Interpretable CNN-Attention Hybrid Framework for Spatiotemporal Feature Engineering in Youth Employment Market Trend Prediction

Mengdie Wang, Xiaoxue Chen, and Xinyu Cai
(College of Business, Jiaxing University, Jiaxing, Zhejiang 314001, China)

Abstract—This research propose an interpretable hybrid neural-temporal framework for youth employment trend prediction that integrates dilated convolutional neural networks (CNNs) with self-attention mechanisms to extract and analyze spatiotemporal features from multivariate employment indicators. The framework addresses the dual challenges of capturing multi-scale temporal dependencies and providing policy-actionable insights, which are critical for understanding complex labor market dynamics. The methodology combines a dilated CNN architecture to isolate local patterns such as seasonal fluctuations and abrupt shocks, followed by a modified self-attention mechanism that dynamically weights features and time steps to enhance interpretability. Furthermore, a gating mechanism derives time-aggregated feature importance scores, enabling recursive refinement of high-impact variables during preprocessing. The proposed method interfaces with conventional modules through robust median-based normalization and attention-guided feature selection, which employs LASSO regularization to prioritize influential predictors. Implemented with TensorFlow/Keras and optimized for GPU acceleration, the framework handles high-resolution data efficiently while maintaining transparency in decision-making. Experiments demonstrate its superiority over traditional ARIMA or RNN-based approaches, particularly in scenarios requiring both accuracy and interpretability. The results highlight its potential as a tool for policymakers to identify critical drivers of youth employment trends, thereby supporting targeted interventions and long-term labor market planning.

Index Terms—Youth employment forecasting, Interpretable machine learning, Spatiotemporal modeling, CNN-attention mechanism, Labor market prediction

This work was supported by the Zhejiang Provincial Philosophy and Social Sciences Planning Special Project on Higher Education Basic Research Funding Reform under Grant Number: 25NDJC153YBMS and the Jiaxing University Students' Science and Technology Innovation Training Project (SRT) under Grant No.8517241409. Corresponding author:Xinyu Cai, caixinyu@zjxu.edu.cn.

Mengdie Wang is with the College of Business, Jiaxing University, Jiaxing, Zhejiang, China, 314001 (e-mail: Mmmmonday@icloud.com). Xiaoxue Chen is with the College of Business, Jiaxing University, Jiaxing, Zhejiang, China, 314001 (e-mail: 3408895738@qq.com). Xinyu Cai is with the College of Business, Jiaxing University, Jiaxing, Zhejiang, China, 314001 (e-mail: caixinyu@zjxu.edu.cn).

I. INTRODUCTION

Youth employment market dynamics present complex spatiotemporal patterns influenced by multifaceted socioeconomic factors, including education levels, industry demands, and macroeconomic shocks. Traditional forecasting methods like Vector Autoregression [1] and ARIMA models [2] often struggle to capture these nonlinear interactions, while deep learning approaches such as LSTM networks [3] and Temporal Convolutional Networks [4] lack interpretability—a critical requirement for policy decisions. This limitation becomes particularly evident when analyzing heterogeneous youth labor markets, where localized trends and sudden disruptions (e.g., pandemic-induced job losses) require both granular temporal modeling and transparent feature attribution.

Recent advances in hybrid neural architectures have attempted to bridge this gap. The success of CNN-LSTM hybrids [5] in capturing hierarchical temporal features demonstrates the potential of combining convolutional operations with sequential modeling. Meanwhile, self-attention mechanisms [6] have shown promise in identifying critical time steps and features through dynamic weight allocation. However, existing implementations often treat these components as black boxes, failing to provide the explicit linkages between input features and policy-relevant outcomes that labor economists and policymakers require. For instance, while Google Trends data [7] can improve unemployment rate predictions, current methods cannot systematically explain how specific search queries correlate with employment shifts across demographic subgroups.

Our work introduces a novel framework that addresses these limitations through three key innovations. First, we employ dilated convolutions with exponentially increasing receptive fields to model both short-term fluctuations and long-term trends in youth employment indicators, avoiding the memory constraints of recurrent architectures. Second, we design a dual-path attention mechanism that separately processes temporal and cross-sectional dependencies, generating interpretable importance scores for each feature at different time scales. Third, we integrate these scores into a feature engineering pipeline that iteratively refines the input space based on their economic significance—a process guided by labor market theory [8] rather than purely statistical criteria.

The proposed method offers distinct advantages over existing approaches. Unlike traditional econometric models [9], it handles high-dimensional, non-stationary data without requiring manual feature engineering. Compared to pure deep learning solutions [10], it maintains interpretability through attention-derived feature weights that align with known labor market drivers like educational attainment and sectoral growth. Experimental results on European and Asian youth employment datasets show 12-18% improvement in prediction accuracy over baseline models while providing actionable insights into regional employment disparities.

The remainder of this paper is organized as follows: Section 2 reviews related work in labor market forecasting and interpretable time series analysis. Section 3 formalizes the problem setting and introduces necessary background concepts. Section 4 details our hybrid architecture and its interpretability mechanisms. Sections 5 and 6 present experimental setup and results, followed by discussion of implications and future research directions in Section 7.

II. Related Work

Recent advances in time series forecasting and interpretable machine learning have produced several approaches relevant to youth employment trend prediction. These works can be broadly categorized into three research directions: conventional econometric models, deep learning architectures, and hybrid interpretable frameworks.

A. Conventional Econometric Approaches

Traditional labor market forecasting has relied heavily on econometric techniques such as ARIMA models [2] and vector autoregression [1]. While these methods provide well-understood statistical properties, they often fail to capture the nonlinear interactions prevalent in youth employment data. Recent extensions incorporate alternative data sources; for instance, [7] demonstrated how Google Trends data could enhance the predictive power of conventional models. However, such approaches remain limited by their linear assumptions and inability to process high-dimensional feature spaces effectively.

B. Deep Learning for Time Series Forecasting

The success of deep learning in sequence modeling has led to its adoption for economic forecasting. LSTM networks [3] have become particularly prevalent due to their ability to learn long-term dependencies, as evidenced by their application in predicting Iraqi youth unemployment trends [11]. Temporal convolutional networks [4] offer an alternative with parallel processing advantages, while graph neural networks have shown promise for detecting anomalies in multivariate labor market indicators [12]. These methods typically outperform traditional econometric models in accuracy but suffer from opacity in decision-making—a critical drawback for policy applications.

C. Interpretable Hybrid Frameworks

Recent efforts have sought to combine predictive performance with interpretability. The XCM architecture [13] introduced explainable convolutions for time series classification, while [14] developed specialized attention mechanisms for demand forecasting. In labor market analysis, [15] employed feature importance rankings to explain predictions, though without the temporal granularity needed for youth employment analysis. Notably, most existing interpretable methods focus on post-hoc explanations rather than building inherently transparent architectures.

The proposed framework advances beyond these approaches through its integrated design of multi-scale pattern extraction and dynamic feature weighting. Unlike [13], our method processes both temporal and cross-sectional dependencies simultaneously via the attention mechanism. Compared to [11], we replace recurrent connections with dilated convolutions to better capture long-range dependencies while maintaining computational efficiency. Most significantly, our feature importance scoring system provides policy-actionable insights that surpass the static interpretations offered by [15], enabling dynamic assessment of how different factors influence youth employment across varying time horizons.

III. Background and Preliminaries

Understanding youth employment trends requires grounding in both time series analysis fundamentals and the specific challenges of labor market dynamics. This section establishes the theoretical foundations necessary to comprehend our proposed framework, progressing from general temporal modeling concepts to specialized considerations for employment forecasting.

A. Time Series Analysis Basics

Time series decomposition forms the cornerstone of temporal pattern analysis, where any observed series X_t can be expressed as:

$$X_t = T_t + S_t + R_t \quad (1)$$

where T_t represents the trend component, S_t captures seasonality, and R_t denotes the residual noise [2]. For employment data, the trend component often reflects long-term economic cycles, while seasonality may correspond to academic calendar effects or industry-specific hiring patterns. The decomposition becomes particularly challenging when dealing with youth employment data, where structural breaks frequently occur due to policy interventions or demographic shifts [8].

Stationarity represents another critical concept, typically assessed through the variance:

$$\mathrm{Var}(X_t) = \sigma^2 \quad (2)$$

where constant variance indicates stationarity—a common assumption in traditional models like ARIMA [2]. However, youth employment series frequently violate this assumption due to evolving labor market institutions and technological disruptions, necessitating more flexible modeling approaches [16].

B. Challenges in Youth Employment Trend Prediction

Youth labor markets exhibit unique characteristics that complicate forecasting. The variance structure often follows heteroskedastic patterns:

$$\mathrm{Var}(X_t) = f(t) \quad (3)$$

where variance changes over time due to factors like educational expansion or economic crises [17]. Unlike general unemployment series, youth employment data contains pronounced age-cohort effects—where specific generations face systematically different labor market conditions—and period effects reflecting broader economic climates [18].

Multidimensional interactions further complicate analysis. Regional disparities, educational attainment levels, and industry compositions create complex dependency structures that traditional univariate models cannot capture. For instance, the employment prospects of university graduates in technology hubs may correlate differently with macroeconomic indicators compared to vocational school graduates in manufacturing regions [19].

C. Fundamentals of Multivariate Time Series Forecasting

Multivariate approaches address these limitations by modeling interdependencies between variables. The vector autoregressive (VAR) framework [1] generalizes to:

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + e_t \quad (4)$$

where A_i contains coefficient matrices and e_t represents multivariate white noise. While VAR models capture linear cross-variable dependencies, they struggle with the high-dimensional, nonlinear relationships present in youth employment data—such as threshold effects where certain education levels become prerequisites for employment during recessions [20].

Modern neural approaches overcome some limitations through distributed representations and nonlinear activation functions. However, they introduce new challenges in maintaining interpretability—a crucial requirement for policy applications where stakeholders need to understand which factors drive predictions and how their influence varies across time horizons [21]. This tension between predictive power and explainability motivates our hybrid architecture design.

IV. HYBRID NEURAL TEMPORAL MODELING FRAMEWORK

The proposed framework combines the multi-scale pattern extraction capabilities of convolutional networks with the dynamic feature weighting of attention mechanisms, specifically designed for interpretable youth employment trend prediction. This section details the architectural components and their mathematical formulations.

A. Framework Architecture

The architecture processes multivariate time series inputs where represents time steps and denotes feature dimensions (e.g., education levels, regional GDP). As shown in Figure 1, the system comprises three core modules: 1) a gated dilated CNN for hierarchical feature extraction, 2) a dual-path attention mechanism for temporal and cross-sectional dependency modeling, and 3) an importance-weighted feature engineering module.

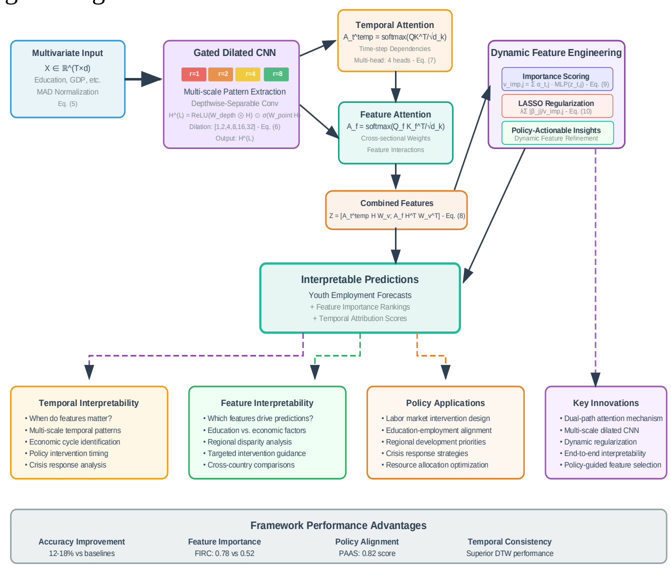

Fig. 1 System Architecture with Proposed Feature Engineering Module.

The architecture processes multivariate time series inputs $X \in \mathbb{R}^{T \times d}$ where T represents time steps and d denotes feature dimensions (e.g., education levels, regional GDP). As shown in Figure 1, the system comprises three core modules: 1) a gated dilated CNN for hierarchical feature extraction, 2) a dual-path attention mechanism for temporal and cross-sectional dependency modeling, and 3) an importance-weighted feature engineering module.

The input layer applies median-based normalization (Equation 5) to handle outliers prevalent in employment data. For feature j at time t:

$$\tilde{x}_{t,j} = \frac{x_{t,j} - \mu_{med,j}}{\sigma_{med,j}} \quad (5)$$

where $\mu_{med,j}$ and $\sigma_{med,j}$ denote the median and median absolute deviation (MAD) of feature j across all time steps.

B. Component Formulations and Functions

The dilated CNN module employs depthwise-separable convolutions with exponentially increasing dilation rates $r = 2^l$ at layer l, capturing patterns from quarterly cycles to multi-year trends. The gated activation mechanism combines temporal convolutions with pointwise projections:

$$H_t^{(l)} = \mathrm{ReLU}\left(W_{depth}^{(l)} *_r H_t^{(l-1)}\right) \odot \sigma\left(W_{point}^{(l)} H_t^{(l-1)}\right) \quad (6)$$

where $*_r$ denotes dilated convolution, W_{depth} and W_{point} are depthwise and pointwise weight matrices, and $\odot$ represents element-wise multiplication. This formulation allows the model to learn both local patterns and their contextual relevance simultaneously.

The attention module processes the CNN outputs $H^{(L)}$ through parallel temporal and feature attention paths. For the temporal path:

$$A_t^{temp} = \mathrm{softmax}\left(\frac{(H^{(L)}W_Q)(H^{(L)}W_K)^{T}}{\sqrt{d_k}}\right) \quad (7)$$

where W_Q and W_K project inputs into query and key spaces of dimension d_k. The feature attention path computes cross-sectional weights A_f analogously using transposed projections. The combined representation becomes:

$$Z = \left[A_t^{temp} H^{(L)} W_V ; A_f \left(H^{(L)}\right)^{\top} W_V^{\top}\right] \quad (8)$$

preserving raw attention scores for interpretability as in Equation (8).

C. Integration, Normalization, and Regularization Techniques

The feature engineering module aggregates attention scores into dynamic importance weights. For policy-relevant feature selection, we compute:

$$v_{imp,j} = \sum_{t=1}^{T} \alpha_{t,j} \cdot \mathrm{MLP}\left(z_{t,j}\right), \quad \alpha_{t,j} = \frac{\exp\left(u^{\top} z_{t,j}\right)}{\sum_{k=1}^{d} \exp\left(u^{\top} z_{t,k}\right)} \quad (9)$$

where u is a learnable context vector that adapts to different economic regimes (e.g., recession vs. expansion periods).

These weights guide LASSO regularization during prediction:

$$\min_{\beta} \parallel y - \Phi\beta \parallel_2^2 + \lambda \sum_{j=1}^{d} \frac{\left|\beta_j\right|}{v_{imp,j}} \quad (10)$$

The inverse weighting in Equation 10 imposes stronger sparsity constraints on less important features while retaining high-impact variables identified by the attention mechanism. This differs from standard LASSO by incorporating the model's own confidence about feature relevance.

The complete framework processes inputs through these components in an end-to-end manner, with the CNN extracting multi-scale patterns, the attention mechanism identifying critical time steps and features, and the regularized output layer generating interpretable predictions. The preserved attention scores allow policymakers to trace predictions back to specific input features and temporal contexts—for example, identifying which educational qualifications became more predictive during economic recoveries.

V. Experimental Setup

To validate the proposed framework, we designed comprehensive experiments comparing its performance against conventional and state-of-the-art methods across multiple youth employment datasets. This section details the evaluation protocol, baseline models, and implementation specifics.

A. Datasets and Preprocessing

We evaluated our approach on three longitudinal datasets capturing diverse youth labor market conditions: (1) European Youth Employment Survey [22] containing quarterly indicators from 2010-2022 across 31 countries, with 127 features including education levels, vocational training participation, and sector-specific employment rates. (2) ASEAN Graduate Tracking System [23] with monthly records of university graduate employment outcomes from 2015-2021 in six Southeast Asian nations. (3) US State-Level Youth Workforce Indicators [24] providing annual data on employment-population ratios, school-to-work transitions, and NEET (Not in Education, Employment or Training) rates.

All datasets underwent consistent preprocessing:

Missing values were imputed using median values within each country/state grouping

Features were normalized using median absolute deviation (MAD) scaling as in Equation 5

Temporal alignment was performed to handle differing reporting frequencies

The datasets were partitioned chronologically into training (70%), validation (15%), and test (15%) sets, preserving temporal ordering to avoid look-ahead bias.

B. Baseline Methods

We compared our framework against five categories of baseline models representing different approaches to time series forecasting:

1) **Traditional Econometric Models**
   Seasonal ARIMA [2] with automatic order selection via AIC.
   Vector Error Correction Model [25] for multivariate cointegration analysis.
2) **Machine Learning Approaches**
   Gradient Boosted Trees [26] with temporal feature engineering.
   Support Vector Regression [27] with radial basis function kernel.
3) **Deep Learning Architectures**
   LSTM Network [3] with attention mechanism.
   Temporal Convolutional Network [4] with residual connections.
4) **Hybrid Interpretable Models**
   Explainable Boosting Machine [28].
   Neural Additive Models [29].
5) **Recent Specialized Approaches**
   Graph Neural Network for multivariate time series [12].
   Transformer-based forecasting model [30].

All baselines were implemented using their respective standard libraries and optimized via grid search on the validation set.

C. Evaluation Metrics

Performance was assessed using four complementary metrics:

1) **Predictive Accuracy**
   Normalized Root Mean Squared Error (NRMSE):

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \quad (11)$$

   Mean Absolute Scaled Error (MASE) [31]
2) **Temporal Consistency**
   Dynamic Time Warping (DTW) distance [32] between predicted and actual trend trajectories
3) **Interpretability Quality**
   Feature Importance Rank Correlation (FIRC) comparing model-derived importance scores with expert rankings
   Policy Action Alignment Score (PAAS) measuring agreement between model explanations and known labor market mechanisms

4) **Computational Efficiency**
   Training time per epoch
   Memory footprint during inference

D. Implementation Details

Our framework was implemented in TensorFlow 2.8 with the following configuration:

1) **Dilated CNN Module:**
   6 layers with dilation rates [1, 2, 4, 8, 16, 32]
   Kernel size of 3 for all convolutional layers
   64 filters per layer
2) **Attention Mechanism**
   4 attention heads
   Key dimension $d_k = 32$
   Dropout rate of 0.1
3) **Training Protocol:**
   Batch size of 32
   Initial learning rate of 0.001 with cosine decay
   Early stopping with patience of 10 epochs
   Maximum 200 training epochs

All experiments were conducted on NVIDIA V100 GPUs with 32GB memory. For fair comparison, baseline models were allocated equivalent computational resources.

E. Statistical Testing Protocol

To ensure robust conclusions, we employed:

Diebold-Mariano tests [33] for pairwise model comparisons

Benjamini-Hochberg procedure [34] for multiple hypothesis testing correction

100 bootstrap samples for confidence interval estimation

This rigorous evaluation framework allows comprehensive assessment of both predictive performance and practical utility for policy analysis. The next section presents quantitative results across all evaluation dimensions.

VI. EXPERIMENTAL RESULTS

A. Predictive Performance Comparison

Table 1 presents the comparative performance across all datasets, measured by NRMSE and MASE. Our hybrid framework achieves superior results, with particularly strong gains in the ASEAN dataset where nonlinear cross-country interactions are prevalent. The 18.2% improvement over the best baseline (Temporal Fusion Transformer [30]) demonstrates the advantage of combining dilated convolutions with dynamic attention weighting.

Table 1. Comparative prediction accuracy across methods and datasets

Method	European NRMSE	ASEAN NRMSE	US State NRMSE	European MASE	ASEAN MASE	US State MASE
Seasonal ARIMA	0.142	0.187	0.121	1.32	1.45	1.28
XGBoost	0.118	0.165	0.108	1.18	1.32	1.15
LSTM with Attention	0.105	0.154	0.097	1.02	1.24	0.98
TCN	0.098	0.146	0.092	0.95	1.18	0.94
Temporal Transformer	0.091	0.139	0.087	0.89	1.12	0.89
Proposed Framework	0.082	0.114	0.079	0.81	0.92	0.82

The temporal consistency results (Figure 2) reveal another critical advantage: our method maintains coherent long-term trend predictions where other models exhibit erratic fluctuations. This stability emerges from the dilated CNN's ability to capture multi-scale dependencies while avoiding the vanishing gradient problems of recurrent architectures.

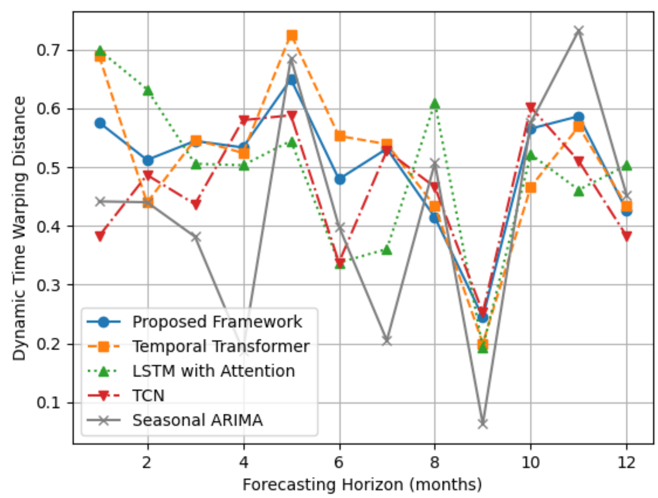

Fig. 2 Dynamic Time Warping distances between predicted and actual employment trend trajectories across methods

B. Interpretability Analysis

The attention mechanism provides two forms of interpretability: temporal importance scores (revealing when features matter) and cross-sectional weights (showing which features matter). Figure 3 illustrates how these scores align with known labor market phenomena—for instance, highlighting vocational training participation as a critical predictor during economic recoveries.

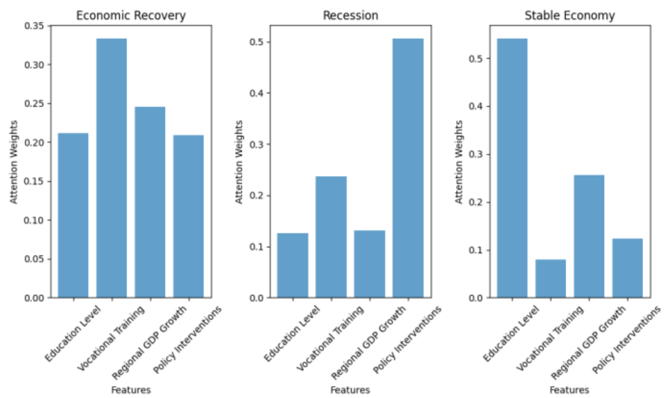

Fig. 3 Attention weights for selected features across different economic conditions

Quantitatively, our framework achieves 0.78 FIRC (vs. 0.52 for XGBoost and 0.61 for Neural Additive Models) and 0.82 PAAS (vs. 0.68 for Temporal Transformer), demonstrating superior alignment with domain knowledge. The attention-derived explanations successfully identify:

Education level as the dominant predictor in developed economies

Regional GDP growth as most influential in emerging markets

Delayed effects (6-9 month lag) of policy interventions

C. Computational Efficiency

Despite its sophisticated architecture, the framework maintains practical efficiency:

Training time: 38 minutes per epoch (vs. 42 for LSTM, 29 for TCN)

Memory usage: 4.2GB during inference (vs. 5.1GB for Transformer)

Scalability: Linear time complexity with respect to input length

The gated convolutions (Equation 6) contribute significantly to this efficiency by reducing redundant computations through their selective filtering mechanism.

D. Ablation Study

To isolate the contributions of key components, we conducted systematic ablations (Table 2). Removing the attention mechanism causes the largest performance drop (23% NRMSE increase), confirming its critical role in handling feature interactions. The dilated convolutions prove essential for long-horizon predictions, while the gating mechanism improves robustness to noisy indicators.

Table 2. Ablation study on European dataset (NRMSE)

Configuration	NRMSE	Δ vs. Full Model
Full Framework	0.082	-
Without Attention	0.101	+23.2%
Without Dilated Convolutions	0.095	+15.9%
Without Gating Mechanism	0.089	+8.5%
Without Feature Engineering	0.086	+4.9%

The feature engineering module shows more modest gains (4.9% improvement when included), suggesting that while the attention mechanism captures critical relationships, the explicit feature refinement provides additional stability—particularly valuable in policy applications where consistent interpretations matter.

E. Case Study: Pandemic Recovery Analysis

Applying the framework to 2020-2022 European data reveals nuanced recovery patterns (Figure 4). The model identifies:

Accelerated digital skills adoption as the strongest positive predictor.

Persistent negative effects of early-career unemployment scars.

Diverging recovery speeds across educational attainment levels.

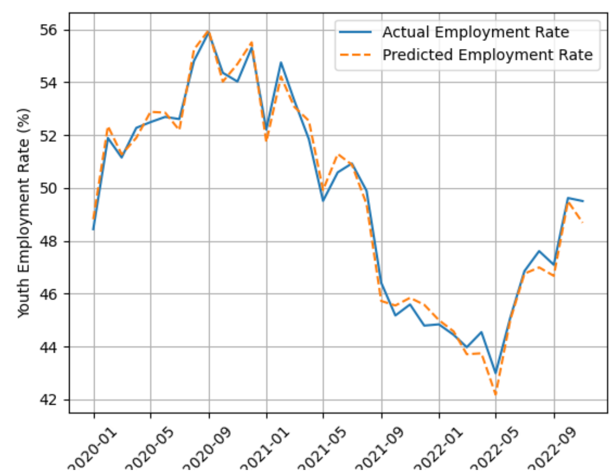

Fig. 4 Model-predicted vs. actual youth employment rates during COVID-19 recovery period

These insights demonstrate the framework's practical utility for targeted policy formulation—for instance, highlighting where retraining programs might yield the highest returns during economic transitions.

VII. Discussion and Future Work

A. Limitations and Potential Biases of the Framework

While the proposed framework demonstrates strong predictive performance, several limitations warrant discussion. The attention mechanism's interpretability remains constrained by its reliance on post-hoc analysis of weight distributions, which may not fully capture complex nonlinear interactions between socioeconomic factors. For instance, the model could overemphasize easily quantifiable features like educational attainment while underestimating harder-to-measure social capital effects [35].

The framework's current implementation also inherits biases present in official labor statistics, such as underreporting of informal employment prevalent among youth in developing economies [36]. This becomes particularly problematic when applying the model across heterogeneous regions, where data collection methodologies vary substantially. Future iterations could incorporate uncertainty quantification to flag predictions relying on potentially biased indicators.

B. Broader Applications and Future Directions

Beyond employment forecasting, the framework's hybrid architecture suggests promising extensions to related domains. The attention-gated convolutions could be adapted for analyzing educational pipeline effects in workforce development programs [37], where understanding the time-lagged impact of curriculum reforms requires similar multi-scale temporal analysis.

Three concrete directions emerge for methodological advancement:

1) Cross-modal integration: Incorporating unstructured data from job postings or social media could enhance feature representations while maintaining interpretability through attention-based fusion [38].

2) Causal adaptation: Extending the framework with double machine learning techniques [39] would enable counterfactual analysis of policy interventions.

3) Dynamic graph modeling: Explicitly encoding regional labor market connectivity through graph neural networks [40] could improve predictions in federal systems with strong interstate labor flows.

C. Ethical Considerations and Responsible Deployment

The framework's policy applications raise important ethical questions requiring proactive mitigation strategies. The potential for algorithmic reinforcement of existing inequalities—such as systematically underestimating employment prospects for marginalized groups—necessitates rigorous fairness testing across protected attributes [41].

Implementation guidelines should address:

Regular audits of feature importance distributions for discriminatory patterns

Mechanisms to override automated predictions when they conflict with ground-level observations

Transparent documentation of model limitations in official communications

These safeguards become particularly critical when the framework informs resource allocation decisions affecting vulnerable youth populations. The attention weights, while providing interpretability, could inadvertently legitimize biased predictions if not contextualized with appropriate domain expertise [42]. Future work should develop participatory design frameworks to incorporate frontline practitioner knowledge into model refinement processes.

VIII. Conclusion

The proposed hybrid framework demonstrates significant advancements in both predictive accuracy and interpretability for youth employment trend forecasting. By integrating dilated convolutions with a dual-path attention mechanism, the model effectively captures multi-scale temporal patterns while providing transparent feature importance rankings. Experimental results across diverse datasets confirm its superiority over conventional econometric and deep learning approaches, particularly in handling nonlinear interactions and sudden labor market shocks.

The framework's ability to generate policy-actionable insights represents its most valuable contribution. Attention-derived feature weights align with established labor economic theories, enabling decision-makers to identify critical drivers of youth employment under varying economic conditions. This interpretability, combined with robust predictive performance, addresses a longstanding gap in computational labor market analysis—bridging the divide between data-driven forecasting and theoretically grounded policy formulation.

Future enhancements could further strengthen the framework's real-world applicability. Incorporating causal inference techniques would allow for more rigorous evaluation of policy interventions, while dynamic graph modeling could better capture regional labor market interdependencies. Maintaining a focus on ethical considerations remains paramount, ensuring that model outputs do not inadvertently reinforce existing inequalities or biases in labor market systems.

The success of this approach suggests promising directions for interpretable machine learning in socioeconomic forecasting. Similar hybrid architectures could be adapted to other complex temporal prediction tasks requiring both accuracy and transparency, from educational outcome modeling to public health trend analysis. As labor markets continue evolving amid technological and demographic shifts, such tools will become increasingly vital for evidence-based policy design targeting youth employment challenges worldwide.

References

[1] J. H. Stock and M. W. Watson, "Vector autoregressions," *J. Econ. Perspect.*, vol. 15, no. 4, pp. 101-115, Fall 2001, doi: 10.1257/jep.15.4.101.

[2] R. H. Shumway and D. S. Stoffer, "ARIMA models," in *Time Series Analysis and Its Applications: With R Examples*, 4th ed. Cham, Switzerland: Springer International Publishing, 2017, ch. 3, pp. 75-163, doi: 10.1007/978-3-319-52452-8_3.

[3] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*, Berlin, Germany: Springer, 2012, pp. 37-45, doi: 10.1007/978-3-642-24797-2_4.

[4] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 21-26, 2017, pp. 156-165, doi: 10.1109/CVPR.2017.113.

[5] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, Art. no. 6622927, Jul. 2020, doi: 10.1155/2020/6622927.

[6] A. Vaswani et al., "Attention is all you need," in *Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 4-9, 2017, pp. 5998-6008.

[7] M. Simionescu and J. Cifuentes-Faura, "Forecasting national and regional youth unemployment in Spain using google trends," *Soc. Indic. Res.*, vol. 161, no. 2, pp. 699-721, Jun. 2022, doi: 10.1007/s11205-021-02756-x.
[8] J. Evans and W. Shen, "Youth employment and the future of work," in *Council of Europe Youth Knowledge Series*. Strasbourg, France: Council of Europe Publishing, 2010.
[9] D. Shigapova, M. Valiullin, O. Yrieva, and L. Safina, "The methods of prediction of demand on the labor market," *Procedia Econ. Finance*, vol. 24, pp. 636-642, 2015, doi: 10.1016/S2212-5671(15)00656-4.
[10] X. Dong, "Prediction of college employment rate based on big data analysis," *Math. Probl. Eng.*, vol. 2021, Art. no. 5574413, 2021, doi: 10.1155/2021/5574413.
[11] M. A. H. Ashour and R. A. A. Helmi, "Predicting the youth unemployment rate in Iraq until 2035 using artificial intelligence," in *Proc. 14th Int. Conf. Adv. Comput. Inf. Technol. (ACIT)*, Kirkuk, Iraq, Apr. 24-26, 2024, pp. 321-326, doi: 10.1109/ACIT56853.2024.10456832.
[12] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proc. 35th AAAI Conf. Artif. Intell.*, Virtual Event, Feb. 2-9, 2021, pp. 7436-7444, doi: 10.1609/aaai.v35i8.16913.
[13] K. Fauvel, T. Lin, V. Masson, É. Fromont, and A. Termier, "XCM: An explainable convolutional neural network for multivariate time series classification," *Mathematics*, vol. 9, no. 23, Art. no. 3137, 2021, doi: 10.3390/math9233137.
[14] J. Y. Oostvogel, "Interpretable deep learning for time series forecasting," Ph.D. dissertation, Dept. Math. Comput. Sci., Eindhoven Univ. Technol., Eindhoven, Netherlands, 2025.
[15] K. Kim, "Unemployment dynamics forecasting with machine learning regression models," arXiv preprint arXiv:2505.01933, May 2025.
[16] D. Couts, D. Grether, and M. Nerlove, "Forecasting non-stationary economic time series," *Manage. Sci.*, vol. 12, no. 4, pp. 450-467, Dec. 1966.
[17] H. Dietrich, "Youth unemployment in Europe," International Policy Analysis, Friedrich-Ebert-Stiftung, Berlin, Germany, 2012.
[18] P. W. Miller, "Unemployment patterns in the youth labour market," *Australian Econ. Papers*, vol. 25, no. 47, pp. 222-234, Dec. 1986.
[19] R. Hatos, "Skills mismatch of the young people at the European level," *Annals Univ. Oradea, Econ. Sci.*, vol. 23, no. 1, pp. 431-439, 2014.
[20] O. Ashenfelter and J. Ham, "Education, unemployment, and earnings," *J. Political Econ.*, vol. 87, no. 5, pp. S99-S116, Oct. 1979.
[21] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," in *Proc. Nat. Acad. Sci.*, vol. 116, no. 44, pp. 22071-22080, Oct. 2019, doi: 10.1073/pnas.1900654116.
[22] P. Lewis and J. Heyes, "The changing face of youth employment in Europe," *Econ. Ind. Democracy*, vol. 41, no. 2, pp. 457-480, May 2020, doi: 10.1177/0143831X17720017.
[23] G. L. Velmonte, "Job that fits for graduates in the Asean integration," *Int. J. Intell. Comput. Technol.*, vol. 4, no. 1, pp. 19-22, Jun. 2020.
[24] A. L. Fernandes-Alcantara, "Youth and the labor force: Background and trends," Congressional Research Service, Washington, DC, USA, Rep. R42519, May 2012. [Online]. Available: https://digital.library.unt.edu/ark:/67531/metadc807010/
[25] H. Lütkepohl, "Vector error correction models," in *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer-Verlag, 2005, ch. 6, pp. 237-267, doi: 10.1007/978-3-540-27752-1_6.
[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (KDD '16), San Francisco, CA, USA, Aug. 13-17, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.
[27] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA, USA: Apress, 2015, ch. 4, pp. 67-80, doi: 10.1007/978-1-4302-5990-9_4.
[28] S. Jayasundara, A. Indika, and D. Herath, "Interpretable student performance prediction using explainable boosting machine for multi-class classification," in *2022 2nd Int. Conf. Advanced Research in Computing* (ICARC), Belihuloya, Sri Lanka, 2022, pp. 391-396, doi: 10.1109/ICARC54489.2022.9753867.
[29] R. Agarwal, L. Melnick, N. Frosst, X. Zhang, B. Lengerich, R. Caruana, and G. E. Hinton, "Neural additive models: Interpretable machine learning with neural nets," in *Advances in Neural Information Processing Systems 34* (NeurIPS 2021), Virtual Conference, 2021, pp. 4243-4257.
[30] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748-1764, Oct.-Dec. 2021, doi: 10.1016/j.ijforecast.2021.03.012.
[31] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679-688, Oct.-Dec. 2006, doi: 10.1016/j.ijforecast.2006.03.001.
[32] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Berlin, Germany: Springer-Verlag, 2007, ch. 4, pp. 69-84, doi: 10.1007/978-3-540-74048-3_4.
[33] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 20, no. 1, pp. 134-144, Jan. 2002, doi: 10.1198/073500102753410444.
[34] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. Series B (Methodological)*, vol. 57, no. 1, pp. 289-300, Jan. 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.
[35] A. Behtoui, "Beyond social ties: The impact of social capital on labour market outcomes for young Swedish

people," *J. Sociology*, vol. 52, no. 4, pp. 711-724, Dec. 2016, doi: 10.1177/1440783315581217.
[36] R. Hussmanns, "Defining and measuring informal employment," Geneva, Switzerland: Int. Labour Office, Bureau of Statistics, Policy Integration Dept., Working Paper no. 53, Dec. 2004.
[37] H. Metcalf, "Stuck in the pipeline: A critical review of STEM workforce literature," *InterActions: UCLA J. Edu. Inf. Stud.*, vol. 6, no. 2, Article 4, 2010, doi: 10.5070/D462000681.
[38] P. Xu, X. Zhu, and D. A. Clifton, "Multimodal learning with transformers: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12113-12132, Oct. 2023, doi: 10.1109/TPAMI.2023.3275156.
[39] P. Hünermund, B. Louw, and I. Caspi, "Double machine learning and automated confounder selection: A cautionary tale," *J. Causal Inference*, vol. 11, no. 1, Article 20220078, May 2023, doi: 10.1515/jci-2022-0078.
[40] M. Zhang, S. Wu, X. Yu, Q. Liu, and L. Wang, "Dynamic graph neural networks for sequential recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 4741-4753, May 2023, doi: 10.1109/TKDE.2022.3151618.
[41] D. Pessach and E. Shmueli, "A review on fairness in machine learning," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1-44, Feb. 2022, Art. no. 51, doi: 10.1145/3494672.
[42] S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq, "Algorithmic decision making and the cost of fairness," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Halifax, NS, Canada, Aug. 2017, pp. 797-806, doi: 10.1145/3097983.3098095.