P-ISSN: xxxx- xxxx | E-ISSN: 3007-9861 | DOI: https://doi.org/10.71285/icpt.v3i2.39

Journal of Innovative Construction and Petrochemical Technologies
https://icpt.ouk.kz/index.php/icpt/index

Research Article

Check for updates

External Audit and Financial Characteristics in Predicting Economic-Sector Membership Using Machine Learning: Evidence from Iraqi Listed Companies

Nadia Talib Salman
Middle Technical University, Administrative Polytechnic College, Baghdad
Email: nadia@mtu.edu.iq

Fatimah Fezea Hadab
Aliraqia University, Economics and Administration College, Baghdad, Iraq
Email: fatimah.f.hadab@aliraqia.edu.iq

Abstract: This study investigates the predictive role of external-audit and financial characteristics in identifying economic-sector membership among Iraqi listed companies using machine-learning techniques. The analysis is based on 263 firm-year observations from 33 companies distributed across six verified economic sectors during 2015–2024. Three predictive specifications are evaluated: financial characteristics only, external-audit characteristics only, and a combined model incorporating both information sets. Logistic regression is adopted as the primary classifier and assessed using leakage-free leave-one-company-out cross-validation with class-preserving nested hyperparameter selection. Additional robustness procedures include alternative machine-learning algorithms, company-level aggregation, permutation testing, stratified company-block bootstrap inference, drop-column analysis, and sample-sensitivity tests. The combined model achieved the highest firm-year macro F1-score of 0.321, compared with 0.266 for the financial-only model and 0.242 for the audit-only model. However, the incremental improvement of 0.055 over the financial-only specification was not statistically conclusive, as the 95% bootstrap confidence interval ranged from −0.024 to 0.129. Firm size emerged as the strongest overall predictor, while modified audit opinion and auditor change were the most influential audit-related characteristics. The combined-model advantage was not consistent at the company level or across all alternative classifiers. The findings indicate that both financial and external-audit characteristics contain sector-related predictive information, while highlighting the need for cautious interpretation in small and imbalanced emerging-market samples.

Keywords: External audit, financial characteristics, economic-sector prediction, machine learning, listed companies, Iraq, audit opinion, sector classification

INTRODUCTION

Machine learning (ML) has grown in use in accounting and auditing research as more and more corporate data has become available and AI has progressed. Predictive modelling is fundamentally different from traditional explanatory analysis since the main aim is to assess how well an observed sample can be generalized to new observations rather than identify causal relationships within the observed sample [1]. In accounting and assurance, ML increasingly has been applied to tasks such as transaction analysis, risk assessment, fraud detection, bankruptcy prediction, and other classification tasks [2]. It has also prompted accounting researchers to look further into the out-of-sample validation of discrete corporate outcomes [3].

ML is especially applicable in situations where the information in the account has complex or nonlinear patterns, and perhaps cannot be detected using traditional statistics. For instance, Bao et al. [4] showed that machine-learning techniques can be used to identify relevant predictive signals from financial-statement data that can be used for fraud detection and that outperform existing benchmark models. These advances are only a few examples of the much larger utility of ML models in discovering economically significant patterns in corporate financial data.

An emerging use is to forecast the sector of the economy to which a firm belongs based on other characteristics of the firm that are observed. The asset structure, capital intensity, the need for financing, profitability, operating cycles and the way in which resources are used will vary from one firm to another, depending on the sector in which they operate. These differences can create a pattern in funds data that is detectable. Van der Heijden [5] proved this by applying 80,790 firm-year observations and showed that publicly available information from financial statements can be used with high out-of-sample accuracy to predict industry membership and that nonlinear classification methods are superior to linear classification methods.

This line of research has been further advanced in recent studies. Waffo Dzuyo et al. [6] integrated numerical and textual data of financial statements to cluster companies by industry and highlighted the importance of the identification of sector-specific financial features in audit planning and for an understanding of the clients' business. Recently, Forradellas et al. [7] studied the profitability, leverage, liquidity, efficiency, interest-coverage, and cash-generation ratios of the S&P 500 firms. They found that, accounting characteristics account for part of the membership, and also they found that there was significant intra-sector financial heterogeneity. All of these studies imply that membership in an economic sector leaves a measurable financial "footprint," but that it is not exactly repackaged in a ratio of just financial ratios.

But corporate financial characteristics are not all the information that is provided externally. The external audit process takes place in the economic and operational environment of the client, and the auditors need to know the business activities, risks, transactions, accounting practices and industry environment of the client. The impact of AI and sophisticated analytics in auditing has thus been gaining research interest. Issa et al. [8] recognized that AI could be a key technology to aid in audit work supplementation and assist more guided audit decisions. Since then, empirical findings have supported the usefulness of data-driven prediction and pattern recognition in modern auditing, as audit firms are seen to invest in AI to enhance their audit quality and efficiency [9].

External-audit characteristics may also be indirectly indicative of industry conditions. Industry knowledge is a characteristic that has been linked to audit-quality outcomes, suggesting that an audit firm's ability to understand an industry's business environment has an impact on the audit process [10]. Likewise, Habib and Bhuiyan [11] reported that the audit-report lags for the firms audited by industry-specialist auditors were shorter, indicating that industry experience of auditor affects audit efficiency. This evidence is the conceptual basis for the expectation that audit-related characteristics, such as modified opinions, auditor changes, and audit-report timing, as well as audit arrangements, will contain information related to differences between audit firms' operating environments. Importantly, it does not mean that the audit characteristics determine a firm's sector, but rather it can include signals that are visible from the audit engagements performed in various economic contexts.

Although related, current works on ML-based sector classification have concentrated on either financial-statement information or textual disclosures that have been issued by corporations [5]–[7]. A lot less work has been done on the question of whether external-audit characteristics have useful predictive information, either alone or in combination with the more traditional financial characteristics, about economic-sector membership. This distinction is relevant because audit attributes could reflect aspects of reporting difficulty, engagement risk, audit timing and auditor-client relationships which are not adequately captured by financial ratios. Meanwhile, any incremental predictive contribution of audit information should be viewed as an empirical question, and not assumed a priori.

The Iraqi capital market is an appropriate context for the analysis of the issue. The Iraq Stock Exchange (ISX) has listed firms in various economic activities such as industry, tourism, agriculture, telecommunications, investment, insurance, banking and services [12]. The population of listed Iraqi companies is relatively small and sectors are not well represented, especially when compared to large datasets of developed markets in the previous sector-prediction studies. This setting is a key methodological problem, as the same companies may be in both training and testing sets, making it possible for overly optimistic prediction results to appear. The present study thus focuses on out-of-sample validation, not on the traditional random splitting method.

This study investigates the external audit and financial characteristics along with their prediction of economic-sector membership of listed companies in Iraq through machine learning. The final analytical sample is made up of 263 firm-year observations from 33 firms across six sectors of economic activity that have been verified over the 2015-2024 period. Three sets of information are examined: external-audit characteristics, financial characteristics and a combined specification. For the audit variables, the ones that are modified are audit opinion, auditor change, audit-report lag, and state-audit status; and for the financial variables, firm size, profit margin, leverage, asset turnover, and return on assets.

Logistic regression is classified as the main classification method and is estimated using leave-one-company-out cross validation and class preserving nested hyperparameter selection, methodologically. All the preprocessing steps are done only in the training folds, so as to avoid information leakage. The robustness and stability of the predictive evidence is tested using alternative machine-learning algorithms, evaluation at the company level, permutation testing, stratified company-block bootstrap inference, drop-column analysis and sample-sensitivity analyses. Similarly to the general principles of predictive modelling [1] and their application in accounting research [3] the focus is on out-of-sample discrimination and not causal inference.

This study has three important contributions to the literature. First, it adds to the existing prediction research on economic-sector performance by using observable external-audit characteristics, in addition to financial characteristics, as inputs for the prediction. Second, it differentiates between "financial only" and "audit only" and "combined" sets of information, allowing for an explicit evaluation of the predictive information in each set without the assumption that the audit variables add anything statistically significant. Third, it offers evidence in a relatively small and sectorly skewed emerging capital market, and uses company independent validation processes that minimize information leakage and overly optimistic performance estimates. In this context, the study brings together the work of accounting prediction based on ML and the external-audit work and examines if the audit and financial features are predictive of companies not included in the model estimation process.

Literature Review and Hypothesis Development

Financial Characteristics and Economic-Sector Prediction

The financial characteristics are well-affiliated to the economic context of the business. Industries vary with regard to production methods, amount of capital invested, operating cycles, need for financing, asset utilization, and profitability cycle. Financial ratios can be systematically different between sectors as a result of these differences. The interrelationships between financial ratios were significantly affected by the industry, and the financial models were multivariate in nature, as in such models ignoring the industry effect could decrease the accuracy of the models, found Martikainen et al. [13]. Similarly Dahlstedt et al. [14] found that firms in traditional industry groups can also be quite financially diverse, highlighting the usefulness and limitations of using financial ratios as a way of identifying economic activity.

The advent of machine-learning techniques has turned this problem into a prediction classification problem, as opposed to the industry comparison problem. Angenent et al. [15] used the large-scale ML technologies to the business-sector prediction and showed that there was enough structure in the corporate financial information to be able to differentiate several economic sectors. Their work now shows how predictive algorithms can be used to provide information related to the sectors from high dimensional data in the financial world.

Evidence of accounting research, which comes later. Van der Heijden [5] demonstrated that information in financial statements can be used to identify industry membership and reported that nonlinear ML models can be more successful than the traditional linear classification models. Further advances in sector classification were made by Waffo Dzuyo et al. [6] who combined numerical and textual financial data and Forradellas et al. [7] who illustrated that there was useful information in the profitability, leverage, liquidity, efficiency and other accounting characteristics. Taken together, this series of findings suggests that there are discernible financial signatures of economic sectors, though there are variations in how different the sectors are.

A variety of ways in which such differences could manifest themselves are represented in the financial characteristics used in this study. The size of the firm is an indicator of the size of its operation and intensity of assets used; leverage is a measure of the structure of the financing; asset turnover is an indicator of how efficiently the asset base generates revenue; profit margin is an indicator of how profitable the firm is; and return on assets is a measure of how much the assets earn. The group patterns of these characteristics may be useful information for discriminating between sectors as their operating environments play a part in forming these characteristics.

Based on this, the first hypothesis is formulated as:

H1: Financial characteristics provide significant out-of-sample predictive information for economic-sector membership among Iraqi listed companies.

External Audit Characteristics as Sector-Related Predictive Signals

External auditing is done in conjunction with the client's operating environment. To plan and perform an audit, it is important that auditors thoroughly understand the nature of the entity, its transactions, accounting systems, regulatory environment, business risks and industry-specific conditions. As a result, observable qualities associated with the audit engagement could be indirectly indicative of aspects of the economic environment in which the firm works.

There has been a large body of audit research that has focused on auditor industry specialization. Balsam et al. [16] discovered that clients of industry-specialist auditors had lower discretionary accruals and higher earnings-response coefficients than clients of nonspecialists, implying that audit knowledge in the industry may impact observable outcomes of reporting. The evidence is not consistent, however. While carefully matching the clients of the specialist and non-specialist auditors, Minutti-Meza [17] did not consistently observe differences in the proxies used to measure audit quality. This finding implies that part of any apparent specialization effects could be due to the characteristics of the clients themselves, not to auditor specializations.

The results of this study are significant as it indicates that there are interdependencies between the audit characteristics and the industry environments, although the direction and strength of the relationship can differ. External-audit information is thus not necessarily auditing-sector driven but may include sector-related information.

A modified audit opinion is one of these attributes. A modified opinion results from the auditor's evaluation of financial reporting conditions, going-concern concerns, scope restrictions or other matters that impact the opinion. Habib [18] conducted a meta-analysis of 73 published studies to identify the audit-related and firm-specific characteristics that influence the decision to modify an audit opinion. He found that auditor specialization, audit-report lag, firm size, leverage and profitability have an influence on whether or not the auditor changes his opinion. This evidence indicates that audit opinions reflect several aspects of client and engagement risk which may vary systematic in the economic environment.

A potential factor is audit-report lag, which refers to the time between the fiscal year-end and the audit-report date. Habib and Bhuiyan [11] reported that the audit-report lags were shorter for the clients of industry-specialist auditors and attributed this to audit efficiency and audit-sector specific knowledge. In a related meta-analysis Habib et al. [19] found that audit-report lag is related to audit opinion, auditor characteristics, firm complexity, profitability, and other governance and firm factors. The timeliness of the audit can thus vary between companies and industry segments.

Another possible informative characteristic of an audit is auditor change. The audit market, governance, auditor specialization, reporting disagreements, audit fees, the need for auditor switching due to client needs, and the audit market changes may be significant reasons for an auditor switch. Cairney and Stewart [20] identified that certain industry characteristics were strongly correlated with the various types of auditor change, such as product-market competition, accounting-standard complexity, industry homogeneity and auditor competition. Their results directly suggest that there is systematic auditor-switching behavior in relation to industry conditions. Ding et al. [21] also explored the context of audit-firm switching using the specialization of the industry. They determined that the switch from nonspecialist to specialist auditors at the engagement partner level was linked to reductions in discretionary accruals, increases in the value relevance and likelihood of modified and going-concern opinions. Such results support the notion that the information in auditor changes or reporting results may include information regarding the sector-specific auditor knowledge and factors associated with the client.

The present study also takes into account the state-audit status. The nature of the auditor–client relationship and selection of auditor may be influenced by institutional characteristics, especially in markets where governments are significantly involved in corporate activities. Wang et al. (2012) [22] found that the state ownership and institutional environments were significantly related to auditor-choice patterns of listed companies in China. While the institutional context of Iraq is distinct from that of China, the evidence shows that, more generally, the government and institutional environment can influence the observable external-audit arrangements.

The results of prior audit research indicate that factors such as modified opinions, auditor changes, audit-report timing and institutional audit arrangements may be indicative of the differences in business complexity, risk, regulation and operating environment. These variables are thus considered to be predicting rather than causally determining membership in a sector.

The second one is therefore formulated as:

H2: External-audit characteristics provide significant out-of-sample predictive information for economic-sector membership among Iraqi listed companies.

Joint Predictive Role of External Audit and Financial Characteristics

The financial sector and external audit attributes are separate information domains, but there is some complementarity between them. Financial variables are variables that directly measure the firm's economic position and operating structure, while audit variables measure characteristics of the independent assurance process over the firm's financial statements.

Existing sector-classification works have demonstrated that there is useful information in financial data related to a sector [5]–[7], [15]. The fact that financial ratios do not perfectly replicate sector classifications, however, puts in question whether more corporate information can help discriminate among sectors. Such information can be provided by external-audit characteristics because the dimensions of reporting risk, complexity, auditor judgment and institutional conditions are not fully captured by the traditional financial ratios, which are present in audit opinions, reporting delays, auditor switching and audit arrangements.

This is possible and the literature on auditor specialization lends support. The interpretation of the auditors' assessment of their clients and the conduct of audit engagements is affected by industry knowledge [10, 16] and the outcomes of audit engagements are also shaped by the characteristics of the audit firm and the audit industry environment [18–21]. Therefore, it is possible that using the combination of audit characteristics and financial information together could provide a more comprehensive set of signals for an ML model, than using either information group individually.

Concurrently, the predictive value of the complementary is not assumed. The results of some audits could well mirror the same financial characteristics that are captured by profitability, leverage, asset use or firm size. Industry-specialist auditors, for instance, as in the case of Minutti-Meza [17] are warned that the relationship might be due in part to differences in client characteristics. Consequently, whether the audit information helps to improve classification over and above the financial information is a question of fact.

We directly test this question by considering the following three predictive specifications: (1) financial characteristics, only; (2) external-audit characteristics, only; and (3) between financial and external-audit characteristics. By incorporating external audit and financial signals, the study will assess the effectiveness of the combination with respect to out-of-sample sector discrimination.

Accordingly:

H3: A combined model of external-audit and financial characteristics provides higher out-of-sample predictive performance for economic-sector membership than a model based on financial characteristics alone.

Research Gap and Study Positioning

Previous studies have shown that the financial statements contain information that identifies companies with different industries [5]–[7] [13]–[15]. However, the studies within the context of sectors have focused mainly on financial ratios, financial-statement variables, or textual corporate information. The inclusion of external-audit characteristics as predictors of economic-sector membership has been rare.

At the same time, the audit literature indicates that the industry characteristics are linked with Auditor Specialization, Audit Opinion, Audit Report Timing, Auditor Switching and Auditor Choice of Institutions [10, 11, 16–22]. These studies, however, tend to look at audit attributes as results of client, auditor, or institutional attributes. They do not directly question if the observable audit characteristics can be used in the reverse predictive sense – to determine the economic environment in which the audited firm performs its operations.

This leaves a gap between two given research streams. The literature in the field of classification shows that company attributes can be used to predict industry, and the literature on the audit process shows that audit results are different across industries depending on the company's characteristics. A few studies, however, do not exist on whether external-audit attributes per se contain sector-specific predictive information or whether they add to the information provided by the traditional financial attributes in out-of-sample sector classification.

However, the present study aims to bridge this gap by combining external audit and financial attributes in a single ML. It does not consider audit variables to be cause variables for sector membership. Instead, it is determined if they are identifiable predictive signals of various economic conditions.

The setting in Iraq is an extension of the current literature in the context of listed companies.The setting in Iraq is an extension of the existing literature in the context of listed companies. The majority of previous works on ML-based sector-classification are based on much larger datasets in developed or international markets [5]–[7] and [15]. The Iraqi market is a more limited environment with a relatively small number of listed companies, and significant imbalance by economic sector. This is a serious issue for traditional random train-test evaluation as it may be possible to get repeated observations of the same company in both sets.

For the purpose of this study, the out-of-sample validation method is used, which is company independent, and financial information, audit information and information from both sets are considered. It does so with the aim of not only establishing if sector classification is feasible, but also which of the information in accounting and audit documents are useful in providing generalisable signals for sector classification when prediction is needed for a company not used for model estimation.

METHOD

Research Design and Sample

In this study, a supervised machine learning classification design is used to explore if the external-audit and financial characteristics have information that is important for predicting the economic-sector membership of Iraqi listed companies. The analysis is performed from 2015 to 2024 and done by following the observations that the firms made in their respective year's reports and data from Iraq Stock Exchange is available in the public domain [12].

There were a total of 340 firm-year observations in the original data set, with 34 firms included in the data. The original sector classifications were analysed on the company level as the prediction target has to be the economic activity instead of the ownership structure. Ownership characteristics labels such as 'Mixed Sector' were not considered to be the economic sectors. Rather, each company was categorized according to its confirmed main economic activity as per its registered company name, historical classification, and sector information available.

One financial-investment company was excluded due to the fact it represented a single company sector, and thus had no meaningful out-of-company classification. Two more firm-year observations were discarded due to known, but uncorrected, financial data inconsistencies, such as unit-scale and accounting-balance issues.

This created the modeling universe of 33 companies in six economic sectors: Manufacturing & Industry (12 companies), Hotels & Tourism (9), Agriculture (4), Real Estate & Transportation (4), Pharmaceuticals (2) and Telecommunications (2). The analytical sample was determined by applying the primary data-availability criteria, leaving 263 firm-year observations of the 33 companies.

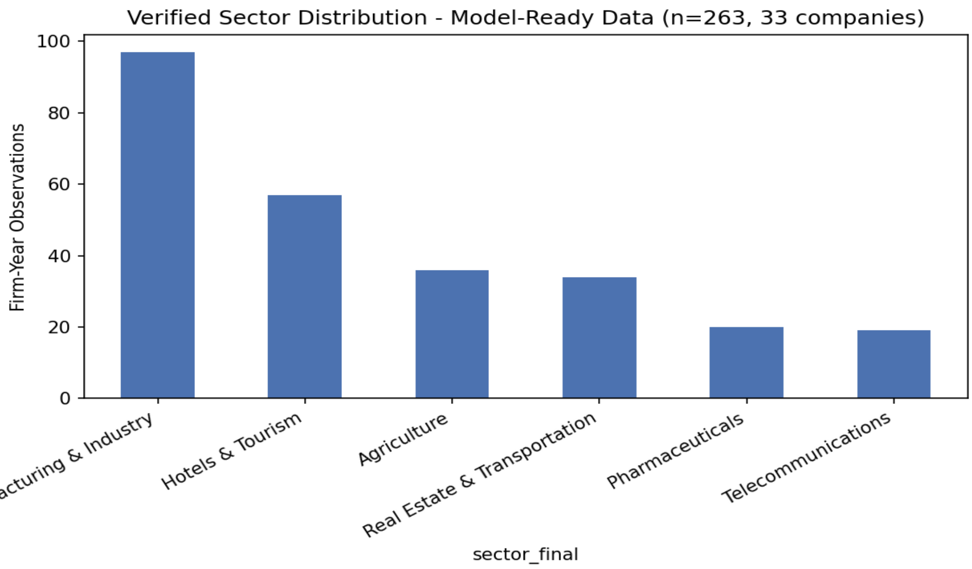

Figure 1. Distribution of firm-year observations across the six verified economic sectors.

The study explicitly focuses on sector imbalance in model estimation and evaluation because the number of companies varied significantly between the different sectors.

Predictor Variables

Two sets of predictors were developed: external-audit characteristics and financial characteristics. The variables selected reflect observable features of audit environment and financial structures of firms, and do not include direct economic-sector-specific information.

The external-audit variables are: modified audit opinion, auditor change, first-year observation of the audit, audit-report lag, and state-audit status. Modified Opinion indicates if auditor has issued a modified opinion, instead of an unmodified opinion. Auditor Change is the information that shows whether or not the auditor has changed from the previously observed year. A First-Year Observed indicator was added as there are no previous auditors for comparison in the first year of each company's observation. Audit Report Lag is calculated based on the number of days between fiscal year-end and the audit-report date and is transformed using the natural log of (1 plus # of days). State Audit identifies observations to which the state-audit arrangement applies. The financial feature consists of profit margin, leverage, asset turnover, asset return (ROA), and size of the firm. These are respectively the variables for profitability, financing structure, operating efficiency, asset profitability and corporate scale.

Table 1. Predictor Variables

Group	Variable	Measurement
Audit	Modified Opinion	1 = modified audit opinion; 0 otherwise
Audit	Auditor Change	1 = auditor changed from previous observed year; 0 otherwise
Audit	First-Year Observed	1 = first available observation for the company; 0 otherwise
Audit	Audit Report Lag	ln(1 + days from fiscal year-end to audit-report date)
Audit	State Audit	1 = state-audit arrangement; 0 otherwise
Financial	Profit Margin	Profit before tax / Revenue
Financial	Leverage	Total liabilities / Total assets
Financial	Asset Turnover	Revenue / Total assets
Financial	ROA	Net profit / Total assets
Financial	Firm Size	ln(Total Assets)

Three predictive specifications were constructed. Model A is a model with financial characteristics, Model B is a model with external-audit characteristics, and Model C is a model with both sets of characteristics. This is the structure that is directly equivalent to H1–H3 and allows to compare the predictive information in each domain.

Data Preprocessing

The test observations were used only in all the preprocessing procedures that can benefit from them. Data was analysed in Python using the scikit-learn framework [23].

Winsorisation of continuous variables at 1st and 99th percentiles was employed to minimize the effect of extreme values. Importantly, thresholds for winsorizing were only estimated on the training data in each of the cross-validation folds. The missing values in modeled variables were then imputed with the training-fold median, and standardized to zero mean and unit variance.

This procedure prevented information from the company that was not included in the analysis from leaking into the winsorization thresholds, imputation values and scaling parameters.

In the primary analysis, observations with missing or improper financial ratios were dropped if it was not possible to reconstruct the financial ratios from the available financial data in a reliable manner. A more liberal specification kept observations with complete absence of financial attributes and dealt with missing financial variables by training-fold imputation as a sensitivity test.

Primary Classification Model and Validation Strategy

Logistic regression was the pre-specified primary classifier due to its parsimony, probabilistic output, interpretability and applicability to a relatively small number of independent companies. To minimize the effect of varying frequencies in the sectors, balanced class weights were used.

The hyperparameter values and performance estimation were derived by using a nested cross validation framework. One point to note is that when model parameters are selected from the same data set as that used for performance evaluation, nested cross-validation is more important since the estimates obtained by conventional cross-validation might be optimistically biased [24].

Leave-One-Company-Out cross validation (LOCO-CV) was used for the outer validation procedure. In both outer iterations, all observations of the firm year of one company were taken off the estimation sample and into the testing sample only. As such, the data of one company would not be seen in the training and test sets at the same time.

This is a significant design aspect for the current study as economic-sector membership is based on the level of the company. Random splitting of firm-year observations may result in observations of the same company being placed in both the training and the testing sets, which can lead to the classifier learning company-specific characteristics instead of patterns in the sector at large.

For each outer-training sample, the logistic-regression regularization parameter was determined using a class-preserving inner cross-validation procedure that was performed in a grouped fashion. For sectors with only one surviving company the companies were kept in the inner-training sample. Stratified grouped splitting was used for the remainder of the companies, but company level was preserved.

The procedure aimed to make sure that all sectors included in an outer training sample were included in each of the inner training partitions. There was no diagnostic failure in the entire analysis, that is, no sector was left out of the training set in any of the 165 inner validation folds. The macro F1 score was used for the hyperparameter selection. The best value for the regularization parameter was determined and the model was re-fitted on the full outer-training sample, and predictions were made for the other company. After repeating the process for each of the 33 companies, we generated out-of-fold predictions for the entire analytical sample, while ensuring that the companies remain independent.

Alternative Machine-Learning Algorithms

To see if the results were sensitive to the logistic-regression specification, four other machine-learning algorithms were used as robustness checks.

The Random Forest method was selected as one of the ensemble tree-based classifiers which can model the nonlinear relationships and interactions [25]. A gradient-boosting technique was used to sequentially fit decision trees to enhance the prediction accuracy, namely XGBoost [26]. Nonlinear decision boundaries were captured by a radial-basis-function Support Vector Machine [27] and the nonparametric distance-based classification approach was the k-Nearest Neighbors [28].

These algorithms were not considered as prime model selection rivals. Extensive hyperparameter tuning for each of the classifiers was therefore not performed, instead fixed settings were implemented. The question was to see if relative predictive power of the financial-only and combined information sets would vary across the different classification approaches.

Performance Evaluation

Due to the uneven distribution of economic sectors, it was decided to use macro F1-score as the main indicator of performance. In Maco F1, the F1 score is computed on a per-sector basis and then all sectors are given equal importance, regardless of the sample frequency. Therefore, it is especially suitable for multiclass classification problem with significant class imbalance.

Other performance measures used as supplements were accuracy, balanced accuracy and weighted F1-score. A baseline of a majority-classifier was added. While the macro F1 is less important than the conventional accuracy for a model with a large class size, it serves as the primary criterion for comparing models.

The performance of the models was also analysed at the company level in addition to the firm-year evaluation. The out of fold predicted probabilities were calculated as the mean for each company that was held out. The predicted sector was then chosen as the company's predicted sector, as it had the highest average predicted probability in the economic sector.

This process offers a more rigorous evaluation of the ability of the models to classify an entire company that has not been included into a given economic sector over the period of observation of its financial and audit history.

Statistical Inference and Robustness Analysis

A few supplementary analyses were performed to consider the statistical uncertainty and robustness of the predictive results.

The incremental performance of the combined model when compared to financial model alone was measured as follows:

$$\Delta \mathrm{Macro}_{\mathrm{F1}} = \mathrm{Macro}_{\mathrm{F1(Combined)}} - \mathrm{Macro}_{\mathrm{F1(Financial)}} \qquad (1)$$

Where ΔMacro-F1 is the difference between the predictive performance of the combined model and the financial-only model. A positive value means the joint specification is doing better.

A stratified company-block bootstrap with 2,000 replications was used to assess the uncertainty of this difference. Various bootstrap methods have been developed for the estimation of the sampling uncertainty, which is a nonparametric method [29].

The sampling of companies was done with replacement at the economic sector level, not at the firm year level. The procedure keeps dependence between the repeated observations from the same company and guarantees that all six sectors will be included in each bootstrap replication. The bootstrap distribution of ΔMacro-F1 was then used to construct a 95% confidence interval. If the zero lies in the interval, it means that the predictive advantage of the combined model cannot be considered statistically significant at the 5% level.

Second, a supplementary fixed-hyperparameter company-level label-permutation test was carried out to check if the classification performance was significantly better than what one would expect by chance if the sector labels were randomly assigned. Permutation testing has been used to assess if a classifier extracts class-related information over and above chance [30].

The observations were kept longitudinal, by randomly re-labeling the sectors at the company level and not at the individual firm-year level. Since this is a supplementary procedure, where the hyperparameters are fixed, the observed macro F1 values were reported independently from the primary nested-CV performance values.

Third, the predictive contribution was analyzed through drop-column. The 10 predictors were removed in turn and the whole nested LOCO-CV method was repeated. The macro F1 was used to quantify the contribution of a predictor by comparing the F1 from each predictor with the F1 obtained when all the predictors were combined into one.

If the macro F1 decreases after the variable's removal, the variable made a positive contribution in the sample observed. Conversely, if an improvement has been made after removal, then there may be some redundancy, overlap and/or model instability. This does not mean that the variable is not considered to be causally significant.

Finally, two more sample based sensitivity checks were performed. The analysis was limited to companies having high-confidence economic-sector classifications in the first case. The second one dropped primary complete case requirement and kept the observations with missing financial variables, which were then filled in using training-fold imputation.

In addition to the alternative-algorithm comparison analysis and company-level analysis, these tests are used to determine if the main conclusions are sensitive to sector verification, missing-data treatment, algorithm selection, and/or to the aggregation level of predictions.

Hypothesis Evaluation

H1 is tested by checking whether the financial-only model has meaningful out-of-sample predictive performance in comparison to the benchmark model, and the additional permutation evidence.

H2 is tested based on the out-of-sample performance of the audit-only model. The goal is to assess whether external audit data which is observable when excluding the financial data has predictive information in the sector.

The evaluation of H3 compares the macro F1 score of the combined model and financial-only model. The quantity of interest observed is the principal quantity ΔMacro-F1. To assess uncertainty about the observed improvement, a 95% CI derived from the stratified company-block bootstrap is utilized because a positive point estimate does not prove to be a statistically reliable incremental PV.

As additional evidence of the robustness of the main conclusions, company-level performance, alternative algorithms, drop-column analysis and sample-based sensitivity analyses are included. Therefore, the evaluation of the hypothesis must be done by the criterion of out-of-sample predictive performance, as this is the study's objective of prediction.

RESULT AND DISCUSSION

Primary Out-of-Sample Classification Performance

The main classification results for firms (by firm-year) are presented in Table 2, which is derived from the nested LOCO-CV procedure. Due to the uneven distribution of the numbers in the sectors, macro F1-score is used as main metric to compare the three information sets.

Table 2. Primary Out-Of-Sample Classification Performance

Model	Information Set	Accuracy	Macro F1	Weighted F1	Balanced Accuracy
Model A	Financial Only	0.240	0.266	0.217	0.340
Model B	Audit Only	0.259	0.242	0.249	0.309
Model C	Combined	0.304	0.321	0.308	0.359
Baseline	Majority Class	0.369	0.090	—	—

The model C had the best macro F1-score of 0.321, whereas the financial-only and the audit-only models had scores of 0.266 and 0.242, respectively. It also had the highest accuracy score and the highest weighted F1 out of the three predictive models. The difference in macro F1 between the two models was about 0.055, suggesting that incorporation of external-audit characteristics led to better primary firm-year point estimate.

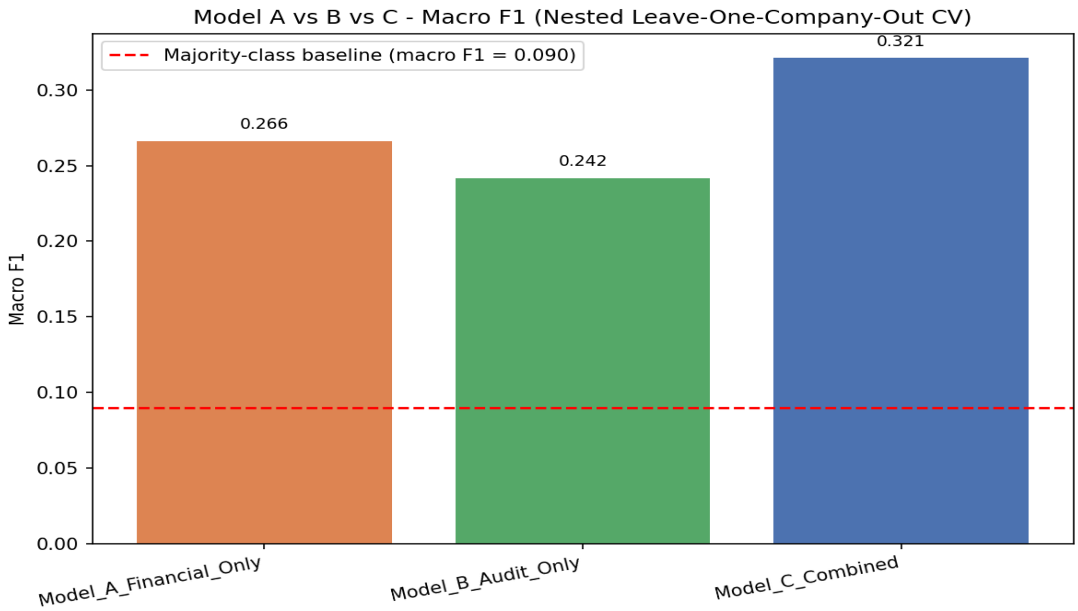

Figure 2. Macro F1 performance of the financial-only, audit-only, and combined models under nested LOCO-CV.

The baseline accuracy for majority class was 0.369, which was higher in number than the accuracy of predictive models. This is however due to the high class imbalance in the sample, since each observation is assigned to the largest sector as baseline. It performed only slightly better than all three ML specifications on macro F1 score with 0.090. This is a good example of why the traditional measure of accuracy is not suitable in this context.

The financial-only model significantly outperformed the baseline with respect to macro F1, supporting the idea that there is some sector-specific predictive information in the financial characteristics. This result is consistent with previous study findings which showed that the characteristics of financial statements can provide some systematic differences between different industries and aid in sector classification [5]–[7] [13]–[15].

The same happened for the audit only model, which obtained a macro F1-score significantly higher than the baseline majority class one. The finding, although it is less than that of the financial-only and combined models, indicates that information related to economic-sector differences can be found in observable external-audit characteristics. This evidence corroborates the previous audit studies on industry conditions' relations with auditor specialization, reporting lag, audit opinion, and auditor rotation [10], [11], [16]–[21].

Incremental Predictive Value of the Combined Model

The key difference between combined and financial only specifications was:

$$\Delta Macro - F1 = 0.321 - 0.266 = +0.055$$

This is a point estimate for the combined specification, but the improvement in the prediction needs to be assessed in conjunction with its sampling uncertainty. The stratified company-block bootstrap yielded the results that are reported in Table 3.

Table 3. Bootstrap Inference for Incremental Predictive Performance

Measure	Result
Observed ΔMacro-F1	+0.055
Bootstrap Replications	2,000
95% CI Lower Bound	−0.024
95% CI Upper Bound	+0.129
CI Excludes Zero	No

The 95% confidence interval is (−0.024, +0.129) which contains zero. Thus, although Model C had the best primary firm-year point performance, there is no evidence that the performance of the incremental improvement over the financial-only specification is statistically better than 5%. It's significant that this is a difference. The findings show that the external-audit variables can provide predictive information but they fail to show that this information may always lead to a better classification in some samples of the listed companies in Iraq. Thus, the results should not be seen as evidence of the superiority or informativeness of financial characteristics over audit characteristics.

Company-Level Classification

As multiple observations are available for each firm, an additional analysis at the company level was carried out, by averaging the predictions of out-of-fold probabilities for the companies across the available years for each firm.

Table 4. Company-Level Classification Performance

Model	Accuracy	Macro F1
Financial Only	0.303	0.325
Audit Only	0.364	0.327
Combined	0.273	0.281

Results are not the same at the company level as they are at the main level of the firm-year results. The company-level macro F1-scores obtained from the audit-only and the financial-only models are 0.327 and 0.325, respectively, which are the highest values, whereas the combined model resulted in an F1-score of 0.281.

The result is that when we combine the model predictions at the firm level, the firm-year advantage does not remain. The finding reiterates the importance for the conservative treatment of audit information. The positive firm-year ΔMacro-F1 ought to be interpreted as a point increase and not as a sustaining strength.

The results of the company-level analyses also emphasize the need for independent-company validation. If longitudinal observations are available, good firm-year outcomes do not necessarily mean that other, entirely new firms are similarly good.

Class-Specific Performance

The results of the primary combined model are given in Table 5, by class.

Table 5. Class-Specific Performance of the Combined Model

Economic Sector	Precision	Recall	F1	Support
Agriculture	0.231	0.250	0.240	36
Hotels and Tourism	0.317	0.333	0.325	57
Manufacturing and Industry	0.524	0.227	0.317	97
Pharmaceuticals	0.029	0.050	0.036	20
Real Estate and Transportation	0.172	0.294	0.217	34
Telecommunications	0.655	1.000	0.792	19

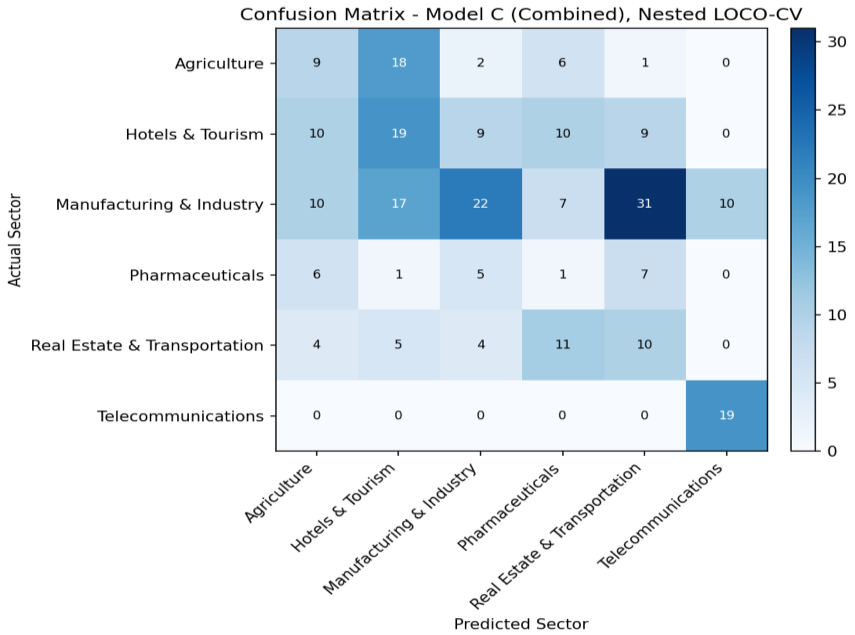

Figure 3. Confusion matrix for the combined model under nested LOCO-CV.

There was a wide degree of difference in performance by economic sector. The telecommunications sector had the highest F1-score of 0.792 and recall of 1.000 among the successfully classified industries. The moderate F1 scores of hotels & tourism (~0.325) and manufacturing & industry (~0.317) were obtained.

Among all the sectors, the Pharmaceuticals sector was the hardest to distinguish with an F1 score of just 0.036. This result indicates that while there may be good average overall performance, there could be significant disparities in the difficulty of the classification tasks across sectors.

The difference is in line with the overall literature that financial characteristics of the sectors are informative, but there is considerable heterogeneity both within and across the sectors [5] [7]. The results also highlight the need for reporting classification results specific to the class and not just overall classification accuracy.

Predictor Contribution

Each individual factor was deleted in turn and the full nested LOCO-CV procedure repeated to evaluate the unique predictive value of each factor. The findings are summarized in Table VI as the percentage change in macro F1 from the complete combined model.

Table 6. Drop-Column Predictor Analysis

Removed Predictor	Group	Macro F1 After Removal	ΔMacro-F1
None—Full Combined Model	—	0.321	0.000
Firm Size	Financial	0.225	−0.096
Modified Opinion	Audit	0.292	−0.029
Auditor Change	Audit	0.307	−0.014
Asset Turnover	Financial	0.308	−0.013
Audit Report Lag	Audit	0.315	−0.007
Leverage	Financial	0.328	+0.007
First-Year Observed	Audit	0.331	+0.010
Profit Margin	Financial	0.335	+0.014
State Audit	Audit	0.347	+0.026
ROA	Financial	0.354	+0.033

Note: ΔMacro-F1 = Macro F1 after predictors are removed − Macro F1 of the full combined model. Removing negative values then will improve predictive performance.

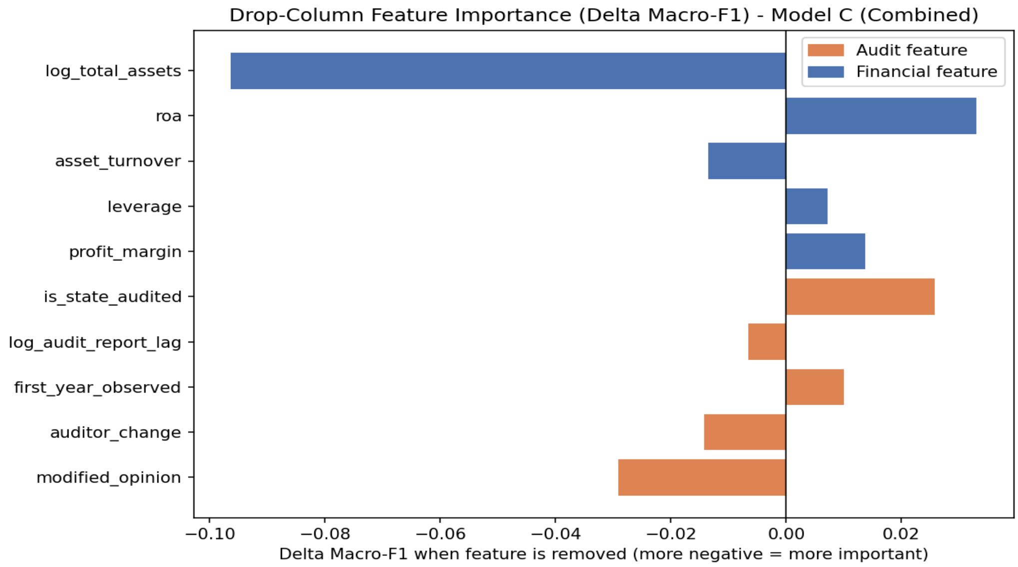

Figure 4. Drop-column predictor contribution for the combined model based on changes in Macro F1.

The most important characteristic overall was firm size (in terms of the log of total assets—ln (Total Assets)). When it was removed, macro F1 went from 0.321 to 0.225, a drop of ~0.096. The result shows that the scale of the company gives the best predictive signal for the sample when it is divided into sectors.

The most significant positive contribution was from the external-audit variables, Modified Opinion. Its elimination led to a decrease in macro F1 of around 0.029. The second strongest audit related predictor was Auditor Change, a decrease of about 0.014. Audit Report Lag contributed slightly, but positively, to the score.

These findings are consistent with previous research that found that audit opinions, auditor switching and audit timing are associated with variations in client conditions, risk, client complexity and industry environment [11], [18]–[21].

Removing the other factors – ROA, State Audit, Profit Margin, First-Year Observed, or Leverage – led to minor gains in macro F1. The results are not necessarily indicative of the fact that these variables are not economically significant. Instead, in the case of a combined predictive specification, they can include redundant information, cause instability or offer less marginal discrimination after other characteristics are specified.

The most important thing to note is that the drop-column analysis validates that the main predictive structure is not just audit variables. As in the external audit group, firm size continues to be the most important characteristic, and Modified Opinion and Auditor Change emerge as the most important incremental signals in the external audit group.

Alternative-Algorithm Robustness

Logistic regression was the primary analysis used. The same three sets of information were used with another four classifiers KNN, Random Forest, SVM-RBF, and XGBoost to see if the combined model advantage depended on the classifier.

Table 7. Robustness across Alternative Classifiers

Classifier	Financial Macro F1	Audit Macro F1	Combined Macro F1	Combined – Financial
KNN	0.338	0.193	0.285	−0.053
Random Forest	0.253	0.165	0.230	−0.023
SVM-RBF	0.281	0.185	0.321	+0.040
XGBoost	0.198	0.159	0.232	+0.034

Combined-model advantage was different for different classifiers. SVM-RBF and XGBoost showed positive differences with values of about +0.040 and +0.034, respectively while the KNN and Random Forest models showed negative difference.

Thus, the added predictive power of the financial and audit characteristics varies from algorithm to algorithm. This finding provides further evidence of a conservative approach to H3 and confirms that it is beneficial to test the alternative model structure instead of a single classifier.

Supplementary Permutation Evidence

A fixed-hyperparameter company level label permutation was performed as an additional piece of evidence. The company-level labels were permuted to create the labels for the sector, and the performance estimates were computed using the fixed-hyperparameter procedure. Since this procedure is not a nested hyperparameter-selection procedure as employed by the main models, the macro F1 values observed here should not be taken as a direct reflection of the main model LOCO-CV values.

The financial-only specification gave the observed macro F1 of 0.289, versus a mean of 0.139 from the permutations (with a null distribution based mean of 0.139 and a permutation p-value of about 0.015). For the combined specification the observed macro F1 value was 0.375 and the null mean value for the macro F1 was 0.134, with a p value of ~ 0.002.

These findings suggest that financial-only and combined information classes show some class-related structure in addition to that which one would expect if the company-sector labels were randomly assigned. The permutation results are used as additional evidence, however, and do not supplant the nested LOCO-CV results.

Sample-Sensitivity Analysis

Two further analyses were conducted to test the sensitivity of the primary results with respect to confidence in the sector labelling and to the lack of financial information.

Table 8. Sample-Sensitivity Results

Sample Specification	Observations	Companies	Financial Macro F1	Combined Macro F1	ΔMacro-F1
Primary Sample	263	33	0.266	0.321	+0.055
High-Confidence Sector Labels	225	29	0.303	0.318	+0.014
Relaxed Complete-Case Sample	328	33	0.282	0.340	+0.058

If the sample is limited to those companies with a high-confidence classification in their sector, the combined model still beats the financial-only model, but the spread is reduced to about +0.014. The advantage of the combined-model grew to about +0.058 when the observations missing values on financial variables were retained and imputation performed during model training.

In both sample-based sensitivity checks, the firm-year combined-model advantage is still positive, therefore. However, the magnitude depends on the construction of the samples, where only high-confidence sector labels are kept.

Hypothesis Assessment

The evidence for the three hypotheses varies.

H1 is supported. The model that predicts only financial data produced an F1 score of 0.266 at the macro level, while the majority class score was 0.090. Additionally, there was also supplementary permutation evidence that financial characteristics carry information beyond chance that is related to the class. This is in line with previous research that showed sector-specific variations in corporate financial information [5]–[7],[13]–[15].

There is predictive support for H2, but less than in the case of H1. The results of the audit-only model were macro F1 score of 0.242, well above the majority-class score of 0.314 and macro F1 score of 0.327 at the company level. The results indicate the presence of signals from external-audits characteristics. The evidence, however, should not be characterised as formal statistical confirmation of the audit-only model, since this was not done separately for Model B.

There is not enough evidence to support H3 at the 5% level. The combined specification led to a slight improvement in the macro F1 of the firm-year model (around 0.055 compared to the financial-only model), but the 95% bootstrap confidence interval of −0.024 to +0.129 contained zero. Furthermore, the result of the combination of the advantages did not remain at the company level and was different for the alternative algorithms. Thus, the evidence shows that there has been a positive firm-year point improvement, but the evidence is not sufficient to conclude that the combination of audit and financial characteristics has a statistically reliable or overall strong incremental effect.

In general, the findings suggest that there is some sector-related predictive information in both financial and external-audit characteristics, but not the same, nor necessarily mutually complementary. The most powerful predictive signal is firm size and the most influential external-audit characteristics are Modified Opinion and Auditor Change. The joint information set generates the best primary firm-year performance, but results must be taken with a pinch of salt as the empirical analyses of the incremental value of the joint information set are not robust and do not provide consistency across all evaluation settings.

CONCLUSION

This study is concerned with the external-audit and financial characteristics' predictive power in external audit member classification of listed companies in the different economic-sectors in Iraq using machine-learning techniques. The analysis comprised 263 firm-year observations from 33 companies across six companies across six verified economic sectors during the 2015–2024 period. Three predictive specifications were tested: a financial characteristics only specification, an external-audit characteristics only specification, and a combined model of both types of information.

The combined model outperforms the financial-only and audit-only models with a firm-year macro F1-score of 0.321, while the financial-only and audit-only models reached 0.266 and 0.242, respectively. This resulted in an observed gains of about 0.055 compared to the financial-only specification. The stratified company-block bootstrap, however, gave a 95% confidence interval of –0.024 to +0.129, which contained the value of zero. Therefore, the additional predictive power of the two models cannot be considered as statistically significant at the 5% level.

These results still suggest that there is predictive information contained in the financial characteristic and the external-audit characteristic for each sector. With regard to the external-audit information set, firm size proved to be the most powerful overall predictor, and auditor change and modified audit opinion were the most significant. The results indicate that observable audit characteristics can include some elements of reporting risk and engagement conditions and institutional environments that are not fully reflected in the traditional accounting numbers.

The robustness analyses still further highlight the need to interpret with caution. This synergistic effect of the combined models did not hold true for the company level and was different for different machine-learning algorithms. Both sample-sensitivity specifications showed the direction of improvement as being positive for the firm year, but the size of the improvement varied greatly among samples. These results suggest that the external-audit information has an informative value which is not consistently strong in all evaluation contexts.

In general, the study advances the field of predictive accounting and auditing research by melding external-audit characteristics with existing financial information in a company-independent machine-learning context. The evidence is supportive of an intermediate and appropriate use of external-audit information in the economic-sector prediction. The characteristics of the audits should therefore be considered as complementary predictive signals, not as causal determinants of sector membership, or as alternatives to using the financial information. Another key finding of the study is that leakage-free validation, uncertainty estimation and robustness testing of machine-learning algorithms used in relatively small and imbalanced longitudinal corporate data sets is crucial.

REFERENCES

[1] G. Shmueli, "To explain or to predict?," Statistical Science, vol. 25, no. 3, pp. 289–310, 2010, doi: 10.1214/10-STS330.

[2] S. Cho, M. A. Vasarhelyi, T. Sun, and C. Zhang, "Learning from machine learning in accounting and assurance," Journal of Emerging Technologies in Accounting, vol. 17, no. 1, pp. 1–10, 2020, doi: 10.2308/jeta-10718.

[3] J. Krupa and M. Minutti-Meza, "Regression and machine learning methods to predict discrete outcomes in accounting research," Journal of Financial Reporting, vol. 7, no. 2, pp. 131–178, 2022, doi: 10.2308/JFR-2021-010.

[4] Y. Bao, B. Ke, B. Li, Y. J. Yu, and J. Zhang, "Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach," Journal of Accounting Research, vol. 58, no. 1, pp. 199–235, 2020, doi: 10.1111/1475-679X.12292.

[5] H. van der Heijden, "Predicting industry sectors from financial statements: An illustration of machine learning in accounting research," The British Accounting Review, vol. 54, no. 5, Art. no. 101096, 2022, doi: 10.1016/j.bar.2022.101096.

[6] G. S. Waffo Dzuyo, G. Guibon, C. Cerisara, and L. Belmar-Letelier, "Linking industry sectors and financial statements: A hybrid approach for company classification," in Proc. AAAI Conf. Artificial Intelligence, vol. 39, no. 16, pp. 16444–16452, 2025, doi: 10.1609/aaai.v39i16.33806.

[7] R. R. Forradellas, D. S. Cabrera, L. M. Garay Gallastegui, and S. L. Náñez Alonso, "Characterization of S&P 500 companies by sector using artificial intelligence: Statistical evidence and machine learning application," The Journal of Finance and Data Science, vol. 12, Art. no. 100193, 2026, doi: 10.1016/j.jfds.2026.100193.

[8] H. Issa, T. Sun, and M. A. Vasarhelyi, "Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation," Journal of Emerging Technologies in Accounting, vol. 13, no. 2, pp. 1–20, 2016, doi: 10.2308/jeta-10511.

[9] A. Fedyk, J. Hodson, N. Khimich, and T. Fedyk, "Is artificial intelligence improving the audit process?," Review of Accounting Studies, vol. 27, pp. 938–985, 2022, doi: 10.1007/s11142-022-09697-x.

[10] K. J. Reichelt and D. Wang, "National and office-specific measures of auditor industry expertise and effects on audit quality," Journal of Accounting Research, vol. 48, no. 3, pp. 647–686, 2010, doi: 10.1111/j.1475-679X.2009.00363.x.

[11] A. Habib and M. B. U. Bhuiyan, "Audit firm industry specialization and the audit report lag," Journal of International Accounting, Auditing and Taxation, vol. 20, no. 1, pp. 32–44, 2011, doi: 10.1016/j.intaccaudtax.2010.12.004.

[12] Iraq Stock Exchange, "Company Guide." [Online]. Accessed: Aug. 20, 2026.

[13] T. Martikainen, K. Puhalainen, and P. Yli-Olli, "On the industry effects on the classification patterns of financial ratios," Scandinavian Journal of Management, vol. 10, no. 1, pp. 59–68, 1994, doi: 10.1016/0956-5221(94)90037-X.

[14] R. Dahlstedt, T. Salmi, M. Luoma, and A. Laakkonen, "On the usefulness of standard industrial classifications in comparative financial statement analysis," European Journal of Operational Research, vol. 79, no. 2, pp. 230–238, 1994, doi: 10.1016/0377-2217(94)90354-9.

[15] M. N. Angenent, A. P. Pereira Barata, and F. W. Takes, "Large-scale machine learning for business sector prediction," in Proc. 35th ACM/SIGAPP Symp. Applied Computing (SAC), pp. 1143–1146, 2020, doi: 10.1145/3341105.3374084.

[16] S. Balsam, J. Krishnan, and J. S. Yang, "Auditor industry specialization and earnings quality," Auditing: A Journal of Practice & Theory, vol. 22, no. 2, pp. 71–97, 2003, doi: 10.2308/aud.2003.22.2.71.

[17] M. Minutti-Meza, "Does auditor industry specialization improve audit quality?," Journal of Accounting Research, vol. 51, no. 4, pp. 779–817, 2013, doi: 10.1111/1475-679X.12017.

[18] A. Habib, "A meta-analysis of the determinants of modified audit opinion decisions," Managerial Auditing Journal, vol. 28, no. 3, pp. 184–216, 2013, doi: 10.1108/02686901311304349.

[19] A. Habib, M. B. U. Bhuiyan, H. J. Huang, and M. S. Miah, "Determinants of audit report lag: A meta-analysis," International Journal of Auditing, vol. 23, no. 1, pp. 20–44, 2019, doi: 10.1111/ijau.12136.

[20] T. Cairney and E. G. Stewart, "Client industry characteristics and auditor changes," Review of Accounting and Finance, vol. 18, no. 2, pp. 245–267, 2019, doi: 10.1108/RAF-01-2017-0017.
[21] F. Ding, Z. Qiao, M. Hu, and M. Lu, "Industry specialization and audit quality: Evidence from audit firm switches in China," Asia-Pacific Journal of Financial Studies, vol. 51, no. 5, pp. 657–681, 2022, doi: 10.1111/ajfs.12365.
[22] Q. Wang, T. J. Wong, and L. Xia, "State ownership, the institutional environment, and auditor choice: Evidence from China," Journal of Accounting and Economics, vol. 46, no. 1, pp. 112–134, 2008, doi: 10.1016/j.jacceco.2008.04.001.
[23] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[24] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," BMC Bioinformatics, vol. 7, Art. no. 91, 2006, doi: 10.1186/1471-2105-7-91.
[25] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
[27] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.
[28] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.
[29] B. Efron, "Bootstrap methods: Another look at the jackknife," The Annals of Statistics, vol. 7, no. 1, pp. 1–26, 1979, doi: 10.1214/aos/1176344552.
[30] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," Journal of Machine Learning Research, vol. 11, pp. 1833–1863, 2010.