Ensemble of Temporal Weighting, Causal Inference, and Hierarchical Attribution towards SHAP Optimization

Archana Salaria[1], **Manik Rakhra**[1], **Nonita Sharma**[2], **Monika Mangla**[3], **Fernando Moreira**[4], **Nishan Singh Bala**[5]

[1] School of Computer Science Engineering, Lovely Professional University [2] Department of Information Technology, Indira Gandhi Delhi Technical University for Women [3] Department of Information Technology, Dwarkadas J. Sanghvi College of Engineering [4] REMIT, IJP, Universidade Portucalense [5] Oakville Trafalgar High School, Halton District School Board

Corresponding Author

Monika Mangla
monika.mangla@djsce.ac.in

Citation

Salaria, A., Rakhra, M., Sharma, N., Mangla, M., Moreira, F., Bala, N.S. Ensemble of Temporal Weighting, Causal Inference, and Hierarchical Attribution towards SHAP Optimization. *J. Vis. Exp.* (225), e69125, doi:10.3791/69125 (2025).

Date Published

November 18, 2025

DOI

10.3791/69125

URL

jove.com/video/69125

Abstract

During the past few years, the need for transparency and interpretability has been intensified owing to significant advancements in data-driven models, leading to the emergence of Explainable Artificial Intelligence (XAI). Several traditional XAI approaches are prevalent; however, these have limited competence in interpreting dynamic relations. The current research aims to address this limitation by proposing a novel Ensemble SHapley Additive exPlanations (SHAP) framework that focuses on temporal weighting, causal inference, hierarchical attribution, and interpretability optimization referred to as TCHSHAP. TCHSHAP prioritizes current information over historical information by temporal weighting through exponential decay. Further, causal inference separates correlation from causality to gain practical insights. Additionally, hierarchical attribution allows insights at granular (region level) and aggregated levels (feature-group impacts). These approaches are integrated to achieve a more interpretable and explainable model. To validate the efficacy of the proposed model, we carry out an experiment on the crop yield dataset collected from Kaggle. Ahead of experimental evaluation, data preprocessing is performed using one-hot encoding. Data normalization is done by min-max scaling, and outliers are removed through the Interquartile range. For the sake of experimental evaluation, the authors used the SHAP XAI model for Random Forest. When assessing the efficacy of the proposed TCHSHAP model, it is observed that while the average prediction for traditional SHAP is 161.137, it escalates to 161.506 after incorporating temporal weighting and causal inference, advocating the effectiveness of employing temporal and causal significance. Additionally, during hierarchical attribution, it is observed that agricultural features have the strongest dominance over the target variable. This dominance is followed by geographical and environmental factors in order. Thus, the obtained results authorize the efficacy of the proposed approach towards enhancing the global and local interpretability, strengthening the user's trust in model predictions. The current work offers ways to improve transparency and interpretability without affecting model performance. The suggested model also enables interpretable and efficient regression modelling in complex, data-driven applications, enabling its widespread application in real-world settings.

Introduction

Regression plays a major role in predictive analytics, encompassing several domains including climate modelling, financial forecasting, healthcare diagnostics, and agricultural yield estimation[1,2]. However, the unexplained results of regression models, especially non-linear and high-capacity models like Gradient Boosting Machines (GBMs), Random Forests, and Deep Neural Networks (DNNs), pose significant challenges, particularly for applications requiring transparency and actionable insights. This interpretability gap can be bridged by employing Explainable Artificial Intelligence (XAI), a technique that is still in its infancy. Currently, conventional XAI frameworks like SHAP and Local Interpretable Model Agnostic Explanation (LIME) give only static feature explanations[3]. These XAI techniques presently fail to explain the intricacies of regression tasks involving temporal dependencies, hierarchical feature structures, and causalities, creating an avenue of research[4,5]. Due to the non-consideration of these factors, XAI may sometimes derive incorrect conclusions. For instance, traditional SHAP only considers the correlation of attributes with the target variable without considering the causality. Let us consider a dataset with two input binary features, namely tobacco consumption and stained teeth. The target variable in this dataset is the probability of cancer. In such a scenario, if correlation is obtained between input features and the target variable, cancer will be highly correlated with both features (tobacco consumption and stained teeth). Now it is quite misleading to conclude that stained teeth cause cancer, as it is caused by regular tobacco consumption. Thus, it can be summarized that despite significant progress in the domain of XAI, the prime issues that need to be addressed in regression contexts include:

Inability to consider temporal dependencies leads to only static explanations that may not work in contexts with dynamic data.

Non-inclusion of causal inference in traditional SHAP adaptations causes false correlations rather than valuable causal insights.

Lack of attention towards hierarchical interpretation leads to insufficient higher-level aggregated explanations.

In order to address these issues, current research work aims to propose a novel XAI framework, TCHSHAP, that encompasses temporal weighting, causal inference, and hierarchical attribution with an aim to enhance the interpretability and computational efficiency of current XAI models. For the same, the proposed framework considers a temporal weighting factor w_t which assigns higher weights to the recent feature contribution in comparison to past values. This weighting factor aims to ensure that the interpretability aligns with the temporal significance of crucial features in dynamic regression problems such as yield forecasting and sales prediction. Further, traditional regression models fail to exhibit causal relationships, leading to ambiguous attributions[6]. To address this challenge, the proposed framework adopts Structural Causal Models (SCMs) to assess the causal impact of a feature F_i on the output. Additionally, the real-time datasets often demonstrate hierarchical relations (spatial or categorical) among features, which are seldom showcased in existing XAI frameworks[4,5,6]. On the contrary, the suggested XAI framework aggregates individual features (F_i) to obtain higher-level features (H_j), to achieve granularity of the explanations.

Apart from incorporating temporal, causal, and hierarchical significance, the proposed ensemble also aims to enhance interpretability and explainability. Such an objective is an effort towards establishing the trade-off between interpretability and accuracy, ensuring that the explanations obtained from TCHSHAP are accurate and interpretable[4]. The suggested XAI framework can be mathematically described as follows, while the description of the used variables is given in **Table 1**. The traditional SHAP value for the feature is calculated as equation (1)[5,6].

$$Shap_i = \sum_{M \subseteq S} \frac{|M|!(T-|M|-1)!}{F_T!}(P(M \cup \{i\}) - P(M)] \quad (1)$$

This traditional SHAP formulation adds up contributions without considering temporal, causal, or hierarchical adjustments.

Variable	Definition
F_i	Feature whose Shap value is being calculated
$Shap_i$	Shap value of the Feature
S	Set of total Features
M	Subset of all features
F_T	Total number of Features
$P(M)$	Model Prediction when only features in M are used
$\lambda > 0$	decay rate
T	Reference Prediction Time
Y	Output

Table 1: Description of variables used. This table describes the significance of various variables used.

This traditional SHAP formulation is modified to handle dynamic dependencies by including temporal weighting. This temporal weighting is incorporated by introducing the weighting coefficient $wt = e^{-\lambda |t - T|}$ where λ represents decay factor ($\lambda > 0$) . As discussed earlier, this exponential weighting gives more weight to recent values compared to past values. This exponential decay is included with an understanding that recent feature values have higher significance than past values. The modified SHAP value (after incorporating temporal significance) is given in equation (2).

$$Shap_{temporal} = w_t . \sum_{M \subseteq S} \frac{|M|!(T-|M|-1)!}{F_T!}(P(M \cup \{i\}) - P(M)] \quad (2)$$

This temporal information is important for dynamic regression tasks like predicting yields or stocks.

Further, in order to make the explanations more robust, SCM is applied to assess the causal impact of the feature F_i on the output Y. The mathematical formulation for causal SHAP is given in equation (3)

$$Shap_{casual} = \sum_{M \subseteq S} \frac{|M|!(T-|M|-1)!}{F_T!} \mathbb{E}[(P(M \cup \{i\})|do(F_i)] - \mathbb{E}[P(M)|do(F_i)]] \quad (3)$$

Here *do*(F_i) represents the feature that has contributed to the change. By using this, the model achieves the competence to explain the causality rather than the correlation. It is worth noting that SCM-based computation of causal effects ensures that only features with direct causal influence are given credit, eliminating false correlations.

Further, the hierarchy is proposed in the suggested model to aggregate attributes across various levels. The mathematical formulation for the hierarchical SHAP is given in equation (4).

$$Shap_{hierarchical} = \sum_{j \epsilon H_j} Shap_{casual_j} \quad (4)$$

Incorporation of hierarchy in the suggested model ensures the achievement of granularity of the explanations. The suggested framework is depicted in **Figure 1**.

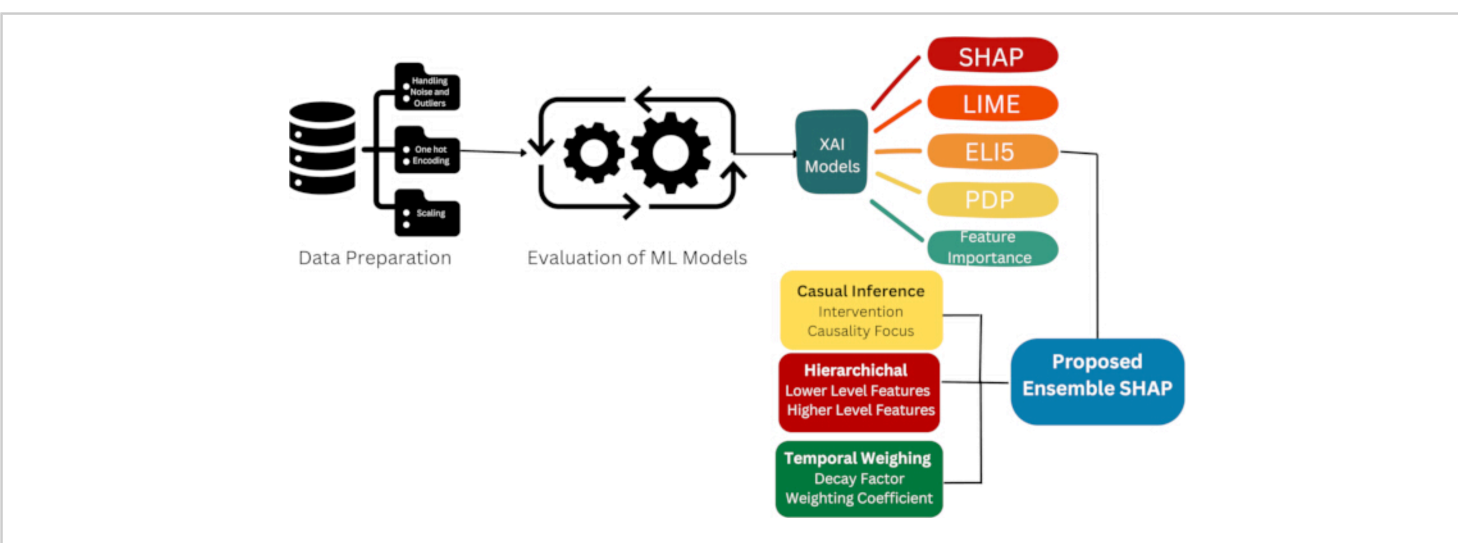

Figure 1: Pictorial illustration of various elements in the proposed framework. This figure gives a pictorial illustration of various elements in the proposed framework TCHSHAP that include temporal weighting, causal inference, and hierarchical attribution. Please click here to view a larger version of this figure.

The novelty of the proposed model TCHSHAP lies in handling these three inconsistencies by employing temporal weighting (to give more weight to current information), Structural Causal Models (SCM; to assess causal effects), and hierarchical aggregation of SHAP values (to gain multi-level insights). Thus, the proposed ensemble expands SHAP beyond its static constraints, providing a regression-optimized explanatory framework.

Proposed TCHSHAP has the potential to achieve great significance for real-time applications in scenarios requiring better decision-making, as it identifies the detailed explanations of the output in regression models, seamlessly connecting modelling with decision-making. In summary, the prime contribution of this manuscript is to design a novel framework that combines temporal weighting, causal inference, and hierarchical relations to enhance the actionable explanations of the traditional SHAP model.

Related work

XAI incorporates different methodologies specially designed for enhancing the transparency and interpretability of different machine learning models. These XAI strategies establish confidence in the results by providing the required insights from the results. For instance, notable XAI methodologies such as LIME and SHAP enhance the interpretability and transparency of otherwise black box models across various domains, namely finance, education, and the healthcare sector[7]. Also, XAI techniques are applied in tasks related to software analytics, clone detection, and Just-in-Time defect prediction[8]. Additionally, Awal and Roy also highlighted the discrepancies and demonstrated lesser reliability in generating the explanations[8]. Here, the authors applied Granular Level Evaluation Metrics for reducing consistency in evaluation. XAI has also been widely employed in the medical domain to understand the most prominent factors affecting the results of the classification task and hence fostering the faith among health practitioners to use AI-based diagnostics[9,10,11,12]. Further, in sectors like Belgian residential rent prediction, XAI approaches facilitate the interpretation of intricate models, hence enhancing quality control and minimizing production errors[13]. Bommer et al. demonstrated that integration of XAI methods with gradients enhances the performance of XAI methods in terms of robustness, faithfulness, complexity, and randomization in climate science[14]. A lot of researchers have attempted to create ensemble and hybrid techniques. In addition to proposing ensemble techniques, researchers also demonstrated the effectiveness of these hybrid techniques towards providing comprehensive insight for complex models[15,16]. A detailed review of the techniques, optimizations, and the various applications are tabulated in **Table 2,** which clearly indicates that the most widely applied techniques are LIME and SHAP. The future research direction points towards exploring optimization in these methods to enhance the interpretability of these models. Continuing the line of research, authors in current research work focus on SHAP optimization by adding causal inference, hierarchical attribution, and temporal weighting.

Citation	XAI Techniques	Optimization	Insights	Applications	Future Research Directions

(Anushree et al., 2024)	LIME SHAP Perturbation mechanisms Attention-based mechanisms	Not provided	Enhancement of interpretability and transparency of black box models across various applications in Finance, education and healthcare sector.	· Trade-off between Accuracy and Interpretability in Finance, Education and healthcare sector	· Need to develop a list of detailed metrics for evaluating various XAI models
(Awal & Roy, 2024)	PyExplainer LIME	Granular Level Evaluation Metrics for reducing consistency in evaluation.	Highlights the discrepancies and demonstrates lesser reliability in generating the explanations.	· Tasks related to Software Analytics · clone detection · Just in Time Defect Prediction	· Advance research methods are required for enhancing the reliability of XAI models for tasks related to software analytics
(Bommer et al., 2024)	Integrated Gradients layer-wise relevance propagation input times gradients sensitivity methods like gradient, SmoothGrad, NoiseGrad, and FusionGrad	Integration of XAI methods with Gradients	Integration of XAI methods with gradients enhances the performance in terms of robustness, faithfulness, complexity, and randomization in climate science.	· Climate Science Prediction · Annual mean Temperature maps prediction	· Focus on task specific evaluation of climate science.
(Bhatnagar & Agrawal, 2024)	LIME SHAP	Ensemble of XAI Techniques	Ensemble provides comprehensive insight and enhance interpretability and explainability of complex models.	· Applied on Moral dataset to enhance the interpretability of the results	· Explore hybrid XAI methods by combining multiple techniques · Customizing various XAI algorithms for specific models catering to different needs of the users
(Dias, 2024)	LIME Eli5	Integration of XAI Techniques	Improved Interpretability and Accuracy	· Toxic Comment Classification	· Enhancing the strength for classification purpose.

(Hajare et al. ,2024)	SHAP	Not Provided	SHAP offers detailed insights into the risk factors in predicting the Acute Coronary Syndrome.	· Acute Coronary Syndrome (ACS)	· To find out novel risk factors in ACS prediction.
(Hamida et al., 2024)	Comprehensive review of various XAI Techniques	Not Provided	Provides insights on XAI applications XAI in the healthcare sector emphasizing the importance of explainability of AI decisions to understand the results provides by AI.	· Actionable Insights in Healthcare sector	· Enhancement of XAI components namely comprehensibility, transparency, interpretability, and explainability. · Custom XAI techniques for healthcare sector for detail insights for healthcare domain
(Ihongbe et al., 2024)	Grad-CAM (Gradient-weighted Class Activation Mapping)	Gradient-weighted Class Activation Mapping	Better accuracy for Pneumonia and COVID-19 diagnosis.	· Improved Accuracy for medical diagnosis	To increase the awareness of the usability of XAI techniques for real life applications
(Lenaers, & De Moor, 2023).	6 XAI Techniques	Ensemble of 6 XAI Techniques on CatBoost Model	Multiple techniques enhance the interpretability of XAI for rent prediction.	· Belgian residential rent prediction	XAI can be utilized for decision making
(Makumbura et al., 2024)	SHAP	Ensemble of RF, LightGBM, XGBoost with SHAP	SHAP reveals the various factors responsible for assessing and predicting the water quality.	· Assessment and prediction of water quality	Can be employed for decision making
(Schlegel & Keim, 2023).	XAI Techniques with Perturbation Analysis	Perturbation Analysis	Can be effectively employed for classification tasks of Time Series Data	· Time Series Data of Finance, Healthcare, Climate science	To combine multiple XAI techniques for enhanced interpretability.
(Silva & Keller, 2023)	XAI Biomolecular Reaction Rate	Not Provided	XAI models have a limitation to explain the results in case of correlated features. Can	· Atmospheric sciences · Atmospheric Chemistry	XAI need to be optimized for correlated features to avoid false explanations.

			be employed for Earth and atmospheric sciences.		
(Y, S., & Challa, M. 2023)	SHAP	Not Provided	Here, XAI models are compared with respect to interpretability and understanding important features and concluded that SHAP and LIME are the most effective.	· Medical Applications	To enhance interpretability for advanced medical applications
	LIME				
	PDP				
	GAM				

Table 2: Comprehensive review of various optimizations done on XAI techniques. This table presents a comprehensive review of different optimization techniques suggested by various researchers.

Protocol

NOTE: This section discusses the proposed ensemble method, including all the proposed amendments as illustrated in **Figure 1**.

Data preparation

In order to validate the effectiveness of the proposed framework, the authors carried out an experiment on the crop yield dataset collected from Kaggle[17]. This dataset consists of Indian agricultural data for various crops from 1997 to 2020 and was accessed in March 2025. This dataset comprises numerous features, namely season, crop_year, fertilizer, pesticide, crop, and yield (target variable), etc. The considered dataset is pre-processed for enhanced efficiency using one-hot encoding to handle categorical values. The categorical features like season, crop, and state are one-hot encoded. MinMax scaling is also employed to scale the numerical features such as area, production, annual_rainfall, fertilizer, and pesticide to the [0,1] range. In order to handle the outliers, we used the Interquartile Range rule with a threshold Q_1 - 1.5 • IQR[17]. A random seed of 42 was applied across all preprocessing steps in order to get the reproducibility of the results. Further, the dataset was split into a training (80%) and a testing dataset (20%) after preprocessing.

Experimental setup

The experiment was carried out using Python 3.10 on an A100 GPU Server with CPU: Dual AMD Rome 7742, 128 cores, 1TB RAM. Libraries imported for simulation are scikit-learn 1.3.0, XGBoost 1.7.6, TensorFlow 2.13, SHAP 0.41.0, LIME 0.2.0.1, and ELI5 0.13.0.

Evaluation of ML models

Here, authors have used various regression models namely LinearRegression(), Ridge(alpha=1,random_state=42), Lasso(alpha=0.1,$random_{state}$ = 42), Decision TreeRegressor(max_depth = 10, random_state = 42), RandomForestRegressor(max_depth = 15random_state = 42), GradientBoostingRegressor(n_estimators = 200, learning_rate = 0.1, random_state = 42), XGBoost, and CNN to determine the impact of various variables on the target variable yield[18]. XGBoost uses 300 estimators and a learning rate of 0.05. In the CNN model, two hidden layers, ReLU activations, and Adam optimizer are used with a learning rate of 0.001. The efficiency of these models is compared using different performance metrics. These models also perform hyperparameter tuning. For instance, Ridge regression and Lasso regression are trained with alpha equals 1.0 and alpha equals 0.1, respectively. In Decision Tree and Random Forest, the maximum depth was set to 10 and 15, respectively. Gradient Boosting uses 200 estimators and a learning rate of 0.1. The performance of these models was evaluated using 5-fold cross-validation in terms of Mean Squared Error (MSE) and R2 square.

Evaluation of XAI models

As discussed, the current work uses various XAI models to compare the results. Here, the authors use *shap.TreeExplainer(model,data = crop_yield)* for tree-based models (Random Forest, Gradient Boosting, and XGBoost). A background dataset of 100 randomly selected cases was chosen from the training dataset to stabilize attributions. Further, *shap.KernelExplainer()* was used for non-tree models (Linear, Ridge, Lasso, and CNN) with 1000 perturbations to get close to SHAP values. Tabular explainer *(lime.lime_tabular.LimeTabularExplainer)* was used with 5000 perturbations per instance to run LIME on the same background dataset to ensure that local explanations are stable. The kernel width was set based on characteristics, and the model's prediction function was used to infer explanations. Local explanations were generated by calling *explainer.explain_instance(x,model.predict)*. ELI5 was utilized as permutation importance with 10 repetitions on the validation split to avoid overfitting of the test data, and was simulated using *eli5.sklearn.PermutationImportance(estimator,random_state* = 42). The significance of mean and variance among repeats was exhibited during the experiment. Finally, the fitted model was used for the validation split to obtain Partial Dependence Plots (PDPs), which were implemented using *PartialDependencyDisplay.from_estimator* (*model,X_val,features* = [*feature*]),

Set up of the TCHSHAP framework

Subsequently, TCHSHAP was implemented by applying various steps in sequence. The first step in the sequence was temporal weighting that uses an exponential decay function with a decay rate of $\lambda = 0.85$, which was selected based on an ablation study across $\lambda \in [0.7.0.95]$. The reference time (T_current) was selected as the recent crop year in the dataset. This ensures that the model explanations consider the dynamics of the agricultural output forecast, where recent features (rainfall, fertilizer use, or pesticide) carry more significance for making decisions. Temporal weighting was followed by SCM to adopt causal inferences. The suggested methodology uses the dot operator to simulate interventions and calculate the Average Causal Effect. The causal inference was applied using python *doWhy* 0.13 library. This helps to alleviate misleading connections due to correlated inputs like production and area, achieving real cause-and-effect links. The final step in the pipeline was hierarchical attribution, where features were organized into higher-level domains. For the current dataset, there are 3 hierarchical features, namely Agricultural Inputs (*Fertiliser, Pesticide*), Geographical Factors (*State*, *Area*, *Crop_Year*), and Environmental Factors (*Annual_Rainfall*, *Season*). By using normalised summation to group SHAP data, explanations at both levels (detailed and general) can be obtained.

In summary, the TCHSHAP pipeline combines three steps- temporal adjustment, causal attribution, and hierarchical grouping into a single process. Here, baseline SHAP first makes explanations, followed by the addition of three improvements. This structured pipeline ensures that TCHSHAP's outputs are consistent, can be reproduced, and give both local and global explanations. The step-by-step description of the proposed framework is given in the pseudocode below.

Input:

X: Dataset with features $X_1, X_2, X_3, \ldots\ldots, X_n$.

t: temporal variable in the dataset

$\mathbf{T_{current}}$: Reference time

λ: Decay rate

G: Feature Groups

Compute Temporal Weights

For each instance, $i \in$ **X**, evaluate $w_t = e^{-\lambda|t-T|}$,

Calculate Shap Values $\mathbf{Shap_i}$

Calculate $\mathbf{Shap}_{Tempral} = \mathbf{Shap_i} * w_t$

Compute Causal Inference

For each Feature, $X_i \in X$

Calculate $\mathbf{Shap}_{Causal} = \mathbf{Shap_i}$ * **Causal Effect**$(X_j \rightarrow Y)$

Hierarchical Attribution

For each Feature, $X_i \in X$

Calculate Shap Values $\mathbf{Shap_i}$

For each $\mathbf{G_i} \in \mathbf{G}$

$$Shap_{hierarchical} = \sum_{j \epsilon HG_i} Shap_{casual_j}$$

Output: Set of Hierarchically aggregated SHAP values

The quantitative checkpoint for facilitating troubleshooting is given below: Random Forest yielding the lowest MSE and highest R^2 across all ML models. The feature importance obtained was Area > Pesticide > Fertilizer, with Season and Crop_Year as minimal. The global baseline SHAP prediction was 161.137. Further, after applying λ = 0.85, the contribution of Crop_year and Season increases. Group-level contributions follow the order Agricultural Inputs>Geographical Factors> Environmental Factors.

Representative Results

This section discusses the results obtained by applying various methods used during the experiment study, comprising various subsections as follows:

Data collection and preprocessing

In order to perform an experimental evaluation of the proposed framework, a dataset regarding crop yield was collected from Kaggle[19]. The collected data was pre-processed, including steps such as label encoding, scaling, and removal of outliers as discussed previously. Based on the desired set objective, the feature significance of the variables was evaluated. The feature importance graph for the considered dataset is illustrated in **Figure 2**, which clearly indicates the high correlation of fertilizer and pesticides to yield production. Further, in order to evaluate the impact of data size on the evaluation of the model, the MAE and R2 score are calculated for varying dataset sizes, and the obtained results are plotted in **Figure 3**.

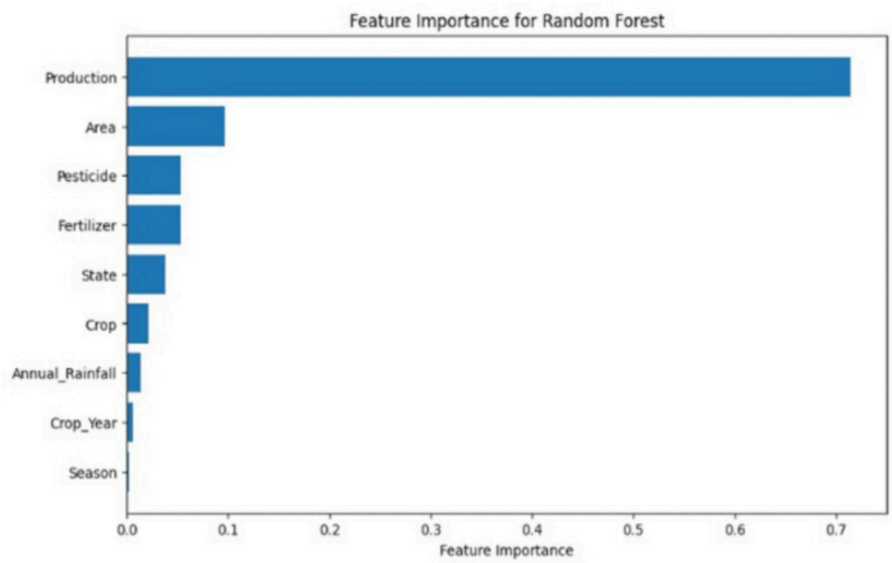

Figure 2: Feature importance. This figure illustrates the feature importance graph, indicating the high correlation of fertilizer and pesticides to yield production. Please click here to view a larger version of this figure.

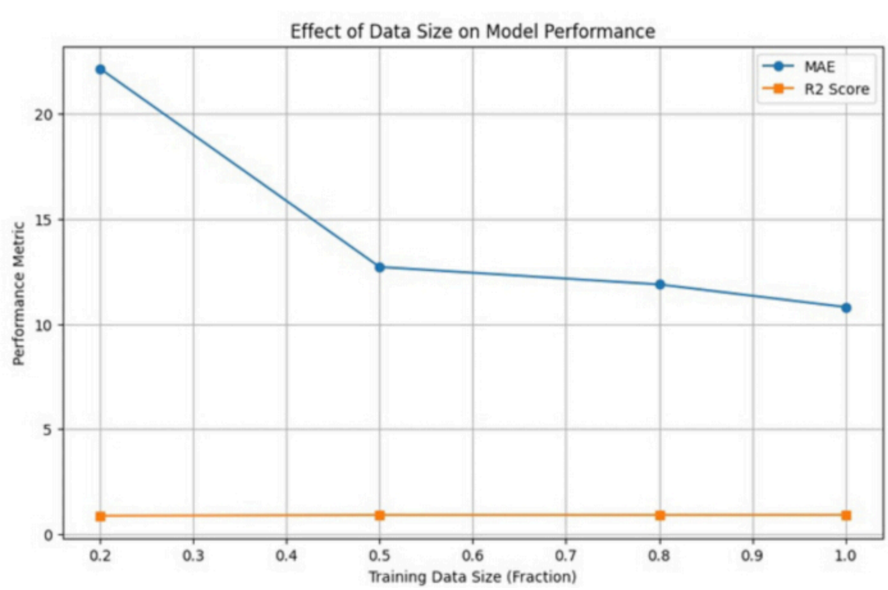

Figure 3: Impact of data size on performance. This figure illustrates the impact of increasing data size on MAE and R2 score. Please click here to view a larger version of this figure.

Evaluation of different models

Data pre-processing is followed by performance evaluation of various models (linear regression, ridge regression, lasso regression, decision tree, random forest, gradient boosting, XGBoost, and CNN) on the dataset. The comparative analysis of these models in terms of Mean Squared Error (MSE) and R^2 is clearly illustrated in **Figure 4**. The figure shows that the random forest model has the lowest value of MSE and the highest value of R^2 score, advocating its preference as the baseline model over other models.

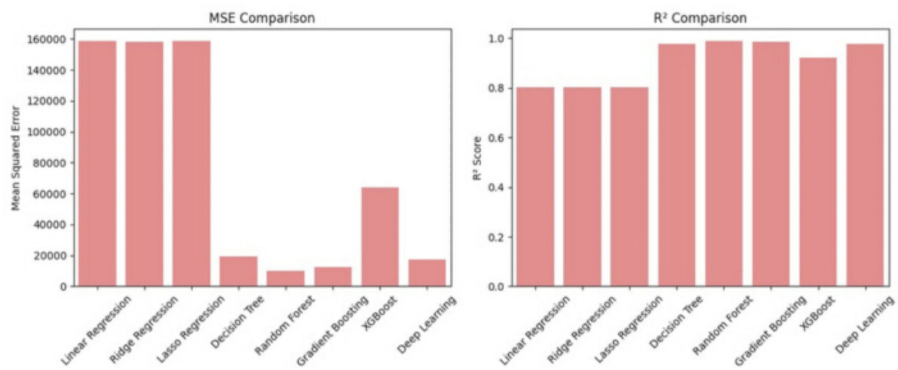

Figure 4: MSE and R2 comparison of various models. This figure illustrates the comparative analysis of Mean Squared Error (MSE) and R^2 for various models. Please click here to view a larger version of this figure.

Evaluation of XAI models

The various steps followed in the execution of the model architecture are elaborated below. The first step was to evaluate the performance of various XAI techniques for the selected ML model, Random Forest, in the current scenario. The explanations obtained from SHAP are illustrated in **Figure 5**. Here, E[F(x)] denotes the average prediction, i.e., 161.137, considering all the samples. For the particular selected instance for explainability, f(x) is 4.655. Features shown in red push the prediction towards the positive side from the average prediction, while blue color pushes it in the opposite direction. From the obtained explanation, it is evident that the features making the major positive impact on the production are area, fertilizer, and pesticide. The insight derived from the SHAP plot suggests that agricultural strategies might focus on fertilizer optimization and expansion of cultivated areas.

Further, **Figure 6** presents feature-wise explanation using ELI5 in the considered dataset. The obtained weights signify the significance of the feature towards the model's prediction and its associated uncertainty. The analysis here is quite evident and straightforward. In line with the obtained results from SHAP, features are ranked based on their importance in descending order, with production having the highest weight, followed by area, fertilizer, and others.

Another visualization plot, namely the Partial Dependence Plot (PDP), is also obtained as illustrated in **Figure 7**. The PDP presents the impact of area on yield, and the plot indicates that an increase in area leads to a decline in partial dependence, indicating the negative relationship between area and yield. From the plot, it is clear that beyond a certain threshold (e.g., Area > 400,000), the marginal effect flattens, suggesting diminishing returns or no additional impact on the crop yield (target variable) for larger areas. In the plot, light blue lines represent decision paths obtained from individual trees in the Random Forest model. Thus, different blue lines indicate that different trees in the ensemble make diverse predictions, especially for lower Area values.

Additionally, the bar chart, as illustrated in **Figure 8**, ranks features based on their importance in the Random Forest model. This importance score indicates the contribution of a feature towards the reduction in the model's error during training. **Figure 8** clearly indicates that area is an important feature contributing around 0.1 to the model's predictions, underscoring its strong relationship with the target. Other features with moderate contributions include Pesticide and Fertilizer. Further, features like Crop_Year and Season have minimal importance, suggesting their little direct impact on the target variable.

The waterfall plot obtained from LIME is demonstrated in **Figure 9**, where orange color depicts the positive contributions while blue color depicts the negative contributions. This waterfall plot also illustrates the contribution of individual features towards the predicted value. It is evident from **Figure 9** that annual rainfall, season, and state contribute negatively to the yield, while other features positively impact the predictor variable.

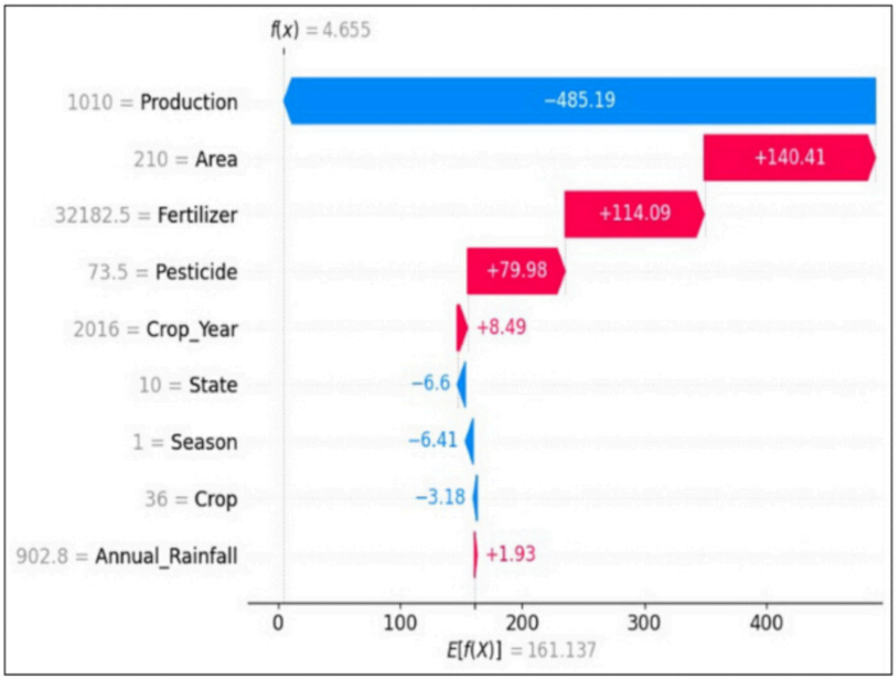

Figure 5: Explanations using SHAP. This figure illustrates the explanations obtained from SHAP, where features shown in red color push the prediction to the positive side, while blue color pushes in the opposite direction. Please click here to view a larger version of this figure.

Weight	Feature
0.7192 ± 0.0821	Production
0.0906 ± 0.1492	Area
0.0550 ± 0.1425	Fertilizer
0.0531 ± 0.1199	Pesticide
0.0371 ± 0.0494	State
0.0225 ± 0.0925	Crop
0.0139 ± 0.0206	Annual_Rainfall
0.0055 ± 0.0077	Crop_Year
0.0031 ± 0.0171	Season

Figure 6: Explanations using eli5. This figure depicts the pictorial visualization of ELI5 representing the feature importance with its weight and associated uncertainty. Please click here to view a larger version of this figure.

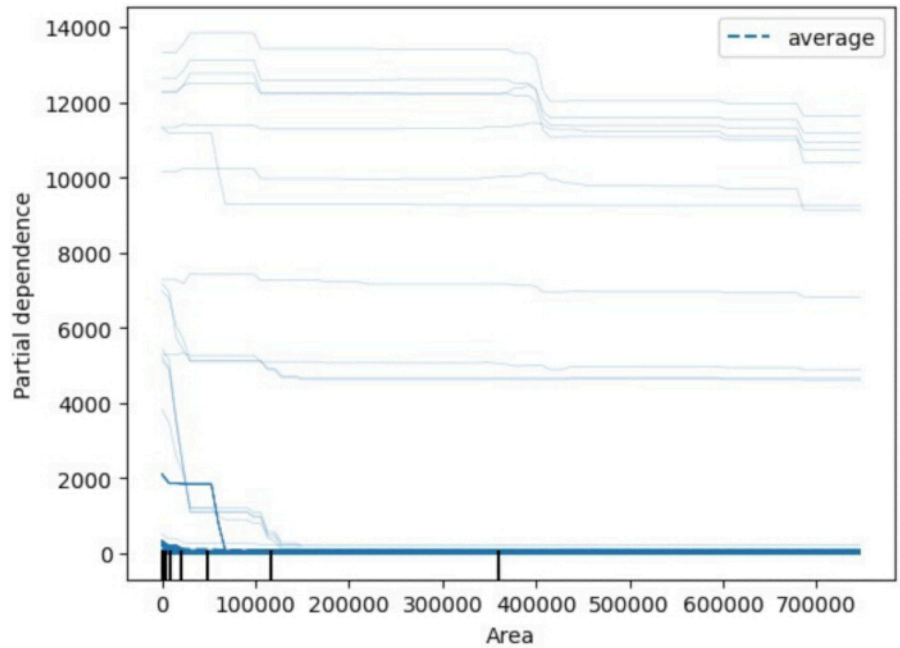

Figure 7: Explanations using PDP. This figure illustrates the Partial Dependence Plot (PDP), indicating the effect of area on yield. Please click here to view a larger version of this figure.

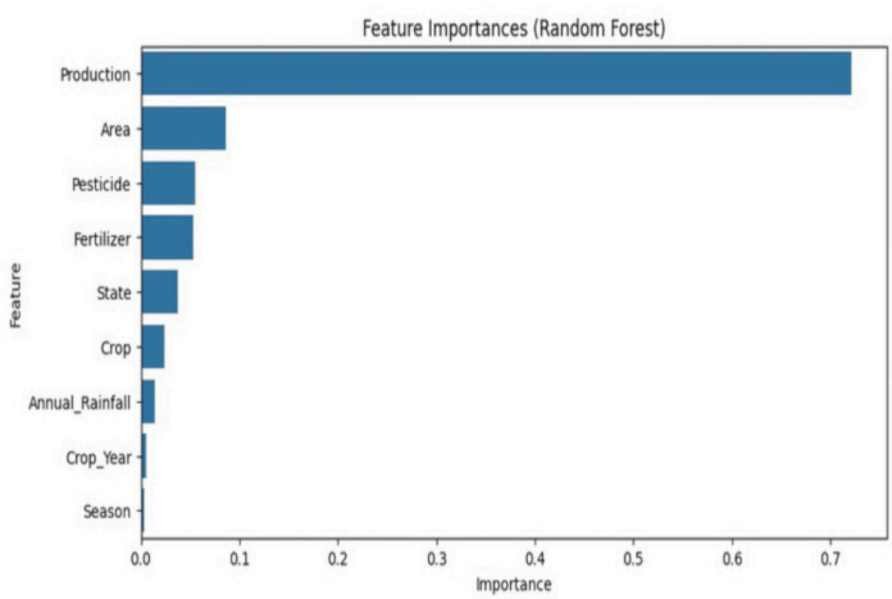

Figure 8: Explanations using feature importance. This figure depicts the bar chart ranking features based on their importance in the Random Forest model. Please click here to view a larger version of this figure.

Figure 9: Explanations using LIME. This figure demonstrates the waterfall plot obtained from LIME, where orange and blue colors depict the positive contributions and negative contributions, respectively. Please click here to view a larger version of this figure.

To summarize the impact of different features as per different XAI methods, the results are tabulated in **Table 3**.

Feature	SHAP	Eli5	PDP	Feature Importance	LIME
Production	-485.19	0.7192 ± 0.0821	Not applicable.	Most important feature.	Major negative impact.
Area	140.41	0.0906 ± 0.1492	Strong negative marginal effect beyond 400,000	Second important feature	Positive contribution.
Pesticide	79.98	0.0531 ± 0.1199	Not applicable	Moderate importance	Positive contribution
Fertilizer	114.09	0.0550 ± 0.1425	Not applicable	Moderate importance	Moderate positive contribution
State	-6.6	0.0371 ± 0.0494	Not applicable	Minor importance	Slight negative impact

Crop_Year, Season, Annual Rainfall	Very small contributions	Minimal importance	Not applicable	Negligible importance	Minimal impact

Table 3: Illustration of the impact of various features. This table gives a quantitative impact of various features using various XAI models.

The comparative analysis clearly indicates that SHAP captures various insights on the local level (instance-specific). SHAP also provides detailed insights and high interpretability at the local level, and hence it is recommended for providing the individual interaction specific to a data point. This also establishes SHAP as the baseline for XAI optimization in current research work.

Experimental analysis: In order to perform comparative analysis, SHAP is considered a baseline, and the different optimizations are added to enhance interpretability. The results obtained are compared based on the following parameters.

With temporal weighting: The waterfall plot in **Figure 10** depicts the visualization of SHAP with temporal weighting. The plot clearly indicates that production still has the highest contribution, although its magnitude is reduced. Further, temporal features like crop year and season have a marginally higher contribution than traditional SHAP, indicating the emphasis on recent data points due to temporal weighting. Also, the contributions from area, fertilizer, and pesticide are adjusted to reflect their relevance over time. It can be safely concluded that SHAP with temporal weighting performs better in terms of reflecting the model's reliance on features that are recently relevant. Thus, it has the potential to augment the temporal interpretability of the explanation.

In conclusion, temporal weighting adjusts the SHAP values using an exponential decay algorithm, aiming to diminish the impact of older data points. Attributes associated with earlier timestamps (e.g., Production for prior years) are downplayed, while recent timestamps are accentuated. The model explanation now incorporates temporal relevance, enhancing its interpretability for dynamic datasets (where time is a critical factor). In the current scenario, temporal features, namely crop_year and season, acquire interpretative relevance, facilitating the identification of trends or patterns that may go unnoticed in static analysis.

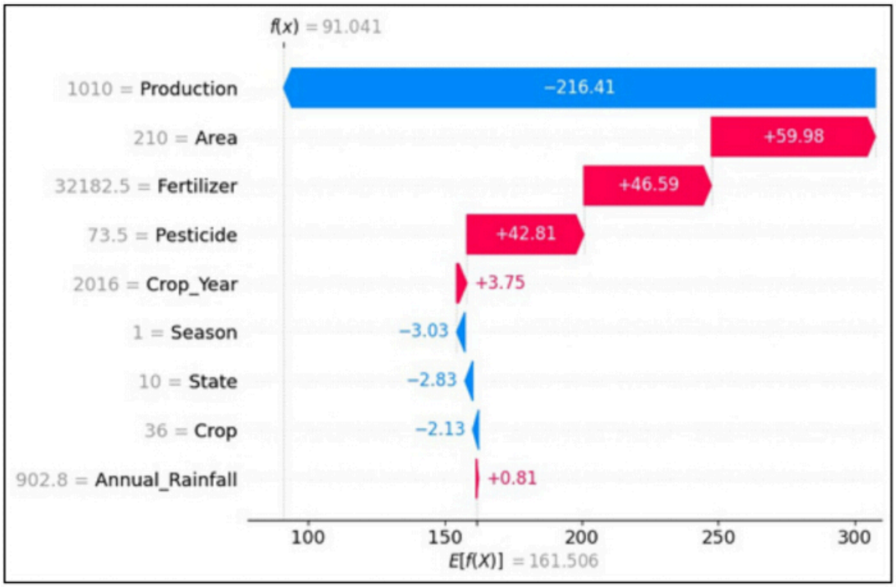

Figure 10: Waterfall plot of SHAP with temporal weighting. This figure depicts the waterfall plot, indicating the visualization of SHAP with temporal weighting. Please click here to view a larger version of this figure.

Causal inference: The waterfall plot depicted in **Figure 11** represents the impact of causal inference on SHAP interpretation. The causal inference plot illustrates that season now dominates the explanation with the largest negative contribution, suggesting a strong causal relationship between season and yield. This strong impact of season is followed by a crop that exhibits a positive impact on the target variable. Further, contrary to the traditional SHAP plot, the causal plot signifies that production, area, fertilizer, and pesticide have no contribution. This no contribution is exhibited as causal inference has alleviated their indirect impact. By adding causal inference, SHAP values now align with the causal structure of the dataset, emphasizing features with direct causal effects on the target variable.

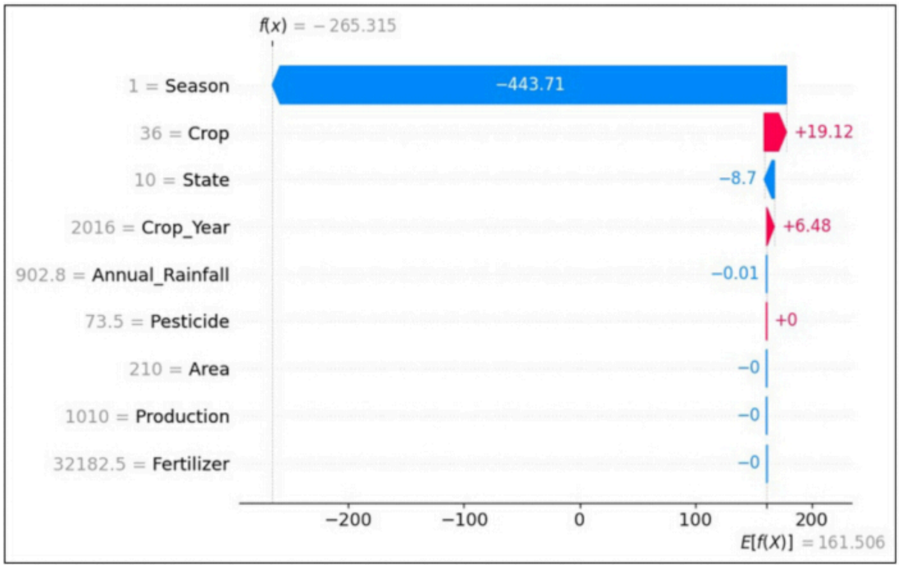

Figure 11: Waterfall plot of SHAP with causal inference. This figure illustrates the waterfall plot representing the impact of causal inference on SHAP interpretation. Please click here to view a larger version of this figure.

Hierarchical attribution: The hierarchical attribution is suggested to add granularity to the features, and the obtained plot is illustrated in **Figure 12**. Here, features of the considered dataset are classified into three main categories, namely agricultural inputs (e.g., fertilizer, pesticide, etc.), geographical factors (e.g., state, area, crop_year), and environmental factors (e.g., annual_rainfall). The obtained hierarchical SHAP contribution (as shown in **Figure 12**) indicates that agricultural inputs have the highest SHAP contribution. Thus, it is clear that input features like fertilizers and pesticides are the most influential for the prediction of this instance. Also, these factors exhibit a positive contribution, suggesting that an increase in these factors leads to an increase in the target variable (yield). Geographical factors also have a significant positive contribution, highlighting the impact of location-specific features on the prediction. On the contrary, environmental Factors demonstrate a negligible contribution in the current scenario, suggesting that environmental conditions have less impact compared to other factors.

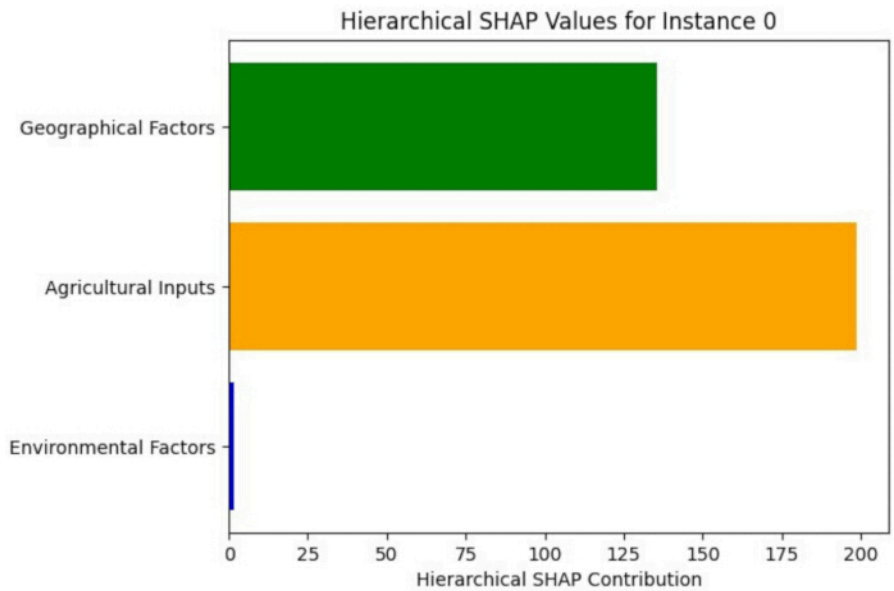

Figure 12: Bar plot of SHAP with hierarchical attribution. This figure illustrates the hierarchical attribution of the features, adding granularity to the features. Please click here to view a larger version of this figure.

Data availability:

The considered dataset is publicly available at Kaggle[19]. The link is https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset.

Discussion

The prime objective of the current study is to incorporate temporal weighting, causal inference, and hierarchical significance in the traditional SHAP model, yielding the TCHSHAP model. The motive behind including these components in XAI techniques is to enhance the interpretability of the results. For temporal weighting, we have considered an exponential decaying technique, which gives more weightage to recent values in comparison to old values. In causal significance, authors try to assess the direct impact of various features on the predictor variable. Further, input parameters are also classified into various hierarchical features to better understand the results. In order to validate the effectiveness of the proposed solution, the authors have considered the crop_yield dataset available on Kaggle. The obtained results clearly indicate that temporal weighting, causal inference, and hierarchical attribution (as suggested in TCHSHAP) lead to interpretability enhancement of traditional SHAP by addressing its limitations[20,21].

However, the proposed methodology also has potential limitations. For instance, currently, the efficacy of the proposed methodology is validated using only a single crop-yield dataset, and the same is found to be motivating. In order to further strengthen the effectiveness, the suggested approach needs to be tested in diverse application domains, including healthcare, finance, and education[22,23]. This positive initiative will strengthen the validation of the suggested approach and, hence, open further avenues for its widespread application.

Conclusion and future work

A significant number of researchers are trying to enhance the traditional XAI models in view of the lack of interpretability. This need is further aggravated owing to the exponential rise in the employment of machine learning models in various domains. In current research, we also aim to propose a model that comprehends temporal behavior, causal inference, and hierarchical relationships among the features in SHAP (TCHSHAP) to enhance the effectiveness and interpretability of XAI models. In order to validate the effectiveness of the proposed model, the authors employ TCHSHAP in an experimental setup considering the crop_yield dataset. During the experimental evaluation, it is observed that temporal weighting adjusts the SHAP values through an exponential decay algorithm that diminishes the impact of older data points. Attributes associated with earlier timestamps (e.g., Production for prior years) are downplayed, whilst those related to more recent timestamps are accentuated. Further, it can be concluded that by adding causal inference, SHAP values align with the causalities in the dataset, emphasizing features with direct causal effects on the target variable. Hierarchical attribution adds hierarchy to the features, making it easier to understand and interpret for users. Future work may be carried out to apply the model to the datasets in various domains and assess its efficacy. The efficacy and supremacy of the proposed framework can be further demonstrated through experimental evaluation prior to its widespread application in real-world scenarios.

Disclosures

The authors declare that there are no conflicts of interest.

Acknowledgments

This work is funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the Programme Contract UID/05105/2025.

References

1. Kaur, B. et al. N-Beats architecture for explainable forecasting of multi-dimensional poultry data. *PloS One*. **20** (4), e0320979 (2025).
2. Sharma, N., Mangla, M., Iqbal, M. M., Mohanty, S. N. Deep Learning Framework for Identification of Skin Lesions. *EAI Endorsed Trans Perv Health Tech*. **9** (2023).
3. Sharma, N., Mohanty, S. N., Mahato, S., Pattanaik, C. R. A novel dataset and local interpretable model-agnostic explanations (LIME) for monkeypox prediction. *Intell Decision Technol*. **17** (4), 1297-1308 (2023).
4. Makumbura, R. K. et al. Advancing Water Quality Assessment and Prediction Using Machine Learning Models, Coupled with Explainable Artificial Intelligence (XAI) Techniques Like Shapley Additive Explanations (SHAP) For Interpreting the Black-Box Nature. *Results Eng*. **23**, 102831 (2024).
5. Schlegel, U., Keim, D. A. A Deep Dive into Perturbations as Evaluation Technique for Time Series XAI. *Communic Comp Informat Sci*. **1903,** (2023).
6. Silva, S. J., Keller, C. A. Limitations of XAI methods for process-level understanding in the atmospheric

sciences. *Artificial Intell Earth Syst*. **3** (1), e230045 (2024).

7. Anushree, G., Madagaonkar, S. B., Ravili, C. H. Unveiling the Black Box: A Comprehensive Review of Explainable AI Techniques. *Int J Sci Res Eng Manag*. **8** (9), 1-11 (2024).
8. Awal, M. A., Roy, C. K. EvaluateXAI: A framework to evaluate the reliability and consistency of rule-based XAI techniques for software analytics tasks. *J Syst Software*. **217**, 112159 (2024).
9. Hajare, S. W., Rewatkar, R., Reddy, K. Design of an iterative method for enhanced early prediction of acute coronary syndrome using XAI analysis. *AIMS Bioeng.* **11** (3), 301-322 (2024).
10. Hamida, S. U., Chowdhury, M. J. M., Chakraborty, N. R., Biswas, K., Sami, S. K. Exploring the landscape of explainable artificial intelligence (XAI): A systematic review of techniques and applications. *Big Data Cognit Comput.* **8** (11), 149 (2024).
11. Ihongbe, I. E., Fouad, S., F. Mahmoud, T., Rajasekaran, A., Bhatia, B. Evaluating Explainable Artificial Intelligence (XAI) techniques in chest radiology imaging through a human-centered Lens. *PloS One*. **19** (10), e0308758 (2024).
12. Swathi, Y., Challa, M. A Comparative Analysis of Explainable AI Techniques for Enhanced Model Interpretability. *3rd Int Conf Pervasive Comput Social Netw.* 229-234 (2023).
13. Lenaers, I., De Moor, L. Exploring XAI techniques for enhancing model transparency and interpretability in real estate rent prediction: a comparative study. *Finance Res Lett.* **58** (Part A), 104306 (2023).
14. Bommer, P. L., Kretschmer, M., Hedström, A. K., Bareeva, D., Höhne, M. Finding the right XAI Method - A Guide for the Evaluation and Ranking of Explainable AI Methods in Climate Science. *Artif Intell Earth Syst.* **3**, e230074 (2024).
15. Bhatnagar, S., Agrawal, R. Understanding explainable artificial intelligence techniques: a comparative analysis for practical application. *Bullet Elect Eng Info*. **13** (6), 4451-4455 (2024).
16. Ennab, M., Mcheick, H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. *Front Robot AI.* **11**, 1444763 (2024).
17. Wan, X., Wang, W., Liu, J., Tong, T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. **14** (1), 135 (2014).
18. Mangla, M., Shinde, S. K., Mehta, V., Sharma, N., Mohanty, S. N. *Handbook of Research on Machine Learning: Foundations and Applications.* CRC Press (2022).
19. Gupta, A. *Agricultural Crop Yield in Indian States Dataset [Data set].* Kaggle. https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset.(2025).
20. Shin, J., Miah, A. S. M., Konnai, S., Hoshitaka, S., Kim, P. Electromyography-Based Gesture Recognition with Explainable AI (XAI): Hierarchical Feature Extraction for Enhanced Spatial-Temporal Dynamics. *IEEE Access.* **13**, 88930-88951 (2025).

21. Das, A., Rad, P. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv.* arXiv:2006.11371 (2020).

22. Srinivasu, P. N., Sandhya, N., Jhaveri, R. H., Raut, R. From blackbox to explainable AI in healthcare: existing tools and case studies. *Mobile Info Syst*. **2022** (1), 8167821 (2022).

23. Fiok, K., Farahani, F. V., Karwowski, W., Ahram, T. Explainable artificial intelligence for education and training. *J Defense Model Simulat.* **19** (2), 133-144 (2022).