



---

*Research article*

## **Bridging financial disclosures and ESG ratings: A data-driven predictive framework**

**Kahyun Lee\***

College of Business Administration, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul, 04066, Republic of Korea

\* **Correspondence:** Email: leekahy@hongik.ac.kr; Tel: +823202159; Fax: +823222293.

**Abstract:** This paper examines the extent to which standardized financial statement disclosures can anticipate both overall environmental, social, and governance (ESG) ratings and their evolution over time for listed firms. Utilizing a longitudinal dataset of 851 companies spanning 11 years, we systematically extract 279 International Financial Reporting Standards (IFRS)-compliant accounts and generate an extensive set of financial ratios to capture the static and dynamic features of corporate performance. Through an empirical comparison of traditional machine learning, gradient boosting (LightGBM, XGBoost), and neural network methods (multilayer perceptrons, convolutional neural networks, and TabNet), the analysis finds that boosting algorithms deliver consistently superior accuracy in ESG prediction tasks involving high-dimensional tabular data. Notably, the overall ESG composite score exhibits the highest level of predictability from financial information, while environmental ratings remain more elusive. Further investigation reveals that balance sheet variables most strongly explain absolute ESG levels, whereas cash flow metrics are pivotal in predicting annual changes in the scores. These findings indicate that both corporate financial structures and resource flows encode substantial amounts of information that is relevant to sustainability evaluations. By linking financial accounting theory with ESG analytics, this study provides a rigorous, data-driven framework that offers practical insights for researchers, policy-makers, and market participants seeking to enhance the reliability and timeliness of ESG evaluations using objective financial data.

**Keywords:** ESG; ESG prediction; financial disclosure; machine learning; deep learning

**JEL Codes:** G32, Q56, C38, C43, C55

---

## 1. Introduction

In recent years, the integration of sustainability considerations into corporate decision-making and capital allocation has grown markedly, reflecting global shifts in investors' priorities and regulatory pressures. Environmental, social, and governance (ESG) ratings now serve as widely adopted proxies for nonfinancial performance, shaping asset prices, stewardship strategies, and access to capital (Orlitzky et al., 2003; Friede et al., 2015; Whelan et al., 2021; Seow, 2025a). However, academic debate persists regarding the accuracy, consistency, and construct validity of ESG measurement, especially their relationship to conventional, audited financial disclosures. Recent literature has demonstrated that ESG scores, unlike earlier corporate social responsibility (CSR)-oriented frameworks, provide a more reliable indicator of corporate sustainability performance and robustly reflect material impacts (Whelan et al., 2021; Seow, 2025b). Despite the proliferation of ESG methodologies and their wide uptake in the field of sustainable finance, rating divergence, methodological opacity, and a lack of universal standards remain issues (Berg et al., 2022; van Zanten, 2025). Contemporary work critiques the continued use of outdated indicators and stresses the importance of integrating up-to-date findings regarding data sources, feature engineering, and machine learning methods for ESG analytics (Seow, 2025a). Rigorous frameworks now advocate for transparent, data-driven approaches that map audited financial accounts to sustainability outcomes, ensuring comparability across firms and contexts.

The motivation for this study arises from several research gaps identified in recent systematic reviews: (1) Extant studies frequently rely on narrow, hand-picked indicators or small feature sets, limiting discovery and risking omitted-variable bias; (2) most analyses draw on restricted samples or short panels, reducing generalizability; (3) the field has predominantly focused on static ESG levels, underexploring the drivers of annual rating changes. Drawing on advances in machine learning and comprehensive feature engineering, this study critically examines to what extent ESG ratings and their annual changes can be predicted from public, International Financial Reporting Standards (IFRS)-compliant financial disclosures for listed firms.

Anchored in signaling theory (Spence, 1973; Carrasco and Vilchez, 2022), stakeholder theory (Freeman, 2010; Kumar, 2023), and financial materiality perspectives, the paper applies robust predictive models—including boosting algorithms (LightGBM, XGBoost) and neural network architectures (multilayer perceptrons (MLP), convolutional neural networks (CNNs), and TabNet)—across a uniquely large panel of 851 firms and 279 financial accounts spanning 11 years. The methodology enables the discovery of novel variables and interpretable benchmarking, assessing absolute ESG ratings and annual upgrades/downgrades while systematically quantifying the respective influence of balance sheet and cash flow metrics.

This study offers four main contributions: (1) Quantifying how far audited financials explain ESG assessments across pillars; (2) systematic feature extraction reducing researcher bias; (3) distinguishing levels vs. changes with the roles of the balance sheet (BS) and cash flow statement (CFS); and (4) actionable insights for rating agencies, investors, and issuers on financial–ESG alignment. The paper proceeds as follows: Section 2 reviews the recent literature and key debates, situating the study within the current academic discourse. Section 3 details the methodology, including the data sources, modeling approaches, training procedures, and evaluation metrics. Section 4 presents the main results and their interpretation, highlighting the key findings. Section 5 provides a discussion, synthesizing the insights and connecting them to broader theoretical and practical implications. Section 6 concludes with a summary of the study's contributions and outlines implications for future research and practice.

## 2. Literature review

### 2.1. Theoretical frameworks

Research on the link between sustainability assessments and firms' observable fundamentals has increasingly coalesced around three complementary lenses that motivate this study: Signaling theory, stakeholder theory, and financial materiality. Early meta-analyses document generally non-negative but heterogeneous associations between corporate responsibility and financial outcomes, suggesting that sustainability policies both reflect and shape the organizational processes that are visible in audited accounts (Margolis and Walsh, 2003; Orlitzky et al., 2003; Friede et al., 2015). Investor-preference channels—such as exclusion of “sin” industries—imply pricing effects that intersect with fundamentals (Hong and Kacperczyk, 2009), while eco-efficiency studies show return differentials consistent with financially material performance improvements (Derwall et al., 2005). These findings support a signaling-based view in which firms with superior sustainability practices undertake costly, visible actions—such as cleaner technologies or governance reforms—that convey information about quality and are partially reflected in profitability, leverage, and cash flow patterns (Spence, 1973; Carrasco and Vilchez, 2022).

Stakeholder theory emphasizes that ESG outcomes arise from a firm's explicit and implicit contracts with stakeholders, which, in turn, shape profitability, liquidity, investment, and risk profiles (Freeman, 2010). From this perspective, ESG ratings aggregate market and societal assessments of how well firms balance stakeholders' interests, and stakeholder-salient financial accounts and ratios (e.g., employment stability, supplier terms, prudent leverage) should correlate with ESG assessments. Financial materiality frameworks, as advanced by the Sustainability Accounting Standards Board (SASB) and subsequent IFRS guidance, argue that sector-specific sustainability issues affect an enterprise's value and should be mapped systematically to financially material accounts in audited statements, reinforcing the need to ground ESG metrics in verifiable financial information (Jeffrey, 2010). Eccles et al. (2014) also highlighted how sustainability reporting and assurance choices shape the reliability and decision-usefulness of nonfinancial information, which is crucial for defining meaningful ESG targets and selecting explanatory variables.

Parallel work has examined the construction and consistency of ESG ratings. Berg et al. (2022) and subsequent studies on ESG rating divergence and sustainable development goal (SDG)-based impact metrics document substantial disagreement across providers and decompose it into differences in scope, measurement, and weighting schemes, underscoring the need to specify the target rating and to be transparent about methodological choices (van Zanten, 2025). Recent evidence also shows that ESG scores, when properly designed, can serve as more reliable indicators of corporate sustainability performance than traditional CSR metrics, but only if they are grounded in financially material information and aligned with stakeholders' expectations. Taken together, these studies support a partial predictability perspective: Some components of ESG ratings should be forecastable from financial statements if the raters encode signals that are economically material to the key stakeholders, whereas residual variation reflects genuinely extra financial factors, methodological choices, and noise.

---

## 2.2. Machine learning in ESG analytics

Building on these theoretical frameworks, an emerging stream of research applies machine learning to ESG analytics, with a particular focus on prediction and interpretability. Seow (2025a) provides a systematic review of machine learning based ESG analytics using bibliometric, thematic, and the theory–context–characteristics–methodology (TCCM) framework, documenting rapidly expanding applications in predicting ESG scores, risk assessment, portfolio construction, and detecting greenwashing, while also identifying gaps in feature engineering, target design, cross-provider robustness, and explainability standards. Studies such as D’Amato et al. (2021) on XGBoost-SHapley Additive exPlanations (SHAP) frameworks show that nonlinear tree-based models can effectively capture the structural relationships between a firm’s characteristics and ESG scores, and that post-hoc interpretability tools (e.g., SHAP, partial dependence plots) help reconnect predictions to economically meaningful drivers.

More recent contributions move closer to the setting of this paper by exploiting accounting information and high-dimensional tabular data for ESG predictions. Raza et al. (2022) and Cini and Ferrari (2025) used BS and income statement (IS) variables (e.g., return on asset (ROA), return on equity (ROE), earnings before interests and taxes (EBIT)) to predict ESG scores via artificial neural networks (ANNs), random forests, and gradient boosting, demonstrating that financial fundamentals carry substantial predictive content for sustainability assessments, though typically over restricted feature sets and relatively short panels. D’Amato et al. (2022) also confirmed that BS structure explains ESG ratings using random forests, emphasizing corporate fundamentals where qualitative disclosures are sparse. Other studies emphasize the value of machine learning based ESG analytics for settings with limited formal ESG reporting, including small to medium enterprises (SMEs) and emerging-market firms, arguing that financial-statement-based models can help “democratize” ESG evaluations when direct ESG disclosures are sparse or inconsistent (Fabijańska et al., 2025). Beyond firm-level ratings, research on sustainable debt markets—such as analyses of green bonds’ spread dynamics around macro shocks—indicates that sustainability labels interact with issuer fundamentals and market conditions, consistent with ESG-salient information being at least partly priced in light of financial risk and performance (Wang et al., 2022).

Across this literature, three limitations are recurrent: The reliance on hand-picked indicators rather than comprehensive feature spaces, modest sample sizes and time horizons that constrain generalizability, and a predominant focus on static ESG levels rather than rating changes, despite theoretical arguments that signaling dynamics and stakeholder adjustments should appear in the flows and deltas as well as in the levels. Positioned against these gaps, we operationalize theoretical lenses using comprehensive IFRS features, benchmark diverse models for levels/changes, and apply SHAP for materiality-grounded interpretability.

**Table 1.** Comparison with prior ESG machine learning studies: Methodological advances.

Study	Features (count)	Firms	Period (years)	Targets	Interpretability
Raza et al. (2022)	Manual (6)	300	13	Levels only	None
D'Amato et al. (2022)	Manual (6)	401	11	Levels only	RF feature importance
Cini and Ferrari (2025)	Manual (20)	NR	6	Levels only	RF feature importance
This study	Systematic (39K)	851	11	Levels + changes	SHAP

Table 1 contrasts our framework with prior financial ESG studies. Beyond the scale (39k vs.  $\leq 20$  features) and coverage (851 firms over 11 years using an emerging market panel), we advance the current understanding by (i) predicting rating changes (which have been underexplored, as per Seow (2025a)), (ii) benchmarking nine algorithms including neural nets, and (iii) SHAP theory alignment mapping BS/CF to signaling/materiality, enabling the causal interpretability that has been absent in prior work.

### 3. Methodology

#### 3.1. ESG rating data

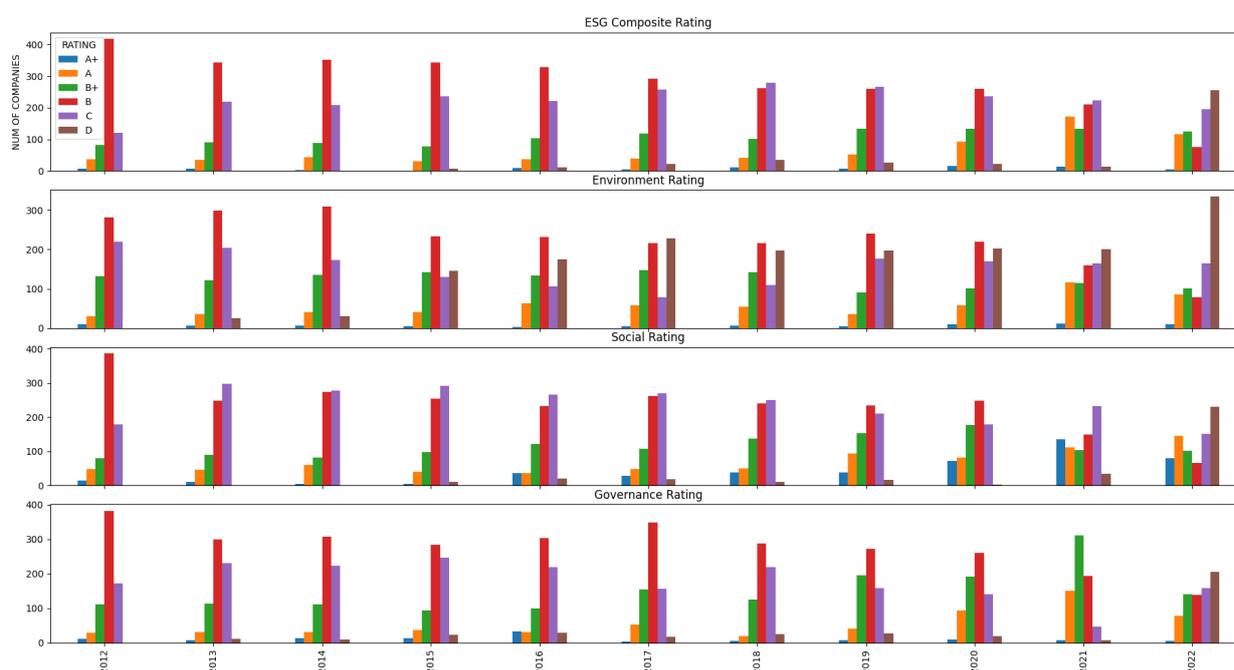
This study utilized ESG rating data provided by the Korea Institute of Corporate Governance and Sustainability (KCGS) for publicly listed companies from 2012 to 2022. The dataset includes four types of ratings: Individual ratings for the E, S, and G components, as well as an overall ESG composite rating (ECR). Each category is evaluated using an absolute grading system, assigning scores across six levels ranging from A+ to D. The highest rating, A+, signifies that a company has established an excellent sustainable management system and consistently demonstrates outstanding performance. Conversely, the lowest rating, D, indicates that a company has a weak sustainable management system requiring significant efforts for improvement. Evaluation models for each category incorporate compliance with the relevant domestic and international regulations (e.g., ISO 14001, Framework Act on Low Carbon and Green Growth, ISO 26000) as well as KCGS's proprietary assessment criteria. These criteria consider factors such as leadership, non-financial risk management, and stakeholder engagement. According to the ECR, the number of evaluated companies steadily increased from 665 in 2012 to 772 in 2022, as detailed in Table 2. The distribution of ratings across the various ESG categories is illustrated in Figure 1. As shown, ESG ratings are not fixed but exhibit significant variability over time. For example, in the environment category, the majority of evaluated companies received a B rating in 2012, whereas by 2022, the most prevalent rating shifted to D. Similarly, in the social category, the proportion of B ratings was initially high; however, as the 2020s progressed, the ratings became more evenly distributed across all levels. This demonstrates the dynamic nature of ESG evaluations over the years.

The choice of KCGS's ESG ratings as the target variable is motivated by three considerations. First, the KCGS provides a comprehensive, longitudinal coverage of Korean listed firms, enabling robust panel analysis with sufficient statistical power to detect financial–ESG relationships—a key gap in prior studies using smaller samples. Second, as a leading domestic provider, KCGS ratings incorporate both global standards (ISO 14001, ISO 26000) and Korea-specific regulatory compliance, making them representative of emerging-market ESG evaluations, where the local context matters. Third, the ratings'

**Table 2.** Number of companies evaluated by year.

Year	Environmental	Social	Governance	ESG composite
2012	673	706	706	665
2013	693	693	693	693
2014	694	694	694	694
2015	696	696	696	696
2016	713	713	713	713
2017	733	733	733	733
2018	727	727	680	727
2019	745	745	698	745
2020	760	760	713	760
2021	765	765	718	765
2022	772	772	726	772

categorical structure (A+ to D) with documented temporal dynamics suits our dual focus on levels and changes, testing predictability under varying granularities.



**Figure 1.** Distribution of ESG ratings by year. Each subplot illustrates the frequency distribution of ESG scores for (a) the overall ESG composite rating, (b) environmental rating, (c) social rating, and (d) governance rating from 2012 to 2022. Bars are grouped by rating level (A+ to D) and colored accordingly to show trends in companies' ESG assessments over time.

### 3.2. Financial statement data

We collected financial statement data for companies subject to ESG ratings during the same evaluation period using the DART (data analysis, retrieval, and transfer system) API (application programming interface). Managed by the Financial Supervisory Service, DART provides public access to the financial disclosures and regulatory filings of audited companies, promoting transparency and ensuring the reliability of financial information. As a result, a total of 198,123 financial statements, comprising four types—BS, CF, CIS (comprehensive income statement), and IS—were collected from 851 companies. South Korea initiated the adoption of IFRS in 2008, with full mandatory implementation for listed companies beginning in 2011. This transition facilitated the use of standardized account categories for electronically disclosed IFRS-compliant financial statements which include BS, CF, CIS, and IS. Notably, financial statements in South Korea are typically presented in a comparative format, allowing for direct year-over-year comparisons to provide greater context and transparency regarding financial performance and trends. While some companies continued to disclose detailed or specialized information using account categories based on the older Korean financial reporting standards, this study focused exclusively on IFRS-compliant account categories to ensure consistency and comparability in the analysis.

The financial statement data used in this study can be categorized into four main types, as shown in Table 3. The first category consists of account-specific amounts for the target year. The second includes account-specific amounts from the prior year. The third category represents the year-over-year ratios of account-specific amounts, comparing the current year with the previous year. Lastly, the fourth category comprises ratios between all accounts within the target year.

**Table 3.** Classification of financial statement data used in this study. The table categorizes the data into four groups: Current-year values, prior-year values, year-over-year ratios, and within-year ratios. Each category is described with its definition and representative examples to clarify.

Category	Description	Example
Current year data	Account-specific amounts for the target year	Revenue, expenses, assets (2022)
Prior year data	Account-specific amounts for the previous year	Revenue, expenses, assets (2021)
Year-over-year ratios	Ratios comparing current year's amounts to the prior year's amounts	Revenue growth rate = (2022 revenue ÷ 2021 revenue)
Current year ratios	Ratios between all accounts within the current year	Debt to equity ratio (2022)

The extracted financial statement accounts from all companies totaled 279 items. Consequently, the first and second categories each contain 279 features, and the third category, representing year-over-year ratios, also comprises 279 features. For the fourth category, which includes all possible ratios between accounts within the same year, the number of features corresponds to the number of two-item combinations from 279 accounts, resulting in 38,781 combinations. Therefore, the total number of features amounts to 39,339 (=279+279+279+38,781).

The selection of the features in this study was driven by the goal of capturing comprehensive and meaningful patterns in the financial data. By incorporating both the absolute values of accounts from the prior year and their year-over-year changes, the feature set allows the model to detect not only the magnitude of financial changes but also the rate and direction of these changes. Additionally, including all possible account-to-account ratios ensures that no potentially meaningful relationships between accounts are overlooked, enabling the identification of latent patterns that may contribute to more accurate predictions and deeper insights. Relying solely on predefined financial ratios deemed significant by humans could hinder the discovery of patterns or relationships that may not be immediately apparent to human perception.

To ensure proper temporal alignment and prevent information leakage, the prediction of ESG ratings for year  $t$  utilized both current-year ( $t$ ) and previous-year ( $t - 1$ ) financial statement data. This design reflects the practical reporting timeline in which ESG ratings are typically assessed after the publication of audited financial statements for the corresponding fiscal year. Thus, the inclusion of both current and lagged data captures the most recent financial position available to evaluators while maintaining temporal validity. A leakage audit was conducted to confirm that no information from future periods ( $t + 1$ ) or beyond was used in training the model or evaluation.

### 3.3. Experiment setting

The experimental targets in this study were divided into two main objectives: (i) Predicting absolute ESG rating, and (ii), predicting changes in the ESG rating, as shown in Table 4. For the prediction of absolute ESG ratings, three approaches were used, based on the level of complexity:

- (Exp. 1) Multiclass classification (four classes): Predicting four ratings (A, B, C, D) without distinguishing the “+” subcategories for A and B.
- (Exp. 2) Multiclass classification (six classes): Predicting six ratings (A+, A, B+, B, C, D), including the “+” subcategories for A and B.
- (Exp. 3) Regression task: Converting each rating into a numerical score (A+ = 6, A = 5, B+ = 4, B = 3, C = 2, D = 1) and predicting the score directly.

For the second target, predicting changes in ESG ratings, the following approaches were implemented:

- (Exp. 4) Multiclass classification (three classes): Predicting whether the ESG rating increased (up = 1), decreased (down = -1) or remained the same (same = 0).
- (Exp. 5) Regression task: Predicting the magnitude of the change in the rating (e.g., a downgrade from A+ to B is represented as -3).

Since each setting involves predicting four categories—environment rating, social rating, governance rating, and ESG composite rating—20 experiments are conducted for each algorithm. Dividing the experimental design into five distinct tasks allows for a comprehensive examination of ESG rating under varying conditions. First, the three approaches for absolute ESG ratings (four classes and six classes classification, and regression) capture different levels of granularity, enabling comparisons between simpler and more detailed categorizations, as well as between classification-based and continuous numeric predictions. Second, the two methods for predicting rating changes (three classes classification and regression) provide insight into whether a rating has shifted—up, down, or the same—and the degree to which it has shifted. By exploring both absolute

ratings and rating changes through multiple modeling strategies, our study ensures a thorough understanding of how the models respond to different representations of the target variable, thus enhancing the overall robustness and interpretability of the findings.

**Table 4.** Experiment settings.

Prediction target	Experiment setting
Absolute rating	(Exp. 1) A multiclass classification task to predict four classes: A, B, C, and D
	(Exp. 2) A multiclass classification task to predict six classes: A+, A, B+, B, C, and D
	(Exp. 3) A regression task where the six classes are converted into numerical scores
Changes in rating	(Exp. 4) A multiclass classification task to predict whether the rating has gone up or down, or remained the same
	(Exp. 5) A regression task to predict the number of steps by which the rating has changed

### 3.4. Machine learning algorithms

In this study, we benchmark a comprehensive suite of models to address both classification and regression tasks, reflecting a finance/machine learning consensus on comparative evaluation. Basic models—ordinary least squares (OLS), support vector machines (SVM), and random forests (RF)—serve as interpretable baselines known for efficiency with structured tabular data. Boosting algorithms, XGBoost (XGB) and LightGBM (LGBM), were selected for their proven ability to handle complex, non-linear patterns with high predictive accuracy, particularly in high-dimensional financial datasets where tree-based methods consistently outperform the alternatives (Grinsztajn et al., 2022). To test deep learning’s potential for capturing intricate feature interactions, we included three neural architectures: MLP, CNNs (O’Shea and Nash, 2015), and TabNet (Arik and Pfister, 2020).

Since most readers are familiar with basic and boosting models, the focus here is on the neural network architectures. MLP, the simplest form, consists of interconnected layers that process tabular inputs through nonlinear transformations, excelling at complex feature interactions without explicit structure. CNNs, adapted from image analysis, treat data as grids to detect local patterns and hierarchical relationships among neighboring financial variables. TabNet, tailored for tabular data, uses attention mechanisms for dynamic feature selection and built-in interpretability via importance scores, aligning with our need for theory-grounded analysis.

Model selection thus benchmarks baselines against state-of-the-art boosting and neural methods, validating superiority for tabular ESG prediction while exploring deep learning limits. Five-fold cross-validation with a grid search ensures robustness across hyperparameter spaces; SHAP provides post-hoc interpretability by quantifying feature contributions consistent with financial materiality frameworks. Temporal alignment—using  $(t)$  and  $(t - 1)$  financials for  $(t)$  ESG ratings—prevents leakage, mirroring real-world rating workflows after an audit disclosure.

### 3.5. Training

During the training phase, each algorithm's performance was assessed using a robust and systematic evaluation methodology. Specifically, a 5-fold cross-validation technique was used to ensure reliable performance estimation and minimize potential biases. In this approach, the dataset was randomly partitioned into five equal folds. For each iteration, one fold served as the test set while the remaining four folds were utilized as the training set, ensuring that every data point was used for both training and testing across the process. To optimize the model's performance, hyperparameter tuning was conducted, focusing on parameters that significantly influence the model's behavior but are not directly learned from the data. A grid search strategy was adopted to systematically explore a range of values for each hyperparameter. Model performance for each hyperparameter configuration was evaluated through the 5-fold cross-validation process, with the goal of identifying the configuration that maximized key performance metrics such as the F1 score and mean squared error (MSE). The final model performance was calculated as the average of the results obtained across all folds, providing a reliable estimate of the algorithm's effectiveness and its potential to generalize to unseen data.

### 3.6. Evaluation

The evaluation of the models varied depending on the type of task and included relevant formulas for clarity. For classification tasks, such as those in experimental settings (Exp. 1, 2, and 4), the models were assessed using the F1 score, which combines precision and recall, as defined in Equations (1)–(3) providing a balanced measure of performance.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

For regression tasks, as in the experimental settings (Exp. 3 and 5), the evaluation was conducted using MSE, a metric that quantifies the average squared difference between the predicted and actual values, as defined in Equation (4).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

Although the ESG ratings are reported as ordered categories, the predictive tasks in this study combine classification and score-based regression, with F1 and MSE serving as our primary evaluation metrics. This choice follows standard practice in recent ESG prediction studies (D'Amato et al., 2022; Cini and Ferrari, 2025; Raza et al., 2022; Sariyer et al., 2024), where F1 summarizes the directional classification performance and MSE captures deviations in the underlying numerical scale used in Exp. 3 and Exp. 5. At the same time, the ordinal structure of the targets is partially reflected in our design, as misclassifications are more heavily penalized in regression-based experiments, where larger step errors (e.g., A vs. C) contribute more to MSE than adjacent deviations (e.g., A vs. A-). Consistent with recent work on ESG prediction and credit-style ratings, we therefore interpret F1 and MSE as

complementary measures of directional accuracy and distance from the true ordered category, while leaving a full exploration of specialized ordinal metrics and calibration diagnostics to future research.

### 3.7. Feature importance

To analyze the contributing factors behind ESG rating predictions, this study uses SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017), a game-theoretic approach widely used to interpret the outputs of machine learning models. Inspired by the Shapley (1953) value, SHAP provides insights into how individual features impact model predictions by assigning each feature an importance value. By applying SHAP to our ESG prediction models, we identify the financial variables that significantly contribute to ESG score predictions, such as profitability ratios, leverage, and liquidity. This analysis quantifies the marginal contribution of each variable to the final ESG rating, offering a deeper understanding of the financial drivers behind ESG performance.

## 4. Results

### 4.1. Prediction performance

Table 5 presents the absolute prediction performance of each algorithm across the three experimental settings. Overall, boosting algorithms achieved superior results compared with the other methods. Out of 12 experiments—comprising four evaluation categories (E, S, G, and ECR) in each of the three settings (Exp. 1, 2, and 3)—LGBM achieved the highest performance in nine instances, XGB did so twice, and TabNet did so once. In Exp. 1, where the ratings were classified into four categories (A, B, C, D), the best F1 scores recorded for E, S, G, and ECR were 0.606, 0.641, 0.678, and 0.652, respectively. When the classification was extended to six categories in Exp. 2 by subdividing A into A+ and A, and B into B+ and B, the performance declined to 0.468, 0.474, 0.459, and 0.521, respectively. Similarly, Exp. 3 also used six categories but each rating was assigned a score ranging from 1 to 6, resulting in MSE values of 0.804, 0.744, 0.630, and 0.538 for E, S, G, and ECR, respectively. Although in rare cases, XGB or TabNet outperformed LGBM in classification accuracy, the margin was minimal. Furthermore, LGBM consistently exhibited the lowest MSE across all tests, indicating it was generally the most effective algorithm for predicting ESG ratings in this study. As for neural network-based models, aside from a single case (the G rating prediction in Exp. 1), their performance was on par with or below that of the basic algorithms. Among the neural network-based algorithms, TabNet achieved the highest performance, followed by CNN, whereas MLP exhibited the lowest performance. Table 6 presents the results for each algorithm's performance in predicting changes in ESG ratings across the different experimental settings. Similar to the findings for absolute rating predictions, the boosting algorithms demonstrated superior performance compared with the other methods. In the context of ESG forecasting based on financial data, neural network-based models—which often demonstrate superior performance in other tasks—underperformed relative to boosting methods in this study. One possible explanation for the relatively lower performance of neural network-based architectures compared to the boosting algorithms lies in the inherent suitability of tree-based models for tabular data characterized by discrete features, missing values, and potential non-linearities (Grinsztajn et al., 2022). In Exp. 4, which involved predicting whether the E, S, G, or ECR ratings would increase, decrease, or remain unchanged, the highest F1 scores achieved were 0.752, 0.620, 0.634, and 0.702, respectively. In Exp. 5, which

---

focused on predicting the degree of rating changes, the MSE values were 0.428, 0.527, 0.569, and 0.391, respectively. Comparing the performance of predicting absolute ratings versus predicting changes in the ratings reveals that forecasting absolute ratings tends to be more challenging. According to the MSE values, the best performance was achieved for ECR among the absolute rating predictions, followed by G, S, and E in descending order. In contrast, when predicting changes in the ratings, the ECR score again was the highest, followed by E, S, and G. Overall, the MSE results indicate that predictions for the ECR rating were more accurate than those for the E, S, and G ratings, whether forecasting the absolute rating or its change.

**Table 5.** Absolute ESG rating prediction performance across different machine learning algorithms. The table reports three evaluation metrics: (Exp. 1) F1 score for binary classification, (Exp. 2) F1 score for multi-class classification, and (Exp. 3) MSE for regression. Each metric is calculated separately for four target variables: environmental (E), social (S), governance (G), and ESG composite rating (ECR). The algorithms are grouped into three categories: Basic (OLS, RF), Boosting (SVMs, XGB, LGBM), and neural networks (MLP, CNN, TabNet). Higher F1 scores and lower MSE values indicate better prediction performance.

Algorithms	(Exp. 1) F1 score				(Exp. 2) F1 score				(Exp. 3) MSE				
	E	S	G	ECR	E	S	G	ECR	E	S	G	ECR	
Basic	OLS	0.443	0.448	0.410	0.315	0.233	0.368	0.334	0.312	3.442	2.746	3.301	3.385
	RF	0.586	0.607	0.648	0.631	0.447	0.446	0.457	0.492	0.821	0.841	0.727	0.639
Boosting	SVMs	0.549	0.538	0.646	0.593	0.351	0.391	0.426	0.424	1.364	0.896	1.144	0.905
	XGB	0.601	0.641	0.657	0.664	0.459	0.475	0.440	0.518	0.887	0.824	0.700	0.598
Neural networks	LGBM	0.606	0.638	0.659	0.672	0.468	0.473	0.459	0.521	0.804	0.744	0.630	0.538
	MLP	0.516	0.528	0.517	0.543	0.343	0.383	0.366	0.426	1.401	0.994	1.005	1.119
	CNN	0.532	0.539	0.618	0.587	0.331	0.366	0.437	0.426	1.130	0.905	0.733	0.897
TabNet	0.582	0.583	0.678	0.608	0.429	0.420	0.458	0.477	0.835	0.847	0.641	0.669	

**Table 6.** Change in ESG rating prediction performance across different machine learning models. The table presents two evaluation metrics: (Exp. 4) F1 score for the classification of rating changes, and (Exp. 5) MSE for regression on the magnitude of rating. Each metric is reported for the E, S, G, and ECR scores. Models are grouped into three categories: Basic (OLS, RF), boosting (SVMs, XGB, LGBM), and neural networks (MLP, CNN, TabNet). Higher F1 scores and lower MSE values reflect better predictive performance.

Algorithms	(Exp. 4) F1 score				(Exp. 5) MSE				
	E	S	G	ECR	E	S	G	ECR	
Basic	OLS	0.527	0.471	0.489	0.499	1.759	2.406	1.978	1.538
	RF	0.712	0.618	0.604	0.682	0.415	0.548	0.533	0.444
	SVMs	0.695	0.602	0.582	0.667	0.446	0.572	0.698	0.472
Boosting	XGB	0.752	0.613	0.626	0.697	0.464	0.593	0.615	0.391
	LGBM	0.750	0.620	0.634	0.702	0.428	0.527	0.569	0.405
Neural networks	MLP	0.715	0.597	0.487	0.630	0.764	0.699	0.987	0.660
	CNN	0.733	0.601	0.560	0.634	0.817	0.625	0.817	0.552
	TabNet	0.745	0.618	0.583	0.669	0.460	0.543	0.682	0.444

## 4.2. Feature importance

Among the boosting algorithms that demonstrated the best predictive performance in the earlier analysis, LGBM was selected for further examination using SHAP. Table 7 presents the five most significant features at the financial statement account level, ranked by importance.

An overall noteworthy observation is the higher prevalence of BS items relative to the other categories. This finding aligns well with the conclusion of D'Amato et al. (2022) that BS information serves as a critical factor in predicting ESG performance. An examination of each ESG category reveals distinct financial drivers underlying corporate sustainability performance. For the environmental rating, 'property, plant, and equipment (PPE)' emerged as a critical factor, likely reflecting the capital-intensive nature of environmentally sensitive operations and the need for substantial investments to meet environmental standards or reduce resource consumption. In the social category, 'current assets' proved to be the most influential, hinting that adequate liquidity may be essential for maintaining stable employment, investing in workforce development, and supporting community engagement initiatives. Meanwhile, 'dividends paid classified as financing activities' stood out within the governance dimension, suggesting that shareholder-oriented capital allocation policies could influence the perceived quality of corporate governance. Finally, the ECR was most affected by 'equity and liabilities', underscoring the importance of sound capital structure and debt management in fostering robust ESG outcomes. These findings highlight how different facets of a firm's BS and CF strategies can uniquely shape stakeholders' views on environmental responsibility, social impact, and governance effectiveness.

To enhance interpretability, the top-ranked financial predictors were further grouped into three broad dimensions: Profitability, liquidity and cash management, and leverage and capital structure. Profitability indicators—such as ROE and operating income—likely capture a firm's ability to generate sufficient internal resources to support sustainability investments and long-term environmental or social programs. Liquidity-related variables, including cash and cash equivalents, and current assets, reflect the short-term financial flexibility that enables firms to respond to stakeholders' expectations and fund ESG-related projects. Finally, leverage and capital structure indicators—such as total liabilities or equity ratio—signal the extent of financial discipline and risk management practices, which are closely tied to governance quality and transparency. This dimensional grouping reveals that firms with stronger profitability, adequate liquidity buffers, and balanced capital structures tend to exhibit higher ESG ratings, consistent with the view that financial resilience underpins credible sustainability performance.

The most critical factors for predicting rating changes are presented in Table 8. A notable observation is the substantial share of CF items. A potential explanation for the prominence of CF items in predicting changes in ESG ratings is that they offer a dynamic perspective on a firm's financial health and strategic priorities. Unlike static BS figures or year-end income statement snapshots, CF data reveals how cash is generated and deployed over time, capturing ongoing operational shifts, liquidity management, and investment activities. This continuous flow of funds can signal whether a company is proactively reallocating resources to improve environmental impact, enhance social initiatives, or strengthen governance mechanisms. It implies a clear distinction between the information used to predict absolute ESG ratings directly from financial statements and the data required to anticipate changes in ESG ratings using financial information.

**Table 7.** Top five most important features for predicting absolute ESG ratings, identified by feature importance scores from the baseline model. Each column corresponds to one ESG dimension: Environmental, social, governance, and ESG composite rating. For each ESG category, the top five ranked features are listed along with their financial statement category (F/S), account name, and source statement: Balance sheet (BS), income statement (IS), cash flow statement (CF), or comprehensive income statement (CIS). This table highlights the most influential financial indicators used in rating prediction models.

Rank	Environmental		Social		Governance		ESG Composite	
	F/S	Account	F/S	Account	F/S	Account	F/S	Account
1	BS	Property plant and equipment	BS	Current asset	CF	Dividends paid as financing activities	BS	Equity and liabilities
2	Ratio (CIS-BS)	Gross Property and equipment	BS	Equity liabilities	BS	Short term borrowings	BS	Assets
3	BS	Non current assets	BS	Current liabilities	BS	Current asset	BS	Non current assets
4	BS	Trade and other current payables	BS	Asset	Ratio (BS-BS)	Deferred liabilities-Liabilities	BS	Current liabilities
5	CIS	Revenue	IS	Gross profit	Ratio (BS-BS)	Investment property-Property and equipment	CIS	Profit loss

**Table 8.** Top five most important features for predicting changes in ESG ratings, based on feature importance scores from the rating change prediction models. The table lists the most influential financial indicators separately for each ESG category—Environmental, social, governance, and ESG composite rating (ECR). Each entry includes the feature’s rank, financial statement category (F/S), account or ratio name, and its source: Balance sheet (BS), income statement (IS), cash flow statement (CF), or comprehensive income statement (CIS). These features were found to be most predictive in capturing year-over-year changes in ESG evaluations.

Rank	Environmental		Social		Governance		ESG Composite	
	F/S	Account	F/S	Account	F/S	Account	F/S	Account
1	Ratio (CF-BS)	Dividends classified as financing activities-Right of use assets	CF	Increase in cash and cash equivalents before effect of exchange rate changes	BS	Inventories decrease	BS	Short term borrowings growth rate
2	Ratio (CF-BS)	Cash flows from used in investing activities-Inventories	Ratio (CIS-BS)	Finance income-Short term borrowings	Ratio (CIS-CF)	Comprehensive income attributable to parent-Dividends paid classified as financing activities	BS	Current liabilities growth rate
3	Ratio (BS-CIS)	Current tax liabilities-gains losses on exchange differences on translation net of tax	Ratio (BS-CIS)	Other financial profit loss	BS	Equity attributable to owners of parent	Ratio (CF-CIS)	Adjustments for increase decrease in trade account payable-Diluted earnings loss per share
4	Ratio (CIS-CF)	Basic earnings loss per share-Cash flows from used in investing activities	Ratio (CF-IS)	Cash flows from used in financing activities Gross profit	BS	Non current provisions growth rate	Ratio (BS-CF)	Current tax liabilities-Increase decrease in cash and cash equivalents
5	Ratio (CF-CIS)	Payments of finance lease liabilities classified as financing activities/Profit loss	CF	Adjustments decrease in trade account receivable	Ratio (CF-BS)	Cash flows used in investing activities/Current liabilities	Ratio (CIS-CIS)	Comprehensive income attributable to noncontrolling interests-Gains losses on exchange differences on translation net of tax

## 5. Discussion

The empirical findings advance our understanding of financial–ESG linkages by confirming the dominance of BS items for absolute ratings, extending prior work on capital structure and profitability (Weber et al., 2008; D’Amato et al., 2022), while uncovering novel drivers such as PPE for environmental ratings that were previously obscured by reliance on hand-picked ratios in smaller feature sets (Seow, 2025a). This BS primacy reflects how long-term capital allocation decisions, including capital expenditures for clean technologies or workforce programs, manifest directly in financial statements and signal sustainability commitments to raters. In contrast, the shift to CF variables for year-over-year rating changes challenges static-focused studies (Cui et al., 2018) and aligns with Cicchiello et al. (2022), who documented flows’ pricing sustainability in green bond markets, capturing real-time strategic reallocations toward green investments, human capital, or governance reforms. The relative predictability—ECR and G exceeding E—validates Berg et al. (2022) on rating divergence, as governance mechanisms embed more transparently in IFRS reporting than environmental externalities like emissions or resource intensity, which often require non-financial data sources.

LGBM’s consistent superiority over neural networks further confirms tree-based models’ efficacy for high-dimensional tabular ESG data (Grinsztajn et al., 2022; D’Amato et al., 2022), operationalizing Seow (2025a)’s recommendations for scalable feature engineering, SHAP interpretability, and financial-statement-based “democratization” of assessments in data-scarce contexts like SMEs (Fabijańska et al., 2025). Theoretically, these BS/CF distinctions operationalize signaling theory (Spence, 1973), where costly sustainability actions leave traces in capital investments; stakeholder theory (Freeman, 2010), linking liquidity buffers to relational contracts; and financial materiality frameworks (Jeffrey, 2010), systematically mapping IFRS accounts to sector-specific issues—revealing ESG’s hybrid composition (ECR MSE=0.538 vs. E=0.804) that reconciles the heterogeneous ESG–financial associations in meta-analyses (Friede et al., 2015).

Practically, the models provide rating agencies with timely, auditable benchmarks to mitigate provider divergence (van Zanten, 2025), enable investors to extract the alpha from BS/CF precursors of ESG trajectories, and guide issuers on resource levers for upgrades; methodologically, systematic feature generations, while LGBM-SHAP enhances transparency in ESG analytics.

Despite the extensive panel, an important limitation of this study is its reliance on a single national rating provider, the KCGS. The rating divergence literature shows that ESG scores can differ substantially across agencies due to differences in scope, measurement, and weighting, implying that the financial predictability documented for one framework may not generalize mechanically to others. In this sense, our results should be interpreted as establishing financial alignment with one coherent, domestically relevant methodology rather than as universal evidence for all ESG ratings. At the same time, the KCGS focus provides a clean testbed where the regulatory context and rating practices are relatively homogeneous, which is valuable for isolating the informational content of audited financial statements. Future research can strengthen the external validity in at least two ways. First, cross-provider benchmarking could replicate our framework on global raters such as Morgan Stanley Capital International (MSCI) and Refinitiv, directly testing how provider-specific design choices affect financial predictability and feature importance profiles. Second, cross-country extensions that apply the same IFRS-based feature engineering to firms in other jurisdictions would allow a systematic assessment of whether the BS-level versus CF-change patterns observed here reflect Korea-specific

---

institutional features or more universal materiality channels. Together, such studies would transform our KCGS-based evidence into a broader comparative research program on when and why financial disclosures predict ESG ratings.

## 6. Conclusions

This study examined the extent to which ESG ratings can be predicted from publicly available financial statement data and found that a meaningful degree of predictability exists despite ESG assessments and financial disclosures originating from distinct domains. Across all model configurations, boosting algorithms consistently outperformed both traditional and neural network-based approaches, indicating that tree-based methods are particularly well suited to capturing the complex, non-linear relationships between financial indicators and ESG performance. Among the targets considered, the composite ECR rating was the most predictable, whereas the environmental pillar remained relatively difficult to infer purely from financials, underscoring the limited visibility of environmental externalities within conventional accounting frameworks. The feature importance analysis further revealed a clear division of roles: BS items are central for explaining absolute ESG levels, while CF variables play a more prominent role in predicting year-over-year rating changes, suggesting that long-term capital structure and short-term liquidity management jointly shape how corporate sustainability is assessed.

At the same time, several limitations of the present design should be acknowledged. First, the analysis relied on categorical ESG ratings rather than continuous scores, which may compress variation and obscure nuances near the rating thresholds. Second, the empirical setting is confined to a single national rating provider, implying that the reported predictability should be interpreted as an alignment with that provider's scoring logic rather than as an objective measure of sustainability. Third, the models focus on financial determinants and do not directly incorporate non-financial or forward-looking information—such as emissions data, governance incidents, or strategic climate commitments—that often drive substantive sustainability outcomes. These constraints delimit the scope of inference but also point toward a clear agenda for future research. Subsequent studies could exploit continuous ESG scores where available, compare results across multiple global rating agencies to assess robustness under rating divergence, and extend the framework to integrate textual disclosures, environmental metrics, and alternative data sources. Pursuing these directions would deepen our understanding of how financial transparency interacts with broader sustainability information and help clarify the conditions under which ESG ratings reflect more than re-labeled accounting fundamentals.

### 6.1. Implications

The findings carry several theoretical, methodological, and practical implications. Theoretically, the documented partial predictability of ESG ratings from audited financial statements supports the view that sustainability attributes are embedded in firms' financial architecture rather than being entirely orthogonal to traditional performance measures. The distinct roles of BS and CF variables reinforce signaling, stakeholder, and financial materiality perspectives by showing that both long-horizon capital allocation and short-term resource reallocation are systematically associated with observed ESG outcomes. At the same time, the lower predictability of the environmental pillar highlights the continued need for non-financial and impact-based metrics, cautioning against interpretations that equate financial predictability with full substantive coverage of sustainability performance.

From a methodological standpoint, the results demonstrate the value of combining gradient boosting algorithms with high-dimensional, IFRS-based feature spaces and post-hoc interpretability tools. The boosting models' consistent superiority over both baseline and neural architectures suggests that, for tabular financial data, carefully tuned tree-based methods remain a strong benchmark for ESG prediction tasks. The SHAP-based importance analysis shows how complex models can be decomposed into intuitive financial drivers, offering a transparent link between machine learning outputs and underlying accounting items. This combination of predictive performance and interpretability provides a blueprint for future data-driven ESG analytics that aims to avoid black-box decision-making.

Practically, the study suggests that financial statement-based prediction models could function as complementary tools within the ESG rating ecosystem rather than as replacements for expert judgment. Rating agencies and auditors might use such models as secondary validation mechanisms to flag potential inconsistencies, detect biases, or prioritize cases for closer review, thereby enhancing the transparency, consistency, and auditability of their assessments. For investors and regulators, the ability to approximate ESG ratings from audited disclosures offers a scalable way to monitor sustainability performance in markets or segments where dedicated ESG coverage is sparse or delayed. At the same time, the risk that ESG scores could drift toward being mere transformations of financial metrics underscores the importance of maintaining a deliberate balance between quantitative financial predictors and qualitative, forward-looking sustainability indicators. Future work on integrating structured financial variables with richer non-financial data will be essential to ensure that ESG evaluation continues to fulfill both its analytical function for capital markets and its broader societal role in directing resources toward genuinely sustainable corporate behavior.

### **Use of AI tools declaration**

The author used AI tools for language editing and grammar review of this manuscript. The author is fully responsible for the content of this publication.

### **Acknowledgments**

This work was supported by 2025 Hongik University Research Fund.

### **Conflict of interest**

The author declares no conflicts of interest in this paper.

### **References**

- Arik SO, Pfister T (2020) TabNet: attentive interpretable tabular learning. *arXiv preprint arXiv:1908.07442*. <https://doi.org/10.48550/arXiv.1908.07442>
- Berg F, Kölbel JF, Rigobon R (2022) Aggregate confusion: the divergence of ESG ratings. *Rev Financ* 26: 1315–1344. <https://doi.org/10.1093/rof/rfac033>

- Carrasco PO, Vilchez VF (2022) Sending corporate social responsibility signals: What organizational characteristics must be met? *Rev Bras Gest Neg* 24: 92–111. <https://doi.org/10.7819/rbgn.v24i1.4146>
- Cicchello AF, Cotugno M, Monferrà S, et al. (2022) Which are the factors influencing green bonds issuance? Evidence from the European bonds market. *Financ Res Lett* 50: 103190. <https://doi.org/10.1016/j.frl.2022.103190>
- Cini F, Ferrari A (2025) Towards the estimation of ESG ratings: A machine learning approach using balance sheet ratios. *Res Int Bus Financ* 73. <https://doi.org/10.1016/j.ribaf.2024.102653>
- Cui J, Jo H, Na H (2018) Does corporate social responsibility affect information asymmetry? *J Bus Ethics* 148: 549–572. <https://doi.org/10.1007/s10551-015-3003-8>
- D’Amato V, D’Ecclesia R, Levantesi S (2021) Fundamental ratios as predictors of ESG scores: a machine learning approach. *Decis Econ Financ* 12: 1087–1110. <https://doi.org/10.1007/s10203-021-00364-5>
- D’Amato V, D’Ecclesia R, Levantesi S (2022) ESG score prediction through random forest algorithm. *Comput Manage Sci* 19: 347–373. <https://doi.org/10.1007/s10287-021-00419-3>
- Derwall J, Guenster N, Bauer R, et al. (2005) The eco-efficiency premium puzzle. *Financ Anal J* 61: 51–63. <https://doi.org/10.2469/faj.v61.n2.2716>
- Eccles RG, Ioannou I, Serafeim G (2014) The impact of corporate sustainability on organizational processes and performance. *Manage Sci* 60: 2835–2857. <https://doi.org/10.1287/mnsc.2014.1984>
- Fabijańska A, Wołczek P, Sikacz H (2025) Can machine learning bring ESG ratings closer to small and medium-sized enterprises? *Oecon Copernic*. <https://doi.org/10.24136/oc.3162>
- Freeman RE (2010) *Strategic management: A stakeholder approach*. Cambridge University Press.
- Friede G, Busch T, Bassen A (2015) ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *J Sustain Financ Invest* 5: 210–233. <https://doi.org/10.1080/20430795.2015.1118917>
- Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on tabular data? *arXiv preprint arXiv:2207.08815*. <https://doi.org/10.48550/arXiv.2207.08815>
- Hong H, Kacperczyk M (2009) The price of sin: the effects of social norms on markets. *J Financ Econ*. <https://doi.org/10.1016/j.jfineco.2008.09.001>
- Jeffrey H (2010) Sustainability Accounting Standards Board (SASB), In: *World Scientific Encyclopedia of Climate Change*, 37–41.
- Kumar S (2023) A review ESG performance as a measure of stakeholders theory. *Acad Market Stud J* 27.
- Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30.

- Margolis JD, Walsh JP (2003) Misery loves companies: rethinking social initiatives by business. *Admin Sci Q* 48: 268–305. <https://doi.org/10.2307/3556659>
- Orlitzky M, Schmidt FL, Rynes SL (2003) Corporate social and financial performance: a meta-analysis. *Organ Stud* 24: 403–441. <https://doi.org/10.1177/0170840603024003910>
- O’Shea K, Nash R (2015) An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*. <https://doi.org/10.48550/arXiv.1511.08458>
- Raza H, Muhammad AK, Mazliham MS, et al. (2022) Applying artificial intelligence techniques for predicting the environment, social, and governance (ESG) pillar score based on balance sheet and income statement data: A case of non-financial companies of USA, UK, and Germany. *Front Environ Sci* 10. <https://doi.org/10.3389/fenvs.2022.975487>
- Sariyer G, Mangla SK, Chowdhury S, et al. (2024) Predictive and prescriptive analytics for ESG performance evaluation: A case of Fortune 500 companies. *J Bus Res* 181: 114742. <https://doi.org/10.1016/j.jbusres.2024.114742>
- Seow RYC (2025a) Transforming ESG analytics with machine learning: a systematic literature review using TCCM framework. *Corp Soc Responsib Environ Manag* 32: 1–20. <https://doi.org/10.1002/csr.70089>
- Seow RYC (2025b) Clarifying CSR and ESG: Causes of conflation, consequences, and pathways to conceptual clarity. *J Environ Manage* 394: 127468. <https://doi.org/10.1016/j.jenvman.2025.127468>
- Shapley LS (1953) A value for n-person games. In: *Contributions to the Theory of Games II*, Princeton University Press, 307–317.
- Spence M (1973) Job market signaling. *Q J Econ* 87: 355–374. <https://doi.org/10.2307/1882010>
- van Zanten JA (2025) Measuring Companies’ Environmental and Social Impacts: An analysis of ESG Ratings and SDG Scores. *Organ Environ* 38: 403–439. <https://doi.org/10.1177/10860266251326895>
- Wang J, Tang J, Guo K (2022) Green Bond Index Prediction Based on CEEMDAN-LSTM. *Front Energy Res* <https://doi.org/10.3389/fenrg.2021.793413>
- Weber O, Koellner T, Habegger D, et al. (2008) The relation between the GRI indicators and the financial performance of firms. *Prog Ind Ecol Int J* 5: 236–254. <https://doi.org/10.1504/PIE.2008.019127>
- Whelan T, Atz U, Holt TV, et al. (2021) ESG and financial performance: uncovering the relationship by aggregating evidence from 1,000 plus studies published between 2015–2020. *NYU Stern Center Sustainable Bus*.



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)