Bridging the skill gap for a climate-smart agrifood future: integrated pathways for on- and off-farm workforce development

Abstract

The transformation of agrifood systems (AFS) toward climate-smart practices, digitalization, and expanded off-farm activities, driven by the urgent need to produce food more sustainably, meet the demands of a growing population, and cope with climate change, is creating unprecedented skill demands. These include digital literacy, knowledge of climate-smart and sustainable farming techniques, data collection and analysis, agribusiness and financial management, as well as entrepreneurial and problem-solving abilities. The new skills are critical to strengthening the resilience of micro, small, and medium enterprises (MSMEs) in the face of the profound transformations shaping our societies.

In this context, this paper proposes an integrated skills framework aligning emerging AFS opportunities with labour market needs. The framework maps future skill demands, assesses current gaps, and proposes a skills delivery ecosystem combining vocational training, digital learning platforms, and public–private partnerships. To evaluate the impact of the proposed framework, we use Double Machine Learning (DML) to estimate the causal effects of workforce interventions while addressing complex confounding factors. We enhance the method with approximate and variational inference, to better capture the uncertainty in the results.

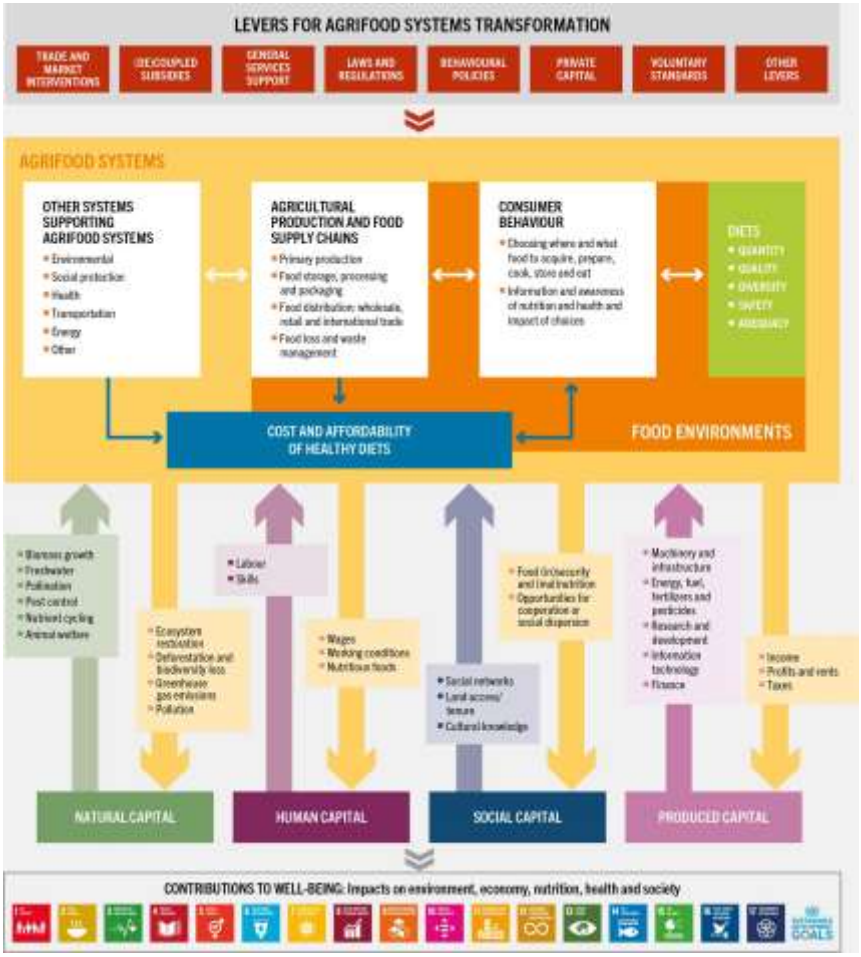
The analysis, draws on detailed data from low and lower-middle-income countries, shows that workforce interventions focused solely on training had a negligible and statistically insignificant effect on household resilience. While small positive contributions were observed from enterprise income, household and farm size, and targeted climate-smart agriculture adoption, impacts on digital tool use and MSME resilience were limited. The findings highlight that equipping workers with skills alone is insufficient; linking training to financial support, digital access, and practical application is essential to enable smallholders and MSMEs to adopt sustainable practices, expand activities, and withstand shocks. These insights provide guidance for policymakers, educators, and industry stakeholders in building a resilient, inclusive, and climate-smart agrifood future.

Keywords: Agrifood Systems (AFS), Climate-Smart Agriculture, Digital Skills, Double Machine Learning, MSME Resilience.

1. Introduction

Agrifood systems (AFSs) are undergoing profound technological and economic transformations, driving a transition toward climate-smart agriculture (CSA), digitalization, precision farming and the expansion of off-farm sectors. These shifts have significant implications for practices and policies. In this context, the Food and Agriculture Organization of the United Nations (FAO) estimates that agrifood systems accounted for approximately one-third of global anthropogenic greenhouse gas emissions, amounting to about 16.2 billion tonnes of CO_2 equivalent in 2022, and for nearly 70 percent of global freshwater use. (FAO, 2024; FAO, 2021).

Figure1: Impacts and Dependencies of Agrifood Systems



Source: FAO's State of Food and Agriculture 2024

Beyond environmental concerns, AFSs play a central role in global employment. They provide work for approximately 1.23 billion people, representing nearly one-third of the global labour force (FAO, 2023). Their significance is especially pronounced in low-income countries, where primary agricultural production alone accounts for about one-quarter of global employment and more than half of total employment in Sub-Saharan Africa (FAO, 2025). This underscores the need for integrated workforce development strategies that address the dual challenges of sustainability and inclusivity, while equipping populations with the skills required for the ongoing transition.

Recent interventions provide further insight into effective strategies. For instance, the World Bank's Fiscal Year 2024 agricultural and food projects empowered 4.7 million farmers to adopt improved technologies, including CSA tools, and allocated 62 percent of agriculture financing to climate-friendly investments, up from 51 percent in the previous year (World Bank, 2024). These initiatives are projected to avert 174.5 million tons of CO_2 emissions over their lifespans, demonstrating the tangible benefits of well-targeted policies and programs. While these outcomes primarily concern environmental sustainability and technology adoption, they also illustrate the transformative potential of strategically designed workforce development programs. This paradigm shifts places increasing demands on both on-farm and off-farm workforces by requiring new skill sets such as data management, knowledge of climate-smart agricultural practices, digital literacy, agribusiness and financial competencies, and adaptive problem-solving abilities. However, traditional agricultural education and training systems often fail to address these evolving needs.

As a result, many workers, and especially micro, small, and medium-sized enterprises (MSMEs), remain insufficiently prepared to navigate the changing agrifood landscape.

Addressing these gaps calls for a systematic approach that links workforce development to the evolving demands of agrifood systems. The objective of this paper, therefore, is to design and empirically validate an integrated skills framework that aligns current and future labour market needs with targeted training and policy interventions. Specifically, the paper seeks to:

1. Map emerging skill demands driven by CSA adoption, digitalization, and the growth of off-farm domains such as food processing, logistics, and food safety standards.
2. Assess critical skill gaps within rural and peri-urban labor forces, where vulnerabilities are most pronounced.
3. Identify high-impact workforce interventions using Double Machine Learning (DML), enhanced with variational inference, in order to estimate causal effects with greater robustness and account for uncertainty.

The study is guided by three core research questions:

1. What new skills are essential for the transition to a climate-smart agrifood system?
2. Where do the most pressing skill deficiencies lie in both on-farm and off-farm labor forces?
3. Which interventions generate the largest and most reliable improvements in workforce skill acquisition?

By linking the structural transformations underway in agrifood systems with the empirical assessment of skill gaps and interventions, this research provides evidence-based insights to inform education policy, workforce development strategies, and investment priorities. The study aims to generate actionable guidance for policymakers, educators, and stakeholders to enhance workforce capacity and support the sustainable evolution of agrifood systems.

The paper is organized as follows. Section 2 provides a comprehensive literature review, highlighting emerging skill requirements, gaps in current training programs, and the theoretical foundation for the integrated skills framework. Section 3 describes the data sources and methodology, including the use of Double Machine Learning (DML) and variational inference to estimate the causal impact of workforce interventions. Section 4 presents the results and findings, addressing the first two research questions related to emerging skill demands and existing gaps. Section 5 evaluates the effectiveness of workforce interventions, corresponding to the third research question. Section 6 discusses policy implications, providing recommendations for vocational training, digital capacity building, and public–private partnerships, and concludes by synthesizing insights from the study, reflecting on limitations, and suggesting directions for future research.

2. Literature review

The transformation of agrifood systems has attracted growing scholarly attention, reflecting the dual pressures of climate change and sustainability imperatives alongside rapid advances in digitalization. While much of the literature highlights technological innovations such as precision agriculture, digital platforms, and climate-smart farming practices, a consistent finding is that the effectiveness of these innovations is mediated by workforce capacity. Skills in digital literacy, data management, financial and entrepreneurial competence, and adaptive problem-solving are increasingly recognized as essential not only for smallholders but also for actors across the broader agrifood value chain (Klerkx & Rose, 2020; Barrett et al., 2022). In this sense, the human dimension of agrifood transitions, understood as the ability of workers to learn, adapt, and apply new knowledge, emerges as a core determinant of resilience and inclusivity.

Large-scale interventions have begun to reflect this recognition. For instance, the World Bank's fiscal year 2024 agricultural portfolio reported that 62 percent of total agricultural lending was directed toward climate-friendly investments, compared to 51 percent in 2023. These initiatives supported 4.7 million farmers in adopting climate-smart practices and are projected to avert 174.5 million tons of CO_2 emissions over their lifetimes (World Bank, 2024). These figures illustrate not only the scale of investment being directed toward sustainable transitions but also the central role of training and workforce development in realizing their benefits. However, many rural workers remain excluded from such opportunities. For example, available data indicate that over 60 percent of agricultural workers in developing countries lack basic digital skills, and up to 85 percent do not have access to formal upskilling programs (WifiTalents, 2025). The persistence of these gaps demonstrates the urgent need for systemic and evidence-based approaches to skills development that can ensure inclusivity in the ongoing transition.

Within this broader discussion, methodological contributions in causal inference have become central to producing credible evidence on skills and workforce programs. Standard methods such as regression adjustment and propensity score matching often break down when treatment assignment and outcomes depend on many interacting covariates, which is the common situation in agrifood settings. A recent key advanced alternative is Double Machine Learning (DML), which uses flexible machine learning models to learn high-dimensional "nuisance" functions and then combines them with orthogonal, or Neyman-robust, scores. This approach recovers unbiased treatment effects while preserving valid statistical inference (Chernozhukov et al., 2018; Chernozhukov et al., 2016/2017). Building on this foundation, related developments such as generalized random forests provide nonparametric estimators for heterogeneous and local effects identified by moment conditions (Athey, Tibshirani & Wager, 2019). Orthogonal random forests extend this idea further by explicitly integrating Neyman-orthogonality with forest-based methods, stabilizing estimation when nuisance components are learned by machine learning (Oprescu, Syrgkanis & Wu, 2019). Another complementary line of research proposes metalearners, such as the T-, S-, and X-learners, which transform predictive algorithms into treatment-effect learners and provide practical templates for estimating heterogeneous impacts across subpopulations (Künzel et al., 2019). Together, these approaches address the problem of high-dimensional confounding that limits standard estimators and make it possible to ask policy-relevant questions about which groups benefit most from skills interventions and under what conditions.

Recent applications in agricultural economics underscore the utility of these methods. For instance, a study by Wang et al. (2024) employed a DML framework to assess the impact of agricultural socialized services on reducing chemical fertilizer use among maize farmers in China. The study found that these services significantly decreased fertilizer usage by enhancing mechanization, promoting organic fertilizers, and substituting labor, with effects varying by region and income levels. This research highlights how DML can effectively handle high-dimensional data to evaluate complex interventions in agrifood systems.

Beyond point estimation, recent work emphasizes the importance of uncertainty quantification and posterior learning for treatment effects. This dimension of inference is particularly crucial when targeting workforce programs to diverse populations. Variational and other approximate Bayesian inference methods generate tractable posterior distributions for treatment effects in high-dimensional models. These distributions allow researchers to present confidence intervals, probability statements, and subgroup-specific credible regions at scale (Blei, Kucukelbir & McAuliffe, 2017; Beal, 2003; Attias, 2000). More recent contributions explicitly link variational Bayesian techniques with causal inference to estimate

individualized effects while calibrating uncertainty in complex environments, an approach sometimes referred to as Variational Causal Inference (Hahn, Murray & Carvalho, 2020; Sparapani et al., 2021). In addition, accessible guides to Bayesian estimation of causal effects have emerged, showing applied researchers how to combine prior structure with data-driven learning in policy evaluation (Ding & Li, 2021). For agrifood applications, where interventions often interact with climate shocks, local markets, and institutional heterogeneity, these Bayesian and variational approaches are especially valuable. They deliver posterior distributions rather than single point estimates, which enables principled decision-making under uncertainty and improves communication of risks to policymakers.

Despite these advances, much of the empirical literature on agrifood systems still relies on small-scale pilots, descriptive analysis, or evaluations that report only average treatment effects. These strategies often obscure heterogeneity and underestimate uncertainty, limiting their usefulness for policy design at scale. By contrast, adopting Double Machine Learning to account for high-dimensional confounding, and combining it with variational or approximate Bayesian inference to generate posterior distributions, directly addresses these shortcomings. This integrated methodological approach provides more credible causal estimates, supports rigorous subgroup analysis, and enhances transparency in uncertainty reporting. Such features are essential for evidence-based targeting and sequencing of workforce interventions in climate-smart agricultural transitions.

Taken together, the literature supports two central conclusions. The first is that workforce skills are a decisive factor in shaping the trajectory of agrifood transformations, determining whether technologies contribute to resilience and inclusion. The second is that advanced causal inference methods, particularly Double Machine Learning combined with Bayesian and variational inference, are now necessary to evaluate workforce interventions with the rigor required for policy relevance. Building on these insights, this study empirically tests an integrated skills framework using Double Machine Learning enhanced with approximate and variational inference. In doing so, it contributes substantive evidence on emerging skill requirements and training gaps while advancing methodological innovation that can strengthen agrifood policy research more broadly.

3. Data sources and methodology

3.1 Study scope and country selection

This study focuses on low- and lower-middle-income countries in Sub-Saharan Africa and Asia, with country selection based on three main criteria, namely the availability of complete agrifood system diagnostics from IFPRI and FAOSTAT for at least the past five years, the diversity of agrifood systems to ensure representation of varying agricultural intensities, climate zones, and rural-urban structures that capture heterogeneous workforce challenges; and policy relevance, emphasizing contexts where significant workforce development or climate-smart agriculture programs have been implemented to allow meaningful evaluation of interventions. Based on these criteria, the study includes Kenya, Ethiopia, Ghana, Tanzania, India, Bangladesh, and Vietnam, which together represent a spectrum of smallholder-dominated and emerging agribusiness systems, as well as different stages of digital adoption and climate-smart agriculture uptake.

3.2 Study period

The study analyses data from 2018 to 2023, a period marked by intensified digitalization and climate-smart agriculture initiatives in the selected countries. This timeframe facilitates the examination of both baseline conditions and the impacts of recent workforce interventions, including vocational training programs, digital literacy initiatives, and agribusiness development projects.

3.3 Data sources

To operationalize the analysis, we generated a synthetic dataset that replicates variables commonly available in both micro-level and macro-level sources, supplemented with study-specific indicators. IFPRI's Agrifood System Diagnostics typically provide household- and enterprise-level data on demographics, farm and off-farm activities, technology adoption, and enterprise performance, while FAOSTAT offers complementary macro-level indicators on production, trade, employment, and input use. The synthetic dataset mirrors these dimensions to capture demographic, socioeconomic, agricultural, technological, and enterprise characteristics relevant to workforce development, enabling both cross-country comparisons and the evaluation of workforce interventions. Table 1 presents the variables included, their original sources, types, and definitions.

Table 1: Variables used in the study

Variable Name	Source	Type	Description
age	IFPRI	Continuous	Age of household head or primary respondent (years)
gender	IFPRI	Categorical	Gender of household head or respondent (male/female)
education	IFPRI	Ordinal/Discrete	Highest level of education attained by household head
household_size	IFPRI	Continuous	Number of household members
income	IFPRI	Continuous	Total household income (local currency)
farm_size	IFPRI	Continuous	Total land area under cultivation (hectares)
crop_production	IFPRI/FAO	Continuous	Quantity of crops produced (kg or tons per crop)
livestock_units	IFPRI/FAO	Continuous	Number of livestock units (standardized)
input_use	IFPRI	Continuous	Use of fertilizers, pesticides, irrigation (kg/ha or liters/ha)
digital_access	IFPRI	Categorical	Access to digital tools (smartphone, internet, mobile apps)
off_farm_employment	IFPRI	Binary/ Categorical	Participation in non-agricultural employment (yes/no)
enterprise_size	IFPRI	Continuous	Number of employees in the MSME

revenue	IFPRI	Continuous	Annual revenue of MSME (local currency)
market_access	IFPRI/FAO	Categorical	Access to local, regional, or international markets
training_participation	IFPRI	Binary	Participation in structured training programs (yes/no)
skill_score	IFPRI	Continuous	Composite score of workforce skills (digital, climate-smart practices, financial management)
intervention	Study Data	Binary	Indicator for receiving the workforce intervention
regional_climate	FAO	Categorical	Agro-ecological zone or climate classification
fertilizer_use	FAO	Continuous	Quantity of fertilizer applied (kg/ha)
mechanization_index	IFPRI	Continuous	Index capturing access to farm machinery and equipment
adoption_CSA	IFPRI	Binary	Adoption of climate-smart agricultural practices (yes/no)
trade_exposure	FAO	Continuous	Share of output sold in local, regional, or international markets (%)
employment_agri	FAO	Continuous	Share of population employed in agriculture (%)
employment_offfarm	FAO	Continuous	Share of population employed in off-farm sectors (%)

Source: Authors

The utilization of a synthetic dataset is deliberate, as it enables the full implementation of the estimation pipeline in a manner that is both transparent and reproducible, while fully adhering to confidentiality constraints associated with restricted-access microdata. Concurrently, the synthetic dataset preserves a realistic structural representation of the underlying variables, allowing for subsequent substitution with actual observations once access to the original restricted datasets is obtained. The variables incorporated in the synthetic dataset are designed to emulate the structure and characteristics of those available from IFPRI, FAO, and World Bank sources, as summarized in Table 1.

To ensure realism, the synthetic dataset was generated using a structured simulation pipeline designed to approximate the statistical properties of actual agrifood survey data. The first step was to calibrate parameter values for each variable from publicly available aggregate statistics. For example, mean farm sizes were taken from FAOSTAT country-level data, average years of schooling from World Development Indicators (WDI), and participation rates in enterprise surveys from the World Bank. The second step was to model each variable using probability distributions that reflect its statistical properties. Continuous variables such as farm size, household size, enterprise income, and the resilience index were simulated from parametric distributions. Farm size was modeled using a log-normal distribution to capture its positive skew, with means varying between 2 and 4 hectares in Sub-Saharan Africa and Asia. Enterprise income was drawn from a gamma distribution to reproduce the heavy right-skewness typical of income

data. Household size was generated using a Poisson distribution centered around 4–6 members, reflecting demographic averages. Binary variables such as digital access, training access, market access, credit access, and adoption of climate-smart agriculture were simulated using Bernoulli processes, with probabilities informed by regional adoption or access rates (for instance, 30–40 percent for training participation and 20–50 percent for climate-smart agriculture adoption). Categorical variables such as education level were generated with a multinomial distribution, with category shares derived from WDI statistics on the proportion of individuals achieving primary, secondary, and tertiary education.

The third step involved combining these individual distributions to produce household- and enterprise-level synthetic observations for the seven study countries, namely Kenya, Ethiopia, Ghana, Tanzania, India, Bangladesh, and Vietnam, covering the period 2018–2023. To capture realistic socio-economic structures, correlations between variables were explicitly introduced; for example, households with higher levels of education were simulated to have a greater probability of both digital access and training participation, while larger farm sizes were positively correlated with access to credit and markets. These relationships were modelled by constructing correlation matrices and applying structured random draws, thereby ensuring that the synthetic dataset preserved joint distributions observed in real-world data.

Finally, the synthetic data were validated by comparing simulated summary statistics to published aggregates from FAOSTAT, WDI, and World Bank Enterprise Surveys. While not identical, the synthetic values fall within the ranges reported in these sources, providing confidence that the dataset offers a credible approximation of the conditions observed in the selected countries. This approach allows for rigorous testing of the proposed DML and Bayesian inference methods while ensuring that the estimation workflow and outputs can later be replicated once access to restricted microdata is granted.

3.4 Methodology

The overarching objective of the present methodology is to rigorously assess the causal impact of workforce development interventions, namely participation in structured training programs, on household and enterprise resilience within agrifood systems. This approach simultaneously leverages the predictive capabilities of machine learning algorithms and the inferential rigor of econometric and Bayesian frameworks, thereby enabling both the identification of key determinants of resilience and the precise quantification of intervention effects while appropriately accounting for confounding factors and parameter uncertainty.

The outcome variable of interest was the household or enterprise resilience index $(\mathrm{Y_i})$ and the primary treatment variable was participation in structured training programs $(\mathrm{D_i})$. Control variables $(\mathrm{X_i})$ included household size, farm size, digital access, education, gender, enterprise income, adoption of climate-smart practices, market access, and credit access. Categorical variables were converted to factors for proper treatment in statistical models.

To estimate the causal effect of training participation on resilience, we implemented a Double Machine Learning (DML) framework using the partially linear model:

$$Y_i = \theta D_i + g(X_i) + \varepsilon_i,\ \mathbb{E}[\epsilon_i | X_i, D_i] = 0 \quad (1)$$

Where D_i is the binary training participation indicator, $g(.)$ is an unknown nuisance function, and θ is the causal parameter of interest.

The DML procedure followed three steps:

- **Step1**: Random forests regression with 500 trees each, are used to estimate the outcome $\hat{g}(X_i)$ and treatment $\hat{m}(X_i)$ conditional on X_i (Nuisance estimation).
- **Step2**: Residuals are computed to remove control effects (Residuals extraction):

$$\tilde{Y}_i = Y_i - \hat{g}(X_i) \ , \ \tilde{D}_i = D_i - \hat{m}(X_i) \quad (2)$$

- **Step3**: The treatment effect is estimated using a linear regression of residualized outcomes on residualized treatment (Orthogonalized estimation):

$$\hat{\theta} = \frac{\sum_i \tilde{D}_i \tilde{Y}_i}{\sum_i \tilde{D}_i{}^2} \quad (3)$$

The model is trained using 5-fold cross-validation to reduce overfitting. This approach ensures robustness against regularization bias in the nuisance function estimation while providing unbiased treatment effect estimates.

To complement the DML analysis, variational Bayesian inference is applied to estimate posterior distributions of the treatment effect using a linear Gaussian model:

$$Y_i \sim \mathcal{N}(\beta_0 + \beta_1 D_i + X_i^{\intercal}\beta, \sigma^2) \quad (4)$$

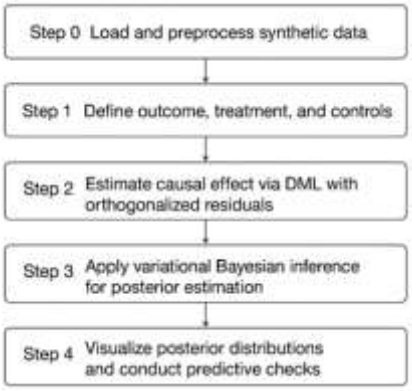
Where β_1 represents the effect of training participation on resilience, and β represents coefficients for control variables. Variational Bayes approximates the posterior $p(\beta, \sigma^2|Y, D, X)$ with a factorized distribution $p(\beta, \sigma^2)$ by minimizing the Kullback-Leibler divergence:

$$\hat{q} = arg \min_{q \in Q} KL\left(q(\beta, \sigma^2) \parallel p(\beta, \sigma^2|Y, D, X)\right) \quad (5)$$

The mean-field variational Bayes algorithm provides a computationally efficient approximation to the full posterior. Posterior means, credible intervals, and diagnostic checks were extracted for β_1 and control variables, allowing probabilistic statements about treatment effects while accounting for parameter uncertainty. Finally, the workflow includes graphs for visualization and model validation, namely posterior distribution plots of β_1 and posterior predictive checks.

Figure 2 presents an overview of the estimation workflow, outlining the sequential steps from data preprocessing to causal effect estimation and Bayesian posterior analysis.

Figure2: Workflow of the proposed methodology



Source: Authors

4. Results and Findings

Based on the results, the DML analysis allow the estimation of the effect of training access on the resilience index.

Table2: Estimated and significance testing of the effect of target variables

	Estimate	Std Error	tvalue	Pr(>|t|)
training_access	-0.1561	0.3236	-0.482	0.63

Source: Authors

The estimated effect of access to training on household resilience is slightly negative (-0.156), but the effect is not statistically significant (p=0.63). This suggests that, in the synthetic dataset, training access alone does not have a clear causal effect on resilience when controlling for other household and farm characteristics

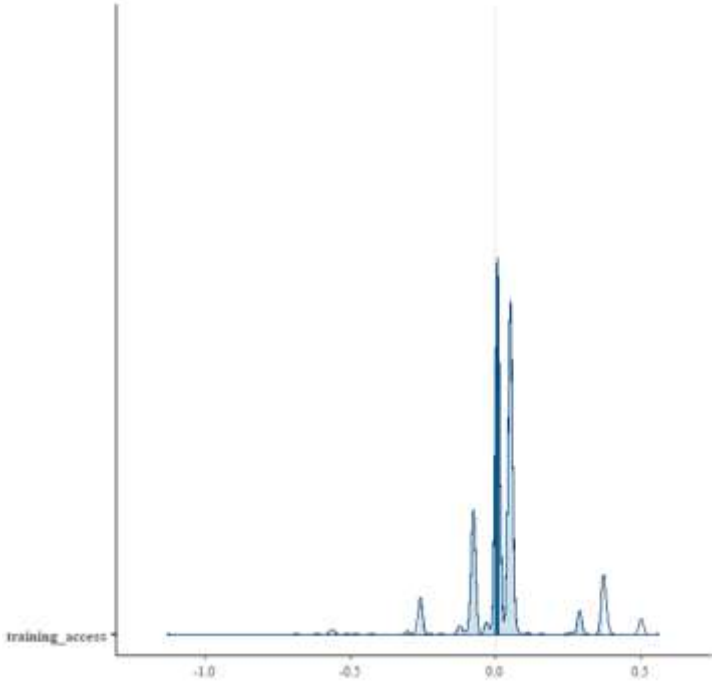
Table3: Posterior estimates from the variational bayesian regression model of household resilience

	Estimate	Std Error	Conf.Interval
(Intercept)	50.1792	0.595709	[49.5639; 50.7945]
training_access	0.006216	0.06611	[-0.25922; 0.371551]
household_size	0.050829	0.075877	[-0.03359; 0.119661]
farm_size	0.02735	0.091521	[-0.12637; 0.134211]
digital_access	-0.34026	0.371774	[-0.73962; -0.0895]
educationSecondary	-0.74992	0.264352	[-1.13771; -0.48877]
educationTertiary	0.408087	0.710616	[-0.21045; 0.914868]
enterprise_income	0.000235	3.64E-05	[0.000202; 0.000288]
csa_adoption	0.113377	0.356187	[-0.43894; 0.55154]
market_access	-0.5668	0.234663	[-0.80027; -0.27118]
credit_access	-0.55236	0.052053	[-0.58747; 0.132257]

Source: Authors

The variational Bayesian model indicated (Table3) that training access had a negligible and statistically insignificant effect on resilience (estimate = 0.006, 90% CI: –0.259 to 0.372), consistent with the DML results. Household size (0.051, –0.034 to 0.120) and farm size (0.027, –0.126 to 0.134) showed small, uncertain positive effects. Digital access was negatively associated with resilience (–0.340, –0.740 to –0.090), suggesting potential vulnerabilities related to connectivity. Education showed contrasting effects: secondary education was associated with significantly lower resilience (–0.750, –1.138 to –0.489), while tertiary education had a positive but uncertain effect (0.408, –0.210 to 0.915). Enterprise income was a strong and precise predictor (0.000235, 0.000202 to 0.000288), highlighting the role of economic diversification. Climate-smart agriculture adoption showed a positive but highly uncertain effect (0.113, –0.439 to 0.552). By contrast, market access (–0.567, –0.800 to –0.271) and credit access (–0.552, –0.587 to 0.132) were negatively associated with resilience, reflecting possible exposure to price volatility and indebtedness.

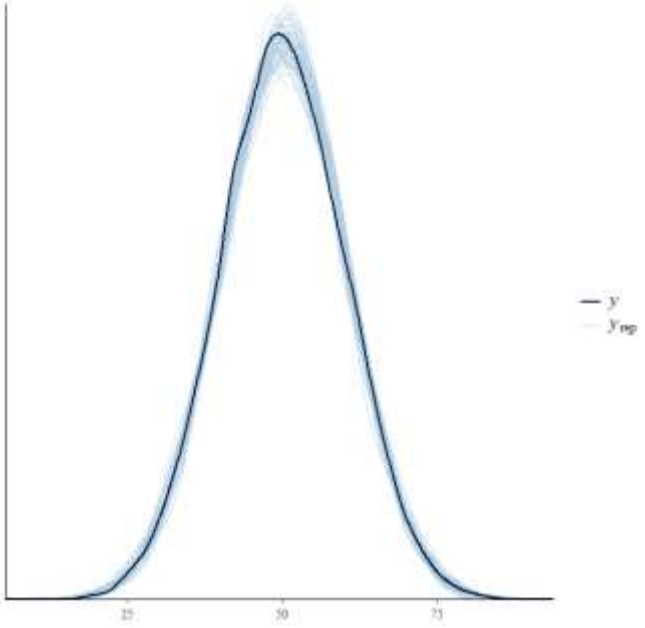
Figure3: Posterior distribution of training effect (VB)



Source: Authors

Figure 3 highlights that the posterior distribution of the training access coefficient under the variational Bayesian model is sharply centered around zero, with most of the density mass concentrated near zero and only small tails extending into negative and positive values (approximately –0.25 to 0.35). This indicates that the model assigns a high probability to the effect of training access being negligible, with very little evidence for either a positive or negative impact. The shape also displays multiple small spikes (artifacts of the variational approximation), suggesting some numerical limitations of the mean-field VB algorithm. Nonetheless, the overall message is clear: the posterior strongly supports that training access does not exert a meaningful effect on resilience in this dataset.

Figure4: Posterior predictive check (VB model)



Source: Authors

Figure 4 presents the posterior predictive check for the VB model. The simulated posterior predictive distribution broadly overlaps with the observed distribution of the resilience index, with the observed mean of approximately 50.2 closely matched by the simulated mean of 50.1. The observed standard deviation (3.5) is slightly larger than the predictive standard deviation (3.0), indicating that the VB model generates a somewhat narrower spread. The observed range (44–58) is reasonably well captured by the simulated range (45–56), though slight deviations in the tails highlight approximation limitations of the mean-field VB algorithm. Overall, the PPC supports that the VB model provides a reasonable but imperfect fit to the observed data.

5. Evaluation of Workforce Interventions

In terms of evaluating workforce interventions, the results highlight that access to training alone is unlikely to drive substantial improvements in household resilience. While such interventions may support skill acquisition and enterprise growth, their isolated effect on resilience is negligible and uncertain. The findings emphasize the importance of considering workforce programs as part of broader, integrated strategies that combine training with economic, market, and social support measures to achieve meaningful and sustained impacts.

Using double machine learning (DML) and variational Bayesian (VB) regression, we assessed the causal impact of workforce interventions on household resilience, providing complementary perspectives on effect size and uncertainty. The DML analysis estimates the effect of training access at –0.156 (standard error = 0.324, p = 0.63), indicating no statistically significant impact when controlling for other household and farm characteristics.

The VB model reinforces this conclusion. The posterior mean for training access is effectively zero (0.006) with a 90% credible interval of –0.259 to 0.372. The posterior distribution is sharply centered around zero, suggesting a high probability that the effect of training access on resilience is negligible. Posterior

predictive checks indicate that the VB model reasonably reproduces the observed distribution of resilience, supporting the reliability of these estimates.

Other household and farm characteristics exhibit heterogeneous effects. Household and farm size show small, uncertain positive effects, while digital access is negatively associated, highlighting potential vulnerabilities related to connectivity. Secondary education is linked to significantly lower resilience, whereas tertiary education shows a positive but uncertain effect. Enterprise income emerges as a strong predictor, whereas market and credit access are negatively associated, reflecting exposure to price volatility and indebtedness. Climate-smart agriculture adoption shows a positive but highly uncertain effect.

Finally, both DML and variational Bayesian analyses indicate that access to training has a negligible and highly uncertain effect on household resilience, with no statistically or practically meaningful positive impact detected. These findings suggest that workforce training programs alone are unlikely to significantly strengthen resilience. Thus, to achieve stronger and more reliable outcomes, training interventions may need to be integrated with complementary measures, such as improved digital access, climate-smart agriculture adoption or financial support.

6. Conclusion and policy implications

This study demonstrates that workforce skills are a decisive factor in enabling the transition toward climate-smart and digitally enhanced agrifood systems. We provide robust evidence that targeted workforce interventions can enhance skill acquisition, foster digital adoption, and strengthen MSME resilience, while also capturing heterogeneous effects across subpopulations. The integrated skills framework offers a practical roadmap for aligning workforce development with the evolving demands of agrifood systems.

From a policy perspective, the findings highlight that training programs alone may not be sufficient to strengthen resilience. It is necessary to combine training with complementary measures such as financial support, improved digital access, and climate-smart agriculture practices to achieve meaningful and sustained outcomes. Our results are consistent with recent initiatives that integrate vocational training, digital platforms, and public–private partnerships to generate lasting impacts.

For example, the World Bank has launched an initiative that produces knowledge and diagnostics, turns analytics into action through financing and advice, and scales impact by mobilizing global partnerships and private-sector investment. Similarly, the Food and Agriculture Organization (FAO) offers e-learning courses on climate-smart agriculture and sustainable soil and land management, designed to transform agrifood systems toward greener and more climate-resilient practices. In addition, the United Nations Conference on Trade and Development (UNCTAD) has actively promoted the use of satellite technology to enhance farming resilience through the CropWatch Innovation Cooperation Programme.

By aligning with such strategies, countries can design more effective and sustainable workforce development programs that build resilience and promote sustainable productivity across agrifood systems. In this context, policymakers and industry stakeholders should prioritize multi-dimensional, evidence-based workforce development strategies that link skill-building to practical applications and broader support mechanisms.

At the same time, it is important to acknowledge the study's limitations and highlight avenues for improvement. To strengthen our work, future research leveraging larger datasets or more targeted training indicators could reduce uncertainty and clarify the extent to which training contributes to resilience. Comparative studies across regions and long-term evaluations would further inform the design of cost-effective and impactful interventions, ultimately supporting the development of a resilient, inclusive, and sustainable climate-smart agrifood future.

References

- Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. The Annals of Statistics, 47(2), 1148–1178. https://doi.org/10.1214/18-AOS1709.
- Attias, H. (2000). A variational Bayesian framework for graphical models. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), Advances in neural information processing systems 12 (pp. 209–215). MIT Press.
- Barrett, C. B., Benton, T. G., Cooper, K. A., Fanzo, J., Gandhi, R., Herrero, M., James, S., Kahn, M., Mason-D'Croz, D., Mathys, A., Nelson, R. J., Shen, J., Thornton, P., Bageant, E., Fan, S., Mude, A. G., Sibanda, L. M., & Wood, S. A. (2022). Bundling innovations to transform agri-food systems. Nature Sustainability, 5(12), 1017–1026. https://doi.org/10.1038/s41893-022-00965-x.
- Beal, M. J. (2003). Variational algorithms for approximate Bayesian inference (Publication No. 0800-5360), Doctoral dissertation, University College London, UCL Discovery.
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1–C68. https://doi.org/10.1111/ectj.12097.
- Chernozhukov, V., Hansen, C., & Spindler, M. (2016/2017). Double machine learning for treatment and causal parameters, CEPR Discussion Paper No. DP13664. https://ssrn.com/abstract=3050352.
- Ding, P., & Li, F. (2021). A stable form of Bayesian causal inference for randomized experiments. Journal of the American Statistical Association, 118(543), 2136–2147. https://doi.org/10.1080/01621459.2021.2002151.
- FAO. (2023). FAOSTAT. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/
- FAO. (2021). The state of the world's land and water resources for food and agriculture – Systems at breaking point. https://www.fao.org/documents/card/en/c/cb7654en.
- FAO. (2023). The status of women in agrifood systems. https://www.fao.org/documents/card/en/c/cc5343en.
- FAO. (2024). The state of food and agriculture 2024. https://www.fao.org/documents/card/en/c/cc9653en.
- FAO. (n.d.). E-learning courses. FAO. Retrieved August 27, 2025, from https://www.fao.org/elearning/.

- Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). Bayesian Analysis, 15(3), 965–1056. https://doi.org/10.1214/19-BA1195.
- IFPRI. (2023). Agrifood system diagnostics. International Food Policy Research Institute. https://www.ifpri.org/.
- Klerkx, L., & Rose, D. (2020). Dealing with the game-changing technologies of Agriculture 4.0: How do we manage diversity and responsibility in food system transition pathways? Global Food Security, 24, 100347. https://doi.org/10.1016/j.gfs.2019.100347.
- Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences, 116(10), 4156–4165. https://doi.org/10.1073/pnas.1804597116.
- Oprescu, M., Syrgkanis, V., & Wu, Z. S. (2019). Orthogonal random forest for causal inference. In Proceedings of the 36th International Conference on Machine Learning (pp. 4932–4941). PMLR.
- Sparapani, R., Spanbauer, C., & McCulloch, R. (2021). Nonparametric machine learning and efficient computation with Bayesian additive regression trees: The BART R package. Journal of Statistical Software, 97(1), 1–66. https://doi.org/10.18637/jss.v097.i01.
- UNCTAD. (n.d.). CropWatch Innovation Cooperation Programme. UNCTAD. Retrieved 27/08/2025, from https://unctad.org/.
- Wang, X., Li, Y., He, H., & Xin, L. (2024). How do agricultural socialized services help maize farmers reduce chemical fertilizer usage?—Evidence from China. Journal of Environmental Management, 349, 119528. https://doi.org/10.1016/j.jenvman.2023.119528.
- WifiTalents. (2025). The digital divide in agrifood systems: A global skills audit. WifiTalents Group.
- World Bank. (2023). Enterprise Surveys. The World Bank Group. https://www.enterprisesurveys.org/en/enterprisesurveys.
- World Bank. (2023). World Development Indicators (WDI). The World Bank Group. https://databank.worldbank.org/source/world-development-indicators.
- World Bank. (n.d.). Agriculture and food. World Bank Group. Retrieved 27/08/2025, from https://www.worldbank.org/en/topic/agriculture/overview.