International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 5, October 2026, pp. 4207~4219
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i5.39427

The impact of vocational training on female labor productivity an evaluation in Vietnam's construction sector

Truc Thai Phuong, Lam Nguyen Son
Faculty of Civil Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Mar 10, 2026
Revised Jul 2, 2026
Accepted Aug 5, 2026

Keywords:

Construction sector
Female labor
Labor productivity
SHAP
Synthetic data augmentation
Vocational training

ABSTRACT

Vietnam's construction sector has experienced strong growth through rapid urbanization and infrastructure investment. However, female workers continue to face significant barriers, including high informal employment rates (60–65%), 15–20% lower labor productivity than males, and limited access to vocational training (only 19–25% hold certificates). This study investigates the impact of the share of females aged 15+ with vocational training certificates (STFW15+) on female construction labor productivity (FCLP). Using a short time-series dataset (2010–2024, n=15) from General Statistics Office (GSO) of Vietnam, International Labor Organization (ILO), and World Bank, the study applied ForestDiffusion synthetic data pipeline-time-series to generate 75 high-quality synthetic samples (total 90 observations). An artificial neural network (ANN) model integrated with SHapley Additive exPlanations (SHAP) was employed to analyze feature importance and nonlinear interactions. Results show that STFW15+ is the strongest predictor of FCLP (mean |SHAP|=0.514). It exerts a strong positive effect and acts as a powerful moderator, amplifying benefits from income and economic growth while mitigating the negative impact of informal employment. The findings highlight the critical role of gender-responsive technical and vocational education and training (TVET). This study recommends targeted vocational training programs, gender integration in public investment, and public-private partnerships to advance gender equality and support sustainable development goals (SDGs) 5 and 8.

Corresponding Author:

Lam Nguyen Son
Faculty of Civil Engineering, Industrial University of Ho Chi Minh City
12 Nguyen Van Bao, Hanh Thong Ward, Ho Chi Minh, Vietnam
Email: nguyensonlam@iuh.edu.vn

1. INTRODUCTION

Vietnam's construction sector has grown strongly due to rapid urbanization and large-scale infrastructure investment, with growth rates of 9–9.6% in 2025 and continued momentum expected in 2026 [1]. However, female workers face major barriers: informal employment rates of 60–65% (above national average), labor productivity 15–20% lower than males, earnings 20–30% lower in manual roles, and limited vocational training access (only 19–25% hold certificates, versus the national rate of 29.2% in 2025 and 29.5% in Q4 2025 per General Statistics Office (GSO) of Vietnam. Technical and vocational education and training (TVET) plays a critical role in advancing gender equality and reducing informal employment in developing countries. International evidence shows that gender-responsive TVET can drive wider gender transformation and support the decent work agenda [2].

Existing studies on female labor in Vietnam have mainly concentrated on textiles/garments and food processing sectors, with very limited focus on construction-specific issues such as physical demands, high seasonality, and project-based employment. Most previous research relies on traditional statistical methods that struggle with small time-series samples and severe multicollinearity. Several studies have applied artificial intelligence (AI) techniques to predict labor productivity in construction processes, including the application of artificial neural network (ANN) methodology and identification of key factors influencing formwork productivity using ANN. Data mining approaches have also been used for labor market forecasting [3]. This study contributes to vocational education and construction education by providing an explainable AI framework to support the design of gender-responsive TVET programs in male-dominated sectors. It addresses critical gaps by employing interpretable ensemble learning methods [4], meta-ensemble models for construction labor productivity [5], and hybrid machine learning solutions for mitigating productivity losses [6]. Advanced forecasting techniques such as slipform labor productivity prediction [7] and ANN-based prediction intervals [8] further demonstrate the potential of these methods, while comparative studies of machine learning regression models highlight their effectiveness in construction-related predictions [9].

Gender-disaggregated construction data in Vietnam are scarce, limited to approximately 15 annual observations from GSO, International Labor Organization (ILO), and World Bank, with severe multicollinearity ($r \approx 0.95$–0.98). To overcome these limitations, this study applies the ForestDiffusion synthetic data pipeline-time-series to generate 75 high-quality synthetic samples (total 90 observations), followed by an ANN integrated with SHapley Additive exPlanations (SHAP) for interpretable feature importance and nonlinear interaction analysis [10]. This work aligns with Vietnam's national strategy on gender equality 2021–2030 as well as global emphasis on gender-responsive TVET for reducing informality and expanding women's opportunities in the construction sector. The novelty of this study lies in being the first to combine ForestDiffusion synthetic data augmentation with an explainable ANN-SHAP framework to investigate the impact of vocational training on female labor productivity in Vietnam's construction sector. This approach effectively addresses the critical challenges of small sample size and severe multicollinearity that have limited previous research. By doing so, the study provides robust, interpretable insights into the moderator role of vocational training for female workers in a male-dominated industry. The main research questions are:

- How does the proportion of females aged 15+ with vocational training certificates (STFW15+) influence female construction labor productivity (FCLP)? (RQ1)
- Is STFW15+ the strongest predictor of FCLP compared to average monthly income in construction (AMICW), gross regional domestic product per capita (GRDPpc), and the share of informal employment in construction (SIEC)? (RQ2)
- What policy measures can improve female labor productivity in the construction sector by reducing informality and strengthening TVET? (RQ3)

2. LITERATURE REVIEW

Global studies highlight the critical role of TVET in advancing gender equality and reducing informal employment in developing countries. Human capital theory posits that investments in skills enhance individual productivity and economic contributions [11]. Gendered labor market theory identifies structural barriers—such as caregiving responsibilities and limited access to specialized skills—that confine women to informal, low-productivity roles [12]. Gender-responsive TVET addresses skill mismatches, counters stereotypes in male-dominated sectors like construction, reduces informality, and enhances income and social protection. Key international reports support this view. UNESCO Global Education Monitoring Report (GEM report) emphasize gender-responsive TVET for closing skill and employment gaps in non-traditional fields for women [13]. ILO's world employment and social outlook: trends 2025 notes that vocational training mitigates informal employment risks for women, builds resilience, and facilitates transitions to formal jobs [14]. The World Bank (2024–2025) underscores TVET's value for green and digital transitions in Vietnam while highlighting shortages of gender-disaggregated data in sectors like construction [15].

In Vietnam's construction sector, informal employment exceeds the national average, with a 15–20% gender productivity and income gap and low female certification rates (19–25% versus approximately 29% in 2025; GSO) [16]. Urbanization and infrastructure growth offer opportunities, yet women remain concentrated in seasonal, project-based roles vulnerable to economic cycles [17]. Existing domestic research on female labor has primarily focused on textiles/garments and food processing sectors, with very limited attention to construction-specific challenges such as physical demands, high seasonality, and project-based employment [18]. In the domain of construction labor productivity, AI and machine learning techniques have been increasingly applied, including ANN for productivity prediction [19]. Ensemble and interpretable methods such as meta-ensemble models [20] and SHAP-based analysis [21] have gained traction. However, such advanced applications remain limited in Vietnam's construction sector, particularly with gender-disaggregated data and small time-series datasets [22]. This creates a significant research gap in both vocational training and construction education literature regarding the use of explainable AI for female workers.

3. METHOD

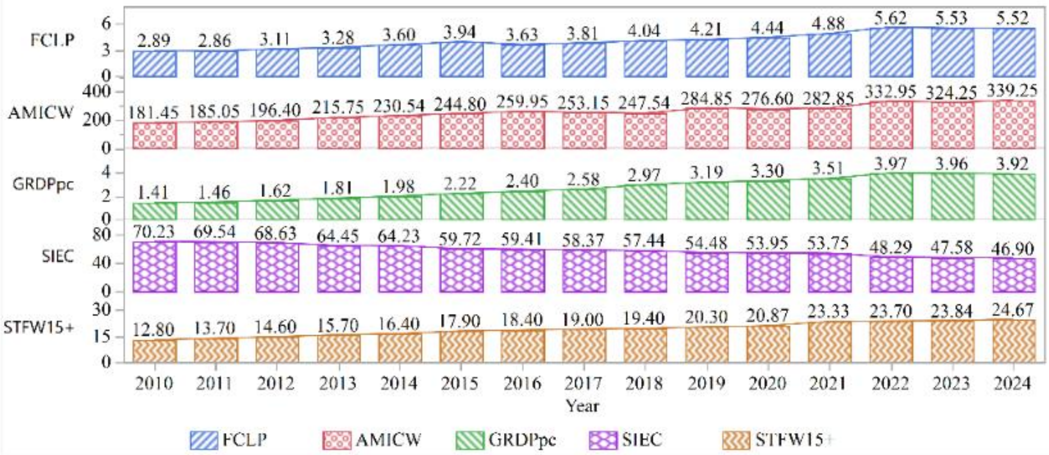
3.1. Data sources, variables and descriptive analysis

This study utilizes an annual time-series dataset from 2010 to 2024 (15 observations) sourced from the GSO of Vietnam, ILO, and World Bank. These sources offer consistent, nationally representative indicators on labor, economic development, and vocational education. Variables are selected based on their theoretical and empirical relevance to female labor productivity in construction, encompassing labor-specific, macroeconomic, and institutional factors. Recent updates from GSO 2025–2026 reports indicate the national trained workforce rate reached 29.2% in 2025 and 29.5% in Q4 2025, with informal employment at ~63.1%. The dependent variable is FCLP, measured in thousand USD per female worker per year, derived from sector value added adjusted by estimated female labor share. The independent variables include: AMICW: (USD/month)–reflects wage levels and incentives. GRDPpc: (thousand USD/person/year)–proxies' economic development and infrastructure intensity. SIEC: (%)–indicates formalization degree and risks (estimated 60–65% in 2025, above national average). STFW15+: share of females aged 15+ with vocational training/certificates (%)–measures skill development access. Data were cross-checked for consistency; any missing values were linearly interpolated, and variables standardized (z-scores) for comparability. Outliers were retained to reflect real economic shocks (2016 downturn and 2022 post-COVID recovery). Descriptive statistics and distribution tests are presented in Table 1. All variables exhibit near-normal distributions (Shapiro-Wilk $p>0.05$), low skewness ($|skew|<0.5$), and suitability for nonlinear modeling. The small sample ($n=15$) and high multicollinearity justify synthetic data augmentation later.

Figure 1 illustrates long-term trends in the key variables in the construction sector from 2010 to 2024. These visualizations reveal strong parallel upward trends for positive variables (FCLP, AMICW, GRDPpc, and STFW15+), while the SIEC shows a steady decline. The patterns demonstrate clear co-movement between vocational training, labor formalization, and productivity gains, highlighting the vulnerability of female workers to economic cycles.

Table 1. Descriptive statistics and distribution tests of variables

Variable	n	Mean	SD	Median	Range (min–max)	IQR (Q1–Q3)	Skewness	p-value (Shapiro-Wilk)	KS statistic	KS p-value	Wasserstein distance
FCLP	15	4.09	0.94	3.94	2.86–5.62	3.28–4.88	0.47	0.19	0.08	1.00	0.05
AMICW	15	257.03	50.78	253.15	181.45–339.25	215.75–284.85	0.14	0.57	0.12	0.99	0.98
GRDPpc	15	2.69	0.93	2.58	1.41–3.97	1.81–3.51	0.07	0.19	0.08	1.00	0.06
SIEC	15	58.46	7.80	58.37	46.90–70.23	53.75–64.45	0.05	0.39	0.08	1.00	0.55
STFW15+	15	18.97	3.84	19.00	12.80–24.67	15.70–23.33	-0.03	0.54	0.11	1.00	0.21



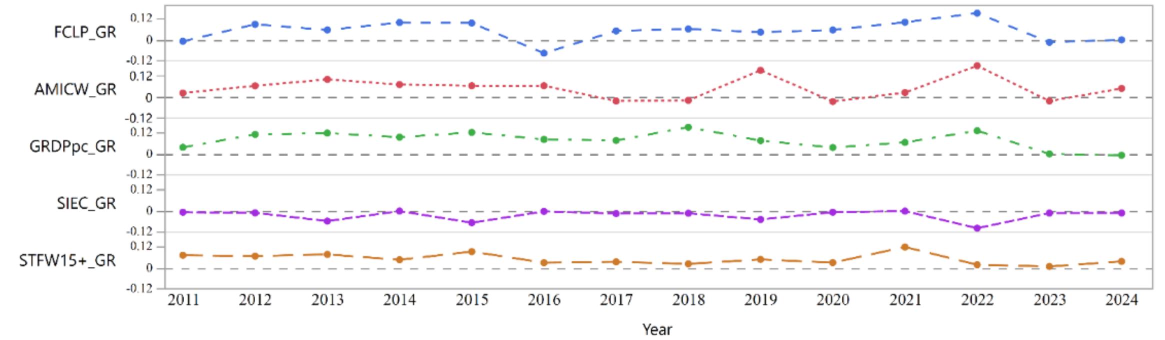
Note: Multi-axis chart combining bars and lines showing fluctuations of the five main variables during 2010-2024. FCLP (light blue hatched bars), AMICW (red dotted bars), GRDPpc (solid green line), SIEC (purple cross-hatched bars), STFW15+ (orange dashed line).

Figure 1. Trends of key variables over time (2010–2024)

Positive variables (FCLP, AMICW, GRDPpc, and STFW15+) exhibit strong parallel upward trends, while SIEC declines steadily. From 2010 to 2024, FCLP increased by 91.0%, GRDPpc by 178.0% (strongest), and SIEC decreased by 33.2%. Key periods include steady growth (2010–2015), a sharp drop in

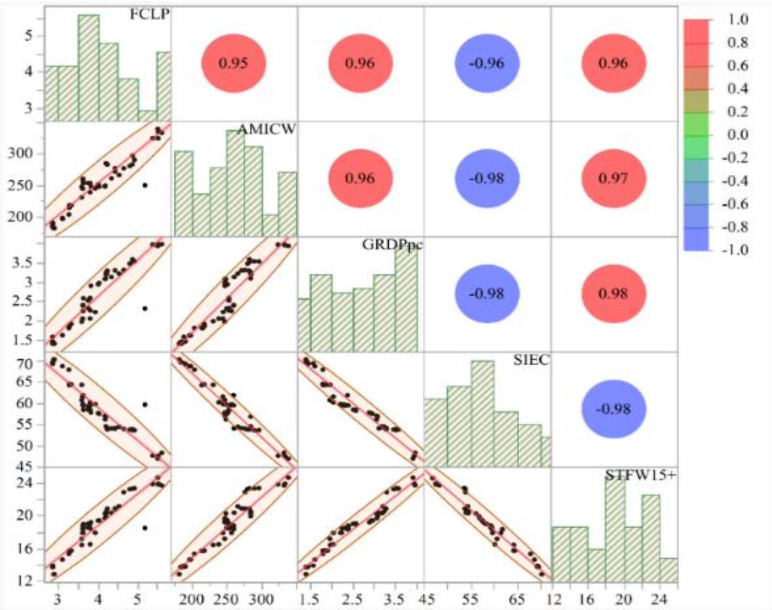
2016 (FCLP –7.87%), rapid recovery (2017–2021), a peak in 2022 (FCLP +15.16%), and stabilization/decline in 2023–2024. These patterns demonstrate co-movement between vocational training, labor formalization, and productivity gains, with female workers (often seasonal/project-based) remaining vulnerable to economic cycles. Figure 2 illustrates annual growth rates of the key variables and periods of strong fluctuation.

Average annual growth rates: GRDPpc_GR highest (+7.68%/year), followed by FCLP_GR (+4.89%), STFW15+_GR (+4.84%), AMICW_GR (+4.75%), and SIEC_GR declining (–2.80%/year). Major outliers include 2016 (FCLP_GR –7.87%, sector downturn) and 2022 (FCLP_GR +15.16%, post-COVID recovery). Growth slowed in 2023–2024, but vocational training sustained positive momentum. This chart complements Figure 1 by highlighting relative fluctuations and the catalytic role of training/formalization during recovery. Figure 3 illustrates the Pearson correlation matrix and severe multicollinearity.



Note: Line chart showing year-over-year growth rates (%) from 2011 to 2024. FCLP_GR (blue dotted line), AMICW_GR (red dashed-dotted line), GRDPpc_GR (solid green line), SIEC_GR (purple line), STFW15+_GR (orange dotted line). Gray dashed horizontal line at 0% serves as reference.

Figure 2. Annual growth rates of key variables (%)



Note: Heatmap of Pearson correlations combined with scatter plots and histograms on the diagonal. Dark red indicates strong positive correlations, dark blue strong negative; circle size proportional to |r|.

Figure 3. Pearson correlation heatmap of variables

Analysis shows very strong correlations: FCLP with STFW15+ (+0.96), AMICW (+0.95), GRDPpc (+0.96), and SIEC (–0.96). Severe multicollinearity among independent variables (GRDPpc–STFW15+, r=+0.98) renders traditional linear regression unstable. Scatter plots confirm linear trends between FCLP and STFW15+, and inverse patterns with SIEC. This multicollinearity is the main reason for selecting ANN combined with SHAP, which handles nonlinear relationships and provides robust feature importance analysis [23]. Figure 3 provides a visual foundation for trends, fluctuations, and inter-variable relationships, while highlighting limitations of small sample size and multicollinearity. This motivates synthetic data augmentation and advanced modeling in subsequent sections. Several studies have employed hybrid AI approaches to predict and optimize construction labor productivity under similar constraints. Data augmentation techniques, such as generative models combined with explainable AI, have also been applied in other domains to improve model performance on limited datasets [24]. Recent GSO 2025-2026 data confirm positive national trends (rising training rates, declining informality), yet persistent gaps in construction—particularly for female workers—underscore the need for deeper, sector-specific analysis.

3.2. Synthetic data augmentation and model specification

Given the limited original observations (n=15 time points), applying complex nonlinear models like ANN risks severe overfitting and unstable predictions. The ForestDiffusion synthetic data pipeline-time-series, adapted from diffusion-based generative models for tabular time-series data, was employed to preserve temporal structure and key economic shocks while expanding the dataset. To preserve temporal structure and macroeconomic features while expanding the dataset, this study employs the ForestDiffusion synthetic data pipeline-time-series, a diffusion-based technique for time-series augmentation. The method adds progressive noise and reverses denoising to generate synthetic samples matching the original distribution, maintaining temporal continuity and reproducing key breakpoints (2016 decline and 2022 surge). From 15 real observations, 75 high-quality synthetic samples were created, yielding a mixed dataset of 90 samples for training and evaluation. Synthetic data quality was validated rigorously: Wasserstein distances: 0.05–0.98 (minimal shift). ACF/PACF matched original at relevant lags. ADF test confirmed preserved stationarity. Major shocks (2016 downturn, 2022 post-COVID surge) faithfully replicated. Ablation study: 75 synthetic samples yielded best generalization (highest test Rsquare, smallest train-test gap). The primary model is a feedforward ANN with 4-3-1 architecture (4 inputs: AMICW, GRDPpc, SIEC, STFW15+; one hidden layer with 3 rectified linear unit (ReLU) neurons; linear output for regression). The prediction formula is: the main model is a feedforward ANN with architecture 4-3-1 (4 inputs, 3 hidden ReLU neurons, and 1 linear output).

Prediction formula (explicitly written):

$$\hat{y} = b_2 + \sum_{i=1}^{3} w_{2,i}\, \sigma\big(b_{1,i} + \sum_{j=1}^{4} w_{1,ij}\, x_j\big) \quad (1)$$

Where, x_j(j = 1…4): input variables (AMICW, GRDPpc, SIEC, STFW15$^+$). $\sigma(\cdot)$: ReLU activation. $w_{1,i,j}$, $b_{1,i}$: hidden layer weights and biases. $w_{2,i}$, b_2: output layer weights and bias. The loss function is mean squared error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

Performance metrics include:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\, ;\, RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

Training: Adam optimizer (initial lr=0.001, adaptive), early stopping (patience=10), batch size=16, dropout (0.2–0.3), L2 regularization. SHAP value formula:

$$\phi_i = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}\,[f_x(S \cup \{i\}) - f_x(S)] \quad (4)$$

Where, N: set of all features. S: subset excluding feature i. $f_x(\cdot)$: model prediction function. SHAP enables global feature importance, local explanations, and visualization of nonlinear interactions (via 3D surface plots).

Model performance and trained parameters are shown in Tables 2 and 3. The ANN model achieved strong performance on the mixed dataset of 90 observations. On the training set, the model recorded an R^2 of 0.873, root mean squared error (RMSE) of 0.290, and mean absolute error (MAE) of 0.206. The validation and testing results remained stable with R^2 values of 0.723 and 0.611 respectively, indicating good generalization ability. The low overfitting score of 0.057 confirms that the combination of synthetic data augmentation, L2 regularization, dropout, and early stopping was highly effective. Test errors ranging from 0.5 to 0.6 thousand USD per worker per year are reasonable when compared to the mean FCLP value of approximately 4.09. In contrast, models trained only on the original 15 observations exhibited significantly lower R^2 values and severe overfitting. These performance metrics demonstrate the robustness of the proposed ANN model in handling small-sample, high-multicollinearity time-series data. The results validate the effectiveness of the ForestDiffusion synthetic data pipeline-time-series augmentation technique adopted in this study. Overall, the model provides reliable predictions and serves as a solid foundation for subsequent SHAP interpretability analysis.

Small model (16 parameters) works well to capture nonlinear interactions. STFW15+ dominates through interactions (later explains SHAP results). This setup overcomes small-data and multicollinearity limitations allowing reliable SHAP analysis and visualization of nonlinear interactions and paves the way for the results. Construction specific applications of hybrid AI optimization methods [25] and hybrid models combining various techniques for labor productivity prediction and optimization [26] have been successfully applied to similar problems. One specific activity, for which case studies show the practical usefulness of forecasting for labor productivity, is formwork in high-rise buildings [27]. Generative models and explainable AI based data augmentation strategies have also been shown effective at improving model robustness on limited or domain specific datasets.

Table 2. Performance metrics of the ANN model

Model	Training			Validation			Testing			Overfitting score
	Rsquare	RMSE	MAE	Rsquare	RMSE	MAE	Rsquare	RMSE	MAE	
ANN	0.873	0.290	0.206	0.723	0.347	0.282	0.611	0.617	0.485	0.057

Table 3. Weights and biases of the trained ANN model

Hidden neurons (i)	Hidden bias (b1i)	Hidden weights (w1, j)				Output weights (w2, i)
		AMICW (j=1)	GRDPpc (j=2)	SIEC (j=3)	STFW15$^+$ (j=4)	
1	-12.91	0.02	0.22	0.20	-0.20	-1.34
2	4.65	-0.01	-1.32	0.08	-0.19	1.04
3	-4.40	0.01	0.54	0.04	-0.03	2.64
Output node (k=1)	Output bias (b2)	Output weights (w2, i)				
		i=1		i=2		i=3
1	4.20	-1.34		1.04		2.64

3.3. Time series diagnostics, baseline comparison, and outlier analysis

To assess the ANN model's validity in a time-series context, diagnostics, baseline comparisons, and outlier analysis were performed. Diagnostics on original (n=15) and mixed (n=90) datasets: Stationarity: Most variables (FCLP, AMICW, GRDPpc, and STFW15+) stationary or stationary after first differencing (ADF $p<0.05$; KPSS $p>0.05$) [28]. Residuals: Ljung-Box test $p>0.05$ (up to lag 10), confirming white-noise and no autocorrelation. These align with recent GSO 2025–2026 data (trained workforce rate 29.2% in 2025, 29.5% in Q4 2025; informal employment ~63.1%; construction growth 9.62%). Baseline comparison on hold-out set (2020–2024): Baselines: OLS/Ridge/Lasso (unstable due to multicollinearity, VIF>10; hold-out Rsquare<0.45); ARIMA/ARIMAX/VAR (Rsquare≈0.50–0.55, miss nonlinearities). ANN+augmentation: hold-out Rsquare≈0.611, RMSE≈0.617, MAE≈0.485—outperforms baselines by 10–20 pp in Rsquare, thanks to nonlinear capability and reduced overfitting. Outlier analysis (key anomalous years): 2016: FCLP –7.87% (sector downturn from stalled investment/banking issues). 2022: FCLP +15.16% (post-COVID recovery, high investment/FDI). 2024: Stalled growth in some variables (income-productivity mismatch). In 2025, construction recovered strongly (9.62% growth; industrial-construction 9.73%; net employment gains ~183–264 thousand vs. 2024), underscoring training's buffering role.

Table 4, which applies the nearest neighbors method on year-over-year growth rates, identifies 2024 as the most prominent outlier with a distance of 2.824824. This outlier is primarily driven by significant deviations in AMICW, SIEC, and FCLP. The analysis highlights the exceptional economic conditions in 2024 compared to previous years. The Augmented ANN model handles these outliers robustly. Synthetic data effectively reproduces the observed fluctuations, while the nonlinear structure combined with regularization techniques minimizes the negative impact of economic shocks. This approach ensures model stability even during abnormal periods [29].

Table 4. Analysis of outliers in annual growth rates (largest outliers from 8 nearest neighbors)

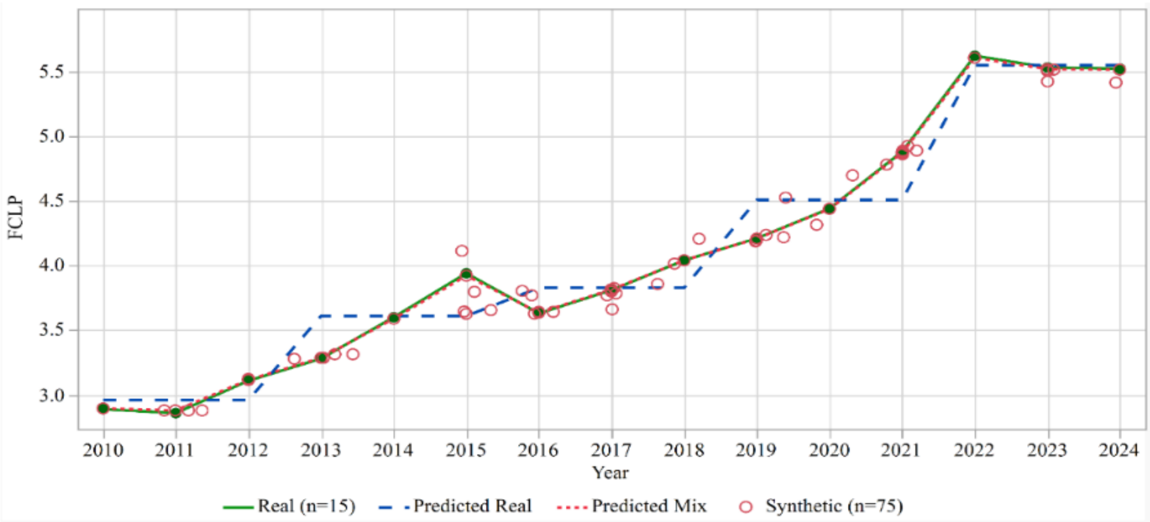
Row	Distance	Near neighbors	Col1	RMS1	Col2	RMS2	Col3	RMS3	Col4	RMS4	Col5	RMS5
15	2.824824	13 14 12 11 10 9 8 7	AMICW	1.1485	SIEC	0.9245	FCLP	0.8784	STFW15+	0.6427	GRDPpc	0.6332

4. RESULTS

4.1. Model prediction fit and augmentation impact

The ANN model on the mixed dataset (90 samples) achieved test Rsquare=0.611, RMSE=0.617, MAE=0.485, as shown in Table 2. Augmentation markedly improved smoothness and accuracy over real-only training (15 samples), better capturing fluctuations and trends. Figure 4 illustrates the effectiveness of synthetic data augmentation in enhancing forecast smoothness and accuracy.

All series show long-term rise (~2.9 in 2010 to ~5.5 in 2022–2024), acceleration 2019–2022. Synthetic samples closely surround real data (Wasserstein 0.05 for FCLP). Predicted Real is step-like, flattens 2016 drop, underestimates 2022 surge [30]. Predicted mix is smoother, tracks real closely, captures key turns (moderated 2016 decline, strong 2022 peak, 2023–2024 stabilization) [31]. Key stages: 2010–2015: near overlap. 2016: mix captures moderated decline; real nearly flat. 2019–2022: mix follows acceleration (peak 5.62). 2023–2024: mix smoother. Figure 4 highlights augmentation's value: fills gaps, learns complex patterns, reduces overfitting (score 0.057), supports reliable SHAP analysis, as shown in Figures 5–7. Mix prediction aligns with 2024–2025 trends (GSO 2025–2026): construction growth 9.62%, industrial-construction 9.73%, trained workforce 29.2% (2025)→29.5% (Q4 2025), reinforcing STFW15+ role amid cycles. Pivotal events: 2016: FCLP –7.87% (downturn, seasonal female workers affected). 2022: FCLP +15.16% (recovery, formal shift boosted productivity). Implications: retraining funds, flexible TVET, gender-integrated investment planning, cycle monitoring. Augmentation ensures model reliability, leading to SHAP feature importance and STFW15+ dominance explanation [32].



Note: Line chart of FCLP (thousand USD/female worker/year, 2010–2024): Real (n=15): solid green with black dots (ground truth). Synthetic (n=75): red circles. Predicted real: blue dashed - - - (trained on 15 reals only). Predicted mix: red dashed - - - (trained on 90 mix samples).

Figure 4. Comparison of actual, synthetic, mixed predicted, and real-only predicted time series

4.2. SHAP based feature importance

SHAP was applied to interpret the ANN model, providing clear assessment of variable contributions to FCLP prediction. STFW15+ emerges as the dominant factor with the strongest positive nonlinear impact. Figure 5 illustrates the ranking of feature importance based on mean absolute SHAP values. STFW15+ has the highest mean |SHAP| (0.514)—nearly double AMICW and 4–6× higher than GRDPpc/SIEC—confirming female vocational training as the most influential driver of FCLP. Despite similar Pearson correlations (~0.96), SHAP highlights STFW15+'s independent nonlinear strength, overcoming multicollinearity. AMICW is secondary; GRDPpc and SIEC show much weaker average effects (SIEC's negative impact mitigated by STFW15+ in later plots). This directly answers RQ1 and RQ2: STFW15+ dominates over AMICW, GRDPpc, SIEC. SHAP outperforms linear methods, linking descriptive trends (Figures 1–4) to deeper interpretability in Figures 6–7, emphasizing STFW15+'s dual role as top predictor and moderator. National alignment: training rate reached 29.5% in Q4 2025 (+0.9–1.0 pp; GSO 2026), but construction female certification remains low (19–25%), urging specialized gender-responsive TVET [33]. Figure 6 illustrates the distribution of impacts (SHAP values) of each independent variable on FCLP predictions.

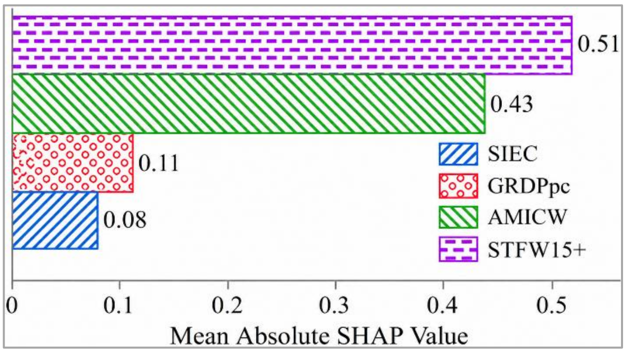
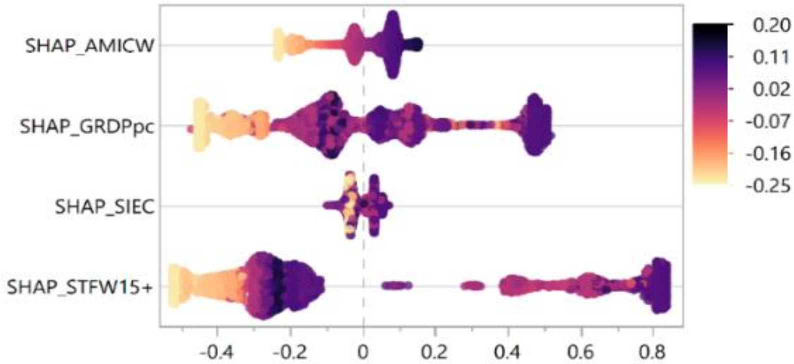


Figure 5. Feature importance ranking based on mean absolute SHAP values



Note: beeswarm shows SHAP distribution across 90 samples: horizontal axis: SHAP value (positive→increase FCLP; negative→decrease). Color: variable value (high: red/yellow; low: purple/blue). Dashed line at 0.

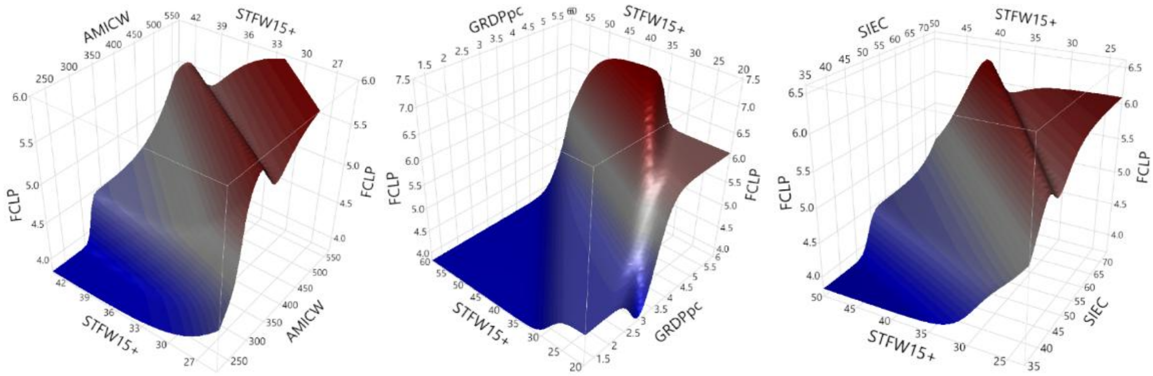
Figure 6. SHAP summary plot (beeswarm) showing feature impacts

Key observations: STFW15+: widest positive skew, dense +0.2 to +0.6, extremes +0.8 at high values (~20–24%)→strongest stable driver (mean |SHAP|=0.514). AMICW: mainly positive (0 to +0.2), smaller impact. GRDPpc: narrow, modest positive (~0 to +0.1), weak lever. SIEC: mostly negative (–0.25 to 0), dense –0.1 to –0.2; training offsets much of negative effect. Figure 6 underscores STFW15+ prominence and signals strong interactions (leading to 3D plots, Figures 7(a)–(c)). Recent GSO 2025–2026 data (training 29.5%, construction growth 9.62%, informal ~63.1%) reinforce STFW15+ as key lever to close gender gaps amid seasonality and skill barriers.

4.3. Nonlinear interaction effects

The three 3D surface plots (arranged left to right: Figure 7(a)–AMICW×STFW15+, Figure 7(b)–GRDPpc×STFW15+, Figure 7(c)–SIEC×STFW15+) illustrate nonlinear interactions between STFW15+ and other variables on predicted FCLP (thousand USD per female worker per year). They complement SHAP results (Figures 5–6), showing STFW15+ as both a strong main driver and powerful moderator. Consistent style: Z-axis=FCLP; color from blue (low) to red (high). To illustrate the nonlinear interactions through 3D surface plots, as seen in Figure 7(a)–(c).

Figure 7(a) (AMICW×STFW15+): Surface low (~4.0, blue) at low AMICW/low STFW15+; rises sharply with STFW15+. Highest peak (~6.0) at high AMICW (>450–500)+high STFW15+(>39–42%). Positive slope with slight waves→strong synergy: training amplifies income's productivity effect. Implication: Trained women leverage higher wages more effectively. Figure 7(b) (GRDPpc×STFW15+): Highest peak (~7.0–7.5) at high GRDPpc (>4.0–5.0)+high STFW15+ (>45–55%). Low region (GRDPpc <2.5, STFW15+<30%) flat (~4.0–4.5, blue). Steep upward slope in high region→sharp red peak. Implication: Strong economic growth multiplies training's productivity impact—clearest evidence to prioritize female training in infrastructure booms. Figure 7(c) (STFW15+×SIEC): lowest (~4.0–4.5, blue) at high SIEC (>60–70%)+low STFW15+(<35%). Surface rises significantly with STFW15+(>45–50%), reaching ~6.0–6.5 even at high SIEC. Local peak at medium SIEC (~45–55%); slight decline at very high SIEC but remains high. Implication: training strongly mitigates informality's negative effect; productivity rises with training despite persistent informality—protective/compensatory role. STFW15+'s moderator role is key in 2025 recovery amid seasonal risks for female workers. GSO 2025–2026 data (industrial-construction +9.73%, construction +9.62%; training rate 29.2% in 2025, 29.5% Q4; informal ~63.1%) confirm raising STFW15+ lifts productivity directly, amplifies GRDPpc/AMICW gains, and compensates SIEC barriers—vital for digital/sustainable construction needing higher skills. The plots confirm STFW15+'s central nonlinear role: amplifying positive drivers (AMICW, GRDPpc) and offsetting informality barriers (SIEC). This transitions to the discussion section for synthesis and policy recommendations on gender equality and female productivity in Vietnam's construction sector [34].



(a) (b) (c)

Note: Horizontal layout for interactions on predicted FCLP; common Z-axis=FCLP; blue-to-red gradient for predicted level.

Figure 7. 3D surface plot: (a) interaction between AMICW and STFW15+ on FCLP, (b) interaction between GRDPpc and STFW15+ on FCLP, and (c) interaction between STFW15+ and SIEC on FCLP

5. DISCUSSION

The findings establish STFW15+ as the strongest predictor of female labor productivity in Vietnam's construction sector (FCLP) over 2010–2024. The mean absolute SHAP value of 0.514 for STFW15+, which is substantially higher than other variables, indicates that vocational training exerts an independent and dominant influence even in the presence of severe multicollinearity [35]. Unlike traditional linear regression or standard ANN models that often fail to disentangle highly correlated predictors, the integration of SHAP with synthetic data augmentation allowed this study to reveal the nonlinear strength and moderator role of STFW15+ [36]. Specifically, the 3D surface plots demonstrate that STFW15+ not only drives direct productivity gains but also amplifies the positive effects of income (AMICW) and economic growth (GRDPpc), while effectively buffering the negative impact of informal employment (SIEC). This moderator effect highlights a critical mechanism: vocational training equips female workers with skills that enable them to better leverage economic opportunities and mitigate structural disadvantages [37]. These results extend human capital theory and gendered labor market theory by providing empirical, explainable evidence in a male-dominated, data-scarce sector [38]. While previous studies have applied ANN to predict construction productivity, they rarely addressed gender-specific dimensions or offered interpretable insights into feature interactions [39]. The current approach therefore fills a methodological gap and strengthens the scientific foundation for gender-responsive TVET research [40]. The significance of these findings goes beyond the construction sector. They offer important ramifications for TVET policy in developing countries, demonstrating that targeted skill development for women can serve as a powerful lever for reducing informality, narrowing gender productivity gaps, and supporting sustainable economic growth.

By aligning with Vietnam's national strategy on gender equality 2021–2030 and the vocational education development strategy, as well as contributing to SDG 5 (gender equality) and SDG 8 (decent work and economic growth), this study provides actionable evidence for advancing inclusive workforce development. From a practical perspective, the findings suggest several pathways for stakeholders. For educators and TVET institutions, there is a need to design flexible, short-term gender-responsive programs that emphasize green construction skills, building information modeling (BIM), occupational safety, and project management tailored to female workers' realities [41]. Policymakers should integrate gender considerations into public investment planning by introducing mandatory training clauses and social insurance requirements in infrastructure contracts, particularly in projects with high female labor participation [42]. Training institutions can adopt similar ANN-SHAP frameworks for regular monitoring and evaluation of TVET program effectiveness, enabling data-driven improvements in curriculum design and outcomes [43]. Evaluations of lean tools and digital technologies have also shown their effectiveness in boosting overall labor productivity in construction, further supporting the integration of advanced skills in TVET programs [44].

Despite the methodological advancements, several limitations should be acknowledged. The original dataset was small (n=15), and although synthetic data augmentation improved model generalization, the results demonstrate strong associations rather than strict causal relationships. National-level data may mask important regional variations in labor markets and training access. Additionally, synthetic data, while rigorously validated, carries inherent risks of introducing artificial patterns. The study also covers data only up to 2024, potentially missing more recent policy impacts in 2025-2026. Future research should address these limitations by utilizing provincial or firm-level datasets to capture heterogeneity, applying causal inference methods such as difference-in-differences (DiD) or instrumental variables (IV) following new TVET policy implementations, and incorporating the effects of climate change, digital automation, and Industry 4.0 technologies on female construction workers [45]. Overall, this study provides robust, interpretable evidence that female vocational training is the pivotal lever for raising productivity, narrowing gender gaps, and supporting sustainable growth in Vietnam's construction sector amid digital and post-COVID transitions.

6. CONCLUSION

This study demonstrates the pivotal role of vocational training in enhancing FCLP in Vietnam's construction sector from 2010 to 2024. By combining ForestDiffusion synthetic data pipeline-time-series augmentation with an ANN and SHAP interpretation, the research successfully overcame the challenges of a small sample size (n=15) and severe multicollinearity, delivering robust and interpretable results. The key contribution lies in identifying STFW15+ as the strongest predictor of FCLP (mean |SHAP|=0.514) and revealing its dual role as both a direct driver and a powerful moderator. It amplifies the benefits of income and economic growth while mitigating the negative effects of informal employment. These findings provide novel, explainable evidence on the effectiveness of gender-responsive TVET in a male-dominated sector. Practically, the results strongly support the following policy recommendations: prioritize specialized, flexible gender-responsive TVET programs focusing on green skills, safety, BIM, and project management; integrate gender considerations into public investment planning with mandatory training and formal contracts in infrastructure projects; and foster public-private partnerships to expand training opportunities for female construction workers. Such measures can help achieve the targets of Vietnam's national gender equality strategy 2021–2030 and contribute meaningfully to SDG 5 and SDG 8. Overall, this study offers a solid foundation for advancing gender equality, improving workforce quality, and promoting SDG in Vietnam's rapidly growing construction sector.

ACKNOWLEDGMENTS

The authors thank Industrial University of Ho Chi Minh City for research facilities, General Statistics Office, International Labor Organization, and World Bank for data support.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Truc Thai Phuong	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓			
Lam Nguyen Son	✓	✓		✓	✓	✓		✓	✓	✓	✓		✓	

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

This study did not involve individuals nor any personal identification information that could require any informed consent.

ETHICAL APPROVAL

This paper does not involve people or animals; no investigation has involved human subjects. Therefore, the authors did not seek approval from any institutional review board.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] National Statistics Office of Vietnam, "Investment and construction." Accessed: Jan. 10, 2026. [Online]. Available: https://www.nso.gov.vn/en/investment-and-construction/

[2] R. Wignall, B. Piquard, and E. Joel, "Up-skilling women or de-skilling patriarchy? How TVET can drive wider gender transformation and the decent work agenda in Sub-Saharan Africa," *International Journal of Educational Development*, vol. 102, p. 102850, Oct. 2023, doi: 10.1016/j.ijedudev.2023.102850.

[3] M. H. Mazlan, S. Puteh, N. L. Sulaiman, K. M. Salleh, and Z. Mohamad, "The domains and components of Industry 4.0 talent for TVET students through work-based learning," *Journal of Technical Education and Training*, vol. 17, no. 3, pp. 106–117, Oct. 2025, doi: 10.30880/jtet.2025.17.03.008.

[4] W. Y. Leong, "Artificial intelligence, automation, and technical and vocational education and training: transforming vocational training in digital era," in *Engineering Proceedings*, Aug. 2025, vol. 103, no. 1, p. 9, doi: 10.3390/engproc2025103009.

[5] Y. Liangyu, K. A. A. Rahman, M. F. Mamat, and S. Hashim, "Beyond the classroom: linking TVET educators' knowledge and skills for Industry 4.0 readiness," *Online Journal for TVET Practitioners*, vol. 9, no. 1, pp. 24–36, Mar. 2024, doi: 10.30880/ojtp.2024.09.01.003.

[6] T. Paramasivam, R. Mustapha, N. Vejaratnam, and S. D. James, "The role of artificial intelligence in the Industrial Revolution 4.0 era and talent development through Malaysian technical and vocational education and training (TVET) programmes," *International Journal of Research and Innovation in Social Science*, vol. 9, no. 3S, pp. 5559–5567, 2025, doi: 10.47772/ijriss.2025.903sedu0404.

[7] R. Parua and W. Yang, "The driving logic of digital transformation in TVET," *Vocation, Technology and Education*, vol. 1, no. 17, pp. 1–11, Jun. 2024, doi: 10.54844/vte.2024.0590.

[8] S. C. David, V. B. Zabala, I. C. Achondo, and S. T. P. Ruales, "Influences on motivation and learning outcomes in home economics: a pathway to enhanced TVET competency development," *Journal of Technical Education and Training*, vol. 17, no. 4, pp. 260–272, Dec. 2025, doi: 10.30880/jtet.2025.17.04.020.

[9] A. Abdeen, S. A. Hoso, and M. Alrefaya, "Dual education and its impact on developing professional skills," *Journal of Arts, Literature, Humanities and Social Sciences*, no. 121, pp. 348–364, Jun. 2025, doi: 10.33193/jalhss.121.2025.1454.

[10] T. Elbashbishy and I. H. El-adaway, "Labor productivity losses across construction trades: a machine learning approach," *Journal of Management in Engineering*, vol. 41, no. 5, Sep. 2025, doi: 10.1061/jmenea.meeng-6671.

[11] C. Dehchar, K. Boudjellal, and M. Bouabaz, "Improvement of productivity in buildings construction," *Selected Scientific Papers-Journal of Civil Engineering*, vol. 18, no. 1, pp. 1–18, Dec. 2023, doi: 10.2478/sspjce-2023-0005.

[12] B. Bayrak, "Prediction of labor productivity in construction processes using artificial intelligence techniques," *Civil Engineering Beyond Limits*, vol. 6, no. 4, pp. 1–7, Oct. 2025, doi: 10.36937/cebel.2025.11075.

[13] B. A. M. Díaz, D. J. L. Peralta, and A. A. Sarmiento, "Application of artificial neural network methodology for the prediction of labor productivity in the construction sector," in *Proceedings of the 22nd LACCEI international Multi-conference for Engineering, Education and Technology*, 2024, pp. 1–8, doi: 10.18687/LACCEI2024.1.1.736.

[14] V. Bhilwade, A. K. Das, P. Nanthagopalan, and K. Modi, "Identification of factors significantly influencing the labour productivity of formwork systems using artificial neural network," in *Proceedings of International Conference on Technological Advancements in Computational Sciences, ICTACS 2022*, Oct. 2022, pp. 632–637, doi: 10.1109/ICTACS56270.2022.9988082.

[15] Y. A. Alsultanny, "Labor market forecasting by using data mining," *Procedia Computer Science*, vol. 18, pp. 1700–1709, 2013, doi: 10.1016/j.procs.2013.05.338.

[16] M. H. Özdemir, B. L. Aylak, C. Cakiroglu, and M. Bağcı, "Prediction of the gender inequality index based on data-driven interpretable ensemble learning methods," *Socio-Economic Planning Sciences*, vol. 103, p. 102366, Feb. 2026, doi: 10.1016/j.seps.2025.102366.

[17] I. Karatas and A. Budak, "Development and comparative of a new meta-ensemble machine learning model in predicting construction labor productivity," *Engineering, Construction and Architectural Management*, vol. 31, no. 3, pp. 1123–1144, Mar. 2024, doi: 10.1108/ECAM-08-2021-0692.

[18] A. Shehadeh, O. Alshboul, and M. F. Tamimi, "Hybrid machine learning solutions for mitigating climate-induced productivity losses in sustainable construction management," *Journal of Cleaner Production*, vol. 538, p. 147285, Jan. 2026, doi: 10.1016/j.jclepro.2025.147285.

[19] Ş. T. Güvel, "Forecasting slipform labor productivity in the construction of reinforced concrete chimneys," *Ain Shams Engineering Journal*, vol. 16, no. 1, p. 103192, Jan. 2025, doi: 10.1016/j.asej.2024.103192.

[20] F. Nasirzadeh, H. M. D. Kabir, M. Akbari, A. Khosravi, S. Nahavandi, and D. G. Carmichael, "ANN-based prediction intervals to forecast labour productivity," *Engineering, Construction and Architectural Management*, vol. 27, no. 9, pp. 2335–2351, May 2020, doi: 10.1108/ECAM-08-2019-0406.
[21] S. Zhang and X. Li, "A comparative study of machine learning regression models for predicting construction duration," *Journal of Asian Architecture and Building Engineering*, vol. 23, no. 6, pp. 1980–1996, Nov. 2024, doi: 10.1080/13467581.2023.2278887.
[22] Q. Zhang, X. Gao, S. U. Khan, and Z. He, "SHAP-ing urban sustainability: Integrating NDDF and lightGBM to uncover coordinated pollution and carbon abatement costs in Chinese cities," *Sustainable Cities and Society*, vol. 130, p. 106542, Jul. 2025, doi: 10.1016/j.scs.2025.106542.
[23] C. Magazzino, M. Mele, and M. Mutascu, "An artificial neural network experiment on the prediction of the unemployment rate," *Journal of Policy Modeling*, vol. 47, no. 3, pp. 471–491, May 2025, doi: 10.1016/j.jpolmod.2024.10.004.
[24] I. S. Pothapragada and R. Sujatha, "GANs for data augmentation with stacked CNN models and XAI for interpretable maize yield prediction," *Smart Agricultural Technology*, vol. 11, p. 100992, Aug. 2025, doi: 10.1016/j.atech.2025.100992.
[25] E. Efekan, T. Celik, and O. B. Tokdemir, "Using hybrid artificial intelligence optimization method to predict construction labour productivity," in *International Symposium of the International Federation for Structural Concrete*, 2023, pp. 1624–1630, doi: 10.1007/978-3-031-32511-3_166.
[26] S. Ebrahimi, A. R. Fayek, and V. Sumati, "Hybrid artificial intelligence HFS-RF-PSO model for construction labor productivity prediction and optimization," *Algorithms*, vol. 14, no. 7, p. 214, Jul. 2021, doi: 10.3390/a14070214.
[27] V. Bhilwade, V. S. K. Delhi, P. Nanthagopalan, A. K. Das, and K. Modi, "Predicting labour productivity for formwork activities in high-rise building construction: a case study," *Asian Journal of Civil Engineering*, vol. 24, no. 4, pp. 959–968, Jun. 2023, doi: 10.1007/s42107-022-00545-6.
[28] E. L. Jacobsen, J. Teizer, S. Wandahl, and I. Brilakis, "Probabilistic forecasting of construction labor productivity metrics," *Journal of Information Technology in Construction*, vol. 24, pp. 58–83, Feb. 2024, doi: 10.36680/j.itcon.2024.004.
[29] A. T. Gurmu, "Econometric analysis of macroeconomic factors influencing construction labour productivity at industry level: evidence from Australia," *Construction Management and Economics*, vol. 44, no. 1, pp. 23–40, Jan. 2026, doi: 10.1080/01446193.2025.2574286.
[30] M. H. Momade, S. Durdyev, S. Dixit, S. Shahid, and A. K. Alkali, "Modeling labor costs using artificial intelligence tools," *International Journal of Building Pathology and Adaptation*, vol. 42, no. 6, pp. 1263–1281, Nov. 2024, doi: 10.1108/IJBPA-05-2022-0084.
[31] P. Goodarzizad, E. M. Golafshani, and M. Arashpour, "Predicting the construction labour productivity using artificial neural network and grasshopper optimisation algorithm," *International Journal of Construction Management*, vol. 23, no. 5, pp. 763–779, Apr. 2023, doi: 10.1080/15623599.2021.1927363.
[32] F. F. Taha, M. A. Ahmed, S. H. R. Aldhamad, H. Imran, L. F. A. Bernardo, and M. C. S. Nepomuceno, "Explainable machine learning-based estimation of labor productivity in rebar-fixing tasks," *Eng*, vol. 6, no. 9, p. 219, Sep. 2025, doi: 10.3390/eng6090219.
[33] P. Das and A. Kashem, "Hybrid machine learning approach to prediction of the compressive and flexural strengths of UHPC and parametric analysis with shapley additive explanations," *Case Studies in Construction Materials*, vol. 20, p. e02723, Jul. 2024, doi: 10.1016/j.cscm.2023.e02723.
[34] N. Alsanabani, K. Al-Gahtani, A. Altuwaim, and A. bin Mahmoud, "A hybrid AI-stochastic framework for predicting dynamic labor productivity in sustainable repetitive construction activities," *Sustainability*, vol. 17, no. 24, p. 11097, Dec. 2025, doi: 10.3390/su172411097.
[35] F. Jian, Q. Liu, C. Feng, Q. Hu, Q. Yu, and Q. Guo, "Critical factors affecting construction labor productivity: a systematic review and meta-analysis," *Buildings*, vol. 15, no. 14, p. 2463, Jul. 2025, doi: 10.3390/buildings15142463.
[36] N. V. Tam, "Impact of self-efficacy on construction labor productivity: the mediating role of work motivation," *Engineering, Construction and Architectural Management*, vol. 32, no. 5, pp. 3407–3431, May 2025, doi: 10.1108/ECAM-11-2023-1114.
[37] A. E. Keser, Ş. Ş. Uçak, M. Kuruoğlu, and O. B. Tokdemir, "Construction labor productivity estimation through machine learning: Performance comparison of regression algorithms," *Journal of Construction Engineering, Management & Innovation*, vol. 8, no. 4, pp. 438–476, Dec. 2025, doi: 10.31462/jcemi.2025.04438476.
[38] H. D. Khanh, S. Y. Kim, and L. Q. Linh, "Construction productivity prediction through Bayesian networks for building projects: case from Vietnam," *Engineering, Construction and Architectural Management*, vol. 30, no. 5, pp. 2075–2100, May 2023, doi: 10.1108/ECAM-07-2021-0602.
[39] K. Y. Kulakov, A. K. Orlov, and V. S. Kankhva, "Evaluation of lean manufacturing tools and digital technologies effectiveness for increasing labour productivity in construction," *Systems*, vol. 11, no. 12, p. 570, Dec. 2023, doi: 10.3390/systems11120570.
[40] A. Kokko and T. T. Nguyen, "Implementing TVET reform in Vietnam: insights from the SOE reform process," *Journal of Southeast Asian Economies*, vol. 42, no. 1, pp. 28–52, 2025, doi: 10.1355/ae42-1b.
[41] R. U. Latief, N. M. Anditiaman, I. R. Rahim, R. Arifuddin, and M. Tumpu, "Labor productivity study in construction projects viewed from influence factors," *Civil Engineering Journal (Iran)*, vol. 9, no. 3, pp. 583–595, Mar. 2023, doi: 10.28991/CEJ-2023-09-03-07.
[42] S. Dabirian, M. Moussazadeh, M. Khanzadi, and S. Abbaspour, "Predicting the effects of congestion on labour productivity in construction projects using agent-based modelling," *International Journal of Construction Management*, vol. 23, no. 4, pp. 606–618, Mar. 2023, doi: 10.1080/15623599.2021.1901330.
[43] O. J. Adebowale and J. N. Agumba, "A meta-analysis of factors affecting construction labour productivity in the Middle East," *Journal of Construction in Developing Countries*, vol. 28, no. 1, pp. 193–220, 2023, doi: 10.21315/jcdc-12-21-0192.
[44] D. A. Nguyen, D. Q. Tran, T. N. Nguyen, and H. H. Tran, "Modeling labor productivity in high-rise building construction projects using neural networks," *Archives of Civil Engineering*, vol. 69, no. 1, pp. 675–692, Apr. 2023, doi: 10.24425/ace.2023.144195.
[45] S. O. Taiwo, B. M. Opeyemi, and S. T. Anike, "Boosting gender equity and female participation in technical vocational education and training in Nigeria: the influence of self-efficacy and social justice awareness," *Gender, Work and Organization*, vol. 33, no. 2, pp. 469–482, Mar. 2026, doi: 10.1111/gwao.70054.

BIOGRAPHIES OF AUTHORS

Truc Thai Phuong is a lecturer at the Faculty of Civil Engineering, Industrial University of Ho Chi Minh City, Vietnam. She holds a Ph.D. in Construction Management from Voronezh State Technical University, Russia (2016). Her research interests include optimization algorithms and machine learning applications in engineering and educational contexts. In this study, she contributed to the data analysis and application of machine learning models to evaluate TVET policies and labor market trends in Vietnam. She can be contacted at email: thaiphuongtruc@iuh.edu.vn.

Lam Nguyen Son is a full-time lecturer at the Faculty of Civil Engineering, Industrial University of Ho Chi Minh City, Vietnam. He holds a Ph.D. in Construction Management from Voronezh State Technical University, Russia (2016). His research focuses on machine learning applications in civil engineering, multidimensional poverty analysis, and human resource development. In this study, he led the development of multi-output regression models and SHAP interpretation to assess barriers to trained labor ratio and propose TVET policy implications in Vietnam. He can be contacted at email: nguyensonlam@iuh.edu.vn.