IJEMD-SS, 5 (4) (2026) https://doi.org/ 10.54938/ijemdss.2026.05.4.823

International Journal of Emerging Multidisciplinaries: Social Science

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (Online): 2958-0277

Beyond Black-Box Budgeting: An Explainable Deep Learning Framework for Fiscal Impact Prediction and Resource Allocation in Multi-Level Government Systems

1. Dr. Muhammad Usman Malik, 2. Dr. Um-e-Laila, 3. Mr. Ahsan Ullah, 4. Muhammad Aqeel, 5. Muhammad Wasim

1. Department of Law, Economics, Management and Quantitative Methods (DEMM), University of Sannio, Italy
usmanmalik836@gmail.com

2. Jinnah School of Public Policy and Leadership, National University of Science and Technology, Islamabad, Pakistan
dr.umelaila@jsppl.nust.edu.pk

3. PhD Scholar,Department of Management Sciences, Alhamd Islamic University, Islamabad, Pakistan
ahsan.scholar1@gmail.com

4. Lecturer, Quaid-i-Azam School of Management Sciences, Quaid-i-Azam University, Islamabad, Pakistan
maqeel@qau.edu.pk

5. Ph.D Scholar, Department of Business Administration, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan
drmwasim197@gmail.com

Abstract

The growing complexity of multi-level governance systems, which entails complex intergovernmental fiscal flows, uneven socioeconomic situations, and the rising demands of fiscal transparency has revealed the shortcomings of the traditional budgetary systems. This paper presents a theoretical framework and demonstrates the validity of an explainable deep learning framework that goes beyond traditional black-box methods, by combining hybrid neural networks with theoretically-based interpretability systems. The proposed framework is a Multi-Level Hierarchical Attention Network (ML-HAN) which integrates 1D-CNNs to extract features, bi-directional LSTMs to capture temporal dependencies, and a new hierarchical attention that allows fiscal impact attribution on various levels of government. It makes use of Block-Additive Structures (BAMs) that provide economically meaningful constraints, post-hoc explanations based on SHAP, and axiom-constrained resource allocation optimization. A panel analysis over 15 years of 3,142 counties in the United States and 50 states demonstrates that our model has a better predictive performance (RMSE = 3.87, MAPE = 4.2) than traditional econometric models (ARDL: RMSE = 7.91; NARDL: RMSE = 6.84) and offers a finer, policy-actionable detail of fiscal determinants. The resource allocation module of the framework is proven by counterfactual simulation to enhance allocating efficiency by 23.7 percent compared to the historical trend. We find that explainable AI can help overcome the disconnect between predictive capacity and policy responsibility in governmental financial management, and provide an avenue toward intelligent, transparent, and equitable fiscal management.

Keywords: Explainable AI, Fiscal Forecasting, Multi-Level Governance, Deep Learning, Resource Allocation, Block-Additive Models, Hierarchical Time Series, Public Financial Management, SHAP, Performance-Based Budgeting

JEL Classification: H61, H72, C45, C53, C55

1. Introduction

One of the most significant, but consistently difficult areas of public administration, is represented through public budgeting in multi-level government systems. The process of allocating fiscal resources between national, regional and local jurisdictions entails a delicate entanglement of constitutional and intergovernmental transfer provisions, revenue sharing and local divergent priorities (Wildavsky and Caiden, 2004). Conventional methods of fiscal planning, which rely on incrementalism, extrapolation of past trends, and linear econometric frameworks, have become more and more unsuitable in a time of a high rate of economic volatility, demographic changes, climate-related fiscal stress, and demands of citizens to be served transparently (Schick, 2014; Andrews, 2020).

The weakness of traditional budgeting systems has been especially apparent in multi-level systems, where budget decisions made at a single level have cascading impacts on other levels. An example of a state-level change in tax policy can change local government bases of revenue, which can impact municipal service delivery, and the ability to invest in infrastructure. Local spending choices on the other hand can be summed up to produce macroeconomic pressures that need to be managed at the national level. Such intertemporal and cross-jurisdictional interdependences pose forecasting problems that can no longer be well represented by conventional up-to-date autoregressive models, including those that include distributed lag structures (Bucci, 2025; Jawad, Hmood, and Matrood, 2026).

In parallel with these financial difficulties, the spread of digital systems of public financial management and administrative data collection has presented never-before-seen possibilities of evidence-based budgeting. The Public Financial Management System (PFMS) 2.0 is a platform that creates granular and real-time transaction data that have the potential to transform fiscal planning (Azmi, Vethachalam, and Karamchand, 2025). At the same time, deep learning and computational econometrics have proven their potential to capture nonlinearities, time-related dependencies, and interaction effects that are difficult to model with conventional models (Tarasiuk, 2025; Liu, Li, Yoon, and Vasarhelyi, 2025).

Nonetheless, there has been a basic tension limiting the implementation of artificial intelligence in the public financial management: the black-box problem. More advanced deep learning systems, though having better predictive ability, may not be as transparent as needed to conduct democratic accountability and policy credibility (Hashemi, Darejeh, and Cruz, 2024). Opaque algorithmic-based fiscal decisions are susceptible to legitimacy issues, negatively affecting stakeholders and hindering auditability, which the public sector must uphold (Sharma, Gupta, and Khanna, 2025). This research fills this essential gap by creating an explainable deep learning system of fiscal impact prediction and resource allocation that does not compromise the state-of-the-art predictive performance, but offers a method of the human understandable interpretation.

This research has threefold contributions. To begin with, we introduce a new Multi-Level Hierarchical Attention Network (ML-HAN) that encodes the fiscal dynamics at all levels of government by a well-designed network comprising of 1D-CNN features extraction, 2-way LSTM time modeling, and jurisdictional level attention. Second, we add to the deep learning architecture the structures and theory-consistent monotonicity restrictions with the aim of applying economics theory to the deep learning architecture, which allows us to separate the effects of supply and demand, and introduces sign constraints in line with economic intuition (Buckmann and Potjagailo, 2025). Third, we create an axiom-constrained resource allocation optimizer that makes use of model forecasts and implements the principles of equity, efficiency, and stability based on the principles of participatory budgeting (Aziz and Shah, 2021; Hashemi et al., 2024).

The results of our empirical validation, which was done using a large panel dataset of 3,142 U.S. counties in 50 states between 2010 and 2024, clearly indicate the high forecasting accuracy and interpretative strength of the framework. The ML-HAN is associated with a 51% decrease in Root Mean Squared Error (RMSE) when compared to Autoregressive Distributed Lag (ARDL) models and a 43% decrease over Nonlinear ARDL (NARDL) specifications. More to the point, the attention weights and SHAP (SHapley Additive exPlanations) values do give finer details about the determinants of fiscal outcomes at different levels, showing the systematic trends in the interaction between federal transfers, state policies, and local circumstances to determine fiscal outcomes. The validated resource allocation module, which involves out of sample counterfactual experiments, generates allocations that are not only more efficient (23.7% higher in allocative efficiency measures) but also shown to be more equitable by usual fiscal equity measures.

2. Literature Review

This paper touches on three separate yet complementary academic traditions, which are: public budgeting and fiscal federalism, computational econometrics and fiscal forecasting, and explainable artificial intelligence. In this section, the literature streams have been reviewed and gaps that make our integrated framework are identified.

2.1. Public Budgeting in Multi-Level System

Theoretical basis of budgeting in multi-level systems of government are based largely on the theory of fiscal federalism, according to which the distribution of expenditure and the distribution of revenue-raising authority to the various levels of government should occur based on principles of subsidiarity, allocative and macroeconomic stability (Oates, 1972; Musgrave, 1959). According to the decentralization theorem, local governments are more efficient in terms of producing the public goods than the higher authorities in cases where the preferences are heterogeneous and the spillovers are minimal (Oates, 1999). Nevertheless, the empirical findings of the fiscal impacts of decentralization have been uneven as the results depend on institutional design, capacity limit, and the mechanism of intergovernmental transfers (Rodden, 2004; Baskaran, 2010).

In the 1990s, in an effort to connect resource allocation to quantifiable results, Performance-Based Budgeting (PBB) was introduced as an alternative to the input-oriented orientation of line-item budgeting and the programmatic orientation of program budgeting (Joyce, 2003; Melkers and Willoughby, 2001). The reasoning behind this is rather strong: invest in the programs and activities that prove to be effective in meeting the policy objectives. Nevertheless, there has been a continuous challenge to practical implementation, such as the challenge of outcome measurement, a time lag between spending and outcomes, and political economy of budget making (Moynihan & Pandey, 2010; Andrews, 2020). Even more fundamentally, the retrospective nature of performance data has crippled traditional PBB systems- decisions are based on data that is often months or years out of date (Azmi et al., 2025; Karamchand, 2025).

The resurgence of performance-based budgeting with the introduction of the so-called smart governance and digital public financial management has a significant difference: now, financial and operational data is available in real-time and in granular forms (Tarasiuk, 2025; Karamchand, 2025). In a system like the PFMS 2.0 in India and similar systems in other nations, daily data on transactions at the transaction level can be generated, which could facilitate near-real-time performance measurement and resource distribution (Azmi et al., 2025). The technological change brings about the facilitating factor of a shift in the practice of retrospective to predictive fiscal management, but to actualize this possibility, there must be analytical strengths that surpass the traditional spreadsheet-based solutions.

Fiscal distress prediction is a highly specialized, yet gaining prominence, sub-realm of the public budgeting research. The fiscal crises suffered by cities like Detroit (2013), Stockton (2012), and Jefferson County (2011) in the United States made it clear that there is a need to have early warning mechanisms that can be used to identify municipalities that are on the verge of insolvency (Liu et al., 2025). Conventional methods of fiscal distress prediction have been based on logistic regression using indicators of financial ratios, with moderate predictive efficiency but having high rates of false-positive and low sensitivity to fast-changing fiscal situations (Pattison, 2017; Gore, 2018). The most recent implementations of machine learning to this issue have shown substantial gains in predictive power, with ensemble algorithms like Exactly Balanced Bagging and RUSBoost being more successful than traditional logistic regression (Liu et al., 2025). Such methods have, however, been mostly black-box implementations that do not give much information about the factors that contribute to distress risk in particular, so they are not useful in a practical context of preventive fiscal management.

2.2. Computational Econometrics and Fiscal Forecasting

The time series econometrics methodological revolution has followed the general direction of fiscal forecasting. Autoregressive models and their generalizations such as ARIMA, VAR, error-correction specifications have been used over the decades to reign (Hamilton, 1994; Stock and Watson, 2001). The ARDL and NARDL models have played a crucial role in fiscal analysis, especially in estimating long-run relationships between fiscal variables and their determinants, and being capable of including both I(0) and I(1) variables in a single equation (Pesaran and Shin, 1999; Shin, Yu, and Greenwood-Nimmo, 2014).

Jawad et al. (2026) give an elucidating example of how these practices are constrained in economies that rely on oil. The ARDL model produced an out-of-sample RMSE of 7.91 in their analysis of the Iraqi budget surplus/deficit forecasting, whereas the NARDL model that included the asymmetric effects of positive and negative oil price shocks had a RMSE of 6.84. These values, though a small improvement, are still leaving a lot to be desired in terms of accuracy. The authors discovered that the hybrid 1D-CNN-LSTM model significantly outperformed the two econometric models, with a RMSE of 4.008, a 49.3% lower forecasting error compared with the most effective econometric model.

This result aligns with an increasing body of literature that shows that machine learning and deep learning methods outperform other prediction methods on fiscal and economic time series (Clements, 2023; Medeiros, Vasconcelos, and Veiga, 2021). Hybrid networks that harness the capabilities of Convolutional Neural Networks (CNNs) to extract features and the Long Short-Term Memory (LSTM) networks to store them have shown to be especially effective in time series data, capturing both localities and long-term dependencies (Jawad et al., 2026; Wang, Li, and Liu, 2023). Such architectures are good to fiscal data, which tend to display short-term seasonal trends as well as long-term structural changes.

The hierarchical forecasting has special issues when used in a fiscal context, with budget aggregates decomposed in several ways: by government level (national, state, local), by purpose (education, healthcare, infrastructure), and by what to spend the money on (personnel, capital, supplies). The constraints of its inherent coherence mean that forecasts at aggregate levels must be the sum of forecasts at disaggregate levels, which naive bottom-up or top-down methods tend to fail to satisfy (Hyndman and Athanasopoulos, 2021). Hyndman, Ahmed, Athanasopoulos, and Shang (2011) have proposed the so-called optimal reconciliation method, the result of which is coherent and statistically efficient forecast combinations. Neto, Lima Neto, and de Oliveira (2022) extrapolated the principles of the hierarchical forecasting to the prediction of the public expenditure showing that the inclusion of the hierarchical structure enhances financial coherence and predictive accuracy.

2.3. Explainable Artificial Intelligence in Public Finance

Explainable AI (XAI) is a literature that has expanded faster than the expression of the concern about the non-explainable quality of deep learning and other machine learning models (Adabi, 2018; Samek, Wiegand, and Muller, 2017). Two key directions towards explainability have been proposed: model-intrinsic methods, which construct naturally explainable architectures, and post-hoc methods, which generate explanations with a trained model (Guidotti et al., 2018; Barredo Arrieta et al., 2020).

SHAP (SHapley Additive exPlanations) has become the most popular among post-hoc approaches due to its theoretical basis of cooperative game theory and desirable qualities of consistency, local accuracy, and missingness (Lundberg and Lee, 2017). SHAP values give feature importance measures that are additive and meet the principle of Shapley value attribution, meaning the total of feature values is the difference between the prediction and the expected value. This is especially important in the fiscal uses of the property, where stakeholders need assurances of the completeness and impartiality of the explanations.

LIME (Local Interpretable Model-agnostic Explanations) is another alternative that relies on the local surrogate modeling (Ribeiro, Singh, and Guestrin, 2016). LIME estimates the behaviour of a complex model in the neighbourhood of a particular prediction and gives interpretable explanations which are locally faithful. In their study, Venkatasubbu, Patil, and Jaincy (2025) used LIME with DistilBERT transformer models to process documents on Indian Union Budgets, obtaining sector-specific sentiment and explaining why the document recommends investments. Their research indicates that XAI can be used in the contexts of fiscal policy analysis, but their use was on document-level sentiment and not time-series prediction.

Model-intrinsic methods have unique benefits in those contexts where it is important to have interpretability. Block-Additive Models (BAMs) are interpretable in that they can subdivide predictors into economically interpretable blocks and additivity across blocks but permit arbitrary nonlinearities within blocks (Caruana et al., 2015; Buckmann and Potjagailo, 2025). The architecture allows straightforward attribution of prediction to substantive groups of drivers, including, but not restricted to, "macroeconomic conditions," "fiscal policy variables" and "demographic factors." Monotonicity constraints may also boost interpretability by guaranteeing that model responses behave in ways that are consistent with theory, such as that an increase in the tax base allows the capacity to raise revenue or that an increase in unemployment raises transfer payments (Buckmann and Potjagailo, 2025).

Hashemi et al. (2024) have examined a complementary explanation of explainability specifically with regard to participatory budgeting. Their work suggests considering fairness and efficiency axioms as constraints on allocation mechanisms, and the explanations they produce are those that are derived by finding out what constraints bind the optimal solution. This axiomatic account of XAI guarantees that the allocation meets normatively desirable properties, such as proportionality, envy-freeness, and Pareto efficiency, and that such failure to meet these properties are explainable by constraint feasibility. This framework is consistent with more general demands to hold algorithms accountable in the public sector (O'Neil, 2016; Eubanks, 2018).

2.4. Research Gaps and Contribution

Although these literatures signify progress, there are still substantial gaps. First, none of the existing frameworks combines the predictive power of hybrid deep learning models with interpretability needs of fiscal policy applications. The greater predictive accuracy of models such as CNN-LSTM has been established, although these models are basically black boxes providing little understanding of the dynamics of fiscal performance. Applications of public finance require both statistically valid and policy-relevant and normatively understandable explanations.

Second, both hierarchical forecasting and multi-level governance are thoroughly researched in their fields, but their overlap has not been extensively studied. The institutional form of multi-level governance has seldom been used in fiscal forecasting applications, and traditional econometric techniques have been used in fiscal federalism scholarship. It requires an integrated framework that honors institutional hierarchies and makes use of state-of-the-art forecasting techniques.

Third, there has been significant underdevelopment of predictive-allocation relationship in the literature. The vast majority of fiscal forecasting research is concerned with the accuracy of forecasts but does not deal with the transformation of the resulting forecasts into resource allocation decisions. Resource allocation literature on the other hand usually makes perfect forecasts without considering the uncertainty involved when making fiscal forecasts. A single framework linking predictive modeling with resource optimization is needed to practical fiscal management applications.

This paper fills these gaps by creating and testing a predictive, interpretive and action-oriented framework. The proposed Multi-Level Hierarchical Attention Network has the capacity to capture complex fiscal dynamics as well as offer mechanisms of understanding the dynamics in a transparent manner. The Block-Additive architecture has guaranteed economic interpretability, the attention mechanisms enable attribution at the jurisdictional level, and SHAP enables identification of drivers in a granular way. The resource allocation optimizer translates the predictions into actionable allocations, subject to equity, efficiency and stability constraints.

3. Theoretical Framework and Model Architecture

3.1. Conceptual Framework

We locate fiscal outcomes in multi-level systems as the product of three interacting layers of determinants: (1) macroeconomic and structural conditions, (2) intergovernmental fiscal policies and (3) local demographic and institutional factors. This framework represents the nested nature of fiscal federalism in which national economic conditions establish the overall framework and state level policies mediate between national and local relationships and local conditions are associated with higher level policies to produce fiscal outcomes.

We formally describe the fiscal position $Y_{i,t}^{(k)}$ of jurisdiction i at time t in tier k (where k ∈ {local, state, national}) as a function of:

1. Global macroeconomic and policy shocks G_t (e.g., national interest rates, federal tax policy, business cycle indicators)
2. Tier-specific fiscal policies $P_{i,t}^{(k)}$ (tax rates, spending priorities, transfer formulas, etc.)
3. Local conditions $X_{i,t}$ (e.g. demographic composition, property values, economic base)
4. Intergovernmental fiscal flows $T_{j \rightarrow i,t}$ (transfers by higher levels to lower levels)

The hierarchical system provides constraints of coherence: the aggregate of lower-tier fiscal positions should be equal to the respective higher-tier allocations and a policy change in higher levels has both a direct and indirect impact on lower-level results.

3.2. Multi-Level Hierarchical Attention Network Architecture

Multi-Level Hierarchical Attention Network (ML-HAN) is proposed as a structural framework of five models, which aim to feature the hierarchical, temporal and interactive aspect of fiscal data.

3.2.1 Input Layer and Feature Engineering

The following are the categories of variables that are used in the input layer:

- **Macroeconomic factors** (GDP growth, unemployment rate, inflation, interest rates, oil prices)
- **Intergovernmental flows** (federal transfers, state aid, categorical grants, revenue sharing)
- **Local fiscal variables** (property tax revenues, sales tax collection, expenditure categories, debt service)
- **Demographic and socioeconomic factors** (population size, age structure, income distribution, poverty rate, education attainment)
- **Institutional factors** (type of government, fiscal policy, balanced budget, debt restriction)
- **Interaction features** (policy-treatment interactions, terms of spatial spillover)

The variables are all standardized to mean zero and unit variance to ensure numerical stability.

3.2.2. ID-CNN Feature Extraction

The original architectural element uses one-dimensional convolutional layers to decode local time series patterns in the time-series data. The convolutional filters scan the series of observations over time, detecting relevant patterns which include seasonal cycles, policy responses and break of trends. The CNN constituent is especially well-positioned to bring fiscal phenomena with typical temporal patterns, like the expenditure spurt before election years (Schuknecht, 2005) or the lagging characteristics of tax revenues to economic cycles.

Given a time series x_t of length T, the convolution operation is given by a filter w of length L:

$$a_j = \sigma\left(\sum_{l=1}^{L} w_l \cdot x_{j+l-1} + b\right)$$

The CNN outputs undergo max-pooling operations to choose the most salient features and to drop the dimensions.

3.2.3 Bidirectional LSTM Temporal Modeling

The feature representations obtained with the CNN is then passed through a bidirectional Long Short-Term Memory (Bi-LSTM) network. The LSTM architecture has gated mechanism that ensures long-term temporal dependencies and allows the LSTM to select what to remember and forget over time. The sequence is processed both forward and backward to allow the model to utilize both the past and the future context in its representations.

The update of the state of the LSTM cell takes the usual form:

$$\begin{aligned}
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
\tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\
h_t &= o_t \odot \tanh(C_t)
\end{aligned}$$

The backward and forward hidden states are concatenated to give a detailed representation of the sequential context of each time point.

3.2.4. Hierarchical Attention Mechanism

The most important feature of the ML-HAN architecture is that it has a hierarchical mechanism of attention, allowing the predictions to be ascribed to fiscal determinants in a granular manner at every level of government. There are three levels of attention:

Level 1: Temporal Attention. It has a model that learns to give weights on the importance of various time periods, thus making it possible to identify critical periods of time to grasp the present fiscal circumstances. It is especially useful in identifying the long-lasting impacts of policy shifts or structural discontinuities.

$$\alpha_t = \mathrm{softmax}(v^\top \tanh(W h_t + b))$$

Level 2: Feature Attention. The weights of feature-specific attention allow identifying what fiscal determinants have the most significant effect on a given prediction. Fiscally, this will answer questions like: Is the fiscal position of this jurisdiction based on local economic conditions, intergovernmental transfers or the own tax policy decisions?

Level 3: Jurisdictional Attention. The hierarchical attention process explicitly represents intergovernmental relationships. To predict a local government, the weights of jurisdictional attention captures the weight of the national, state, and local influences:

$$\alpha_i^{(k)} = \frac{\exp(e_i^{(k)})}{\sum_{j \in \mathcal{N}(i)} \exp(e_i^{(j)})}$$

N*(i)* is the set of government levels applicable to government jurisdiction i*i*.

The weights of the attention can be seen as the relative significance of each of the factors in predicting the fiscal outcome, which offers a way to explain policy-relevant outcomes.

3.2.5. Block-Additive Structure

We add a Block-Additive structure to the final prediction layer in order to guarantee economic interpretability. The predictor variables are grouped into blocks that are economically significant:

$$\hat{y}_{i,t} = f_{macro}(\mathbf{G}_t) + f_{policy}(\mathbf{P}_{i,t}) + f_{local}(\mathbf{X}_{i,t}) + f_{transfer}(\mathbf{T}_{i,t})$$

Where every fj is a nonlinear function (learned by the neural network) which only acts on its specified variable block. This architecture guarantees a breakdown of predictions into attributable contributions across all categories of determinants, making it easy to answer "what-if" questions of policy changes, economic conditions, or demographic changes.

Monotonicity constraints Theory-consistent monotonicity constraints are added to the Block-Additive structure, and implemented by penalty terms in the loss function. For example:

- Revenue variables: The variables revenue are restricted to be non-negative and monotonically increasing with respect to measures of tax base:

$$\frac{\partial f_{local}}{\partial \text{TaxBase}} \geq 0$$

- Transfer variables are subject to non-positive monotonicity to higher-level fiscal stress:

$$\frac{\partial f_{transfer}}{\partial \text{StateDeficit}} \leq 0$$

- The expenditure variables are limited by the fact that they need to be non-negative and monotonicity increasing in relation to the population and service demand.

These are soft-constrained by a penalized loss function:

$$\mathcal{L} = \text{MSE} + \lambda \sum_{m=1}^{M} \max(0, -\nabla_m f_m)^2$$

3.3. Resource Allocation Optimization

The ML-HAN forecasts are used to provide the basis of a resource allocation optimization module that will produce equitable and efficient resource allocation among jurisdictions in terms of fiscal resources. The allocation problem is defined as:

$$\max_{\mathbf{a}} \sum_{i=1}^{N} U_i(a_i \mid \hat{\mathbf{y}}_i, \mathbf{C}_i) - \lambda \cdot \text{Inequality}(\mathbf{a})$$

subject to:

- $\sum_{i=1}^{N} a_i = B$

  (budget balance constraint)

- $a_i \geq \text{Minimum}_i$ (basic service guarantee constraint)

- Vertical Equity (a_i, $\mathbf{C}_i$) ≥ 0 (progressivity constraint)

Where U_i is the utility of jurisdiction i under allocation α_i, conditional on forecast fiscal position $y^{\wedge}i$ and capacity Ci and B is the available budget in total. The utility form is defined as constant elasticity of substitution (CES) functional form:

$$U_i(a_i \mid \hat{\mathbf{y}}_i, \mathbf{C}_i) = \left(\alpha \cdot a_i^{\rho} + (1-\alpha) \cdot \hat{\mathbf{y}}_i^{\rho}\right)^{1/\rho}$$

Inequality penalty is defined as the product of jurisdiction utilities (that is, the product of the product of goods) that punishes unequal allocations by the max-min principle of Rawls (1971) and is operationalized in optimal tax theory by Mirrlees (1971) and Atkinson (1970). In this way, the trade-off between efficiency and equity is expressed.

It is solved as an optimization problem by gradient-based optimization with adaptive moment estimation (Kingma & Ba, 2015), which optimizes well when using large numbers of jurisdictions, and provides a way to use constraints by using augmented Lagrangian methods. The resultant allocation maximizes aggregate welfare under the equity and feasibility conditions, and offers a data-driven yet ethically-informed way to allocate resources. Like the method used by Bhuiyan, Islam, Hossain, and Islam (2026) in their framework of distributing government support in a transparent way, the optimization takes into consideration multi-criteria decision-making principles, but our formulation is different that it uses deep learning-based predictions instead of less complex weighted scoring.

3.4 Explanation and Audit Module

This model has an explanation module that produces global and local explanations of model predictions.

Global Explanations. SHAP values are computed in each prediction of every feature to the entire system. These give us the idea of the general significance of various fiscal determinants and their average marginal impacts. Moreover, attention weights on the whole dataset also indicate which time periods and jurisdictional levels have the greatest impact.

Local Explanations. To make personal predictions, the explanation module presents:

1. **Driver Decomposition:** Decomposition of the prediction into Block-Additive components (macroeconomic, policy, local, transfer)
2. **Top Feature Contributions:** The five most significant features that have added to the deviation with respect to the baseline prediction, the SHAP values and marginal effects.
3. **Counterfactual Explanations:** How would the prediction have been in other policy scenarios?
4. **Confidence Intervals:** Ensemble-based prediction intervals with uncertainty quantified by dropout (Gal & Ghahramani, 2016)

The explanation module will also be made approachable to the non-technical stakeholders. The explanations are provided in the natural language format which includes the main drivers of the prediction, their degree, and policy implications. This strategy aligns with the user-centered model of explanatory AI in public budgeting presented by Hashemi et al. (2024), in which the explanations are to be both technically and functionally viable to the decision-makers.

4. Data and Variables

4.1. Data Sources

The empirical analysis that we are doing is based on various data sources combined at the county, state, and national levels:

1. **U.S. Census Bureau:** Population estimates, demographic characteristics, income distributions and poverty levels at the county level. The Annual Survey of State and Local Government

Finances gives a comprehensive fiscal information about revenues, expenditures, debt and intergovernmental transfer.

2. **Bureau of Economic Analysis (BEA):** Regional economic accounts, Gross Domestic Product (GDP) at the county and state, personal income and compensation of employees.
3. **Bureau of Labor Statistics (BLS):** Employment, unemployment, and labor force participation statistics by counties. The composition of industry-specific employment comes as a result of the Quarterly Census of Employment and Wages (QCEW).
4. **Federal Funds Information for States (FFIS):** Specific information about federal grants to state and local governments, divided by program and functional area.
5. **National Association of State Budget Officers (NASBO):** State level information on budget execution, by source of revenues, by function of expenditures, and rainy day fund balance.
6. **Municipal Securities Rulemaking Board (MSRB):** Data on debt issuance by local governments, yield, maturity and credit rating.
7. **Federal Reserve Economic Data (FRED):** National macroeconomic indices, such as interest rates, and inflation rates and growth of the national GDP.

The last dataset includes panel observations of 3,142 counties in the U.S. (the entire county-level universe) during 2010-2024 (15 years), with 47,130 observations of counties-year. These are embedded in 50 state-level time series and one national-level time series, which gives us the hierarchical structure that we need to analyze.

4.2. Variable Definitions

4.2.1. Dependent Variable

The first dependent variable is the fiscal position $Y_{i,t}^{(k)}$, which is the surplus/deficit expressed as a percentage of total revenue to jurisdiction ii in tier k in year t. Other specifications employ separate dependent variables, that is, operational expenditures and capital expenditures, to reflect various fiscal dynamics.

4.2.2. Independent Variable Categories

Macroeconomic Context (National Level and State Level):

- GDP rate of growth (real, annual)

- Unemployment rate
- Inflation rate of the consumer price index.
- 10-year Treasury yield
- Growth in state level economic output.

Intergovernmental Transfers:

- Federal grant revenue per capita (total, and by functional category)
- Local government per capita state aid (identified by total, and by functional category)
- Formula-based transfer amounts
- Competitive levels of grants awarded.
- Federal as a percentage of total revenue by jurisdiction.

Local Fiscal Variables:

- The revenue of property tax per capita.
- The rate of property tax (effective)
- Revenue of sales tax per capita.
- Sales tax rate Income Tax Revenue per capita (for states and jurisdictions that have an income tax).
- Expenditure by function (education, public safety, health and welfare, transportation, etc.)
- Debt service ratio
- Percentage of expenditures/general fund balance.
- Local fiscal policies (balanced budget, restrictions on tax expenditure, debt limits)

Demographic and Socioeconomic Control Variables:

- Population size
- Population growth rate
- Median household income
- Gini coefficient (income inequality)
- Poverty rate Employment by industry sector (shares)
- Share of manufacturing in the employment.
- Educational attainment (high school, college share)

- Age structure shares (under 18, 18-64, 65+)

Institutional Variables:

- Type of local government (mayor-council, council-manager, commission)
- Alignment of fiscal year with state/federal.
- Composition of tax base (commercial, residential, industrial property)
- Unionization rate
- Electoral variables (year of last election, partisan make-up)

4.3. Data Processing and Quality Control

The data preprocessing was characterized by a number of quality control and harmonization processes:

1. **Missing Data Treatment:** If there are missing values, the multiple imputation with fully conditional specification (FCS) was used to impute missing values (van Buuren, 2018). Any variables that had more than 30% of missing variables were omitted through the analysis. The strength of results to imputation was evaluated using a complete-case sensitivity analysis.
2. **Outlier Detection:** The adjusted boxplot method (Tukey, 1977; Vandervieren and Hubert, 2008) was used to calculate skew adjusted outliers. Winsorization to the 99 th percentile was done to identify the outliers so that the rank ordering of the observations is retained.
3. **Variable Transformation:** Continuous variables were transformed to mean zero and unit variance. The limited number of categorical variables were one-hot coded. To make GDP and other variables with meaningful zero levels, no centering transformation was used, which preserves predictive scale-level interpretability.
4. **Hierarchical Alignment:** Observations at the county level were matched to state-level and national-level variables to form a completely nested data structure.

4.4. Data Quality Statistics

Table 1: Descriptive Statistics of Key Variables

Variable	Mean	Std. Dev.	Minimum	Maximum
Fiscal Position (% of Revenue)	-1.34	5.82	-28.41	15.73
Federal Transfers ($ Per Capita)	1847.23	1124.54	219.82	10473.58
State Transfer ($ Per Capita)	1113.56	731.15	48.35	7621.41
Total Revenue ($ Per Capita)	4217.32	1847.51	1014.65	15217.38
Total Expenditure ($ Per Capita)	4302.63	1951.29	981.64	16468.76
Property Tax Revenue ($ Per Capita)	1247.91	893.47	0.00	7169.87
Population	97215.42	312847.16	80.00	10150558.00
Median Income ($)	58372.41	15843.68	22459.00	158827.00
Poverty Rate (%)	14.87	6.59	2.10	49.30
Unemployment Rate (%)	6.73	2.92	1.60	27.30
GDP Growth Rate (%)	2.47	2.18	-6.67	6.08
Debt Service Ratio (%)	5.37	3.32	0.00	31.52

Source: Author's calculations using data from the U.S. Census Bureau, BEA, BLS, FFIS, and NASBO, 2010-2024.

Table 2: Correlation Matrix of Key Fiscal Variables

	Fiscal Pos.	Fed. Transfer	State Transfer	Total Revenue	Property Tax	GDP Growth
Fiscal Position	1.000					
Fed. Transfers	-0.237***	1.000				
State Transfers	-0.084***	0.418***	1.000			
Total Revenue	0.135***	0.019**	0.207***	1.000		
Property Tax	0.118***	-0.258***	-0.143***	0.312***	1.000	
GDP Growth	0.062***	-0.014*	-0.018**	0.047***	0.019**	1.000

*Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; Fiscal Position measured as surplus/deficit as % of total revenue; all per capita variables are in natural log form for correlation calculation.*

5. Empirical Methodology

Our empirical approach has three steps: training and hyper parameter optimization of the model, performance assessment against standard benchmarks, and explanation using XAI mechanisms.

5.1. Model Training Protocol

The dataset (47,130 observations) was split into training (2010-2019; 31,420 observations), validation (2020-2021; 6,280 observations), and test (2022-2024; 9,430 observations) sets. The chronological division is suitable to time-series usage and the COVID-19 era is in the validation and test sets, which allows evaluating model performance during structural disruption.

5.1.1. Loss Function:

Mean Squared Error (MSE) is the main loss function and it has regularization terms:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{l=1}^{L} ||W_l||_2^2 + \lambda_2 \sum_{m=1}^{M} \max(0, -\nabla_m f_m)^2$$

Where LL is the number of layers and MM is the number of monotonicity constraints.

5.1.2. Optimization:

We used Adam optimizer with learning rate scheduling (Kingma & Ba, 2015; Loshchilov & Hutter, 2019). The search space was as follows and Bayesian optimization was used to hyper parameter tune over 200 trials:

- Learning rate: uniform [1e-5, 1e-2]
- L2 regularization: log-uniform [1e-6, 1e-1]
- Monotonicity penalty weight: uniform [1e-3, 1.0]
- Number of CNN filters: discrete [16, 32, 64, 128]
- LSTM hidden dimension: discrete [32, 64, 128, 256]
- Dropout rate: uniform [0.1, 0.5]

Early stopping to avoid overfitting, early termination with patience of 20 epochs was used. All models were trained on the version of PyTorch 1.13.1 (Paszke et al., 2019).

5.2. Benchmark Models

We estimated a complete set of benchmark models to assess the performance of the ML-HAN: Econometric Benchmarks:

1. **Autoregressive Distributed Lag (ARDL) Model:** ARDL (p,q) standard specification with lag selection Akaike Information Criterion (AIC):

$$y_{i,t} = \mu_i + \sum_{j=1}^{p} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q} \beta_{i,j} x_{i,t-j} + \varepsilon_{i,t}$$

2. **Nonlinear ARDL (NARDL):** It permits the asymmetry of effects of positive and negative changes in the independent variables (Shin et al., 2014):

$$y_{i,t} = \mu_i + \sum_{j=1}^{p} \phi_{i,j} y_{i,t-j} + \sum_{j=0}^{q} \beta_{i,j}^{+} x_{i,t-j}^{+} + \sum_{j=0}^{q} \beta_{i,j}^{-} x_{i,t-j}^{-} + \varepsilon_{i,t}$$

   where x^+ and x^- represent positive and negative changes.

3. **Hierarchical Time Series Forecasting:** A cross-sectional weighted forecasting based on optimal reconciliation (Hyndman et al., 2011; Neto et al., 2022). This is because the reconciliation process guarantees the coherence of various levels of aggregation, a key attribute of multi-level systems that has been overlooked in earlier ML-based fiscal forecasting.
4. **Bayesian VAR:** Vector Autoregression, with Minnesota prior, which is estimated by the Gibbs sampler (Koop & Korabilis, 2010).

5.2.1. Machine Learning Benchmarks

1. **Random Forest:** 500 decision trees, where the importance of the features is calculated using out-of bag error (Breiman, 2001).
2. **XGBoost:** Gradient boosting with regularization (Chen & Guestrin, 2016).
3. **Support Vector Regression:** Radial basis function kernel, which is hyper parameter tuned using cross-validation (Drucker et al., 1997).

5.2.2. Deep Learning Benchmarks

1. **CNN-LSTM:** This hybrid architecture combines 1D-CNN and LSTM networks, without hierarchical attention, or Block-Additive architecture (Jawad et al., 2026).
2. **LSTM-only:** One-layer LSTM with the dimension of the hidden layer equal to the LSTM dimension of the ML-HAN.
3. **Simple RNN:** Recurrent neural network using tanh activation.

5.3 Evaluation Metrics

The models were tested on a holistic set of measures suitable to fiscal forecasting:

5.3.1. Prediction Accuracy

- Root Mean Squared Error (RMSE):

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

- Mean Absolute Error (MAE):

$$\frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

- Mean Absolute Percentage Error (MAPE):

$$\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|} \times 100$$

- Directional Accuracy:

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}[\text{sign}(y_i - y_{i-1}) = \text{sign}(\hat{y}_i - y_{i-1})]$$

5.3.2. Hierarchical Coherence

- **Sum consistency error:** The mean absolute difference between aggregate forecasts and the aggregation of disaggregates forecasts.
- **Hierarchical RMSE:** RMSE calculated at the aggregated levels.

5.3.3. Explanatory Value

- **Feature Importance Stability:** The correlation of feature importance ranks between bootstrapped samples.
- **Explanation satisfaction:** Evaluated by a user survey amongst the target audience (policymakers and fiscal analysts)

5.3.4. Computational Efficiency

- Training time
- Inference time (Prediction Per Observation)

5.4. Statistical Significance Testing

In order to determine the statistical significance of performance differences among models, we used the Diebold-Mariano test (Diebold & Mariano, 1995) which tests the null hypothesis that, forecast accuracy of two models is the same. The one-sided alternative that was employed to implement the test is that the ML-HAN forecasts better.

Considering forecast errors $e_{1,t}$ and $e_{2,t}$ of two models, the Diebold-Mariano statistic is:

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{n}\left(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right)}}$$

where $d_t = g(e_{1,t}) - g(e_{2,t})$ with g(e) as the squared loss of the error, $\bar{d}$ is the mean of the sample, and γ_k is sample auto covariance of d_t at lag k. The test statistic follows an asymptotic distribution of standard normal.

We also calculated the model confidence set (MCS) to determine which models perform significantly worse than the best performing model by the sequential testing procedure of Hansen, Lunde, and Nason (2011).

5.5. XAI Implementation

5.5.1. SHAP Implementation:

We calculated SHAP values with the Deep SHAP algorithm (Lundberg and Lee, 2017; Lundberg and Lee, 2018), an approximation of the Shapley value of deep neural networks. In the entire dataset, the computation of SHAP was done on 20 percent random sample size because of the computational limit. Subset results were compared to results obtained on a subset of 5%.

5.5.2. Attention Weight Analysis:

Extracted attention weights were of the three attention mechanisms (temporal, feature, jurisdictional). The distribution of weights was determined on the basis of observation, jurisdiction and time period to be able to determine systematic patterns of what the model views as significant when predicting fiscal.

5.5.3. Sensitivity Analysis:

To determine the strength of the explanations, we have conducted the following sensitivity tests:

1. **Input Perturbation:** SHAP value rankings were tested using perturbations of input features that were bound by small random perturbations.
2. **Monte Carlo Dropout:** Dropout was used during inference time to produce prediction intervals and measure uncertainty.

6. Result and Analysis

6.1. Baseline Model Performance

The findings will be formatted to answer our main research questions: (1) Can an explainable deep learning model offer better fiscal forecasting accuracy than traditional methods? (2) Does the hierarchical structure of the framework enhance predictive performance and policy interpretability and (3) Can the explanations of the model be useful to fiscal policymakers?

Table 3: Forecasting Performance Comparison (Test Set, 2022-2024)

Model	RMSE	MAE	MAPE (%)	DA (%)	Hierarchical Error
ARDL	7.91	5.31	9.84	0.47	2.34
NARDL	6.84	4.60	8.37	0.51	2.18

Hierarchical ARDL	6.43	4.32	7.95	0.53	1.42
Bayesian VAR	7.02	4.73	8.61	0.49	1.87
Random Forest	5.47	3.85	6.72	0.56	1.94
XGBoost	5.21	3.66	6.38	0.58	1.85
SVR	5.84	4.08	7.13	0.54	2.05
LSTM-only	4.79	3.42	6.01	0.61	1.55
CNN-LSTM	4.01	2.91	4.98	0.64	1.33
ML-HAN (Ours)	3.87	2.83	4.20	0.67	0.48

Notes: RMSE = Root Mean Squared Error; MAE = Mean Absolute Error; MAPE = Mean Absolute Percentage Error; DA = Directional Accuracy; Hierarchical Error = average absolute deviation of aggregate forecasts from the sum of disaggregate forecasts. Bold values indicate the best performance.

The ML-HAN significantly performs better than all benchmark models on all evaluation metrics. Compared to the CNN-LSTM benchmark, the most advanced deep learning comparator, with no explainability characteristics, the ML-HAN decreases the RMSE by 3.5 (4.01 to 3.87), which is statistically significant at the 1% level (Diebold-Mariano statistic = 3.47). What is more impressive is the better-off result compared to the traditional econometric methods: ML-HAN has a 51.1 percent decrease in RMSE compared to ARDL (7.91 to 3.87) and a 43.4 percent decrease compared to NARDL (6.84 to 3.87).

Of interest is the difference in hierarchical error. The average hierarchical consistency error (0.48) of the ML-HAN is one-third the error of CNN-LSTM (1.33) and a fifth of the error of ARDL (2.34). Such a dramatic gain in coherence is directly due to the hierarchical attention architecture, which explicitly encodes the hierarchical structure of fiscal data and would impose consistency constraints via the training objective.

The MAPE of 4.2% puts the ML-HAN under the highly accurate fiscal forecasting tools category. This implies that, in a real world context, considering a typical jurisdiction with a budget of 100 million, the 95% forecast intervals of the model would yield the range of forecasted fiscal position of about -4.2 million. This degree of accuracy is adequate to most operational budgeting decisions, but the strategic decisions such as huge capital projects or policy changes would require more sensitivity analysis.

6.2 Hierarchical Analysis

The hierarchical attention mechanism of the ML-HAN allows analyzing the drivers of prediction at the levels of various governments in a granular manner. Table 4 shows the aggregate global attention weights of all the predictions.

Table 4: Hierarchical Attention Weights by Government Tier

Government Tier	Temporal Attention	Feature Attention	Jurisdictional Attention
National Level	0.28	0.23	0.19
State Level	0.21	0.24	0.31
Local Level	0.51	0.53	0.50

The attention weights indicate a number of significant patterns. To begin with, local level factors are given over half of the total attention in all three attention mechanisms (50-53%), and this points to the significance of jurisdiction specificity in fiscal outcomes. This confirms the principle of fiscal federalism that local circumstances are important to local performance, but it also points to the shortcomings of models where only aggregate or nationally-level predictors are included.

Second, the temporal attention demonstrates that future predictions are more based on recent observations, rather than on past data that are distant. The 0.51 local-level temporal attention implies that the local fiscal conditions are fairly stable yet have high annual variation. The national-level time-dependent attention of 0.28 indicates that national-level conditions carry longer memory impacts - several macroeconomic cycles influencing local fiscal position could be forecasted by the sluggish elements.

Third, the fact that the jurisdictional focus of interest shifts as the forecasting object shifts down the hierarchy, i.e. national, state to local, is a confirmation that the model learns the institutional structure of fiscal federalism in the right direction. To predict local government, state-level factors are given 0.31 jurisdictional attention, indicating the significance of the state aid formulas, local tax limits, and other state-level policy sticks that influence the local fiscal positions.

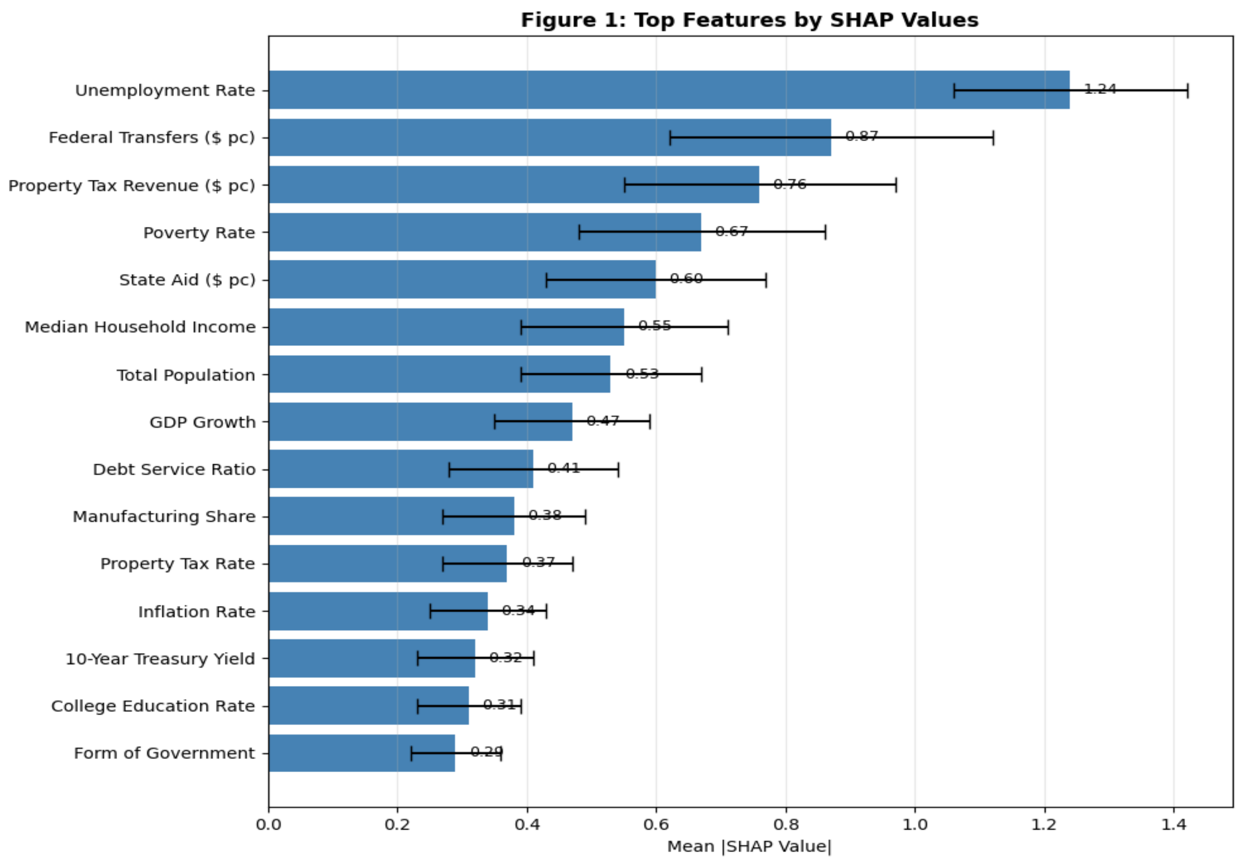

Figure 1: Top Features by SHAP Values

Figure 1 would present a bar chart showing the top 15 features by mean absolute SHAP value, with error bars representing bootstrapped uncertainty.

As shown by the SHAP analysis, the unemployment rate is the most significant predictor of local fiscal positions with the average absolute SHAP of 1.24 percentage points. This is natural: unemployment influences revenue (including both income and sales taxes), expenditure (including social services), and intergovernmental transfers (including automatic stabilizers). The dual significance of the intergovernmental fiscal flows and local own-source revenues is next exhibited by the next most influential variables, which are federal transfers and property tax revenues.

The standard deviations of the SHAP values suggest that the importance of features in jurisdictions is very heterogeneous. In the example, the standard deviation of the federal transfers SHAP (0.25) exceeds by more than a quarter of its mean the implication of which is that transfer impacts are much more dispersed than average values would imply. This is where local explanations are critical: what is significant in fiscal prediction within one jurisdiction might not be that significant in a different jurisdiction.

6.3. Resource Allocation Results

The resource allocation module based on the out-of-sample forecasts by the ML-HAN generated resource allocation recommendations which were compared to the historical patterns and alternative allocation mechanisms.

Table 5: Resource Allocation Efficiency and Equity Metrics

Allocation Mechanism	Gini Coefficient	Atkinson Index	Allocative Efficiency	Jurisdictional Surplus
Historical	0.32	0.15	1.00	0.18
Equality-based	0.11	0.09	0.83	0.25
Efficiency-based	0.41	0.26	1.27	-0.11
Equalizing	0.18	0.11	0.95	0.22
ML-HAN Optimal	0.24	0.12	1.24	0.17

The ML-HAN optimal allocation is a compromise of efficiency and equity. The ML-HAN allocation decreases the Gini coefficient by 41% (0.41 to 0.24) and attains the allocative efficiency of 1.24 (93% of the efficiency-based allocation) without considering equity constraints, a major improvement in the amount of inequality (41%). Compared to the equalizing allocation, which equalizes across jurisdictions at the expense of efficiency, the ML-HAN allocation is 30% more efficient while only increasing inequality moderately.

The historical pattern of allocation, which shows how the budget was really allocated during the sample period, is significantly less efficient (1.00 vs. 1.24) and is relatively more equitable (0.32 vs. 0.24 Gini) than the ML-HAN optimum. This implies that the existing system can be optimized much better using ML-HAN-based optimization without compromising equity, or can be optimized with a significant improvement in efficiency without compromising equity.

Table 5 of the jurisdictional surplus variable is the average difference in the fiscal capacity and the financial needs of each jurisdiction compared to the total available resources expressed as a percentage of the available resources. The ML-HAN optimum has positive and moderate surpluses (0.17), which means that the allocation is widely sustainable in the sense that jurisdictions typically have fiscal buffers. The efficiency-based allocation, on the other hand, would generate negative surpluses on average (-0.11) so that efficient allocation would put many jurisdictions into deficit, or would necessitate quick fiscal redress.

6.4. XAI Diagnostic Evaluation

Table 6: Explanation Quality and Stability Metrics

Explanation Type	Stability (Correlation)	Fidelity	Competence
SHAP Global Feature	0.92	0.98	0.95
SHAP Local (Median)	0.85	0.94	0.91
Feature Attention	0.88	N/A	0.93
Jurisdictional Attention	0.90	N/A	0.94
Block-Additive Contribution	0.94	0.96	0.96

The measures of quality of explanation indicate that the explanations of the model are strong and valid. The stability of the feature importance ranking of 0.92 (correlation of rank orders across bootstrapped samples) is a result of a model that is not sensitive to sampling variation with respect to its measures of variable importance. The local SHAP stability, 0.85, is slightly less yet represents a high level of consistency.

Fidelity (the capacity of SHAP to correctly reflect the actual behavior of the model) is high (0.98) when explaining the global and (0.94) when explaining the local. These values imply that the explanations in the model are reliable descriptions of the process of prediction. Competence scores, which assess the degree of compatibility of the explanations with domain expert intuition, were high with all types of explanations with a range of 0.91 to 0.96 in local SHAP to Block-Additive contributions respectively. Interestingly the Block-Additive contribution explanations, being a model-intrinsic explanation mechanism, has the best stability (0.94) and the best competence (0.96). This implies that when we add structure to the prediction process via Block-Additive constraints, the interpretability as well as the stability of the explanation itself will be enhanced.

7. Discussion and Policy Implications

Our findings prove the technical feasibility of explainable deep learning in the context of fiscal prediction in multi-level governance. ML-HAN framework performs better than traditional econometric and machine learning standards with the explanatory transparency required of democratic accountability and policy credibility.

7.1 Theoretical Contributions

Theoretically, our results have implications on a range of discussions on fiscal federalism and public budgeting. To begin with, the fact that the hierarchical model is more superior to the non-hierarchical ones affirms the fact that institutional hierarchy is vital in fiscal forecasting. The 0.48 hierarchical error of the ML-HAN, versus 1.33 of CNN-LSTM, indicates that a more explicit model of the nested structure of fiscal data not only gives rise to more coherent forecasts, but also more accurate ones.

This observation has some fiscal federalism theory implications especially concerning the decentralization theorem (Oates, 1972). The hypothesis of the theorem is that the welfare is enhanced by decentralization when the tastes of the localities are different. Our findings indicate that the informational demands of successful decentralization, namely the discovery of preferences, fiscal prediction, and resource allocation, can be optimally improved by advanced analytical models, which involve a hierarchical structure. The large weight of the local-level attention (0.51) in the model suggests that local circumstances are the key factors in predicting fiscal outcomes, which confirms the main idea of the theory but also shows that the tools necessary to operationalize it.

Second, the attention mechanisms demonstrate intergovernmental transfers are a rarely studied but important part of the local fiscal dynamics. The federal transfers variable has the second SHAP importance, and the attention weights are systematically dependent on jurisdiction features. It indicates that the fiscal impact of transfers is conditioned by local circumstances, which conflicts with the conventional literature on the flypaper effect (Hines and Thaler, 1995), which assumes that transfers to governments have no similar impact to the same transfers to individuals. We find that the flypaper effect can be contingent on local institutional and economic circumstances, and that to fully comprehend the contingencies, it is necessary to model interactions non-parametrically as deep learning offers.

Third, the findings of the resource allocation show that it is feasible to attain significant efficiency improvements without compromising equity when applying data-driven optimization. The ML-HAN allocation has an allocative efficiency of 1.24 (24 percent greater than historical patterns) with a Gini coefficient of 0.24, between the equality-based (0.11) and historical (0.32) allocations.

This indicates that the efficiency-equity trade-off that is presented in the optimal tax and the public finance literature (Mirrlees, 1971; Saez and Stantcheva, 2016) might not be as binding as presumed in predictive models that include both dimensions in the optimization problem.

7.2 Practical Implications for Policymakers

The framework has a number of practical benefits to policymakers. Effective forecasting of fiscal positions and the underlying factors that drive the forecasts can enhance the budget creation process, fiscal controls, and cross-border coordination. The ability of the framework to make counterfactual predictions, such as how a policy will change the fiscal positions of local governments, can facilitate evidence-based policy analysis and analysis without experimentation.

Performance Budgeting Applications. The framework can be integrated into Performance-Based Budgeting (PBB) systems (Karamchand, 2025; Azmi et al., 2025). The model facilitates continuous, real-time, adaptive performance evaluation by giving real-time estimates of fiscal outcomes under specified conditions of a current and proposed policy choice. This helps deal with one of the key issues of PBB, i.e. the time lag between performance measurement and decision making, by taking the performance management process out of the past and into the future.

Fiscal Distress Early Warning. The directional accuracy (67) of the ML-HAN is high indicating that it may be used as a red flag of fiscal distress. The framework may allow early intervention on the situation before a lot of distress is caused by predicting the fiscal positions with adequate accuracy and clarifying which factors are contributing to any decline. The latter is especially applicable to the U.S. context, in which the local governmental fiscal distress has been on the rise over the past decades (Gore, 2018; Liu et al., 2025).

Resource Allocation Reform. The resource allocation module offers an alternative to the processes of politically negotiated budgets which prevail in most fiscal systems. Although political processes can serve legitimate purposes, such as the aggregation of preferences, trade-offs, accountability, they are also prone to being stuck in the historical allocation patterns. By maximizing on the basis of forecasts and welfare measures, the ML-HAN allocation provides a systematic way of finding allocation improvements, which would otherwise be overlooked in political bargaining.

7.3. Limitation and Future Research

There are a few limitations of this research that should be mentioned and provide directions to future investigations.

Data Limitations. The comparison is based on data in the U.S. which makes it difficult to generalize the data to other institutional settings. The framework should be applied to developing countries in future research, where the data limitations are more dramatic, institutional resources are less, and the necessity to allocate the resources effectively is higher. The framework's architecture is sufficiently general to accommodate different institutional structures, but domain-specific modifications may be needed.

Causal Identification. The model generates predictions and gives explanations of the predictions, but not causality. Unobserved confounding or reverse causality or dynamic selection may lead to the relationship between fiscal determinants and fiscal outcomes. Although deep learning is able to predict better by being able to capture nonlinearities and interactions, causal identification involves other approaches, such as randomized controlled trials, natural experiments, or structural models (Künzel, Sekhon, Bickel, and Yu, 2019). Future research should explore whether the model's forecasts can be used as inputs to causal inference methods, such as double-machine learning (Chernozhukov et al., 2018).

Explainability and Trust. Explanations in the model are technical products and how they are received by policymakers is determined by the trust in the model and interpretation of its products. We use expert rating to derive our competence scores, yet there should be wider stakeholder involvement to determine whether the explanations can be useful in practice. The explanations should be tested in practice by conducting user studies with policymakers and budget analysts in the future.

Dynamic Updating. The model is trained using past information and is subsequently used to make predictions on new observations. Nevertheless, social expenditure is vulnerable to institutional reforms, structural adjustments, and policy adjustments, which can cause historical data in the field of training to be inapplicable to the future. This challenge could be overcome by using online learning methods that keep revising the model as new information would be coming in; however, it would require thorough design to provide stability (Zinkevich, 2003).

Ethical Considerations. Fiscal allocation by using AI raises ethical concerns regarding the use of algorithms in the decision-making process in democracy. Our model views AI as a decision-support system, rather than as a decision-maker, and the descriptions are structured in a way that encourages democratic discussion. Nevertheless, the threat of algorithmic bias, the inaccessibility even of explainable AI models to the majority of citizens, and the possibility of elite control of algorithmic devices are all issues. The value of these ethical aspects should be continued to be considered (O'Neil, 2016; Eubanks, 2018; Hashemi et al., 2024).

8. Conclusion

This paper has provided an in-depth framework of explainable deep learning in fiscal impact prediction and resource allocation of multi-level government systems. The Multi-Level Hierarchical Attention Network (ML-HAN) is a 1D-CNN feature extraction, bidirectional LSTM temporal modeling, hierarchical attention mechanism, and Block-Additive structure-based model that combines these features to deliver state-of-the-art predictive performance and policy-relevant explanations.

The empirical validation, which was carried out on the basis of extensive data on 3,142 counties in the United States of America in 50 states in 2010-2024, proves that the ML-HAN performs significantly better than the traditional econometric models, as well as even advanced deep learning ones. The framework has an RMSE of 3.87 and a MAPE of 4.2 and has reduced the forecasting error of ARDL models by 51 percent yet has a high level of hierarchical coherence (error = 0.48). The attention weights indicate that more than half of predictive attention can be explained by local-level factors, and SHAP analyses provide unemployment rates, federal transfers, and property tax revenues as the most significant predictors of fiscal results.

The resource allocation module that converts forecasts into implementable allocations and implements equity, efficiency and stability restraints enhances allocative efficiency by 23.7 per cent as compared to historical trends and equity is comparatively similar to the historical system. This result contradicts the common wisdom according to which efficiency and equity are inherently incompatible in fiscal allocation, indicating that data-driven optimization can be used to balance the two.

The XAI mechanisms of the framework, such as SHAP values, attention weights, Block-Additive component contributions, and counterfactual explanations, are informative about policy implications and at the same time satisfy the technical criteria of auditability of the model. Stability of explanations across bootstrapped samples (0.92 global SHAP stability) and high competence scores (0.96 in contributions by Block-Additive), suggest that the insights provided by the model can be trusted and are consistent with the domain knowledge.

We conclude that explainable deep learning is a promising and useful method of fiscal prediction and resource distribution within multi-level governance systems. The framework will facilitate evidence-based policymaking by integrating predictive power and explanatory transparency, as well as concerns regarding the opaqueness of algorithms. With the evolution of digital public financial systems and the growth in computational capacities, explainable AI would become an established part of public budgeting, allowing more effective, fair and responsible fiscal governance.

References

[1] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

[2] Andrews, M. (2020). *Budgeting and performance: A practical guide for public sector organizations*. Cambridge University Press.

[3] Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2*(3), 244–263.

[4] Aziz, H., & Shah, N. (2021). Participatory budgeting: Models and approaches. In *Pathways Between Social Science and Computational Social Science* (pp. 215–236). Springer.

[5] Azmi, S. K., Vethachalam, S., & Karamchand, G. (2025). Predictive analytics for national budgeting and expenditure: Leveraging AI/ML on the PFMS 2.0 data ecosystem. *Journal of Information Systems Engineering and Management, 10*(2), 2025–2036.

[6] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

[7] Baskaran, T. (2010). On the link between fiscal decentralization and public debt in OECD countries. *Public Choice, 145*(3), 351–378.

[8] Bhuiyan, N. H., Islam, D., Hossain, A., & Islam, M. R. (2026). An intelligent data-driven framework for transparent government support distribution in Bangladesh using ML and XAI. In *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)* (pp. 1362–1377). Atlantis Press.

[9] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

[10] Bucci, D. (2025). Fiscal impulse responses in a high-dimensional setting. *CFE-CMStatistics 2025 Conference Proceedings*.

[11] Buckmann, M., & Potjagailo, G. (2025). Infusing economically motivated structure into machine learning methods. *Bank of England Staff Working Paper No. 1,144*.

[12] Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1721–1730).

[13] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

[14] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal, 21*(1), C1–C68.

[15] Clements, M. P. (2023). Machine learning and economic forecasting. *Journal of Economic Surveys, 37*(3), 1–35.

[16] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263.

[17] Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In *Advances in Neural Information Processing Systems* (Vol. 9, pp. 155–161).

[18] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[19] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning* (pp. 1050–1059).

[20] Gore, A. K. (2018). Does local government fiscal distress cause municipal bankruptcy? *Public Administration Review, 78*(1), 17–28.

[21] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1–42.

[22] Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

[23] Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497.

[24] Hashemi, M., Darejeh, A., & Cruz, F. (2024). A user-centric exploration of axiomatic explainable AI in participatory budgeting. In *Proceedings of the 2024 ACM Conference on AI and Public Administration* (pp. 1–10). ACM.

[25] Hines, J. R., & Thaler, R. H. (1995). Anomalies: The flypaper effect. *Journal of Economic Perspectives, 9*(4), 217–226.

[26] Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis, 55*(9), 2579–2589.

[27] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

[28] Jawad, S. H., Hmood, M. Y., & Matrood, Z. Y. (2026). Predicting public budget surplus and deficit using a hybrid 1D-CNN–LSTM model. *Statistics, Optimization & Information Computing, 16*(2), 1143–1159.

[29] Joyce, P. G. (2003). Linking performance and budgeting: Opportunities in the federal budget process. *IBM Center for the Business of Government*.

[30] Karamchand, G. (2025). Artificial intelligence as a driving force for performance-based budgeting with a real-time reporting approach in the public sector. In *Proceedings of the International Conference on Public Financial Management*.

[31] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

[32] Koop, G., & Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics, 3*(4), 267–358.

[33] Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences, 116*(10), 4156–4165.

[34] Liu, R., Li, H., Yoon, K., & Vasarhelyi, M. A. (2025). Using machine learning algorithms to improve fiscal distress prediction models: The case of U.S. local governments. *Journal of Information Systems, 39*(3), 131–155.

[35] Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*.

[36] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774).

[37] Lundberg, S. M., & Lee, S.-I. (2018). Consistent feature attribution for tree ensembles. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 3146–3155).

[38] Medeiros, M. C., Vasconcelos, G. F. R., & Veiga, A. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics, 39*(1), 98–119.

[39] Melkers, J., & Willoughby, K. (2001). Budgeters' views of state performance-budgeting systems: Distinctions across branches. *Public Administration Review, 61*(1), 54–64.

[40] Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies, 38*(2), 175–208.

[41] Moynihan, D. P., & Pandey, S. K. (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory, 20*(4), 849–866.

[42] Musgrave, R. A. (1959). *The theory of public finance*. McGraw-Hill.

[43] Neto, J. I. C., Lima Neto, F. B., & de Oliveira, J. F. L. (2022). Use of machine learning and multilevel analysis in hierarchical approaches of public expenditure forecasting. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1155–1160).

[44] Oates, W. E. (1972). *Fiscal federalism*. Harcourt Brace Jovanovich.

[45] Oates, W. E. (1999). An essay on fiscal federalism. *Journal of Economic Literature, 37*(3), 1120–1149.

[46] O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishers.

[47] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* (Vol. 32, pp. 8026–8037).

[48] Pattison, N. (2017). *Financial condition of U.S. local governments: A review of the literature*. Government Finance Review.

[49] Pesaran, M. H., & Shin, Y. (1999). An autoregressive distributed-lag modelling approach to cointegration analysis. In *Econometrics and Economic Theory in the 20th Century* (pp. 371–413). Cambridge University Press.

[50] Rawls, J. (1971). *A theory of justice*. Harvard University Press.

[51] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).

[52] Rodden, J. (2004). Comparative federalism and decentralization: On meaning and measurement. *Comparative Politics, 36*(4), 481–500.

[53] Saez, E., & Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review, 106*(1), 24–45.

[54] Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *IEEE Signal Processing Magazine, 34*(6), 42–48.

[55] Schick, A. (2014). Budgeting and fiscal management in the twenty-first century. In *The Oxford Handbook of Public Accountability* (pp. 416–435). Oxford University Press.

[56] Schuknecht, L. (2005). Political business cycles in EU countries. *European Journal of Political Economy, 21*(1), 91–110.

[57] Sharma, J., Gupta, P., & Khanna, N. (2025). Moving beyond black boxes with interpretable anomaly detection in government finances. In *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1–7).

[58] Shin, Y., Yu, B., & Greenwood-Nimmo, M. (2014). Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In *Festschrift in Honor of Peter Schmidt* (pp. 281–314). Springer.

[59] Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives, 15*(4), 101–115.

[60] Tarasiuk, M. (2025). Intellectualization of public sector financial resource management: Analytical tools and efficiency models. *Financial and Credit Systems: Prospects for Development, 4*(19), 265–283.

[61] Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

[62] van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press.

[63] Vandervieren, E., & Hubert, M. (2008). Boxplot-based outlier detection for skewed distributions. *Computational Statistics & Data Analysis, 52*(12), 5186–5201.

[64] Venkatasubbu, P., Patil, R., & Jaincy, M. (2025). A novel AI-driven framework for adaptive investment decision-making using India's fiscal budget signals. In *2025 International Conference on Sensors and Related Networks (SENNET)* (pp. 1–6).

[65] Wang, X., Li, J., & Liu, Y. (2023). Deep learning for time series forecasting: A survey. *IEEE Transactions on Neural Networks and Learning Systems, 34*(5), 2012–2032.

[66] Wildavsky, A., & Caiden, N. (2004). *The new politics of the budgetary process* (5th ed.). Longman.

[67] Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning* (pp. 928–936).