $L \in \{12, 18, 24, 36\}$. Lower metres values indicate more accurate predict. “T” indicates the name of Time. “*” in Transformers indicates the name of former.

integration of both public and confidential domain knowledge, making TimeMKG well-suited for MTSD tasks with clearly defined header names.

Long and Short-term Forecasting Tasks. As shown in Table 1, (1) TimeMKG consistently outperforms all baselines across datasets, achieving the best performance in 38/48 in Long-term sub-tasks and all short-term average tasks. (2) LLMs-based methods outperform conventional deep learning models, highlighting the benefits of semantic-aware multimodal modeling for MTSD tasks. (3) LLMs-powered methods (e.g., TimeMKG) surpass LLMs-backboned ones (e.g., TimeLLM and UniTime), with average improvements of 15% and 7.3% in MSE, respectively, indicating that the pretrained LLMs are better suited for textual modalities than serving as backbones. (4) By integrating MKG and causal knowledge, TimeMKG achieves superior performance on datasets with clearer variable semantics (e.g., Weather and ICL vs. ETT and IoTFlow), with an average improvement of 4% and 8.9% in SOTA metrics compared to the second-best methods, respectively.

Classification Tasks. In Figure 4, in MTSD classification datasets with clearly defined header variables, TimeMKG achieves the best average predictive accuracy of 71.0%, outperforming classical methods such as XGBoost (69.7%) and deep learning models like TimesNet (68.2%). Notably, TimeMKG demonstrates a substantial improvement over PatchTST (65.2%), which performs poorly in classification tasks due to its channel-independent design that neglects inter-variable dependencies. In contrast, TimeMKG explicitly models both causal and statistical relationships among variables, equipping the model with a deeper understanding of complex multivariate entanglement in MTSD.



Şekil 3: ETT Veri Setleri için ÇBKB Oluşumuna Bir Örnek. <> her değişkene özgü, benzersiz bir tanımlayıcıyı temsil eder.

Yöntem	TMKG	TXer	Patch*	iTrans*	FED*	TsNet	DLinear	
ICL	SMAPE	38.79	43.77	42.59	44.04	71.26	47.81	72.24
	MASE	0.353	0.422	0.391	0.411	0.725	0.461	0.790
	OWA	0.456	0.527	0.502	0.505	0.880	0.577	0.921
IoTFlow	SMAPE	84.78	85.15	84.91	84.82	90.92	84.85	88.22
	MASE	1.053	1.056	1.055	1.181	1.058	1.316	
	OWA	1.542	1.548	1.543	1.675	1.545	1.695	
Nasdq	SMAPE	22.51	23.55	22.99	23.06	26.44	25.32	27.86
	MASE	3.136	3.349	3.214	3.221	3.986	3.763	4.503
	OWA	1.116	1.184	1.143	1.146	1.380	1.307	1.509
Internet	SMAPE	62.24	62.88	63.25	62.37	64.04	62.78	66.21
	MASE	0.838	0.843	0.855	0.854	0.860	0.845	0.898
	OWA	1.017	1.027	1.036	1.028	1.046	1.027	1.086
Batarya	SMAPE	3.898	4.712	3.907	4.611	12.15	4.156	19.44
	MASE	7.186	8.922	7.232	8.739	22.14	7.752	32.77
	OWA	0.958	1.197	0.959	1.167	3.167	1.039	5.075

Tablo 2: Kısa vadeli tahmin görevi. Geçmiş dizi uzunluğu $T=24$ olarak ayarlanmıştır. Tüm sonuçlar, 4 farklı tahmin uzunluğunun $L \in \{12, 18, 24, 36\}$ ortalamasından elde edilmiştir. Daha düşük ölçüt değerleri, daha doğru tahminleri gösterir. “T” Zamanın adını belirtir. Transformers'daki “*” ise öne eklenen adını belirtir.

hem kamusal hem de gizli alan bilgisinin entegrasyonunu sağlayarak, TimeMKG'yi açıkça tanımlanmış başlık adlarına sahip MTSD görevleri için oldukça uygun hale getirir.

Uzun ve Kısa Vadeli Tahmin Görevleri. Tablo 1'de gösterildiği gibi, (1) TimeMKG, tüm veri kümelerindeki temel modellerden tutarlı bir şekilde daha iyi performans göstermiş; uzun vadeli alt görevlerde 48'de 38'inde ve tüm kısa vadeli ortalama görevlerde en iyi performansı elde etmiştir. (2) BBM'lere dayalı yöntemler, geleneksel derin öğrenme modellerinden daha iyi performans göstererek, Çok Değişkenli Zaman Serisi Keşfi (MTSD) görevleri için semantik farkındalıkla sahip çok modlu modellemenin faydalalarını vurgulamaktadır. (3) BBM deseteki yöntemler (örn. TimeMKG), BBM omurgalı olanları (örn. TimeLLM ve UniTime) geride bırakarak, sırasıyla %15 ve %7.1'lik ortalama MSE iyileştirmeleri sağlamış; bu da önceden eğitilmiş BBM'lerin omurga olarak hizmet etmekten ziyade metinsel modaliteler için daha uygun olduğunu göstermektedir. (4) ÇBKB ve nedensel bilgiyi entegre ederek, TimeMKG, daha net değişken semantигine sahip veri kümelerinde (örn. Hava Durumu ve ICL vs. ETT ve IoTFlow), ikinci en iyi yöntemlere kıyasla sırasıyla SOTA metriklerinde %4 ve %8.9'lık ortalama bir iyileşme ile üstün performans sergilemektedir.

Sınıflandırma Görevleri. Şekil 4'te, açıkça tanımlanmış başlık değişkenlerine sahip MTSD sınıflandırma veri kümelerinde TimeMKG, %71.0'lık en iyi ortalama tahmini doğruluğa ulaşarak, XGBoost (%69.7) gibi klasik yöntemleri ve TimesNet (%68.2) gibi derin öğrenme modellerini geride bırakmaktadır. Özellikle TimeMKG, değişkenler arası bağımlılıkları göz ardı eden kanal bağımsız tasarımını nedeniyle sınıflandırma görevlerinde kötü performans gösteren PatchTST'ye (%65.2) kıyasla önemli bir iyileşme sergilemektedir. Buna karşılık, TimeMKG, değişkenler arasındaki hem nedensel hem de istatistiksel ilişkileri açıkça modelleyerek, modeli ÇZSZ'deki karmaşık çok değişkenli etkileşimi daha derinlemesine anlamasıyla donatır.

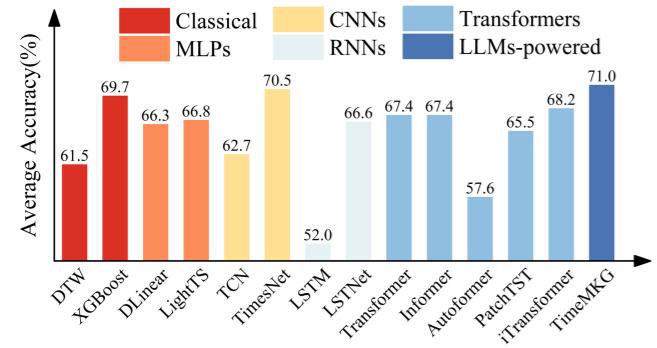


Figure 4: Classification Task. The results are the average accuracy.

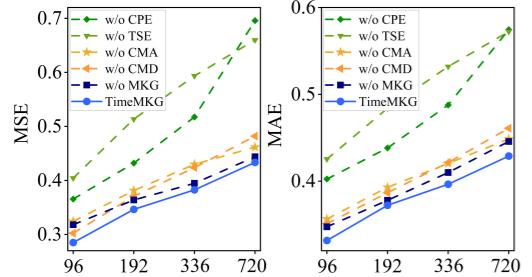
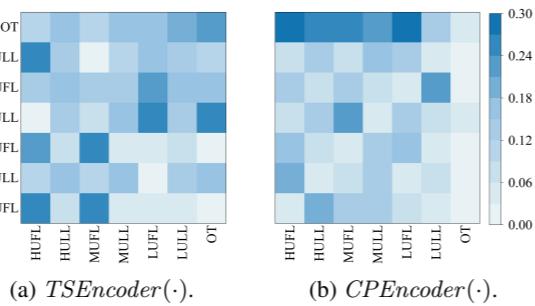


Figure 5: Ablation study of model structure. The historical sequence length is fixed at $T = 96$, and results are averaged over four subtasks in the long-term forecasting setting with prediction lengths $L \in \{96, 192, 336, 720\}$.

Ablation Study. Figure 5 shows the ablation results on the ETT dataset for the long-term forecasting task. The reported scores are averaged over different prediction lengths and masking ratios. The variant w/o MKG component queries the LLMs using knowledge as prompts, thereby bypassing the graph construction process. Results indicate that removing explicit knowledge retrieval significantly impairs the model’s ability to capture causal relationships. The removal of encoder variants (w/o CPE and w/o TSE) leads to the most severe performance decline. This is because these variants directly feed the embeddings into the CMA module, resulting in the loss of learnable parameters essential for representation learning. In the variant w/o CMA, the outputs of the dual-branch encoders are simply concatenated. In contrast, the variant w/o CMD removes the decoder from the prediction module. The performance drop caused by removing the decoder is less severe than that from eliminating the cross-modal fusion structure, underscoring the crucial role of dual-modality interaction.

Attention Visualization. We visualize the attention scores of different encoders in Figures 6(a) and 6(b). Although both the $TSEncoder(\cdot)$ and the $CP Encoder(\cdot)$ adopt the same Pre-LN Transformer architecture, they capture complementary inter-variable dependencies. Specifically, $TSEncoder(\cdot)$ learns sample-specific dynamic correlations, reflecting statistical relationships within the data, while



(a) $TSEncoder(\cdot)$. (b) $CP Encoder(\cdot)$.

Figure 6: Visualization of the attention score matrix S .

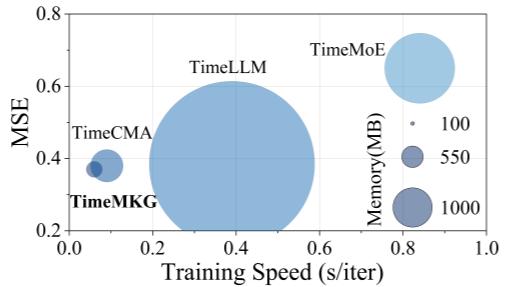


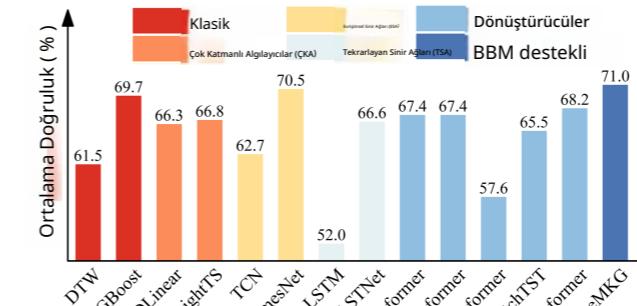
Figure 7: Efficiency analysis of LLM-based methods.

$CP Encoder(\cdot)$ captures domain-informed static dependencies, which represent prior causal knowledge. Notably, the attention score matrix of $CP Encoder(\cdot)$ closely resembles the MKG-based causal structure shown in Figure 3, highlighting strong associations between load variables and the target variable. This consistency suggests that $CP Encoder(\cdot)$ effectively incorporates and utilizes prior knowledge embedded in the MKG.

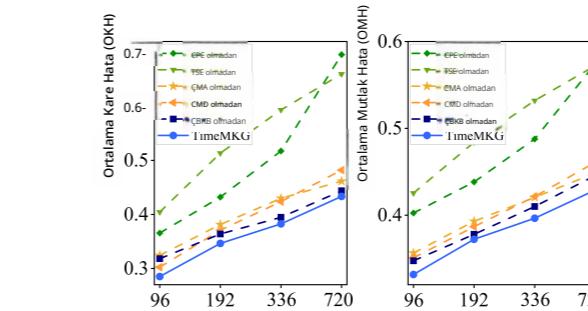
Efficiency analysis. Figure 7 presents an efficiency analysis of LLM-based methods. To ensure a fair comparison, all models are evaluated using the same batch size. Time-MoE follows its original pretraining speed and model scale settings. The results demonstrate that TimeMKG, built upon an LLMs-powered architecture, achieves significant advantages in both model size and training speed compared to LLM-backboned TimeLLM and the foundational LLM model Time-MoE. Compared to TimeCMA, TimeMKG adopts a strategy of pre-storing causal prompts, enabling faster loading of semantic representations and avoiding repeated LLM queries, which greatly reduces both runtime and memory consumption.

Conclusion

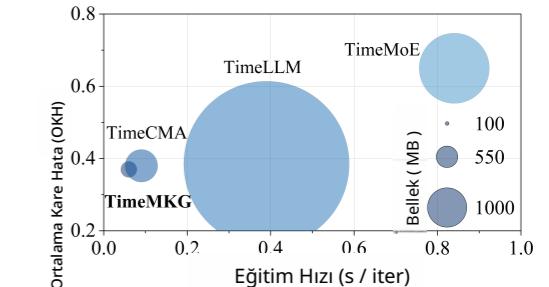
This paper proposes TimeMKG, an LLMs-powered framework for multimodal modeling of MTSD. TimeMKG leverages LLMs to extract domain-specific knowledge from textual modalities and construct a knowledge graph encoding causal relationships. A dual-modality encoder framework models both the causal representations from prompts and the statistical features from numerical time series. Cross-modality attention enables variable-level alignment and fu-



Şekil 4: Sınıflandırma Görevi. Sonuçlar ortalama doğruluktur.



Şekil 6: Dikkat skoru matrisi S ’nin görselleştirilmesi.



Şekil 7: BBM tabanlı yöntemlerin verimlilik analizi.

Şekil 5: Model yapısının ablasyon çalışması. Geçmiş dizi uzunluğu $T = 96$ olarak sabitlenmiş olup, sonuçlar tahmin uzunlukları $L \in \{96, 192, 336, 720\}$ olan uzun vadeli tahmin ayarında dört alt görev üzerinden ortalaması alınmıştır.

Ablasyon Çalışması. Şekil 5, uzun vadeli tahmin görevi için ETT veri kümesi üzerindeki ablasyon sonuçlarını göstermektedir. Bildirilen skorlar, farklı tahmin uzunlukları ve maskeleme oranları üzerinden ortalaması alınmıştır. $\setminus o$ MKG bileşenine sahip w varyantı, bilgiyi istemler olarak kullanarak BBM’leri sorgular ve böylece grafik oluşturma sürecini atlar. Sonuçlar, açık bilgi erişiminin kaldırılmasının, modelin nedensel ilişkileri yakalama yeteneğini önemli ölçüde bozduğunu göstermektedir. Kodlayıcı varyantlarının ($w \setminus o$ CPE ve $w \setminus o$ TSE) kaldırılması en ciddi performans düşüşüne yol açmaktadır. Bunun nedeni, bu varyantların gömülüleri doğrudan CMA modülüne beslemesi ve gösterim öğrenimi için esas olan öğrenilebilir parametrelerin kaybına yol açmasıdır. $\setminus o$ CMA varyantında, çift dallı kodlayıcıların çıktıları basitçe birleştirilmiştir. Buna karşılık, $\setminus o$ CMD varyantı tahmin modülünden kod çözücüyü kaldırmaktadır. Kod çözücüünün çıkarılmasından kaynaklanan performans düşüşü, çapraz-modal füzyon yapısının ortadan kaldırılmasından kaynaklanandan daha az şiddetlidir; bu durum, çift-modal etkileşimin kritik rolünü vurgulamaktadır.

Dikkat Görselleştirmesi. Şekil 6(a) ve 6(b)’de farklı kodlayıcıların dikkat skorlarını görselleştirmektedir. Hem $TSEncoder(\cdot)$ hem de $CP Encoder(\cdot)$ ayrı Pre-LN Transformer mimarisini benimsenmiş olsalar da, tamamlayıcı değişkenler arası bağımlılıkları yakalarlar. Özellikle, $TSEncoder(\cdot)$ örneğe özgü dinamik korelasyonları öğrenir ve veriler içindeki istatistiksel ilişkileri yansıtırken,

$CP Encoder(\cdot)$ alana özgü statik bağımlılıkları yakalar; bu bağımlılıklar, önceden var olan nedensel bilgiyi temsil eder. Özellikle, $CP Encoder(\cdot)$ ’nin dikkat skoru matrisi, Şekil 3’te gösterilen ÇBKB tabanlı nedensel yapıya oldukça benzemekte olup, yük değişkenleri ile hedef değişken arasındaki güçlü ilişkileri vurgulamaktadır. Bu tutarlılık, $CP Encoder(\cdot)$ ’nın ÇBKB’ye gömülü ön bilgiyi etkin bir şekilde dâhil ettiğini ve kullandığını göstermektedir.

Verimlilik analizi. Şekil 7, BBM tabanlı yöntemlerin verimlilik analizini sunmaktadır. Adil bir karşılaştırma sağlamak amacıyla tüm modeller aynı yığın boyutu kullanılarak değerlendirilmiştir. TimeMoE, orijinal ön eğitim hızı ve model ölçüği ayarlarını takip etmektedir. Sonuçlar, BBM’ler destekli TimeLLM ve temel BBM modeli Time-MoE’ye kıyasla hem model boyutu hem de eğitim hızında önemli avantajlar elde ettiğini göstermektedir. TimeCMA ile karşılaştırıldığında, TimeMKG, nedensel istemleri önceden depolama stratejisini benimseyerek, semantik gösterimlerin daha hızlı yüklenmesini sağlamaktır ve tekrarlanan BBM sorgularından kaçınarak hem çalışma süresini hem de bellek tüketimini büyük ölçüde azaltmaktadır.

Sonuç

Bu makale, çok değişkenli zaman serisi verilerinin çok modlu modellenmesi için BBM destekli bir çerçeveye TimeMKG’yi önermektedir. TimeMKG, metinsel modalitelerden alana özgü bilgiyi çıkarmak ve nedensel ilişkileri kodlayan bir bilgi grafiği oluşturmak üzere BBM’lerden yararlanır. Çift-modalite kodlayıcı çerçevesi, hem istemlerden gelen nedensel temsilleri hem de sayısal zaman serilerinden elde edilen istatistiksel özelliklerini modeller. Çapraz-modalite dikkat mekanizması, değişken düzeyinde hızlama ve füzyon imkanı sunar.

sion between statistical time-series patterns and encoded domain knowledge for downstream tasks. Extensive experiments demonstrate the effectiveness of incorporating inter-variable causal embeddings in improving model reasoning.

References

- Bai, S.; Kolter, J. Z.; and Koltun, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- Berndt, D. J.; and Clifford, J. 1994. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd international conference on knowledge discovery and data mining*, 359–370.
- Chen, T.; and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, 785–794. New York, NY, USA: Association for Computing Machinery. ISBN 9781450342322.
- Das, A.; Kong, W.; Leach, A.; Mathur, S.; Sen, R.; and Yu, R. 2023. Long-term forecasting with tide: Time-series dense encoder. *arXiv preprint arXiv:2304.08424*.
- Guo, Z.; Xia, L.; Yu, Y.; Ao, T.; and Huang, C. 2024. Light-trag: Simple and fast retrieval-augmented generation. *arXiv preprint arXiv:2410.05779*.
- Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.
- Huang, Y.-H.; Xu, C.; Wu, Y.; Li, W.-J.; and Bian, J. 2025. TimeDP: Learning to Generate Multi-Domain Time Series with Domain Prompts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(17): 17520–17527.
- Huguet Cabot, P.-L.; and Navigli, R. 2021. REBEL: Relation Extraction By End-to-end Language generation. In Moens, M.-F.; Huang, X.; Specia, L.; and Yih, S. W.-t., eds., *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2370–2381. Punta Cana, Dominican Republic: Association for Computational Linguistics.
- Jin, M.; Wang, S.; Ma, L.; Chu, Z.; Zhang, J. Y.; Shi, X.; Chen, P.-Y.; Liang, Y.; Li, Y.-F.; Pan, S.; et al. 2023a. Time-ilm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*.
- Jin, Z.; Chen, Y.; Leeb, F.; Gresele, L.; Kamal, O.; Lyu, Z.; Blin, K.; Gonzalez Adauto, F.; Kleiman-Weiner, M.; Sachan, M.; et al. 2023b. Cladder: Assessing causal reasoning in language models. *Advances in Neural Information Processing Systems*, 36: 31038–31065.
- Kim, Y.; Kang, E.; Kim, J.; and Huang, H. H. 2024. Causal reasoning in large language models: A knowledge graph approach. *arXiv preprint arXiv:2410.11588*.
- Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR ’18, 95–104. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356572.
- Li, X.; Dong, Y.; Chang, L.; Chen, L.; Wang, G.; Zhuang, Y.; and Yan, X. 2023. Dynamic hybrid modeling of fuel ethanol fermentation process by integrating biomass concentration XGBoost model and kinetic parameter artificial neural network model into mechanism model. *Renewable Energy*, 205: 574–582.
- Liu, C.; Xu, Q.; Miao, H.; Yang, S.; Zhang, L.; Long, C.; Li, Z.; and Zhao, R. 2024a. Timecma: Towards llm-empowered time series forecasting via cross-modality alignment. *arXiv preprint arXiv:2406.01638*.
- Liu, J.; Meng, S.; Gao, Y.; Mao, S.; Cai, P.; Yan, G.; Chen, Y.; Bian, Z.; Shi, B.; and Wang, D. 2025. Aligning vision to language: Text-free multimodal knowledge graph construction for enhanced llms reasoning. *arXiv preprint arXiv:2503.12972*.
- Liu, X.; Hu, J.; Li, Y.; Diao, S.; Liang, Y.; Hooi, B.; and Zimmermann, R. 2024b. Unitime: A language-empowered unified model for cross-domain time series forecasting. In *Proceedings of the ACM Web Conference 2024*, 4095–4106.
- Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2023. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.
- Mameche, S.; Cornanguer, L.; Ninad, U.; and Vreeken, J. 2025. SPACETIME: Causal Discovery from Non-Stationary Time Series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(18): 19405–19413.
- Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.
- Oreshkin, B. N.; Carpol, D.; Chapados, N.; and Bengio, Y. 2019. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.
- Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; and Cottrell, G. 2017. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*.
- Salinas, D.; Flunkert, V.; Gasthaus, J.; and Januschowski, T. 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, 36(3): 1181–1191.
- Seong, E.; Lee, H.; and Chae, D.-K. 2024. Self-supervised framework based on subject-wise clustering for human subject time series data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 22341–22349.
- Sezer, O. B.; Gudelek, M. U.; and Ozbayoglu, A. M. 2020. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, 90: 106181.
- Shi, X.; Wang, S.; Nie, Y.; Li, D.; Ye, Z.; Wen, Q.; and Jin, M. 2024. Time-moe: Billion-scale time series foundation models with mixture of experts. *arXiv preprint arXiv:2409.16040*.
- ististiksel zaman serisi örüntüleri ile aşağı akış görevleri için kodlanmış alan bilgisi arasındaki ilişki. Kapsamlı deneyler, değişkenler arası nedensel gömülerin model akıl yürütme etkinliğini göstermektedir.
- Referanslar**
- Bai, S.; Kolter, J. Z.; ve Koltun, V. 2018. Sira modellemesi için genel evrişimsel ve tekrarlayan ağların ampirik bir değerlendirmesi. *arXiv ön baskısı arXiv:1803.01271*.
- Berndt, D. J.; ve Clifford, J. 1994. *Zaman serilerinde örüntü bulmak için dinamik zaman bükme kullanımı. Bilgi keşfi ve veri madenciliği üzerine 3. uluslararası konferansın bildirilerinde*, 359–370.
- Chen, T.; ve Guestrin, C. 2016. *XGBoost: Ölçeklenebilir Bir Ağczę Arırma Sistemi. Bilgi Keşfi ve Veri Madenciliği Üzerine 22. ACM SIGKDD Uluslararası Konferansı Bildirileri içinde*, KDD ’16, 785–794. New York, NY, ABD: *Hesaplama Makineleri Birliği*. ISBN 9781450342322.
- Das, A.; Kong, W.; Leach, A.; Mathur, S.; Sen, R.; ve Yu, R. 2023. Tide ile uzun vadeli tahmin: Zaman serisi yoğun kodlayıcı. *arXiv ön baskısı arXiv:2304.08424*.
- Guo, Z.; Xia, L.; Yu, Y.; Ao, T.; ve Huang, C. 2024. Light-trag: Basit ve hızlı erişim destekli üretim. *arXiv ön baskısı arXiv:2410.05779*.
- Hochreiter, S.; ve Schmidhuber, J. 1997. Uzun kısa süreli bellek. *Neural computation*, 9(8): 1735–1780.
- Huang, Y.-H.; Xu, C.; Wu, Y.; Li, W.-J.; ve Bian, J. 2025. TimeDP: Alan İstekleriyle Çok Alanlı Zaman Serileri Üretmeyi Öğrenmek. *AAAI Yapay Zeka Konferansı Bildirileri*, 39(17): 17520–17527.
- Huguet Cabot, P.-L.; ve Navigli, R. 2021. *REBEL: Uçtan Uca Dil Üretimi ile İlişki Çıkarması*. Moens, M.-F.; Huang, X.; Specia, L.; ve Yih, S. W.-t., editörler, *Hesaplama Dilbilim Derneği Bulguları: EMNLP 2021*, 2370–2381. Punta Cana, Dominican Republic: Hesaplama Dilbilim Derneği.
- Jin, M.; Wang, S.; Ma, L.; Chu, Z.; Zhang, J. Y.; Shi, X.; Chen, P.-Y.; Liang, Y.; Li, Y.-F.; Pan, S.; vd. 2023a. Time-LLM : Büyük dil modellerini yeniden programlayarak zaman serisi tahmini. *arXiv ön baskısı arXiv:2310.01728*.
- Jin, Z.; Chen, Y.; Leeb, F.; Gresele, L.; Kamal, O.; Lyu, Z.; Blin, K.; Gonzalez Adauto, F.; Kleiman-Weiner, M.; Sachan, M.; vd. 2023b. Cladder: Dil modellerinde nedensel akıl yürütmenin değerlendirilmesi. *Advances in Neural Information Processing Systems*, 36: 31038–31065.
- Kim, Y.; Kang, E.; Kim, J.; ve Huang, H. H. 2024. Büyük dil modellerinde nedensel akıl yürütme: Bir bilgi grafiği yaklaşımı. *arXiv ön baskısı arXiv:2410.11588*.
- Lai, G.; Chang, W.-C.; Yang, Y.; ve Liu, H. 2018. Derin Sinir Ağları ile Uzun ve Kısa Vadeli Zamansal Desenlerin Modellemesi. *İçinde Bilgi Erişimi Araştırma ve Geliştirme Üzerine 41. Uluslararası ACM SIGIR Konferansı*, SIGIR ’18, 95–104. New York, NY, ABD: *Hesaplama Makineleri Birliği*. ISBN 9781450356572.
- Li, X.; Dong, Y.; Chang, L.; Chen, L.; Wang, G.; Zhuang, Y.; ve Yan, X. 2023. Biyokütle konsantrasyonu XGBoost modeli ve kinetik parametre yapay sinir ağı modelinin mekanizma modeline entegrasyonu ile yakıt etanol fermantasyon sürecinin dinamik hibrit modelenmesi. *Yenilenebilir Enerji*, 205: 574–582.
- Liu, C.; Xu, Q.; Miao, H.; Yang, S.; Zhang, L.; Long, C.; Li, Z.; ve Zhao, R. 2024a. Timecma: Çapraz-modalite hizalaması yoluyla BBM destekli zaman serisi tahmini. *arXiv ön baskısı arXiv:2406.01638*.
- Liu, J.; Meng, S.; Gao, Y.; Mao, S.; Cai, P.; Yan, G.; Chen, Y.; Bian, Z.; Shi, B.; ve Wang, D. 2025. Görüşü diley hizalama: Gelişmiş BBM akıl yürütmesi için metinsiz çok modlu bilgi grafiği oluşturma. *arXiv ön baskısı arXiv:2503.12972*.
- Liu, X.; Hu, J.; Li, Y.; Diao, S.; Liang, Y.; Hooi, B.; ve Zimmermann, R. 2024b. Unitime: Alanlar arası zaman serisi tahmini için dil destekli birleşik bir model. *ACM Web Konferansı 2024 Bildiriler Kitabı*’nda, 4095–4106.
- Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; ve Long, M. 2023. itransformer: Ters çevrilmiş transformatörler zaman serisi tahmini için etkilidir. *arXiv ön baskısı arXiv:2310.06625*.
- Mameche, S.; Cornanguer, L.; Ninad, U.; ve Vreeken, J. 2025. SPACETIME: Durağan Olmayan Zaman Serilerinden Nedensel Keşif. *AAAI Yapay Zeka Konferansı Bildirileri*, 39(18): 19405–19413.
- Nie, Y.; Nguyen, N. H.; Sinthong, P.; ve Kalagnanam, J. 2022. Bir zaman serisi 64 kelime değerlendirildi: Transformatörlerle uzun vadeli tahmin. *arXiv ön baskısı arXiv:2211.14730*.
- Oreshkin, B. N.; Carpol, D.; Chapados, N.; ve Bengio, Y. 2019. N-BEATS: Yorumlanabilir zaman serisi tahmini için sınırsız taban genişletme analizi. *arXiv ön baskısı arXiv:1905.10437*.
- Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; ve Cottrell, G. 2017. Zaman serisi tahmini için çift aşamalı dikkat tabanlı bir tekrarlayan sinir ağı. *arXiv ön baskısı arXiv:1704.02971*.
- Salinas, D.; Flunkert, V.; Gasthaus, J.; ve Januschowski, T. 2020. DeepAR: Otoregresif tekrarlayan ağlarla olasılıksal tahmin. *International Journal of Forecasting*, 36(3): 1181–1191.
- Seong, E.; Lee, H.; ve Chae, D.-K. 2024. İnsan denek zaman serisi verileri için denek bazlı kümelemeye dayalı, kendi kendini denetleyen bir çerçeve. In *AAAI Yapay Zeka Konferansı Bildirileri*, cilt 38, 22341–22349.
- Sezer, O. B.; Güdelek, M. U.; ve Özbayraklı, A. M. 2020. Derin öğrenme ile finansal zaman serisi tahmini: 2005–2019 yılları için sistematik bir literatür taraması. *Applied Soft Computing*, 90: 106181.
- Shi, X.; Wang, S.; Nie, Y.; Li, D.; Ye, Z.; Wen, Q.; ve Jin, M. 2024. Time-moe: Uzman karışımı ile milyar ölçekli zaman serisi temel modelleri. *arXiv ön baskısı arXiv:2409.16040*.

- Susanti, Y.; and Färber, M. 2024. Knowledge graph structure as prompt: improving small language models capabilities for knowledge-based causal discovery. In *International Semantic Web Conference*, 87–106. Springer.
- Tan, M.; Merrill, M.; Gupta, V.; Althoff, T.; and Hartvigsen, T. 2024. Are language models actually useful for time series forecasting? *Advances in Neural Information Processing Systems*, 37: 60162–60191.
- Wang, C.; Qi, Q.; Wang, J.; Sun, H.; Zhuang, Z.; Wu, J.; Zhang, L.; and Liao, J. 2025. ChatTime: A Unified Multimodal Time Series Foundation Model Bridging Numerical and Textual Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(12): 12694–12702.
- Wang, P.-F.; Zhu, Q.-X.; and He, Y.-L. 2024. Novel Multiscale Trend Decomposition LSTM Based on Feature Selection for Industrial Soft Sensing. *IEEE Transactions on Industrial Informatics*, 20(12): 14249–14256.
- Wang, Y.; Wu, H.; Dong, J.; Liu, Y.; Qiu, Y.; Zhang, H.; Wang, J.; and Long, M. 2024a. TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables. *ArXiv*, abs/2402.19072.
- Wang, Y.; Xu, Y.; Yang, J.; Wu, M.; Li, X.; Xie, L.; and Chen, Z. 2024b. Fully-connected spatial-temporal graph for multivariate time-series data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 15715–15724.
- Woo, G.; Liu, C.; Kumar, A.; Xiong, C.; Savarese, S.; and Sahoo, D. 2024. Unified Training of Universal Time Series Forecasting Transformers. *ArXiv*, abs/2402.02592.
- Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.
- Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; et al. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.
- Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11121–11128.
- Zhang, T.; Zhang, Y.; Cao, W.; Bian, J.; Yi, X.; Zheng, S.; and Li, J. 2022. Less Is More: Fast Multivariate Time Series Forecasting with Light Sampling-oriented MLP Structures. *ArXiv*, abs/2207.01186.
- Zhao, Z.; Wang, P.; Wen, H.; Wang, S.; Yu, L.; and Wang, Y. 2025. STEM-LTS: Integrating Semantic-Temporal Dynamics in LLM-driven Time Series Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(21): 22858–22866.
- Zheng, C.; Fan, X.; Wang, C.; and Qi, J. 2020. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 1234–1241.
- Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.
- Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, 27268–27286. PMLR.
- Susanti, Y.; ve Färber, M. 2024. Bilgi grafiği yapısı bilgi istemi olarak: Bilgi tabanlı nedensel keşif için küçük dil modellerinin yeteneklerini geliştirme. In *Uluslararası Anımsal Web Konferansı*, 87–106. Springer.
- Tan, M.; Merrill, M.; Gupta, V.; Althoff, T.; ve Hartvigsen, T. 2024. Dil modelleri zaman serisi tahmini için gerçekten faydalı mı? *Advances in Neural Information Processing Systems*, 37: 60162–60191.
- Wang, C.; Qi, Q.; Wang, J.; Sun, H.; Zhuang, Z.; Wu, J.; Zhang, L.; ve Liao, J. 2025. ChatTime: Sayısal ve Metinsel Verileri Birleştiren Birleşik Çok Modlu Zaman Serisi Temel Modeli. *AAAI Yapay Zeka Konferansı Bildirileri*, 39(12): 12694–12702.
- Wang, P.-F.; Zhu, Q.-X.; ve He, Y.-L. 2024. Endüstriyel Yumuşak Algılama İçin Özelliğin Seçimine Dayalı Yeni Çok Ölçekli Eğilim Ayırıştırma LSTM. *IEEE Transactions on Industrial Informatics*, 20(12): 14249–14256.
- Wang, Y.; Wu, H.; Dong, J.; Liu, Y.; Qiu, Y.; Zhang, H.; Wang, J.; ve Long, M. 2024a. TimeXer: Dışsal Değişkenlerle Zaman Serisi Tahmini İçin Transformatörleri Güçlendirmek. *ArXiv*, abs/2402.19072.
- Wang, Y.; Xu, Y.; Yang, J.; Wu, M.; Li, X.; Xie, L.; ve Chen, Z. 2024b. Çok değişkenli zaman serisi verileri için tam bağlı uzamsal-zamansal çizge. In *AAAI Yapay Zeka Konferansı Bildirileri*, cilt 38, 15715–15724.
- Woo, G.; Liu, C.; Kumar, A.; Xiong, C.; Savarese, S.; ve Sahoo, D. 2024. Evrensel Zaman Serisi Tahmin Transformatörlerinin Birleşik Eğitimi. *ArXiv*, abs/2402.02592.
- Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; ve Long, M. 2022. Timesnet: Genel zaman serisi analizi için zamansal 2D-varyasyon modellemesi. *arXiv önbaskısı arXiv:2210.02186*.
- Yang, A.; Li, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Gao, C.; Huang, C.; Lv, C.; ve diğerleri. 2025. Qwen3 teknik raporu. *arXiv önbaskısı arXiv:2505.09388*.
- Zeng, A.; Chen, M.; Zhang, L.; ve Xu, Q. 2023. Transformatörler zaman serisi tahmini için etkili midir? In *AAAI Yapay Zeka Konferansı Bildirileri*, cilt 37, 11121–11128.
- Zhang, T.; Zhang, Y.; Cao, W.; Bian, J.; Yi, X.; Zheng, S.; ve Li, J. 2022. Az Çoktur: Hafif Örneklemme Odaklı ÇKGA Yapıları ile Hızlı Çok Değişkenli Zaman Serisi Tahmini. *ArXiv*, abs/2207.01186.
- Zhao, Z.; Wang, P.; Wen, H.; Wang, S.; Yu, L.; ve Wang, Y. 2025. STEM-LTS: BBM Odaklı Zaman Serisi Analizinde Semantik-Zamansal Dinamikleri Entegre Etme. *AAAI Yapay Zeka Konferansı Bildirileri*, 39(21): 22858–22866.
- Zheng, C.; Fan, X.; Wang, C.; ve Qi, J. 2020. Gman: Trafik tahmini için bir çizge çoklu dikkat ağı. İçinde *AAAI yapay zeka konferansı bildirileri*, cilt 34, 1234–1241.
- Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; ve Zhang, W. 2021. Informer: Uzun Dizili Zaman Serisi Tahmini için Verimli Transformatörün Ötesinde. In *AAAI Yapay Zeka Konferansı Bildirileri*, cilt 35, 11106–11115.
- Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; ve Jin, R. 2022. Fedformer: Uzun Vadeli Seriler için Frekansla Zenginleştirilmiş Ayırıştırılmış Transformatör Tahmini. In *Uluslararası Makine Öğrenimi Konferansı*, 27268–27286. PMLR.