From Cost Structure to ESG Sustainability: Structural and Cross-Country Evidence from the EU Livestock Sector

Ayagoz Orazbayeva[1,2], Maira Bauer[1,2] and Aliya Ismailova[2]
[1]*Humboldt Innovation GmbH, Humboldt University of Berlin, Unter den Linden 6, 10117 Berlin, Germany*
[2]*Institute of Business and Digital Technologies, Saken Seifullin Kazakh Agrotechnical Research University, Zhenis Avenue 62, 010011 Astana, Kazakhstan*
a.orazbaeva@kazatu.edu.kz, mairak@bk.ru, ismailova61@mail.ru

Keywords: Cost Structure, ESG Sustainability, Livestock Sector, Cross-Country Analysis, Sustainability Assessment, Energy Intensity.

Abstract: The transition toward climate-neutral development in the European Union has intensified the need for structured and data-driven approaches to sustainability assessment in agriculture. Despite extensive research on environmental performance in livestock systems, empirical models explicitly linking managerial cost structure to ESG sustainability outcomes at a cross-country level remain limited. This study develops an integrated analytical framework to examine whether cost allocation patterns and energy structure systematically determine sustainability profiles in the EU livestock sector. The empirical analysis is based on a harmonized panel dataset covering Germany, France, the Netherlands, Spain, and Poland for the period 2000-2023. A multi-stage modelling approach combining k-means and fuzzy clustering, regression analysis, and Random Forest classification was implemented in MATLAB. The clustering stage identified three stable ESG structural profiles. Regression results indicate that energy share and environmental investment allocation significantly explain variation in GHG intensity ($R^2 = 0.9173$). Machine learning classification further confirms the structural importance of energy-related cost components in differentiating sustainability classes. The findings demonstrate that sustainability performance in livestock systems is not random but structurally embedded in managerial cost architecture. The study contributes to sustainability accounting by integrating ESG metrics with sector-level cost indicators and provides a scalable analytical framework for sustainability benchmarking and policy evaluation in the EU agri-food sector.

1 INTRODUCTION

The transition of the European Union (EU) toward climate-neutral development has significantly increased the demand for quantitative tools capable of assessing sustainability performance in agriculture. Livestock production remains one of the most emission-intensive branches of the agricultural sector and is closely associated with methane emissions, energy consumption, and resource allocation efficiency [1], [2]. At the same time, EU climate policy and decarbonization strategies exert growing pressure on agricultural systems to reduce greenhouse gas (GHG) intensity while preserving economic viability [3].

Existing approaches to agricultural sustainability assessment are predominantly based on multi-criteria and composite index frameworks, including systems such as IDEA4 [4], [5]. While these methodologies provide structured evaluation tools, they often rely on predefined weighting schemes and do not fully integrate managerial cost architecture into sustainability analysis. Research in environmental and carbon management accounting demonstrates that operational cost components — such as feed expenditure, fossil energy use, veterinary services, and environmental investments — directly influence environmental performance indicators [6]. However, empirical cross-country models explicitly linking sector-level cost structure to carbon intensity and ESG differentiation remain limited.

Recent advances in machine learning (ML) enable the identification of nonlinear patterns and latent structural relationships in sustainability data. ML-based ESG analytics have been widely applied in corporate finance and environmental performance prediction [7], [8], whereas in agriculture, ML applications predominantly focus on yield forecasting or production monitoring [9]. The integration of clustering techniques and ensemble learning methods into cross-country livestock sustainability modelling remains insufficiently explored.

Furthermore, carbon footprint and emission intensity studies confirm that energy structure and management practices play a decisive role in environmental outcomes [1], [10]. Nevertheless, these analyses rarely incorporate managerial accounting variables into predictive sustainability frameworks, thereby limiting their ability to connect cost allocation decisions with ESG performance.

This study addresses this methodological gap by developing an integrated ML-based analytical framework for assessing ESG sustainability in the livestock sector across five EU countries — Germany, France, the Netherlands, Spain, and Poland — for the period 2000-2023. The proposed framework combines clustering techniques to identify ESG structural profiles, regression modelling to quantify the influence of cost allocation patterns on GHG intensity, and Random Forest classification to determine sustainability classes and assess predictor importance.

The contribution of this research is threefold. First, it integrates ESG sustainability metrics with managerial accounting indicators at the sectoral level. Second, it applies a multi-stage analytical framework to cross-country panel data, enhancing methodological robustness. Third, it provides an interpretable structural linkage between cost architecture, energy dependence, and sustainability differentiation.

The remainder of the paper is structured as follows. Section 2 describes the dataset and methodological approach. Section 3 presents the empirical results. Section 4 discusses implications, and Section 5 concludes.

2 DATA AND METHODS

2.1 Data Description

The empirical analysis is based on a harmonized cross-country panel dataset covering five EU member states — Germany, France, the Netherlands, Spain, and Poland — for the period 2000-2023. These countries were selected to ensure structural heterogeneity within the sample, representing both highly industrialized livestock systems (Germany, the Netherlands, France) and structurally transitional systems (Spain, Poland). Together, they account for a substantial share of EU livestock production and provide variation in technological development, energy dependence, and sustainability investment patterns, enabling meaningful cross-country structural comparison. The dataset integrates production, environmental, and managerial cost indicators derived from FAOSTAT, Eurostat, and national statistical sources. After data harmonization and cleaning, the constructed panel comprises annual observations for each country over the study period.

The analytical variables include sector-level cost structure indicators (share_feed, share_energy, share_vet_medicine, share_labor_wages, share_env_taxes_fees, share_env_investments), environmental performance indicators (GHG_intensity, energy_intensity), and country dummy variables to capture structural heterogeneity across national livestock systems.

To ensure cross-country comparability and eliminate scale effects, all continuous variables were standardized prior to modelling using z-score normalization. This procedure allows consistent estimation within clustering and machine learning algorithms while preserving relative structural differences among observations.

Variables with insufficient observations or structural breaks over the observation period were excluded from subsequent modelling stages to maintain statistical consistency and avoid bias from incomplete time series. The resulting dataset provides a coherent analytical basis for examining the structural linkage between cost allocation patterns and ESG sustainability outcomes in the EU livestock sector.

2.2 Clustering Analysis

To identify latent ESG structural profiles within the EU livestock sector, clustering analysis was conducted using a multi-algorithm approach implemented in MATLAB. Prior to clustering, all continuous variables were standardized using z-score normalization to ensure comparability across indicators and prevent dominance of scale-sensitive variables.

Three complementary clustering techniques were applied:

- K-means clustering, providing compact partitioning of observations by minimizing within-cluster variance;
- Fuzzy C-means (FCM) clustering, allowing partial membership and identification of transitional sustainability trajectories;

- Hierarchical clustering (Ward linkage), enabling structural validation of grouping patterns.

The optimal number of clusters was determined as k = 3 based on within-cluster sum of squares analysis, cluster interpretability, and stability across repeated random initializations. The selection of k = 3 was guided by within-cluster variance reduction and structural interpretability of sustainability profiles. Given the moderate sample size and the panel nature of the dataset, emphasis was placed on stability across repeated initializations and cross-method consistency (k-means, fuzzy C-means, hierarchical clustering) rather than reliance on a single cluster validity metric. Multiple runs of the k-means algorithm confirmed convergence stability and consistent partitioning outcomes.

The clustering input vector included the following standardized variables:

- GHG_intensity;
- share_energy;
- share_feed;
- energy_intensity;
- share_env_investments.

This five-dimensional feature space integrates managerial cost allocation and environmental performance indicators, allowing identification of structurally differentiated sustainability configurations. The combined use of hard (k-means) and soft (FCM) clustering enhances robustness by distinguishing stable structural profiles from transitional observations.

2.3 Regression Analysis of Carbon Intensity

To evaluate the structural relationship between managerial cost allocation and environmental performance, regression analysis was conducted in MATLAB using a linear modelling framework. Greenhouse gas intensity (GHG_intensity) was specified as the dependent variable representing sector-level environmental outcomes across countries and years.

The explanatory variables included energy cost share (share_energy), environmental investment share (share_env_investments), operational energy intensity (energy_intensity), and country dummy variables. The inclusion of country dummies allows control for structural heterogeneity across national livestock systems. Given the limited cross-sectional dimension of the dataset (five countries), a parsimonious linear modelling framework was adopted to preserve interpretability while accounting for cross-country fixed structural effects.

All continuous variables were standardized prior to estimation to ensure comparability across indicators and to avoid scale-driven distortions. The regression was estimated using the full available panel sample within MATLAB's linear modelling environment. Model performance was evaluated using goodness-of-fit indicators, including R^2, root mean squared error (RMSE), and mean absolute error (MAE).

The objective of this regression stage is not to establish strict causal inference, but rather to assess the explanatory capacity of managerial cost architecture in accounting for variation in GHG intensity across countries and years within the proposed sustainability framework.

A potential limitation of the regression specification is the possibility of endogeneity between explanatory variables and GHG intensity. In particular, environmental investment allocation may partly reflect a response to previously observed emission levels rather than acting solely as an exogenous determinant. Given the scope of the study and the aggregate nature of the panel data, the model is interpreted as capturing structural associations rather than strict causal relationships. Addressing endogeneity through instrumental variable approaches or dynamic panel techniques remains an important direction for future research.

2.4 Random Forest ESG Classification

To further evaluate sustainability differentiation, a Random Forest classification model was implemented in MATLAB. The objective of this stage was not predictive forecasting, but structural validation of sustainability profiles and identification of dominant explanatory variables.

Since the constructed regression panel did not contain a predefined ESG class label, the classification target variable was derived from the k-means clustering results (k = 3). The resulting clusters were ordered into sustainability levels, where Class A represents the most sustainable structural profile and Class C represents the least sustainable configuration.

The classification input variables included:

- share_feed;
- share_energy,
- share_env_investments;
- GHG_intensity;
- energy_intensity;
- country dummy variables.

The Random Forest model was trained using MATLAB's ensemble learning framework. Model performance was evaluated using out-of-bag (OOB) validation, which provides an internal cross-validation mechanism and reduces the risk of overfitting.

Permutation-based feature importance was calculated to assess the relative contribution of each predictor to classification accuracy. This procedure enables identification of structurally dominant sustainability drivers within the integrated analytical framework.

The classification stage complements clustering and regression analysis by providing an independent validation of sustainability differentiation and highlighting the explanatory relevance of managerial cost architecture.

2.5 Integrated ML Framework

The proposed methodological architecture integrates clustering, regression analysis, and ensemble classification into a unified analytical system for ESG sustainability assessment in the livestock sector. Rather than treating these methods as isolated analytical tools, the framework applies them sequentially to construct a structurally consistent interpretation of sustainability differentiation.

The analytical sequence operates as follows. First, clustering techniques identify latent ESG structural profiles based on multidimensional cost-emission indicators. This stage reveals endogenous sustainability groupings without imposing predefined classifications. Second, regression analysis quantifies the structural relationship between managerial cost allocation and GHG intensity, providing explanatory insight into emission dynamics across countries and years. Third, Random Forest classification validates sustainability differentiation and determines the relative importance of cost and energy variables in predicting structural sustainability profiles.

The integration of these stages ensures methodological triangulation: clustering uncovers structural patterns, regression explains cost-emission relationships, and machine learning evaluates classification robustness and predictor dominance. This layered approach strengthens internal consistency and reduces reliance on any single modelling technique.

Overall, the framework enables a transition from descriptive sustainability measurement toward structurally interpretable ESG-oriented analytical control. By linking cost architecture, energy dependence, and environmental performance within a coherent modelling system, the proposed approach advances sustainability accounting methodologies at the sectoral level.

3 RESULTS

3.1 Clustering Results

The clustering analysis identified three stable ESG structural profiles across the five EU countries over the period 2000-2023. As illustrated in Figure 1, the observations form three clearly distinguishable groups within the multidimensional cost-emission space. The figure presents a three-dimensional projection of the five-dimensional clustering solution, enabling visualization of structural differentiation across sustainability profiles.

The k-means algorithm (k = 3) demonstrated robust partitioning, confirmed by repeated random initialization and convergence stability across iterations. The resulting clusters represent structurally distinct combinations of managerial cost allocation, energy dependence, and environmental performance indicators.

Cluster 1 corresponds to a resource-intensive structural configuration characterized by a high energy cost share, elevated GHG intensity, and relatively limited environmental investment allocation. This profile is associated with higher energy dependence and comparatively lower ESG sustainability levels. Similar structural contrasts between energy-intensive and environmentally modernized agricultural systems have been identified in prior sustainability assessment research [4], [5].

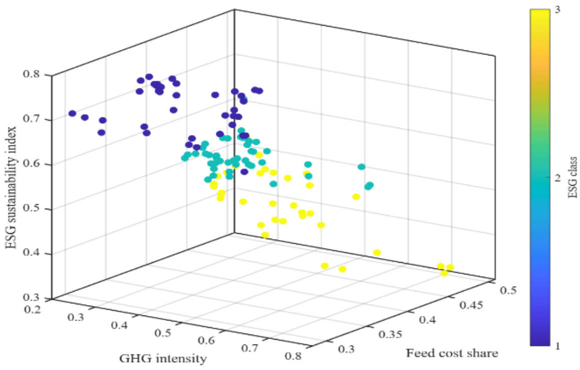

Figure 1: Three-dimensional projection of k-means clustering results (k = 3) showing ESG structural profiles in the EU livestock sector (2000-2023).

Cluster 2 represents a transitional sustainability profile, characterized by moderate indicator values and more balanced cost allocation patterns. Observations within this group demonstrate intermediate energy intensity and moderate environmental investment shares, suggesting ongoing structural adjustment.

Cluster 3 reflects an environmentally modernized structural configuration, characterized by lower GHG intensity, reduced energy intensity, and higher environmental investment allocation. Countries such as Germany and the Netherlands are more frequently associated with this profile in the later years of the observation period, indicating progressive sustainability transformation.

Fuzzy C-means clustering further confirmed the presence of transitional trajectories, particularly for Spain and Poland, suggesting gradual structural evolution rather than abrupt regime shifts. Hierarchical clustering using Ward linkage produced consistent grouping patterns, reinforcing the robustness and stability of the identified ESG structural profiles.

3.2 Regression Analysis of Carbon Intensity

The regression analysis reveals a strong structural association between managerial cost allocation and environmental performance in the EU livestock sector. The estimated model explains a substantial share of variation in GHG intensity, achieving an in-sample goodness-of-fit of $R^2 = 0.9173$.

This high explanatory power reflects the structural interdependence between energy-related cost components and emission intensity at the sectoral level. In particular, energy cost share and operational energy intensity demonstrate systematic positive associations with GHG intensity, indicating that greater dependence on energy inputs translates into higher emission outcomes. Conversely, environmental investment allocation exhibits a mitigating relationship with emission intensity, suggesting that targeted sustainability-oriented expenditures contribute to improved environmental performance.

Country dummy variables capture cross-country heterogeneity in production structures, technological development, and institutional settings. Their inclusion ensures that the identified relationships reflect structural cost-emission linkages rather than purely national differences.

The relationship between observed and predicted values, presented in Figure 2, confirms the consistency of the regression fit across the sample period. The close alignment between actual and predicted GHG intensity values indicates that cost architecture and energy dependence constitute dominant explanatory dimensions within the proposed sustainability framework.

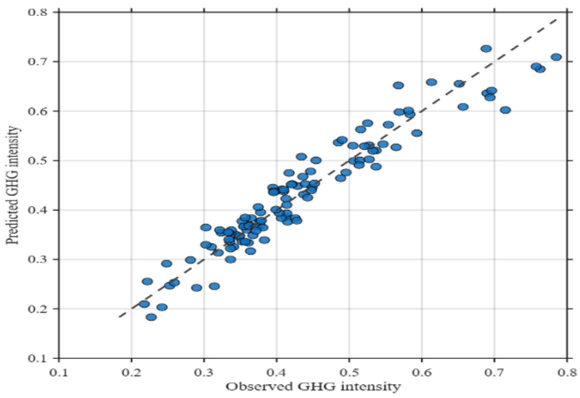

Figure 2: Actual versus predicted GHG intensity values obtained from the regression model ($R^2 = 0.9173$).

It is important to emphasize that the regression stage identifies structural associations rather than strict causal effects. The results demonstrate that variation in emission intensity across countries and years can be systematically explained by managerial allocation patterns, reinforcing the conceptual premise of the study: sustainability differentiation is embedded in cost structure dynamics.

3.3 Random Forest ESG Classification

Random Forest classification was applied to differentiate three sustainability classes (A-C) derived from the k-means clustering stage. Since the regression panel did not contain an externally predefined ESG class label, the clustering output (k = 3) was used to construct the target variable. The clusters were subsequently ordered according to sustainability performance, where Class A represents the most sustainable structural configuration and Class C represents the least sustainable profile.

The model was evaluated using out-of-bag (OOB) validation, ensuring internal consistency without requiring external sample partitioning. Permutation-based feature importance analysis was conducted to quantify the relative contribution of each explanatory variable to classification performance. The resulting importance ranking is presented in Figure 3.

The results indicate that energy share and environmental investment allocation consistently emerge as dominant predictors of sustainability class.

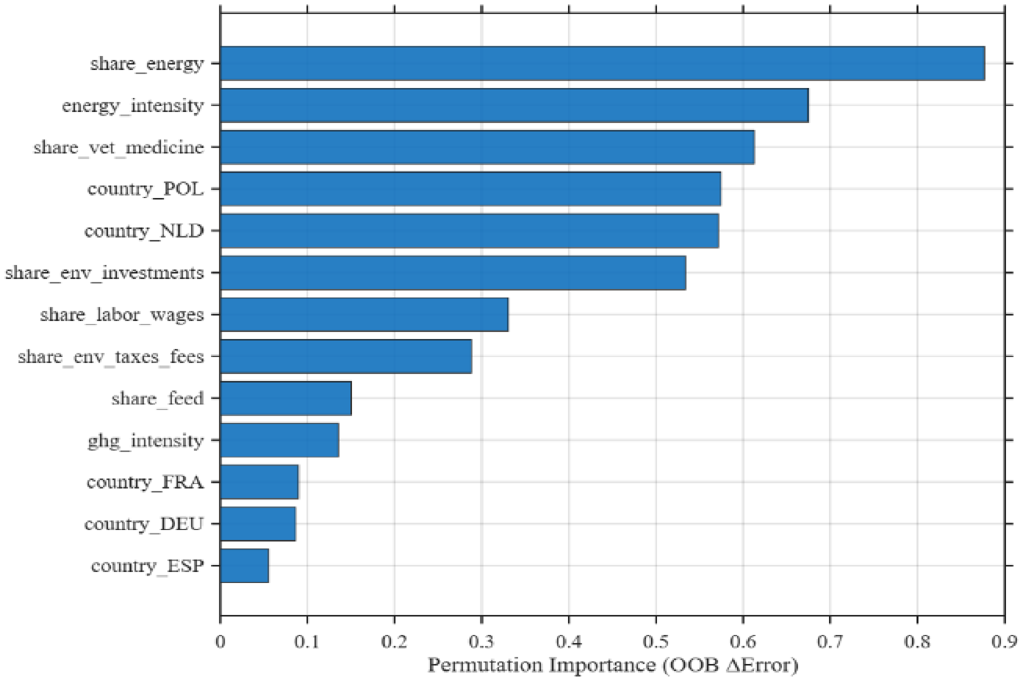

Figure 3: Permutation importance of predictors in the Random Forest ESG classification model.

This finding reinforces the regression results, where energy-related variables explain a substantial share of GHG intensity variation. The convergence between regression coefficients and permutation importance confirms that managerial cost structure is not only associated with emission intensity but also structurally determines sustainability differentiation.

Operational energy intensity further strengthens classification performance, suggesting that efficiency-related indicators act as transmission mechanisms between cost allocation and environmental outcomes. In contrast, country dummy variables demonstrate comparatively lower importance, indicating that sustainability differentiation is primarily driven by internal cost architecture rather than purely national institutional factors.

The alignment between clustering structure, regression associations, and classification importance provides methodological triangulation and strengthens the robustness of the proposed framework. Together, these results support the central argument of the study: ESG sustainability profiles in the EU livestock sector are structurally embedded in managerial allocation patterns and energy dependence dynamics.

4 DISCUSSION

The findings of this study contribute to the growing body of research on ESG-oriented managerial analytics by demonstrating that sustainability differentiation in livestock systems can be structurally derived from cost allocation patterns. Unlike conventional ESG assessments based on composite indices or externally assigned sustainability ratings, the proposed framework embeds sustainability evaluation within sector-level managerial accounting variables.

Although the analysis focuses on five EU member states, the selected sample captures substantial structural diversity within the EU livestock sector. The modelling approach prioritizes interpretability and structural consistency over methodological complexity, ensuring robustness within the constraints of sector-level panel data.

The convergence of clustering, regression, and machine learning classification results strengthens the internal consistency of the modelling approach. Across all analytical stages, energy-related cost components and environmental investment allocation emerge as structurally dominant drivers of sustainability differentiation. This consistency suggests that ESG performance is not an exogenous attribute imposed by institutional conditions, but rather an endogenous outcome of managerial resource allocation decisions.

From a theoretical perspective, the results extend sustainability accounting research by operationalizing the linkage between cost architecture and environmental performance at the sectoral level. The framework moves beyond descriptive sustainability measurement toward structurally interpretable ESG analytics, where cost monitoring functions as an instrument of environmental control. This supports the conceptual transition from traditional cost accounting toward ESG-integrated controlling systems.

From a policy perspective, the findings imply that decarbonization strategies in the livestock sector may be more effective when aligned with internal cost reallocation mechanisms. Targeted incentives for environmental investment and energy efficiency improvements could produce measurable emission reductions by reshaping cost composition rather than relying solely on regulatory constraints.

Nevertheless, several limitations should be acknowledged. The constructed panel relies on macro-level sectoral data, which may not fully capture firm-level managerial heterogeneity. Additionally, the regression framework identifies structural associations rather than causal effects. Future research may extend the model by incorporating micro-level accounting data, dynamic panel techniques, or scenario-based simulation to examine long-term sustainability transitions.

Overall, the study demonstrates that ESG sustainability profiles in the EU livestock sector are structurally embedded in managerial cost allocation and energy dependence dynamics, providing a scalable analytical foundation for sustainability-oriented controlling systems.

5 CONCLUSIONS

This study developed and empirically validated an integrated analytical framework linking managerial cost structure to ESG sustainability differentiation in the EU livestock sector. Using harmonized panel data for five EU countries (2000-2023), the research combined clustering techniques, regression analysis, and Random Forest classification to examine how cost allocation and energy dependence shape sustainability outcomes.

The results demonstrate that GHG intensity is strongly explainable within a cost-based analytical structure, with regression analysis achieving substantial explanatory power ($R^2 = 0.9173$). Machine learning classification further confirms that energy share and environmental investment allocation are dominant determinants of sustainability class differentiation. The consistency across clustering, regression, and classification stages reinforces the structural validity of the proposed framework.

The central contribution of the study lies in demonstrating that ESG sustainability performance in livestock systems is not an externally imposed attribute but a structurally embedded outcome of managerial allocation patterns. By integrating accounting indicators with environmental performance metrics, the research advances ESG-oriented sustainability analytics at the sectoral level.

The proposed framework provides a scalable methodological basis for sustainability benchmarking, strategic investment planning, and ESG-integrated controlling in the agri-food sector. Future research may expand this approach by incorporating firm-level accounting data, dynamic modelling techniques, and broader geographical comparisons to further refine structural sustainability assessment.

ACKNOWLEDGMENTS

The authors acknowledge the academic support provided during the research internship at Humboldt Innovation GmbH, Berlin. The authors also appreciate the institutional support of Saken Seifullin Kazakh Agrotechnical Research University.

REFERENCES

[1] K. R. Lucas and E. Kebreab, “Modeling livestock ecosystem services through carbon footprint,” Ecological Indicators, vol. 181, art. no. 114406, 2025, [Online]. Available: https://doi.org/10.1016/j.ecolind.2025.114406.

[2] X. Kang et al., “Livestock methane emissions in China: Spatiotemporal dynamics and mitigation strategies,” Resources, Conservation and Recycling, vol. 226, art. no. 108695, 2026, [Online]. Available: https://doi.org/10.1016/j.resconrec.2025.108695.

[3] C. Duffy, C. Doedens, R. Moriarty, H. Daly, D. Styles, and M. Meinshausen, “National temperature neutrality, agricultural methane and climate policy: Reinforcing inequality in the global food system,” Environmental Research Letters, vol. 20, no. 9, art. no. 094019, 2025, [Online]. Available: https://doi.org/10.1088/1748-9326/adf12d.

[4] F. Zahm et al., “Assessing farm sustainability: The IDEA4 method, a conceptual framework combining dimensions and properties of sustainability,” Cahiers Agricultures, vol. 33, art. no. 10, 2024, [Online]. Available: https://doi.org/10.1051/cagri/2024001.

[5] T. Deng and Z. F. Mohamad, “Context-sensitive agricultural sustainability assessment: A systematic review of frameworks and local adaptation criteria,” Agricultural Systems, vol. 231, art. no. 104532, 2026, [Online]. Available: https://doi.org/10.1016/j.agsy.2025.104532.

[6] S. R. Ndemewah, K. Menges, and M. R. Hiebl, “Management accounting research on farms: What is known and what needs knowing?,” Journal of Accounting & Organizational Change, vol. 15, no. 1, pp. 58-86, 2019, [Online]. Available: https://doi.org/10.1108/JAOC-05-2018-0044.

[7] G. Sariyer, M. Sarioglu, and B. Ramtiyal, “Machine learning-driven clustering based environmental, social, and governance performance prediction model,” in The Unified International Conference on Emerging Technologies in Cyber-Physical Systems and Industrial AI, Cham: Springer Nature Switzerland, 2024, pp. 661-675, [Online]. Available: https://doi.org/10.1007/978-3-031-95963-9_46.

[8] J. Zhang and Z. Zhao, “Corporate ESG rating prediction based on XGBoost-SHAP interpretable machine learning model,” Expert Systems with Applications, vol. 295, art. no. 128809, 2026, [Online]. Available: https://doi.org/10.1016/j.eswa.2025.128809.

[9] A. Yenkikar, V. P. Mishra, M. Bali, and T. Ara, “An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction,” MethodsX, art. no. 103442, 2025, [Online]. Available: https://doi.org/10.1016/j.mex.2025.103442.

[10] Y. W. Li, “Spatiotemporal investigation into water, land, energy, and carbon footprints of broiler and correlations among them in China,” Discover Sustainability, vol. 6, no. 1, art. no. 794, 2025, [Online]. Available: https://doi.org/10.1007/s43621-025-01760-2.