Improving the prediction of firm performance using nonfinancial disclosures: A machine learning approach

Sufi, Usman ; Hasan, Arshad ; Hussainey, Khaled

Journal of Accounting in Emerging Economies

DOI:
10.1108/JAEE-07-2023-0205

E-pub ahead of print: 24/06/2024

Peer reviewed version

Improving the prediction of firm performance using nonfinancial disclosures: a machine learning approach

Usman Sufi
Innovation Technology Centre,
Lahore School of Economics, Lahore, Pakistan
usman.sufi@lahoreschool.edu.pk

Arshad Hasan*
Faculty of Business,
Lahore School of Economics, Lahore, Pakistan
arshad@lahoreschool.edu.pk

Khaled Hussainey
Bangor Business School
Bangor University, Bangor, UK
k.hussainey@bangor.ac.uk

* Corresponding author

Improving the prediction of firm performance using nonfinancial disclosures: a machine learning approach

ABSTRACT

Purpose: The purpose of this study is to test whether the prediction of firm performance can be enhanced by incorporating nonfinancial disclosures, such as narrative disclosure tone and corporate governance indicators, into financial predictive models.

Design/Methodology/Approach: Three predictive models are developed, each with a different set of predictors. This study utilises two machine learning techniques, random forest and stochastic gradient boosting, for prediction via the three models. The data are collected from a sample of 1250 annual reports of 125 nonfinancial firms in Pakistan for the period 2011-2020.

Findings: Our results indicate that both narrative disclosure tone and corporate governance indicators significantly add to the accuracy of financial predictive models of firm performance.

Practical implications: Our results offer implications for the restoration of investor confidence in the highly uncertain Pakistani market by establishing nonfinancial disclosures as reliable predictors of future firm performance. Accordingly, they encourage investors to pay more attention to these disclosures while making investment decisions. In addition, they urge regulators to promote and strengthen the reporting of such nonfinancial information.

Originality: This study addresses the neglect of nonfinancial disclosures in the prediction of firm performance and the scarcity of corporate governance literature relevant to the use of machine learning techniques.

Keywords: Firm Performance, Machine Learning, Random Forest, Stochastic Gradient Boosting, Narrative Disclosure Tone, Corporate Governance.

1. Introduction

The ability to predict a firm's performance with ever-improved accuracy is imperative, especially in highly uncertain environments (Yang *et al.,* 2019). In this regard, Hoang and Wiegratz (2023) contend that machine learning (ML) techniques are becoming increasingly popular in finance because of their superiority over traditional econometric techniques. For instance, they establish that ML techniques significantly reduce prediction errors relative to econometric techniques. Accordingly, several studies have reported encouraging results using ML techniques to predict financial outcomes (Van Binsbergen *et al.,* 2023). In addition, Rundo *et al.* (2019) reported that ML techniques can track complex interdependencies within high-dimensional data, rendering them relatively more robust. However, most of the empirical research on firm performance prediction employs traditional regression techniques (Chang *et al.,* 2015; Ibhagui and Olokoyo, 2018). While most of these discussions about firm performance investigate financial indicators as predictors, there have been suggestions that nonfinancial disclosures, such as narrative disclosure tone, have an essential role in this regard (Saha and Kabra, 2022; Saha, 2024). For instance, Beretta *et al.* (2021) elaborate on incremental information theory's stance that narrative disclosure tone (NDT) provides additional information relative to financial disclosures. In a similar context, El-Deeb *et al.* (2022) use signalling theory to establish that managers utilise NDT to signal investors about a firm's future performance.

Despite this, Iqbal and Riaz (2021) contend that empirical research on the prediction of financial outcomes using nonfinancial disclosures, such as NDT, is scarce. Accordingly, only a limited number of studies have investigated the association between NDT and firm performance (Beretta *et al.,* 2021; Iqbal and Riaz, 2021; Mousa *et al.,* 2022). Therefore, Iqbal and Riaz (2021) suggest that the neglect of NDT regarding its ability to predict firm performance is an evident gap in the literature. Another form of nonfinancial disclosures that can predict firm performance are corporate governance indicators (CGIs) (Puni and Anlesinya, 2020).

From an academic standpoint, the relationship between CGIs and firm performance is grounded in multiple theories, such as agency and stakeholder theories (Jensen and Meckling, 1976; Arora and Sharma, 2016). Agency theory posits that better corporate governance mechanisms reduce informational asymmetry and enhance performance (Jensen and Meckling, 1976). From a stakeholder theory perspective, Arora and Sharma (2016) suggest that corporate governance has an imperative role in balancing the diverse interests of all stakeholders, ultimately leading to better performance. While this notion has received sufficient empirical attention, Di Vito and Trottier (2022) contend that most of it employs conventional techniques. Consequently, they suggest that using machine learning algorithms in the corporate governance literature is rare yet in high demand. Furthermore, Yang *et al.* (2019) suggest that an additional source of inconsistency stems from the contrasting roles of market and accounting-based performance (Yang *et al.,* 2019). Accordingly, they posit that focusing on a single facet of performance may lead to individual bias. Finally, Mousa *et al.* (2022) deem that identifying predictors of firm performance is more vital in an emerging economy, as most research in this regard focuses on developed markets.

Accordingly, this study aims to fill the abovementioned gaps by utilising machine learning algorithms to examine whether nonfinancial disclosures improve the prediction of firm performance in Pakistan.

In this regard, Pakistan is an apt setting because it is experiencing unprecedented economic uncertainty and diminishing investor confidence (Rashid *et al.,* 2022). Consequently, investors in such a setting rely on nonfinancial disclosures to foresee future performance (Aly *et al.,* 2018; Shahid and Abbas, 2019). Therefore, establishing the reliability of such disclosures in this context is imperative for restoring investor confidence and the stability of the business environment.

This study uses two widely popular ML techniques for prediction: random forest (RF) and stochastic gradient boosting (SGB). The data are collected from the annual reports of 125 nonfinancial firms in Pakistan spanning ten years from 2011-2020. Sentiment analysis is performed for the operationalisation of NDT, while CGIs are taken directly from the annual reports. Firm performance is proxied by two accounting-based (ROA and ROE) and two market-based estimates (Tobin's Q and MTB). Finally, three predictive models are developed, each containing a different set of predictors. Model 1 contains a set of financial disclosures as predictors, while Model 2 contains NDTs and financial disclosures as predictors. Finally, Model 3 contains CGIs and financial disclosures as predictors. Prediction using each of these models is then performed by utilising the two ML techniques. Accordingly, the performance of Models 2 and 3 is compared with that of Model 1. This is specifically to test whether the addition of nonfinancial disclosures to strictly financial predictive models of firm performance improves accuracy.

Our results show that both NDT and CGIs significantly enhance the accuracy of predictive models based on financial predictors alone. CGIs improve the prediction of all four performance proxies employed in the study, while NDT is deemed relatively imperative for the prediction of market-based performance.

By conducting this research, we contribute to the literature in four ways. First, we address the scarcity of related research relevant to the use of NDT as a predictive tool by establishing its ability to accurately predict firm performance (Mousa *et al.,* 2022). Second, our results also contribute by utilising ML algorithms to establish the reliability of CGIs as predictors of firm performance. In doing so, we respond to the call of Di Vito and Trottier (2022), as they contend that corporate governance needs to be amalgamated with the ML literature. Third, as Hoang and Wiegratz (2023) suggest, we contribute to the limited literature on ML in finance and accounting. This approach is especially relevant because ML techniques are becoming increasingly popular in finance due to their superiority over traditional regression techniques (Rundo *et al.,* 2019; Van Binsbergen *et al.,* 2023). Finally, this study contributes by examining the predictive ability of nonfinancial disclosures in an emerging economy, Pakistan. This is crucial, as most related research has focused on developed markets and is not generalisable to emerging economies (Iqbal and Riaz, 2021; Mousa *et al.,* 2022). Moreover, as Rashid *et al.* (2022) suggest, Pakistan is characterised by heightened economic uncertainty. Therefore, identifying ways to improve the predictability of firm performance is imperative for all stakeholders in such a setting.

The remainder of this paper is structured as follows. In Section 2, we focus on the literature review and the development of hypotheses. Section 3 describes the data and methodology. Section 4 delineates the empirical framework. The results are described and discussed in Section 5. Finally, Section 6 concludes the study.

2. Literature Review and Hypotheses

2.1. Theoretical Framework

The overarching theoretical framework of the current study is based primarily on incremental information, signalling, agency and stakeholder theories (Arora and Sharma, 2016; Beretta *et al.,* 2021). While incremental information, signalling and agency theories explain the relationship between NDT and firm performance, agency and stakeholder theories justify the relationship between CGIs and performance.

The main premise of incremental information theory is that companies utilise NDT to reduce informational asymmetry and provide value-relevant information to future investors (Beretta *et al.,* 2021). In this manner, incremental information theory suggests that NDT in annual reports provides incremental information about a firm's performance relative to financial disclosures. In line with this, signalling theory posits that managers use NDT to signal investors about the firm's future (El-Deeb *et al.,* 2022; Mousa *et al.,* 2022). In addition, agency theory suggests that information asymmetries rise due to agency conflicts between managers and shareholders, ultimately compromising performance (Jensen and Meckling, 1976). Interestingly, Saha and Kabra (2022) suggest that voluntary disclosures, such as narratives in annual reports, can effectively reduce informational asymmetries and agency costs, thereby enhancing performance. In addition, they posit that good corporate governance can complement NDT in the process.

Therefore, a possible solution grounded in agency theory is to employ a better corporate governance framework, as it can potentially eliminate agency conflicts and informational asymmetries (Jensen and Meckling, 1976; Puni and Anlesinya, 2020). Another theoretical perspective regarding the relationship between corporate governance and firm performance comes from the stakeholder theory (Arora and Sharma, 2016; Adedeji *et al.,* 2020). Interestingly, stakeholder theory has two further branches, namely, normative and instrumental stakeholder theories (Ayuso *et al.,* 2014). According to Ayuso *et al.* (2014), normative stakeholder theory prioritises ethics over shareholder wealth maximisation, suggesting that firms are responsible to society at large. In contrast, instrumental stakeholder theory deems that better firm performance is in the interest of all stakeholders of a firm in the long term. Accordingly, instrumental stakeholder theory suggests balancing the interests of the firm's stakeholders to enhance long-term performance (Adedeji *et al.,* 2020). In this context, Arora and Sharma (2016) suggest that better corporate governance is crucial.

Below, we utilise the theoretical framework of the study discussed above to discern relevant literature regarding the prediction of firm performance.

2.2. Review of relevant literature

Most literature on firm performance prediction employs financial disclosures as predictors (Delen *et al.,* 2013; Ibhagui and Olokoyo, 2018; Lambey *et al.,* 2021). For instance, Delen *et al.* (2013) employ financial ratios in their study and suggest that liquidity and leverage ratios have predictive ability regarding the forecasting of future firm performance. Interestingly, firm size is also critical (Ibhagui and Olokoyo, 2018). As such, Ibhagui and Olokoyo (2018) predict that the negative effect of leverage on performance is more imminent in smaller firms. Furthermore, Lambey *et al.* (2021) suggest that older firms signal more experience and are associated with high performance. In addition, financial indicators such as operating cash flows and firm risk are also deemed crucial in this regard (Chang *et al.,* 2015). While there is ample empirical evidence that financial disclosures can effectively predict firm performance, the utilisation of nonfinancial disclosures such as NDTs in this regard remains relatively unexplored (Beretta *et al.,* 2021; Mousa *et al.,* 2022). Despite this, the literature has acknowledged the relationship between nonfinancial voluntary disclosures and firm performance. For instance, Saha and Kabra (2020) suggest that voluntary disclosures are essential for communicating firm performance to potential investors. From an empirical standpoint, Saha (2024) finds a positive relationship between voluntary disclosures and firm value. One form of such disclosures are narrative in nature (El-Deeb *et al.,* 2022).

According to Iqbal and Riaz (2021), narrative disclosures within annual reports offer a detailed description of how the management views the firm. Furthermore, they complement financial disclosures and constitute an important part of annual reports. In an interesting empirical study, Beretta *et al.* (2021) utilise NDT as a determinant of future firm performance. However, their study was limited to the automotive industry. Similarly, Aly *et al.* (2018) utilise NDT as a predictor of financial performance using traditional regression techniques, whereas more advanced methods of analysis are recommended (Mousa *et al.,* 2022). In a comprehensive analysis specific to Vietnamese listed firms, Tran *et al.* (2023) investigate whether linguistic tone in annual reports can predict a firm's future performance. However, similar to most other studies in this regard, their study was also limited to traditional regression techniques. In contrast, Mousa *et al.* (2022) utilise three machine learning algorithms to determine whether the performance of financial predictive models is improved by incorporating NDT in them. Despite the use of machine learning techniques, their study is limited to banking institutions and has a relatively small sample size. Accordingly, they suggest extending their study by incorporating a more extensive and diverse sample of nonfinancial firms. Furthermore, they encourage the use of certain CGIs in the prediction of performance.

CGIs constitute an important part of overall nonfinancial disclosures. Interestingly, the notion that better CGIs enhance firm performance has received ample empirical attention (Ramadan and Hassan, 2022; Saha, 2024). For instance, Alodat *et al.* (2022) investigate the impact of CGIs on firm performance using a sample of listed firms in the Amman stock exchange. Moreover, in the Ghanaian context, Puni and Anlesinya (2020) also investigate the relationship between corporate governance and firm performance. However, their study failed to incorporate important control variables due to the large number of main predictors. This limitation could be attributed to the use of traditional regression techniques and their inability to deal with high-dimensional data (Rundo *et al.,* 2019). While most of these studies analysing the corporate governance-firm performance nexus employ traditional regression techniques, using ML algorithms in this context is scarce and imperative (Di Vito and Trottier, 2022). This is especially important due to the ability of ML algorithms to address high-dimensional data, trace complex interdependencies within the data, and significantly reduce prediction errors (Hoang and Wiegratz, 2023). Therefore, our review of the relevant literature identifies a myriad of prominent gaps regarding the prediction of firm performance.

First, most related research in this regard is restricted to financial disclosures as predictors, while the use of nonfinancial disclosures is limited (Iqbal and Riaz, 2021). The current study responds to this by employing NDT and CGIs as predictors. Second, the limited literature focusing on NDT as a predictor of performance employs traditional regression techniques and is mainly limited to a specific industry (Aly *et al.,* 2018; Beretta *et al.,* 2021). The current study addresses this issue by employing ML algorithms and a sample comprising nonfinancial firms representing various sectors for prediction purposes. Third, the existing corporate governance literature is also restricted to traditional regression techniques, whereas the use of ML algorithms in this context has explicitly been advocated (Di Vito and Trottier, 2022). As mentioned above, ML algorithms have shown superior performance over traditional techniques, especially in reducing prediction errors (Hoang and Wiegratz, 2023). Therefore, the current study answers the call of Di Vito and Trottier (2022) by amalgamating the corporate governance literature with ML. Below, we delineate specific empirical literature regarding the development of hypotheses.

2.2.1. Narrative disclosure tone (NDT)

NDT as a significant predictor of firm performance has recently achieved empirical significance (Mousa *et al.,* 2022; Saha, 2024). For instance, Iqbal and Riaz (2021) provide empirical evidence that NDT predicts future firm performance in 58 banks from 16 emerging economies. Similarly, Beretta *et al.* (2021) empirically demonstrate that the positive tone in narrative disclosures of the top 10 automotive companies worldwide predicts their ESG performance. In addition, Tran *et al.* (2023) find that the linguistic tone of narrative disclosures in the annual reports of Vietnamese firms can predict firm performance one year ahead. They also highlight important implications for policymakers to revise regulations regarding these disclosures and for investors to pay more attention to them for investment decision-making. Furthermore, Saha (2024) finds a significant relationship between overall voluntary disclosures and firm value. They also establish that voluntary disclosures mediate the relationship between corporate governance and firm performance. In another similar analysis specific to Egyptian firms, El-Deeb *et al.* (2022) find that the NDT in annual reports significantly impacts firm value. In line with Saha (2024), they also find that voluntary disclosures, such as NDT, mediate the impact of corporate governance on firm value. Finally, Mousa *et al.* (2022) provide empirical evidence supported by machine learning algorithms that NDT enhances the ability of financial disclosures to predict firm performance. This notion is theoretically grounded in incremental information theory, agency and signalling theories (Beretta *et al.,* 2021).

As mentioned above, incremental information theory posits that narrative disclosures contain value-relevant incremental information about the future of firm performance that cannot be captured by financial disclosures alone (Beretta *et al.,* 2021). From an agency theory perspective, Saha and Kabra (2022) posit that voluntary disclosures are essential tools for reducing informational asymmetries and agency costs, eventually leading to better performance. Finally, both of these theoretical perspectives align with signalling theory, which posits that managers signal the firm's future performance through NDT (El-Deeb *et al.,* 2022; Mousa *et al.,* 2022). Despite these theoretical perspectives and relevant empirical literature cited above, using narrative disclosures and their tone as predictors of firm performance is still a relatively unexplored area of research, especially in emerging economies (Mouse *et al.,* 2022).

In this context, Pakistan provides the unique setting of an emerging economy characterised by heightened economic uncertainty (Shahid and Abbas, 2019). In such a setting, the business environment suffers as investor confidence diminishes. Consequently, most investors in such an environment rely on narrative disclosures for decision-making as they are important tools through which information about firm performance is communicated (Aly *et al.,* 2018). Accordingly, Saha and Kabra (2022) posit that voluntary disclosures are imperative for the restoration of investor confidence in environments of uncertainty. Therefore, based on theoretical, empirical and contextual motivation, we develop the following hypothesis:

H1: Narrative disclosure tone improves the ability of financial disclosures to predict firm performance.

2.2.2. Corporate governance indicators (CGIs)

CGIs have always been empirically relevant to firm performance (Saha, 2024). For instance, Ramadan and Hassan (2022) find a positive impact of board size and institutional ownership on the performance of Egyptian listed firms. Furthermore, Saha (2024) contends that corporate governance mechanisms complement voluntary disclosures to significantly improve firm value. In addition, Alodat *et al.* (2022) find empirical evidence of a significant relationship between CGIs, such as board and audit committee composition, and firm performance. Furthermore, they highlight that this has a plethora of implications for developing more stringent regulations in emerging economies. Puni and Anlesinya (2020) establish an empirical relationship between CGIs and firm performance in another developing economy setting. Moreover, they urge regulatory authorities in emerging economies to establish a better corporate governance code. From a strictly academic standpoint, CGIs and their impact on firm performance is grounded in agency theory and stakeholder theories (Jensen and Meckling, 1976; Arora and Sharma, 2016; Adedeji *et al.,* 2020).

As previously mentioned, the central premise of agency theory is that informational asymmetries caused by internal agency conflicts hinder performance. Agency theory further suggests that better governance mechanisms can reduce this effect and eventually enhance performance. Furthermore, Arora and Sharma (2016) posit from a stakeholder theory perspective that good corporate governance structures balance the interests of all stakeholders of a firm, leading to better long-term performance. Interestingly, this approach becomes more relevant in emerging economies with extreme economic uncertainty and rapidly diminishing investor confidence, such as Pakistan (Shahid and Abbas, 2019).

Therefore, it is imperative for regulators and policymakers to restore the confidence of investors in such settings (Shahid and Abbas, 2019). Similar to voluntary disclosures, Saha and Kabra (2022) also identify CGIs as essential communication tools for mitigating information asymmetry and promoting investor confidence. In this context, establishing the reliability of CGIs as predictors of firm performance is more relevant in settings such as Pakistan. Therefore, we develop the following hypothesis based on the above discussion:

H2: Corporate governance indicators improve the ability of financial disclosures to predict firm performance**.**

2.2.3. Market and accounting-based performance measures

The literature relevant to firm performance is subject to inconsistencies between its multiple dimensions, primarily market and accounting-based performance, making it important to capture a holistic performance perspective when predicting it (Yang *et al.,* 2019).

According to Yang *et al*. (2019), one major difference where the literature converges is that accounting-based estimates of firm performance reflect the past, whereas market-based measures reflect the future. Furthermore, they suggest that market-based measures consider market factors and reflect investors' future growth expectations. Interestingly, Davis *et al.* (2012) find that nonfinancial disclosures, such as the tone of earnings press releases, are also more market-oriented. They provide empirical evidence and deem that net optimistic tone in earnings press releases is positively associated with market-based future performance. In another interesting analysis, Dalwai *et al.* (2021) deem that annual reports that are easier to read are associated with a higher market-based performance and thus reflect the confidence of investors. The contradictory role of market-based and accounting-based measures is also evident through their association with corporate governance variables (Elvin and Bt Abdul Hamid, 2016; Ramadan and Hassan, 2022).

For instance, Elvin and Bt Abdul Hamid (2016) empirically prove that corporate governance and ownership structure variables are more relevant to market-based performance. They explain this by suggesting that governance mechanisms have evolved to be more market-centric and are focused on futuristic value creation. Given the rising investor uncertainty in Pakistan, it would be interesting to test whether the market responds relatively more to nonfinancial disclosures, as suggested by literature (Dalwai *et al.,* 2021). Therefore, we develop the following hypotheses:

H3: Narrative disclosure tone improves the ability of financial disclosures to predict market-based measures of firm performance relatively more than accounting-based measures.

H4: Corporate governance indicators improve the ability of financial disclosures to predict market-based measures of firm performance relatively more than accounting-based measures.

3. Data and Methodology

3.1. Data collection and sample

The data are extracted from a population of all publicly listed nonfinancial firms in Pakistan. The use of nonfinancial companies is suitable because the literature on the prediction of firm performance within financial companies is well established (Aly *et al.,* 2018; Mousa *et al.,* 2022). First, we exclude firms from this population for which relevant data are incomplete or not publicly available. Second, we eliminate firms whose annual reports are not machine-readable and cannot be converted into one. This is crucial, as we operationalise NDT via the automated natural language processing technique using the software R. Therefore, after excluding these firms, our final sample consists of annual reports of 125 nonfinancial firms listed on the Pakistan Stock Exchange.

In this regard, Pakistan provides a suitable setting because it is experiencing severe social, political and economic uncertainty (Shahid and Abbas, 2019; Rashid *et al.,* 2022). In such settings, investor confidence is strongly shattered and the business environment suffers as the capital provided by investors trickles down (Shahid and Abbas, 2019). In particular, Aly *et al.* (2018) deem that investors rely on narrative disclosures in such settings for performance-based investment decision-making. Furthermore, Saha (2024) finds that CGIs are also important channels through which firms communicate performance and are also imperative for the restoration of investor confidence in highly uncertain environments. For this reason, it is crucial to establish the reliability of these disclosures in predicting firm performance within the unique Pakistani context, which is characterised by an unstable business environment.

Furthermore, most studies relevant to the prediction of firm performance utilise a sample limited to one sector (Beretta *et al.,* 2021; Iqbal and Riaz, 2021; Mousa *et al.,* 2022). Therefore, Mousa *et al.* (2022) suggest utilising a sample of firms covering diverse sectors. Accordingly, our sample covers firms from various sectors, which are summarised along with the sampling procedures in Table 1. Furthermore, the data span ten years from 2011-2020. This period is suitable because it marks the beginning of the post global financial crisis era. As Harakeh *et al.* (2023) suggested, global financial markets suffered a significant loss of investor confidence in the market. Accordingly, identifying ways to improve the prediction of firm performance during such a time is crucial. Furthermore, relevant data at the time of collection were publicly available until 2020. Therefore, the final sample consists of 1250 annual reports, constituting a relatively appropriate sample size for most emerging economies.

[Insert Table 1 here]

The predictor variables used in the study are divided into two broad categories: nonfinancial and financial disclosures. Nonfinancial disclosures have two further sub-categories, namely, NDT and CGIs.

3.2. Nonfinancial disclosures as predictor variables

3.2.1. Narrative disclosure tone (NDT)

The first form of nonfinancial disclosures used as predictor variables in the study are represented by different disclosure tones operationalised through a sentiment analysis of annual reports. For this purpose, we utilise the lexicon of Loughran and McDonald (2011), which is a widely popular and reliable resource for performing such an analysis using financial text. This lexicon is highly specific to financial research because it contains words commonly occurring in annual reports or financial jargon, whereas other alternatives are more generalised (Mousa *et al.,* 2022). Accordingly, Goel and Uzuner (2016) suggest that this leads to fewer misclassifications and a more accurate representation of sentiment, rendering the Loughran-McDonald (LM) dictionary valid in the financial context. In addition, it is highly comprehensive, as Loughran and McDonald (2011) developed this dictionary by analysing a comprehensive sample of both forms of annual reports (annual and quarterly) from 1994-2008. Consequently, it comprises six comprehensive lists of words, each representing a particular tone commonly associated with the financial context (Mousa *et al.*, 2022). Specifically, it contains 2355 negative, 354 positive, 297 uncertain, 904 litigious, 56 superfluous and 184 constraining words. Other lexicons, such as the Harvard IV dictionary, contain only two broad tone categories: negative and positive. Therefore, the degrees of sentiments between the negative and positive extremes are largely ignored, leading to the misrepresentation of words in either of these extremes (Goel and Uzuner, 2016). With a wide variety of six tones, the LM dictionary ensures that words capture what they intend to capture, enhancing their validity (Goel and Uzuner, 2016; Mousa *et al.,* 2022). The LM dictionary can be retrieved via the *tidytext* package in the software R.

To further ensure the reliability of these scores, we follow Goel and Uzuner (2016) and count the number of words representing a particular tone using the automated and computerised process of natural language processing in R. We perform this automated word counting technique using the *tidytext* and *tidyverse* packages in R, which are widely recognised as reliable for this purpose (Fay, 2018; Mucko, 2021). Using these packages, R automatically reads an annual report and counts the words representing each tone based on the LM dictionary. We employ this specific automated technique for three reasons. First, Goel and Uzuner (2016) posit that the annual report is a large text and that manual counting of words in this context is unfeasible. Second, they deem this automatic recognition less prone to errors and personal bias, rendering it more reliable. Third, the availability of a comprehensive and context-specific lexicon, such as the LM dictionary, integrated in R makes this feasible (Loughran and McDonald, 2011; Goel and Uzuner, 2016; Beretta *et al.,* 2021; Mousa *et al.,* 2022).

Finally, we follow Mousa *et al.* (2022) and compute a score for each of the six tones within the LM dictionary based on the frequency of words representing each tone in a particular annual report. At the end of this process, we have a score for each of the six categories of tones in the LM dictionary, resulting in six predictor variables in the form of NDTs, namely, negative (NEG), positive (POS), uncertain (UNC), litigious (LIT), superfluous (SUP) and constraining (CON).

3.2.2. Corporate governance indicators (CGIs)

The CGIs used as predictor variables in this study were chosen after a thorough analysis of literature, as previously discussed (Puni and Anlesinya, 2020; Alodat *et al.,* 2022; Ramadan and Hassan, 2022; Saha, 2024). In total, we utilise twelve CGIs associated with firm performance in empirical literature, namely, board size (BSIZE), board independence (BI), board gender diversity (BGD), board meetings (BM), audit committee size (ACSIZE), audit committee independence (ACI), audit committee gender diversity (ACGD), audit committee meetings (ACM), institutional (IOWN), foreign (FOWN), managerial (MOWN) and concentrated ownership (COWN). The data for operationalising these variables are taken directly from the firm's annual reports. These twelve CGIs form our second set of nonfinancial predictor variables.

3.3. Financial disclosures as predictor variables

We utilise a total of six financial disclosures as predictor variables in this study based on the discussion of literature above, namely, firm size (SIZE), leverage (LEV), firm risk (BETA), cash flow from operating activities (CFO), firm age (AGE) and liquidity (LIQ). All of these variables, barring firm risk, are sourced from the firms' annual reports. In summary, 24 predictor variables are utilised in the study as predictors of our target variables.

3.4. Target Variables

For the target variables, two accounting-based and two market-based estimates of firm performance are utilised. ROA and ROE represent the accounting-based estimates, while Tobin's Q and market-to-book value are the market-based estimates. Once we have operationalised these variables, the next step is to form classes of each target variable. Specifically, we follow Mousa *et al.* (2022), who classify a single target variable into three classes based on the upper quartile, the interquartile range and the lower quartile.

For instance, the data points lying within the upper quartile of a particular target variable are labelled TOP. The observations within the interquartile range are MID; similarly, observations in the lower quartile are labelled LOW. This process is repeated for all four target variables. Consequently, we have three classes for each target variable. The operationalisation and source of all variables used in the study are summarised in Table 2.

[Insert Table 2 here]

3.5. Preparing the best-fit model using optimal feature selection

One of the most critical steps of machine learning classification techniques is feature selection (Xiaomao *et al.,* 2019). Features are another word for predictor variables in the ML literature. Feature selection works by filtering out irrelevant features for a particular target variable.

According to Yeh and Chen (2020), this approach avoids overfitting. In addition, this approach improves the simplicity and interpretability of the model (Xiaomao *et al.,* 2019). Simply put, feature selection aims to improve the accuracy of predictive models by identifying the most relevant predictors. After checking all the features in the study for multicollinearity via a variance inflation factor (VIF) test and generating a correlation matrix, we conducted feature selection via the Boruta algorithm.

The Boruta algorithm uses the random forest classifier and performs several iterations on the overall list of features (Mousa *et al.,* 2022). It eliminates features that are relatively inconsequential for classification of the target variable at every iteration. Given that we have four target variables, we run the Boruta algorithm for each target variable separately. At the end of this process, the most relevant features pertaining to the prediction of each target variable are obtained. We perform the Boruta algorithm using the *boruta* package in R. Next, we proceed to the training and testing split.

3.6. Splitting the dataset into two subsets – Training and Testing

After performing the Boruta algorithm and optimally selecting features, we split the data into training and testing data. As in all ML prediction problems, splitting data into training and testing data is crucial (Yeh and Chen, 2020). The training data are a subset of the entire dataset that the machine learning algorithm uses to learn patterns and consequently, applies them to the test dataset for prediction. Accordingly, the test dataset must be a part of the dataset that is never seen by the algorithm (Yeh and Chen, 2020). For analyses specific to the prediction of future data from past historical data, the training data always precede the testing data with respect to time (Moghaddam *et al.,* 2016). Therefore, we follow Mousa *et al.* (2022) in splitting our 10-year datasets into a 2011-2019 subsample as training data and data within 2020 as test data. This process is performed separately for each target variable and its most relevant predictors.

Finally, we apply suitable ML algorithms to the training data to train them and then apply them to the testing data for prediction. As mentioned above, Hoang and Wiegratz (2023) suggest that ML algorithms significantly reduce prediction errors relative to conventional techniques. Furthermore, they can track hidden complex relationships within the data, rendering them more robust (Rundo *et al.,* 2019). Finally, they can deal with high-dimensional data better than conventional techniques (Hoang and Wiegratz, 2023). Therefore, two ML algorithms are utilised for prediction. They are described in the empirical framework below, along with different models for the testing of hypotheses.

4. Empirical Framework

4.1. Algorithms

4.1.1. Random forest (RF)

The first supervised learning method we employ is random forest (RF). Chen *et al.* (2020) contend that RF is a popular technique for classification and maximising purity. In addition, they state that RF builds a myriad of randomised decision trees using the training data. Accordingly, it works by partitioning the feature space of a decision tree at each node using various tests. This process is continued until all decision tree nodes contain samples of a single class. This is how the RF algorithm learns. Accordingly, it can identify the output class given a set of inputs by utilising what it has learned (Chen *et al.,* 2020; Petropoulos *et al.,* 2020). Thus, it can predict the outputs in the test data by identifying the most commonly predicted class for a given set of inputs across decision trees during the training phase. Moreover, RF prevents overfitting and is robust to missing data (Halteh *et al.,* 2018; Petropoulos *et al.,* 2020). The RF algorithm is run using the *randomForest* package and library in R.

4.1.2. Stochastic gradient boosting (SGB)

Stochastic gradient boosting (SGB) represents a powerful ensemble prediction method (Halteh *et al*. 2018). Unlike the RF method, it generates numerous decision trees in a more sequential manner. The output is subsequently aggregated to produce the most accurate model. Furthermore, as Halteh *et al.* (2018) suggested, SGB is robust to relatively inaccurate measurements of the target variable. The sequential nature of tree building in SGB allows it to learn extra information with the addition of each new tree (Sadorsky, 2021). This helps SGB build an aggregate model with the highest accuracy. Several tuning parameters can be adjusted in an SGB model to find the optimal model. The SGB algorithm is run using *xgBoost* and the *caret* packages and libraries in R

4.2. Models

To test our hypothesis, we form different models for prediction with each model distinguished by the set of features in it.

4.2.1. Model 1 (Financial features)

Model 1 utilises only a set of financial features to predict each of the four target variables via both algorithms.

4.2.2. Model 2 (Financial and NDT features)

Model 2 utilises both NDT and financial features to predict each of the four target variables via both algorithms.

The comparison of Model 2 and Model 1´s prediction performance is then utilised to test H1. By comparing Model 2's prediction with Model 1's prediction, we test whether adding disclosure tone features to a model containing financial features alone improves the prediction of firm performance. Consequently, if Model 2 is a better predictive model than Model 1 concerning a particular target variable and algorithm, H1 is supported for that comparison. Note that predictions of Model 2 and Model 1 run using the same algorithm and for a particular target variable must be significantly different for the comparison to be valid. For that reason, following Mousa *et al*. (2022), a *t-test* is employed.

4.2.3. Model 3 (Financial and CGI features)

Model 3 contains CGIs and financial variables as features for predicting each target variable via both algorithms. The comparison of Model 3 and Model 1´s prediction performance is utilised to test H2. If Model 3 is a better predictive model than Model 1 for a particular target variable using a particular algorithm, H2 is supported for that comparison. Note that predictions of Model 3 and Model 1 run using the same algorithm and for a particular target variable must be significantly different for the comparison to be valid. For that reason, following Mousa *et al.* (2022), a *t-test* is employed.

4.2.4. The comparison of market-based and accounting-based measures of performance

To test hypothesis 3, we compare Model 2's prediction of market-based performance with its prediction of accounting-based performance. As mentioned above, Model 2 contains NDT and financial features utilised to predict each target variable separately. The prediction of each market-based target variable is compared to the prediction of each accounting-based target variable using the features in Model 2. These comparisons are repeated for each algorithm separately. For a particular comparison within the same algorithm, Model 2's performance with the two target variables being compared must be significantly different in order for the comparison to be valid. This difference is identified using a *t-test*. H3 is supported if Model 2 shows a greater improvement in the prediction of market-based performance measures relative to accounting-based measures. This process is repeated with the features in Model 3 to test H4.

4.3. Parameters for comparing the predictive models

To determine the predictive power of these algorithms, we utilise several parameters commonly used in the literature (Petropoulos *et al.,* 2020; Mousa *et al.*, 2022). Below, we briefly describe each of these metrics.

We employ accuracy and the Kappa coefficient to evaluate a single model as a whole (Mousa *et al.,* 2022). Specifically, accuracy measures the proportion of correct classifications and predictions. In contrast, the Kappa coefficient measures how frequently the model performs when it is compared by chance or its reliability. These two measures are employed by Mousa *et al.* (2022) when they assess the performance of banking institutions through machine learning algorithms. In addition, the model's significance is also monitored using a statistical test. The null hypothesis for this test is that accuracy is equal to the no information rate (the highest proportion of the observed classes), while the alternate hypothesis is that accuracy is greater than the no information rate. Accordingly, if the null hypothesis of this statistical test is rejected, the model is significant. In addition, we utilise certain class-specific metrics to analyse the performance of the classes individually.

First, sensitivity and specificity are especially employed. According to Petropoulos *et al.* (2020), these measures eliminate any doubt of misinterpretation of model performance. Consequently, they utilise these measures to evaluate the performance of machine learning algorithms in predicting bank insolvency. As explained by Mousa *et al.* (2022), for a given class, sensitivity reflects the percentage of acceptance of a correct classification, while specificity reflects the percentage of rejection of an incorrect classification. Having covered specific measures for classification, we also utilise measures for the prediction performance of each class. Accordingly, we gauge prediction performance by employing positive predicted value (PPV) and negative predictive value (NPV). PPV measures the percentage of acceptance of a correct prediction, while NPV measures the percentage of rejection of an incorrect prediction (Mousa *et al.,* 2022). Furthermore, as suggested by Petropoulos *et al.* (2020), balanced accuracy, which is the mean of sensitivity and specificity, is also employed.

Finally, certain variable importance measures are utilised to determine the most important features for each prediction. Following Sadorsky (2021), we employ the mean decreased Gini metric for random forest. This metric is generated using the *VarImp* function in the *randomForest* package. Specifically, for SGB, we use the same function in the *Caret* package and generate a metric of relative importance (Halteh *et al.,* 2018).

5. Results and Discussion

5.1. Descriptive Statistics

The descriptive statistics of the overall sample are presented in Table 3. Interestingly, NEG tone is the most dominant across the sample with an average of 427 words in an annual report. Interestingly, it also has the most standard deviation. All other tones, barring SUP, are not far behind. This implies that our sample has expressed various sentiments through narrative disclosures.

[Insert Table 3 here]

In addition, the sample has an institutional setting, with a mean of 59% institutional ownership. This is interesting because, in such settings, the overall effectiveness of CGIs, such as independent directors, decreases as agency problems increase (Saha, 2024). Therefore, the results regarding corporate governance would be interesting in this context.

5.2. Feature selection using the Boruta algorithm

5.2.1. ROA and ROE

The results of the Boruta algorithm for the prediction of ROA and ROE are shown in Figure 1.

[Insert Figure 1 here]

As depicted by the colour green, 23 attributes are confirmed to be significant predictors of both ROA and ROE. However, ACM for both is depicted in yellow. This means that the Boruta algorithm does not have the desired confidence in the importance of this feature, and its decision is tentative. Therefore, we follow Kursa and Rudnicki (2010) and use the *TentativeRoughFix* of the *boruta* package function to decide on this feature. After performing this function, ACM is deemed unimportant for the prediction of both ROA and ROE and is eliminated.

5.2.3. Tobin's Q and MTB

The results of the Boruta algorithm for the prediction of Tobin's Q and MTB are shown in Figure 2. All 24 features are depicted in green for both target variables and are confirmed to be significant predictors of Tobin's Q and MTB. Therefore, after identifying the most relevant features for all four target variables, we proceed to their prediction using RF and SGB.

[Insert Figure 2 here]

5.3. Performance comparison of models using the RF and SGB algorithms

Below, we separately delineate the results regarding the prediction of all four target variables and test our hypotheses.

5.3.1. ROA

Table 4 shows the overall performance of each model with metrics relevant to the prediction of ROA. RF's prediction of ROA using Model 1 achieves 68% accuracy. However, the p-value for Model 1 is 0.15, which indicates that the model is insignificant. Model 2 performs worse relative to Model 1, as it achieves 65% accuracy and is insignificant at 0.39. Interestingly, Model 3 achieves 72% accuracy with a Kappa coefficient of 43% and a p-value of 0.02. As Model 3 performs better than Model 1, H2 is supported. This finding implies that adding CGI features to financial features improves the prediction of firm performance when proxied by ROA. The results regarding SGB's prediction of ROA follow a similar trend as predictions using both Models 1 and 2 are insignificant, with a p-value above 0.1. In contrast, Model 3 achieves 70% accuracy, with a Kappa coefficient of 42% and is significant at 0.06. Therefore, our results regarding the prediction of ROA using SGB also support H2.

[Insert Table 4 here]

The class-specific characteristics are shown in Table 5. In Model 1's prediction using RF, the Mid class performs best regarding sensitivity alone, while the Low class performs best regarding specificity and PPV. Finally, the top class performs best regarding NPV and balanced accuracy. Furthermore, Models 2 and 3, achieve parallel results using the RF algorithm, barring a few exceptions. The class-specific metrics obtained using SGB for all three models also follow a similar pattern.

[Insert Table 5 here]

The most important variables for the prediction of ROA in all three models using the RF and SGB algorithms are shown in Figure 3. Regarding RF, LEV is the most important variable in all three models, followed by CFO and LIQ. However, in Model 2, POS outranks AGE; in Model 3, all ownership structure variables outrank AGE, while IOWN outranks SIZE. This implies the importance of certain NDTs and CGIs over certain financial variables.

[Insert Figure 3 here]

These results are similar for the SGB algorithm, barring certain exceptions, as CFO is the most important predictor of ROA in all three models, while NEG outranks AGE as an important predictor in Model 2.

5.3.2. ROE

As evident from the overall metrics presented in Table 6, Models 2 and 3 significantly outperform Model 1 regarding accuracy, Kappa coefficient and significance when ROE is predicted using the RF algorithm. This lends support to both H1 and H2.

[Insert Table 6 here]

Specifically, adding NDT to financial variables improves accuracy from 62% to 65% and the Kappa coefficient from 33% to 38%, as is evident from the comparison between models 1 and 2. Moreover, adding CGIs to financial variables improves accuracy from 62% to 67% and the Kappa coefficient from 33% to 43%, as is evident from the comparison between models 1 and

3. Finally, Models 2 and 3 are significant, while Model 1 is insignificant. Therefore, it is clear from our results using the RF algorithm that both NDT and CGIs significantly improve the prediction of firm performance as proxied by ROE. However, when the prediction of ROE is performed using the SGB algorithm, Models 1 and 2 are insignificant, as their p values are greater than 0.1. However, Model 3 is significant with a p-value of 0.06 and performs relatively better with respect to accuracy and the Kappa coefficient. This finding lends support to H2 and implies that CGIs improve the prediction of ROE. Therefore, the prediction of ROE using the RF algorithm supports both H1 and H2, while its prediction using the SGB algorithm supports H2 alone. Finally, the class-specific characteristics are shown in Table 7.

[Insert Table 7 here]

Regarding the variable importance in these models, Figure 4 shows that CFO is the most important feature in the prediction of ROE for all three models using both algorithms.

[Insert Figure 4 here]

However, ownership structure variables show their importance in the prediction of ROE, as they outrank certain financial features in Model 3 using both algorithms. In addition, POS outranks AGE using the RF algorithm, while both NEG and POS outrank AGE using SGB. This finding implies that in the prediction of ROE, certain NDTs and CGIs are more important than certain financial features.

5.3.3. Tobin's Q

Table 8 shows the overall metrics pertaining to all three models for the prediction of Tobin's Q. All three models using both algorithms are significant. Model 1 achieves 57% accuracy using the RF algorithm, while Models 2 and 3 achieve accuracies of 62% and 66%, respectively. This trend of improvement can also be observed in the Kappa coefficient. Consistent with this, the results of the SGB algorithm show that Models 2 and 3 are 60% and 70% accurate, respectively, while Model 1 is 57% accurate. This pattern is also evident in the kappa coefficient. Therefore, the results obtained using both algorithms pertaining to the prediction of Tobin's Q lend support to both H1 and H2, providing evidence of the predictive ability of both NDTs and CGIs. The class-specific characteristics are presented in Table 9.

[Insert Table 8 here]

[Insert Table 9 here]

Finally, the most important variables for these models are shown in Figure 5. LEV is consistently the most important feature in all models predicting Tobin's Q using the RF algorithm. Interestingly, POS outranks BETA in Model 2, while all ownership structure variables, barring MOWN, overlap certain financial variables in Model 3. The results for the SGB algorithm are similar, except for Model 2, where LIQ outranks LEV as the most important predictor.

[Insert Figure 5 here]

5.3.4. MTB

The overall metrics for the prediction of MTB are shown in Table 10. All three models using both algorithms are significant. Models 2 and 3 use the RF algorithm to achieve 70% and 75% accuracy, respectively, relative to Model 1's 68%. The Kappa coefficient follows a parallel trend.

[Insert Table 10 here]

Similarly, using the SGB algorithm, Models 2 and 3 achieve 69% and 76% accuracy, respectively, relative to Model 1's 62%. Regarding the Kappa coefficient, Models 2 and 3 achieve 47% and 60%, respectively, relative to Model 1's 34%. Therefore, the results of both algorithms support H1 and H2, providing evidence of the improved prediction of MTB using both NDT and CGIs. The class-specific characteristics for the prediction of MTB using both algorithms are shown in Table 11.

[Insert Table 11 here]

Finally, the most important variables for all three models relevant to the predictions of MTB utilising both algorithms are shown in Figure 6. AGE is the most important variable for the prediction of MTB in Models 1 and 3 using the RF algorithm, while CFO outranks AGE in Model 2. However, POS outranks a financial variable (LIQ) as an important predictor of MTB, while all ownership structure variables appear to be important corporate governance features, as they outrank certain financial features.

[Insert Figure 6 here]

However, when the SGB algorithm is used, CFO is the most important variable for the prediction of MTB in Models 1 and 2. In addition, COWN is the most important predictor of MTB in Model 3, followed by FOWN and IOWN.

5.4. Comparison of accounting and market-based performance estimates

To test H3, Model 2's predictions of accounting-based estimates are compared with its predictions of market-based estimates using both RF and SGB algorithms.

Specifically, for the RF algorithm, Model 2 is insignificant when it predicts ROA, while Model 2's predictions of both Tobin's Q and MTB are highly significant and achieve accuracies of 62% and 70%, respectively. Therefore, this finding supports H3, which states that by adding NDT to financial predictive models, market-based estimates of firm performance are better predicted than accounting-based estimates. Interestingly, Model 2's prediction of ROE is significant with 65% accuracy and consequently outperforms its prediction of Tobin's Q. This contradicts H3. However, Model 2's prediction of MTB performs best when compared to its predictions of both ROE and ROA, as it is highly significant with an accuracy of 70%, lending further support to H3. This pattern is also evident in our results using the SGB algorithm, as Model 2's prediction of both ROA and ROE are insignificant. In contrast, its predictions of both Tobin's Q and MTB are highly significant. Accordingly, these results support H3, implying that NDT adds to the prediction of market-based estimates relatively more than accounting-based estimates.

Similarly, for H4, we compare Model 3's prediction of both accounting-based estimates with that of both market-based estimates. Using the RF algorithm, comparing Model 3's prediction of ROA and Tobin's Q provides evidence against H4. Specifically, the prediction of ROA is 72% accurate and has a Kappa coefficient of 43%. However, the prediction of Tobin's Q is only 66% accurate, with a Kappa coefficient of 42%. A similar pattern is evident when the prediction of Tobin's Q is compared with that of ROE, providing further evidence in contradiction to H4. Furthermore, Model 3's prediction of MTB using the RF algorithm performs best in terms of all overall metrics relative to its predictions of both ROA and ROE. These results lend support to H4. The results using the SGB algorithm also support H4, as Model 3's predictions of both market-based estimates outperform its predictions of both accounting-based estimates in terms of accuracy, the kappa coefficient and significance. This implies that like NDT, CGIs also improve the prediction of market-based estimates of performance relatively more than accounting-based estimates.

5.5. Summary and discussion of findings

In summary, our overall results indicate that CGIs improve the prediction of all four proxies of firm performance using both algorithms. However, NDT mainly improves the prediction of market-based performance. These results are consistent with theoretical and empirical literature (Aly *et al.,* 2018; El-Deeb *et al.,* 2022; Mousa *et al.,* 2022; Ramadan and Hassan, 2022; Saha, 2024).

First, the results highlight the imperativeness of NDT in improving the prediction of firm performance. In doing so, they confirm incremental information theory's stance that NDT provides value-relevant information about the firm's future, in addition to what financial disclosures depict (Beretta *et al.,* 2021). Furthermore, these results are also rooted in the agency theory perspective. For instance, Saha and Kabra (2022) suggest that voluntary disclosures, such as narrative disclosures, are important tools through which managers reduce agency costs and communicate information about firm performance. This also aligns with signalling theory, as it suggests that managers signal information about the firm's future performance via narrative disclosures (El-Deeb *et al.,* 2022). Moreover, these results are also supported by empirical literature. For instance, Beretta *et al.* (2021) empirically prove that disclosure tone captures incremental information about a firm's ESG performance in the context of the global automotive industry. They explain this by suggesting that firms globally are now more aware that misreporting can have negative consequences. In addition, Saha (2024) provides empirical evidence that overall voluntary disclosures within firms listed on the Bombay Stock Exchange impact firm value. Similarly, El-Deeb *et al.* (2022) find consistent results in the Egyptian context, where they analyse NDT and its impact on firm value. Finally, Mousa *et al.* (2022) use ML algorithms to empirically provide evidence that NDT improves the prediction of a firm's future performance specific to banking institutions in emerging markets. They justify this by suggesting that NDT captures value-relevant information regarding a firm's performance. Therefore, our results regarding NDT contribute to the literature by confirming its increased importance in a developing economy characterised by heightened economic and financial instability. Moreover, our results are not limited to firms within a specific sector (refer to Table 1), thereby indicating that narrative disclosure tone predicts firm performance in a diversity of firms.

Second, our results regarding CGIs can be explained by agency and stakeholder theories (Jensen and Meckling, 1976; Arora and Sharma, 2016; Adedeji *et al.,* 2020). As mentioned above, agency theory posits that better CGIs reduce informational asymmetry and agency costs, enhancing performance (Adedeji *et al.,* 2020). Furthermore, Ayuso *et al.* (2014) argue from a stakeholder theory perspective that firm performance benefits all stakeholders of a firm.

Therefore, good corporate governance functions to balance the interests of a diverse set of stakeholders, which leads to better long-term performance (Arora and Sharma, 2016). In addition, our results are supported empirically by a myriad of studies that have established a significant link between corporate governance and firm performance (Puni and Anlesinya, 2020; Ramadan and Hassan, 2022; Saha, 2024). For instance, Puni and Anlesinya (2020) and Saha (2024) provide empirical evidence of CGIs' significant impact on firm value in the Ghanaian and Indian contexts, respectively. Interestingly, Saha (2024) suggests that certain CGIs are less effective in a sample dominated by an institutional setting. As is evident from our descriptive statistics (refer to Table 3), our sample is characterised by 59% institutional ownership. Despite this, our results indicate that CGIs improve the prediction of all four performance proxies, highlighting their magnifying impact in the Pakistani context. Furthermore, Alodat *et al.* (2022) and Ramadan and Hassan (2022) also confirm that CGIs significantly impact firm performance. However, most of these studies rely on traditional regression techniques. Therefore, Di Vito and Trottier (2022) point out an increasing need to establish the reliability of CGIs as predictive tools of firm performance by utilising ML algorithms. This is because ML techniques are considered more robust than traditional techniques as they are less prone to prediction errors, have the ability to handle high-dimensional data, and trace complex interdependencies within the data (Hoang and Wiegratz, 2023). To this end, our results contribute by demonstrating the ability of corporate governance mechanisms to improve the prediction of firm performance, based on ML techniques. Furthermore, our results provide valuable implications, especially in a setting plagued by heightened financial instability (Aly *et al.,* 2018; Shahid and Abbas, 2019; Saha and Kabra, 2020).

As mentioned above, Pakistan is experiencing a dark period with ensuing economic uncertainty coupled with diminishing investor confidence (Rashid *et al.,* 2022). This compromises the stability of the business environment as investment is halted (Shahid and Abbas, 2019; Rashid *et al.,* 2022). In such settings, nonfinancial disclosures, such as NDT and CGIs, become more crucial as investors rely on these disclosures heavily for investment decision-making (Aly *et al.,* 2018; Shahid and Abbas, 2019; Saha and Kabra, 2022). Therefore, it is essential to establish the reliability of such disclosures to restore investor confidence and the stability of the business environment, especially in Pakistan.

Accordingly, our results provide valuable insights by utilizing ML algorithms to prove that both NDT and CGIs can significantly enhance the prediction of firm performance in Pakistan. In doing so, they provide practical implications for investors in Pakistan by suggesting that they can safely rely on such disclosures for investment decision-making. Consequently, investors in Pakistan should pay more attention to nonfinancial disclosures in annual reports to make better investment decisions. Accordingly, this study provides implications for the restoration of investor confidence and the eventual stability of the business environment. In addition, the results also provide implications for regulators and policymakers to promote and strengthen such disclosures by implementing stringent regulations and policies in this context.

6. Conclusion

The present study utilises two widely popular machine learning (ML) algorithms, namely, random forest (RF) and stochastic gradient boosting (SGB), to test whether nonfinancial disclosures such as corporate governance indicators (CGIs) and narrative disclosure tone (NDT) improve the prediction of firm performance. In addition to nonfinancial variables, financial variables are also used as predictors of firm performance. Firm performance is proxied by two accounting-based measures (ROA and ROE) and two market-based measures (Tobin's Q and MTB). The data are collected from the annual reports of 1250 nonfinancial firms in the emerging economy of Pakistan. Different predictive models are created and compared for hypothesis testing. Model 1 contains financial variables only; Model 2 contains financial and NDT variables and Model 3 contains financial and CGIs as predictors. Our results indicate that both NDT and CGIs significantly improve the prediction of firm performance, predominantly market-based firm performance.

This study contributes to the literature by first addressing the neglect of NDT regarding the prediction of firm performance (Beretta *et al.,* 2021; Mousa *et al.,* 2022). Second, the study contributes by amalgamating corporate governance with machine learning literature, which is a rarity (Di Vito and Trottier, 2022). In doing so, we establish the importance of CGIs for predicting firm performance. Third, by using machine learning algorithms, we contribute to the scant ML literature in the realm of accounting and finance, consequently adding to the reliability of these techniques in the context (Rundo *et al.,* 2019; Mousa *et al.,* 2022; Hoang and Wiegratz, 2023). Fourth, the study is carried out in Pakistan's unique emerging economy setting, whereas most studies regarding the prediction of firm performance are concentrated towards developed economies (Iqbal and Riaz, 2021; Mousa *et al.,* 2022).

Our results provide valuable insights and significant implications for investors, managers and policymakers of Pakistani firms. First, the study's results suggest that investors can use NDT and CGIs in annual reports as vital information to gauge where the firm is headed. Therefore, this study outlines the imperativeness of these disclosures for investors in making better investment decisions. Accordingly, this study has implications for the restoration of investor confidence and a stable business environment. Furthermore, the study encourages regulators and policymakers to focus on strengthening the disclosure of nonfinancial information in annual reports. This is especially important for Pakistan and other emerging economies with heightened economic uncertainty where investors rely on such nonfinancial information.

In addition, the results provide some implications for research, as they add to the reliability of ML algorithms as predictive tools of firm performance. Therefore, researchers are encouraged to use these algorithms and the study's framework for the prediction of other financial outcomes such as bankruptcies, insolvencies, and crises. Furthermore, the study's results strongly validate incremental information, signalling and agency theory perspectives regarding NDT and firm performance. In addition, it also confirms agency and stakeholder theory's stance regarding corporate governance and firm performance. Moreover, the study validates these theories in the unique setting of an emerging economy, Pakistan. Despite having substantial implications for both research and practice, the study is not without its limitations.

First, the study is limited to only one emerging economy due to a lack of available data. Future studies could incorporate more emerging economies into their analysis. Second, the study is restricted to narrative disclosures found in annual reports alone, whereas they are not the only mediums through which firms disclose textual information. Future studies could use other sources of content, such as earnings press releases, for the operationalisation of disclosure tone. Finally, the study is limited to board and audit committee characteristics and some ownership structure variables. Future studies could employ other CGIs, such as those relevant to the risk committee, in their analysis.

REFERENCES

Adedeji, B.S., Ong, T.S., Uzir, M.U.H. and Abdul Hamid, A.B. (2020), "Corporate governance and performance of medium-sized firms in Nigeria: does sustainability initiative matter?", Corporate Governance, Vol. 20 No. 3, pp. 401-427. https://doi.org/10.1108/CG-09-2019-0291

Alodat, A. Y., Salleh, Z., Hashim, H. A. and Sulong, F. (2022), "Corporate governance and firm performance: empirical evidence from Jordan", *Journal of Financial Reporting and Accounting*, Vol 20 No. 5, pp. 866–896, https://doi.org/10.1108/JFRA-12-2020-0361

Aly, D., El-Halaby, S. and Hussainey, K. (2018), "Tone disclosure and financial performance: evidence from Egypt", *Accounting Research Journal*, Vol. 31 No. 1, pp. 63–74. https://doi.org/10.1108/ARJ-09-2016-0123

Arora, A. and Sharma, C. (2016), "Corporate governance and firm performance in developing countries: evidence from India", Corporate Governance, Vol. 16 No. 2, pp. 420-436.

Ayuso, S., Rodríguez, M. A., García-Castro, R. and Ariño, M. A. (2014), " Stakeholders' Interests: An Empirical Analysis of the Stakeholder Approach to Corporate Governance", *Business and Society*, Vol. 53 No. 3, pp. 414–439. https://doi.org/10.1177/0007650311433122

Beretta, V., Demartini, M. C., Lico, L. and Trucco, S. (2021), "A tone analysis of the disclosure in the automotive industry", *Sustainability*, Vol. 13 No. 4, pp. 1–16, https://doi.org/10.3390/su13042132

Chang, C. S., Yu, S. W. and Hung, C. H. (2015), "Firm risk and performance: the role of corporate governance", *Review of Managerial Science*, Vol. 9 No. 1, pp. 141–173, https://doi.org/10.1007/s11846-014-0132-x

Dalwai, T., Chinnasamy, G. and Mohammadi, S. S. (2021), "Annual report readability, agency costs, firm performance: an investigation of Oman's financial sector", *Journal of Accounting in Emerging Economies*, Vol. 11 No. 2, pp. 247–277.

Davis, A. K., Piger, J. M. and Sedor, L. M. (2012), "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language", *Contemporary Accounting Research*, Vol. 29 No. 3, pp. 845–868. https://doi.org/10.1111/j.1911-3846.2011.01130.x

Delen, D., Kuzey, C. and Uyar, A. (2013), "Measuring firm performance using financial ratios: A decision tree approach", *Expert Systems with Applications*, Vol. 40 No. 10, pp. 3970–3983, https://doi.org/https://doi.org/10.1016/j.eswa.2013.01.012

Di Vito, J. and Trottier, K., (2022), "A literature review on corporate governance mechanisms: past, present, and future", *Accounting Perspectives*, Vol. 21 No. 2, pp. 207-235.

El-Deeb, M. S., Halim, Y. T. and Elbayoumi, A. F. (2022), "Disclosure tone, corporate governance and firm value: evidence from Egypt", *Asia-Pacific Journal of Accounting and Economics*, Vol. 29 No. 3, pp. 793–814, https://doi.org/10.1080/16081625.2021.1934708

Elvin, P. and Bt Abdul Hamid, N. I. N. (2016), "Ownership structure, corporate governance and firm performance", *International Journal of Economics and Financial Issues*, Vol. 6 No. 3, pp. 99–108.

Fay, C. (2018), "Text mining with R: a tidy approach", *Journal of Statistical Software*, Vol. 83, pp. 1-3, doi: 10.18637/jss.v083.b01.

Goel, S. and Uzuner, O. (2016), "Do Sentiments Matter in Fraud Detection? Estimating

Semantic Orientation of Annual Reports", *Intelligent Systems in Accounting, Finance and Management*, Vol. 23 No. 3, pp. 215–239, https://doi.org/10.1002/isaf.1392

Halteh, K., Kumar, K. and Gepp, A. (2018), "Financial distress prediction of Islamic banks using tree-based stochastic techniques", *Managerial Finance*, Vol. 44 No. 6, pp. 759–773, https://doi.org/10.1108/MF-12-2016-0372

Harakeh, M., Leventis, S., El Masri, T. and Tsileponis, N. (2023), "The moderating role of board gender diversity on the relationship between firm opacity and stock returns", *The British Accounting Review*, Vol. 55 No. 4, pp. 101145, https://doi.org/https://doi.org/10.1016/j.bar.2022.101145

Hoang, D. and Wiegratz, K. (2023), "Machine learning methods in finance: Recent applications and prospects", *European Financial Management*, Vol. 29 No. 5, pp. 1657–1701, https://doi.org/10.1111/eufm.12408

Ibhagui, O. W. and Olokoyo, F. O. (2018), "Leverage and firm performance: New evidence on the role of firm size", *The North American Journal of Economics and Finance*, Vol. 45, pp. 57–82, https://doi.org/10.1016/J.NAJEF.2018.02.002

Iqbal, J. and Riaz, K. (2021), "Predicting future financial performance of banks from management's tone in the textual disclosures", *Quality and Quantity*, Vol. 56, pp. 2691-2722, https://doi.org/10.1007/s11135-021-01216-5

Jensen, M. C. and Meckling, W. H. (1976), "Theory of the firm: Managerial behavior, agency costs and ownership structure", *Journal of Financial Economics*, Vol. 3 No. 4, pp. 305–360.

Kursa, M. B., & Rudnicki, W. R. (2010), "Feature selection with the Boruta package.", *Journal* *of Statistical Software*, Vol. 36 No. 11, pp. 1-13.

Lambey, R., Tewal, B., Sondakh, J. J. and Manganta, M. (2021), "The effect of profitability, firm size, equity ownership and firm age on firm value (leverage basis): evidence from the Indonesian manufacturer companies", *Archives of Business Review,* Vol. 9 No. 1, pp. 128-129.

Loughran, T. and McDonald, B. (2011), "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *The Journal of Finance*, Vol. 36 No. 1, pp. 35–65.

Moghaddam, A. H., Moghaddam, M. H. and Esfandyari, M. (2016), "Stock market index prediction using artificial neural network", *Journal of Economics, Finance and Administrative Science*, Vol. 21 No. 41, pp. 89–93, https://doi.org/10.1016/j.jefas.2016.07.002

Mousa, G. A., Elamir, E. A. H. and Hussainey, K. (2022), "Using machine learning methods to predict financial performance: Does disclosure tone matter?", *International Journal of Disclosure and Governance*, Vol. 19 No. 1, pp. 93–112.

Petropoulos, A. Siakoulis, V., Stavroulakis, E., and Vlachogiannakis, N. E. (2020), "Predicting bank insolvencies using machine learning techniques", *International Journal of Forecasting*, Vol. 36 No. 3, pp. 1092–1113.

Puni, A. and Anlesinya, A. (2020), "Corporate governance mechanisms and firm performance in a developing country", *International Journal of Law and Management*, Vol. 62 No. 2, pp. 147–169, https://doi.org/10.1108/IJLMA-03-2019-0076

Ramadan, M. M. and Hassan, M. K. (2022), "Board gender diversity, governance and Egyptian listed firms’ performance", *Journal of Accounting in Emerging Economies*, Vol. 12 No.

2, pp. 279–299.

Rashid, A., Nasimi, A. N. and Nasimi, R. N. (2022), "The uncertainty–investment relationship: scrutinizing the role of firm size", *International Journal of Emerging Markets*, Vol. 17 No. 10, pp. 2605–2635, https://doi.org/10.1108/IJOEM-09-2019-0698

Rundo, F., Trenta, F., di Stallo, A. L. and Battiato, S. (2019), "Machine learning for quantitative finance applications: A survey", *Applied Sciences*, Vol. 9 No. 24, pp. 5574.

Sadorsky, P. (2021), "A random forests approach to predicting clean energy stock prices", *Journal of Risk and Financial Management*, Vol 14 No. 2, pp. 48.

Saha, R. (2024), "Corporate governance, voluntary disclosure and firm valuation relationship: evidence from top listed Indian firms", Journal of Accounting in Emerging Economies, Vol. 14 No. 1, pp. 187-219, https://doi.org/10.1108/JAEE-09-2021-0288

Saha, R. and Kabra, K. C. (2020), "Corporate Governance and Voluntary Disclosure: A Synthesis of Empirical Studies" *Business Perspectives and Research*, Vol. 8 No. 2, pp. 117–138, https://doi.org/10.1177/2278533719886998

Saha, R. and Kabra, K. C. (2022), "Corporate governance and voluntary disclosure: evidence from India", *Journal of Financial Reporting and Accounting*, Vol. 20 No. 1, pp. 127–160, https://doi.org/10.1108/JFRA-03-2020-0079

Shahid, M. S. and Abbas, M. (2019), "Does corporate governance play any role in investor confidence, corporate investment decisions relationship? Evidence from Pakistan and India", *Journal of Economics and Business*, Vol. 105, pp. 105839.

Tran, L. T. H., Tu, T. T. K., Nguyen, T. T. H., Nguyen, H. T. L. and Vo, X. V. (2023), "Annual report narrative disclosures, information asymmetry and future firm performance:

evidence from Vietnam", *International Journal of Emerging Markets*, Vol. 18 No. 2, pp. 351–375, https://doi.org/10.1108/IJOEM-08-2020-0925

Van Binsbergen, J. H., Han, X. and Lopez-Lira, A. (2023), "Man versus Machine Learning: The Term Structure of Earnings Expectations and Conditional Biases", *Review of Financial Studies*, Vol. 36 No. 6, pp. 2361–2396, https://doi.org/10.1093/rfs/hhac085

Xiaomao, X., Xudong, Z. and Yuanfang, W. (2019), "A comparison of feature selection methodology for solving classification problems in finance", *Journal of Physics: Conference Series*, Vol. 1284 No. 1, pp. 12026.

Yang, M., Bento, P. and Akbar, A. (2019), "Does CSR influence firm performance indicators? Evidence from Chinese pharmaceutical enterprises", *Sustainability*, Vol. 11 No. 20, pp. 1–18, https://doi.org/10.3390/su11205656

Yeh, J.-Y. and Chen, C.-H. (2022), "A machine learning approach to predict the success of crowdfunding fintech project", *Journal of Enterprise Information Management*, Vol. 35 No. 6, pp. 1678-1696, https://doi.org/10.1108/JEIM-01-2019-0017.

Table 1: Sampling Process

Particulars	Number of Companies
Panel A: Sampling process	
Total PSX population	551
Less: Financial, investment and banking companies	(129)
Less: Relevant data missing or incomplete	(272)
Less: Firms with annual reports not machine readable or not convertible	(25)
Final sample	125
Total number of firm-year observations (125*10)	1250
Panel B: Sample by sector	
Oil, gas, mining and refineries	14
Technology and communication	6
Power generation, distribution, cable and electric goods	11
Chemical and fertilizers	11
Construction, engineering and property	14
Food, sugar and personal care	15
Textile composite, spinning and weaving	20
Pharmaceuticals	5
Automobile parts, assemblers and transportation	9
Glass, ceramics, paper and board	5
Miscellaneous	15
Total	**125**

Sampling process and sector wise breakdown of final sample

Table 2: Variables, Operationalization and Source

Symbol	Definition	Operationalisation	Source
Panel A: Target variables			
ROA	Return on Assets	Net Income/Total Assets	Annual Reports
ROE	Return on Equity	Net Income/Total Equity	Annual Reports
Tobin´s Q	Tobin´s Q	Market Value of Total Assets/Total Assets Replacement Cost	Annual Reports
MTB	Market to Book Value	Market Value of Equity/Book Value of Equity	Annual Reports
Panel B: Financial predictor variables			
SIZE	Firm Size	Natural log of total assets	Annual Reports
LIQ	Liquidity	Current assets/Current liabilities	Annual Reports
LEV	Leverage	Total liabilities/Total assets	Annual Reports
AGE	Firm Age	The number of years the since the firm was formed	Annual Reports
CFO	Cash flow from operations	Net cash flow generated from operating activities	Annual Reports
BETA	Firm Risk	Covariance of the stock´s returns with the market return/ Market return	Pakistan Stock Exchange
Panel C: Non-financial disclosures			
Narrative disclosure tone			
POS	Positive sentiment	The number of positive words in the annual reports	Annual Reports
NEG	Negative sentiment	The number of negative words in the annual reports	Annual Reports
UNC	Uncertain sentiment	The number of uncertain words in the annual reports	Annual Reports
LIT	Litigious sentiment	The number of litigious words in the annual reports	Annual Reports
SUP	Superfluous sentiment	The number of superfluous words in the annual reports	Annual Reports
CON	Constraining sentiment	The number of constraining words in the annual reports	Annual Reports
Corporate governance mechanisms			
BSIZE	Board Size	The number of directors on the board	Annual Reports
BI	Board Independence	The proportion of independent directors on the board	Annual Reports
BGD	Board Gender Diversity	The proportion of female directors on the board	Annual Reports
BM	Board Meetings	The number of times the board meets in a year	Annual Reports
ACSIZE	Audit Committee Size	The number of directors on the audit committee	Annual Reports
ACI	Audit Committee Independence	The proportion of independent directors on the audit committee	Annual Reports
ACGD	Audit Committee Gender Diversity	The proportion of female directors on the audit committee	Annual Reports
ACM	Audit Committee Meetings	The number of times the audit committee meets annually	Annual Reports
IOWN	Institutional Ownership	The percentage of shares owned by institutions	Annual Reports
MOWN	Managerial Ownership	The percentage of shares owned by managers	Annual Reports
FOWN	Foreign Ownership	The percentage of shares owned by foreigners	Annual Reports
COWN	Concentrated Ownership	The percentage of shared owned by shareholders having 5% or more shares	Annual Reports

Table 3: Descriptive Statistics

Variable	Mean	Median	SD*	Min*	Max*	1st Quartile	3rd Quartile	VIF*
SIZE	16.46	16.52	1.8	11.52	20.68	15.44	17.47	1.68
LIQ	1.16	0.9	1.17	0.01	14.29	0.51	1.38	1.18
LEV	0.97	0.47	2.93	0.00	25.05	0.04	1.10	1.07
AGE	40.46	37	18.9	4	107	25	56	1.13
CFO	4.03	0.72	1.54	-37.3	37.9	0.02	3.2	1.30
BETA	0.88	0.94	1.11	-25.38	8.79	0.45	1.35	1.04
NEG	426.9	337	288.7	28	2093	211	570.8	12.5
POS	290.8	225.5	214.7	29	1704	132	398.2	4.01
UNC	253	205	154.4	16	893	133.2	336	12.98
LIT	248.4	193	169.9	24	1267	128	324.8	5.9
SUP	5.875	4	8.18	0	89	1	7	1.64
CON	183.3	148	112.4	9	759	101.2	237	12.2
BSIZE	8.38	8	8.38	5	17	7	9	1.41
BI	0.18	0.14	0.13	0	0.86	0.11	0.28	1.62
BGD	0.08	0	0.12	0	0.71	0	0.14	2.22
BM	5.56	5	2.34	2	22	4	6	1.3
ACSIZE	3.6	3	0.86	3	9	3	4	1.5
ACI	0.31	0.33	0.21	0	1.33	0.25	0.33	2.5
ACM	4.38	4	0.85	4	10	4	5	1.16
ACGD	0.08	0	0.15	0	0.75	0	0	2.02
IOWN	0.59	0.69	0.30	0.0	0.99	0.31	0.84	7
MOWN	0.18	0.06	0.24	0.0	0.89	0.0	0.29	5.27
FOWN	0.17	0.02	0.27	0.0	0.98	0.0	0.23	1.46
COWN	0.64	0.68	0.20	0.0	0.98	0.50	0.78	2.13

*SD: Standard Deviation. Min: Minimum. Max: Maximum. VIF: Variance Inflation Factor.

Table 4: Overall metrics for the prediction of ROA

Random Forest	Stochastic Gradient Boosting

	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Accuracy	0.68	0.65	0.72	0.65	0.66	0.70
95% CI*	(0.59, 0.76)	(0.56, 0.73)	(0.63, 0.80)	(0.56, 0.73)	(0.57, 0.74)	(0.62, 0.78)
NIR*	0.632	0.632	0.632	0.632	0.632	0.632
p-value	0.15	0.39	0.02	0.39	0.32	0.06
Kappa coefficient	0.37	0.27	0.43	0.32	0.32	0.42
McNemar's p-value	0.01	NA	NA	0.00	0.09	NA

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of ROA using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

Table 5: Class specific metrics for the prediction of ROA

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.4	0.8	0.63	0.23	0.81	0.63	0.5	0.85	0.5
Specificity	0.96	0.5	0.88	0.94	0.37	0.92	0.94	0.5	0.93
PPV*	0.75	0.73	0.43	0.54	0.69	0.53	0.71	0.74	0.5
NPV*	0.84	0.59	0.94	0.79	0.53	0.94	0.86	0.62	0.93
Balanced Accuracy	0.7	0.67	0.76	0.59	0.59	0.77	0.72	0.66	0.72
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.33	0.78	0.69	0.37	0.78	0.56	0.53	0.80	0.56
Specificity	0.94	0.5	0.87	0.93	0.46	0.90	0.95	0.50	0.94
PPV*	0.67	0.73	0.44	0.61	0.71	0.45	0.75	0.74	0.53
NPV*	0.82	0.58	0.95	0.82	0.55	0.93	0.86	0.66	0.93
Balanced Accuracy	0.68	0.65	0.75	0.65	0.62	0.73	0.72	0.67	0.72

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of ROA using RF and SGB algorithms.

Table 6: Overall metrics for the prediction of ROE

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.62	0.65	0.67	0.61	0.62	0.64
95% CI*	(0.53, 0.71)	(0.56, 0.73)	(0.58, 0.75)	(0.52, 0.69)	(0.52, 0.70)	(0.55, 0.72)
NIR*	0.568	0.568	568	0.568	0.568	0.568
p-value	0.12	0.04	0.01	0.21	0.16	0.06
Kappa coefficient	0.33	0.38	0.43	0.31	0.32	0.40
McNemar's p-value	0.61	0.69	0.26	0.61	0.71	0.27

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of ROE using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

Table 7: Class specific metrics for the prediction of ROE

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.42	0.76	0.46	0.5	0.78	0.46	0.50	0.75	0.64
Specificity	0.89	0.54	0.89	0.89	0.57	0.90	0.88	0.63	0.91
PPV*	0.5	0.68	0.54	0.54	0.71	0.57	0.52	0.73	0.67
NPV*	0.85	0.63	0.85	0.87	0.66	0.85	0.87	0.65	0.90
Balanced Accuracy	0.66	0.65	0.68	0.69	0.68	0.68	0.69	0.69	0.78
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.35	0.77	0.39	0.42	0.75	0.46	0.62	0.68	0.57
Specificity	0.91	0.5	0.86	0.87	0.54	0.9	0.82	0.70	0.89
PPV*	0.5	0.67	0.44	0.46	0.68	0.57	0.47	0.75	0.59
NPV*	0.84	0.63	0.83	0.85	0.62	0.85	0.89	0.62	0.88
Balanced Accuracy	0.63	0.64	0.62	0.65	0.64	0.68	0.72	0.69	0.73

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of ROE using RF and SGB algorithms.

Table 8: Overall metrics for the prediction of TOBIN´S Q

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.57	0.62	0.66	0.57	0.60	0.70
95% CI*	(0.48, 0.66)	(0.52, 0.70)	(0.57, 0.74)	(0.48, 0.66)	(0.51, 0.69)	(0.61, 0.78)
NIR*	0.48	0.48	0.48	0.48	0.48	0.48
p-value	0.03	0.00	0.00	0.03	0.00	0.00
Kappa coefficient	0.28	0.35	0.42	0.27	0.33	0.50
McNemar's p-value	0.00	0.00	0.00	0.00	0.01	0.03

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of Tobin´s Q using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

Table 9: Class specific metrics for the prediction of TOBIN´S Q

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.35	0.78	0.4	0.40	0.85	0.40	0.43	0.85	0.56
Specificity	0.94	0.45	0.87	0.94	0.46	0.92	0.95	0.49	0.94
PPV*	0.74	0.57	0.43	0.76	0.59	0.56	0.81	0.61	0.71
NPV*	0.75	0.69	0.85	0.77	0.77	0.86	0.78	0.78	0.9
Balanced Accuracy	0.65	0.61	0.64	0.66	0.66	0.7	0.69	0.67	0.75
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.25	0.82	0.48	0.43	0.80	0.4	0.60	0.82	0.56
Specificity	0.95	0.43	0.87	0.89	0.49	0.92	0.93	0.65	0.91
PPV*	0.71	0.57	0.48	0.69	0.59	0.56	0.80	0.68	0.61
NPV*	0.73	0.72	0.87	0.77	0.73	0.86	0.83	0.79	0.89
Balanced Accuracy	0.6	0.62	0.68	0.66	0.65	0.66	0.76	0.73	0.74

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of Tobin´s Q using RF and SGB algorithms.

Table 10: Overall metrics for the prediction of MTB

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.68	0.70	0.75	0.62	0.69	0.76
95% CI*	(0.59, 0.76)	(0.61, 0.78)	(0.67, 0.82)	(0.53, 0.71)	(0.60, 0.77)	(0.68, 0.83)
NIR*	0.52	0.52	0.52	0.52	0.52	0.52
p-value	0.00	0.00	0.00	0.01	0.00	0.00
Kappa coefficient	0.43	0.45	0.57	0.34	0.47	0.60
McNemar's p-value	0.00	0.00	0.00	0.04	0.08	0.09

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of MTB using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

Table 11: Class specific metrics for the prediction of MTB

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.44	0.91	0.42	0.47	0.94	0.38	0.62	0.91	0.54
Specificity	0.92	0.5	0.97	0.95	0.50	0.97	0.96	0.60	0.97
PPV*	0.68	0.66	0.79	0.76	0.67	0.77	0.84	0.71	0.82
NPV*	0.82	0.83	0.86	0.83	0.88	0.86	0.87	0.86	0.89
Balanced Accuracy	0.68	0.7	0.7	0.71	0.72	0.68	0.79	0.75	0.75
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.35	0.93	0.46	0.59	0.83	0.46	0.68	0.85	0.65
Specificity	0.89	0.52	0.92	0.87	0.62	0.96	0.91	0.70	0.96
PPV*	0.55	0.65	0.6	0.63	0.70	0.75	0.74	0.75	0.81
NPV*	0.79	0.74	0.87	0.85	0.77	0.87	0.88	0.81	0.91
Balanced Accuracy	0.62	0.67	0.69	0.73	0.72	0.72	0.79	0.77	0.81

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of MTB using RF and SGB algorithms.

Figure 1 – Feature Selection using Boruta algorithm – ROA and ROE

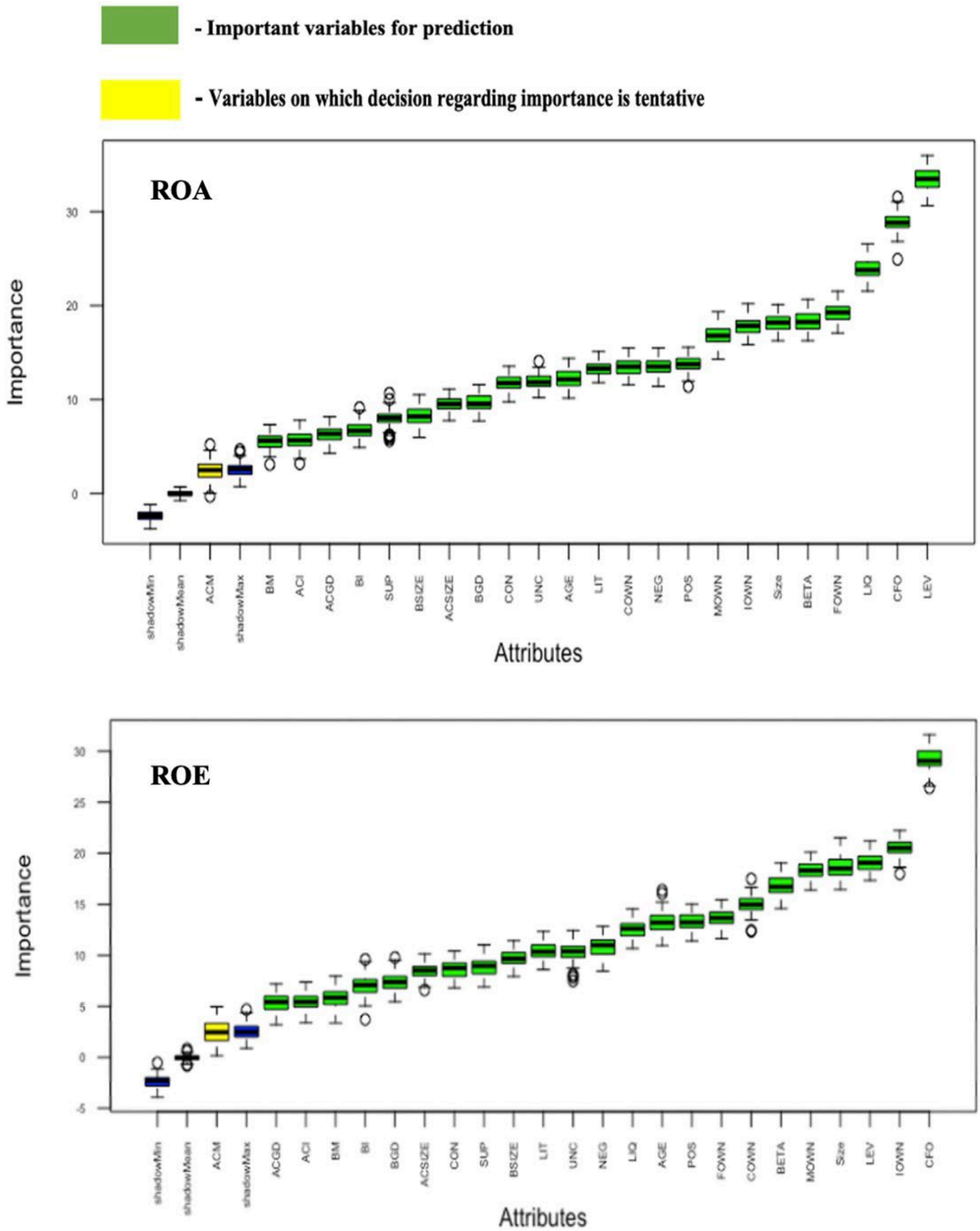

Figure 2 – Feature Selection using Boruta algorithm – TOBIN´S Q and MTB

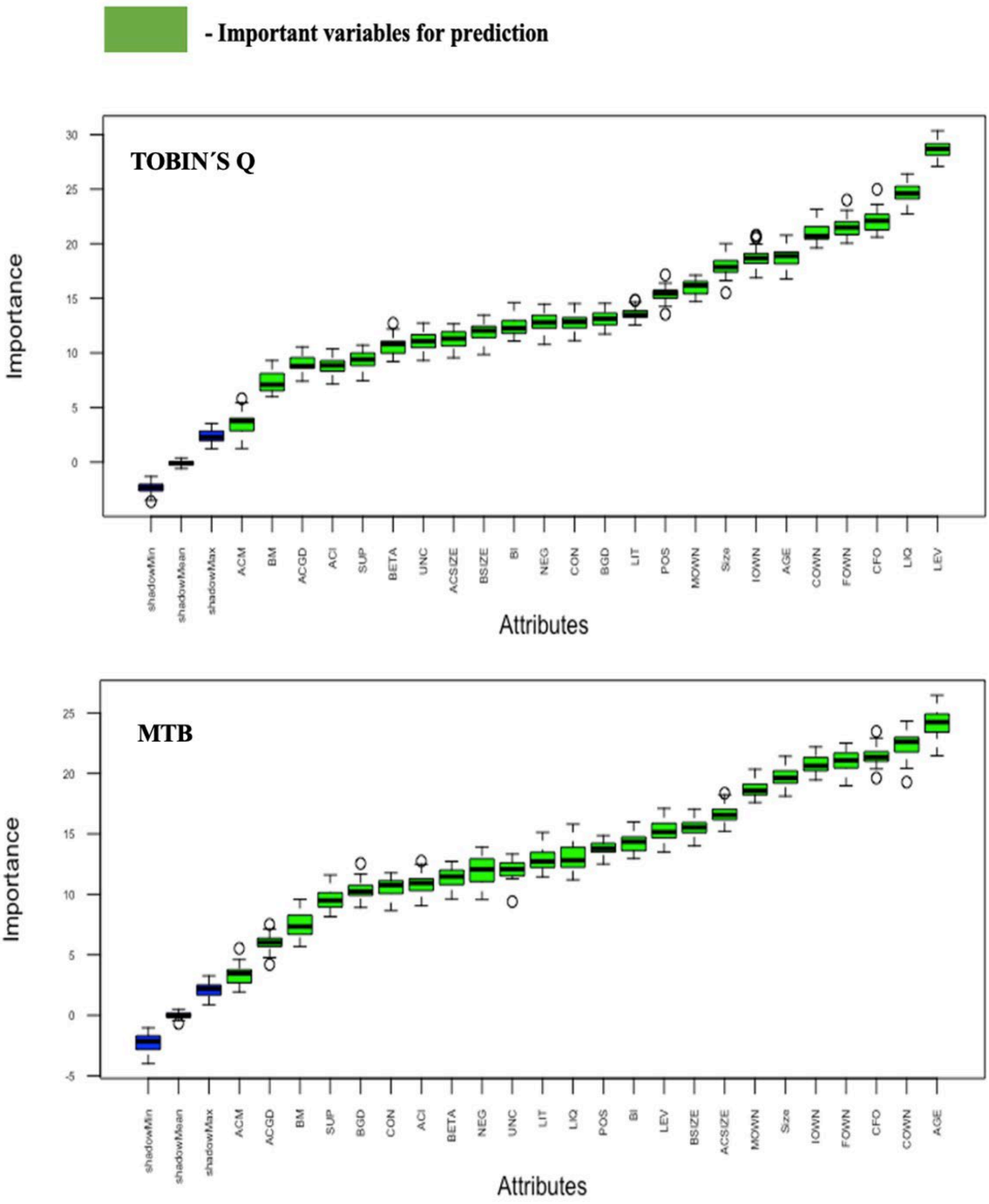

Figure 3: Variable Importance – ROA

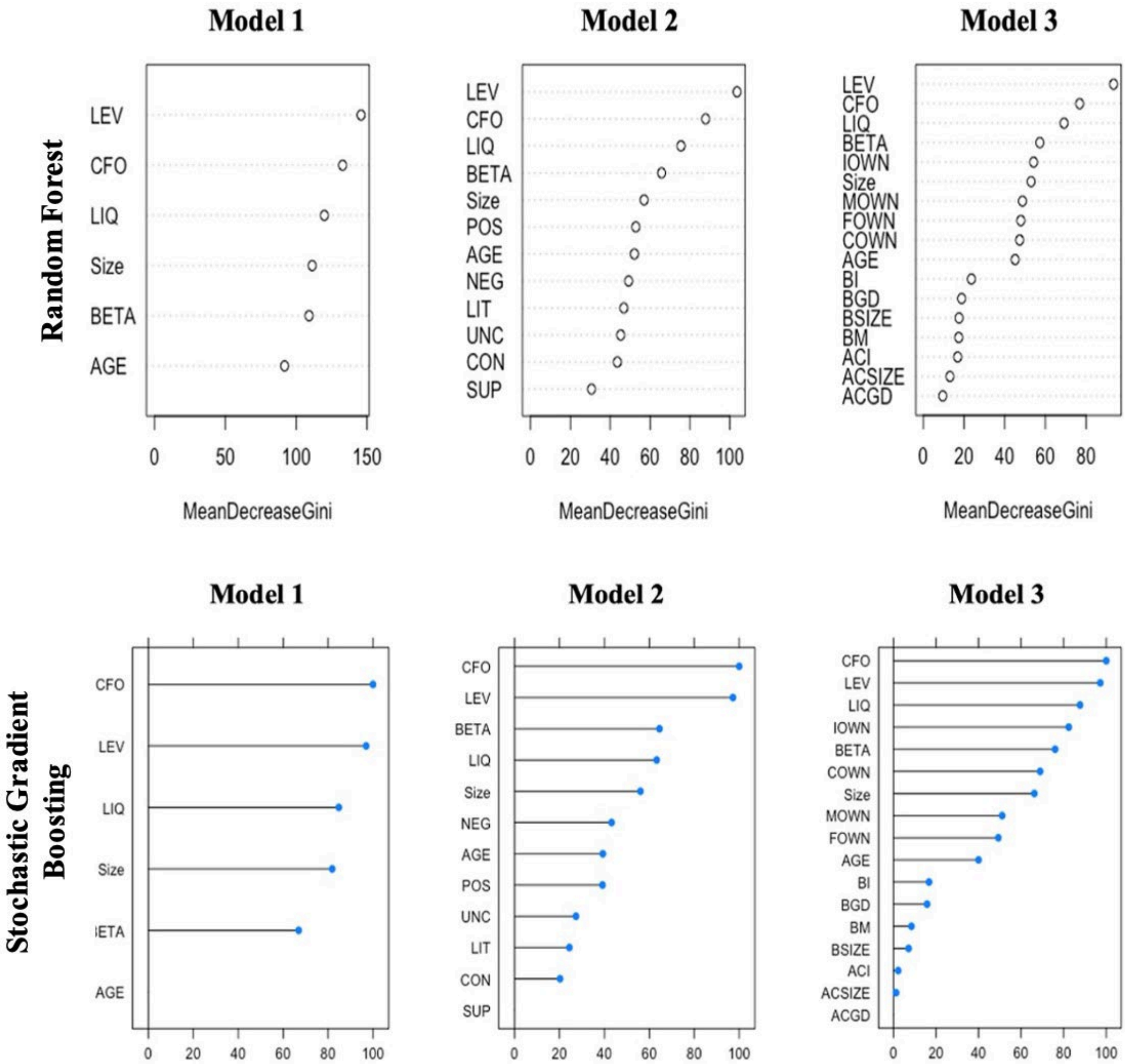

Figure 4: Variable Importance – ROE

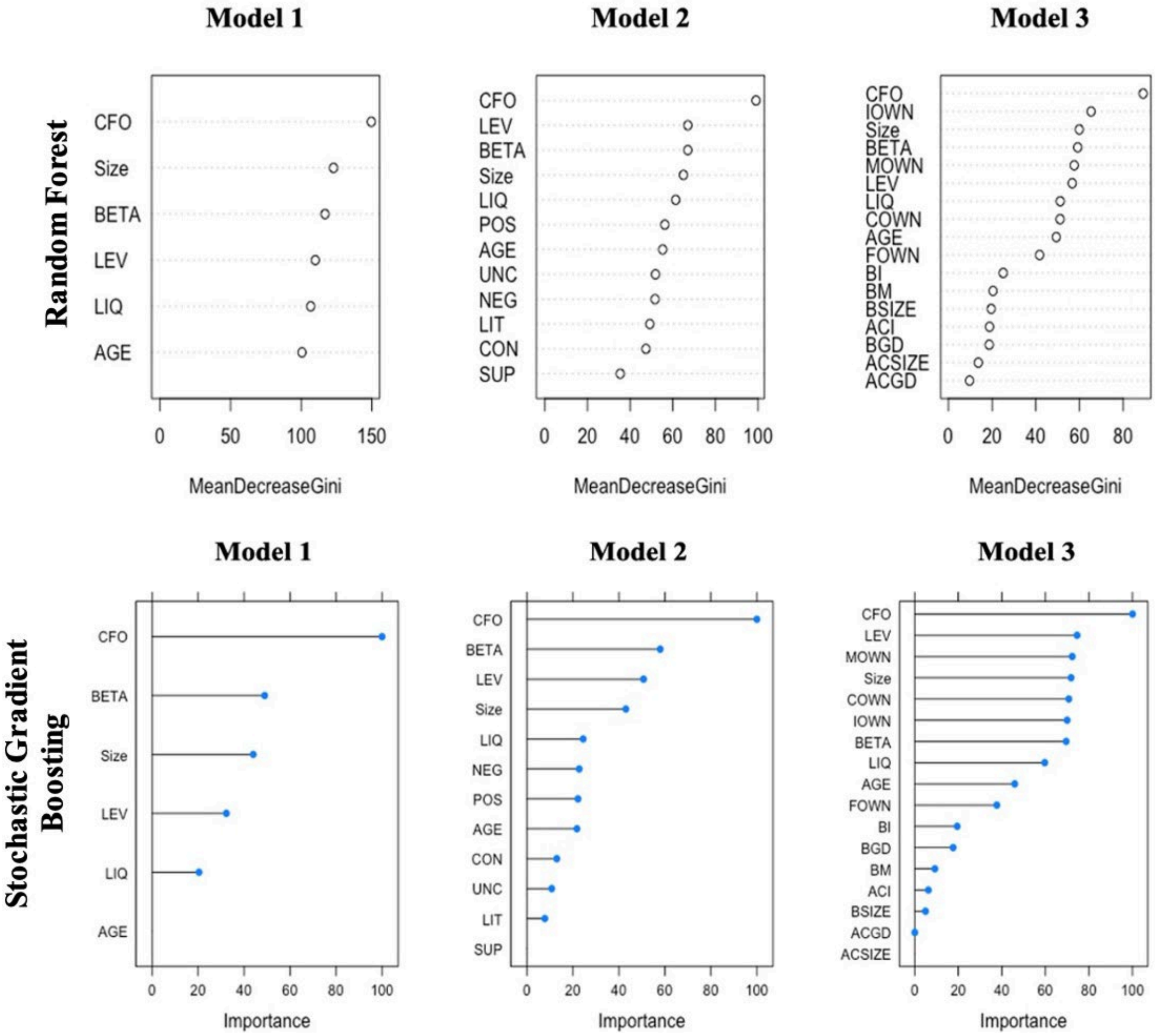

Figure 5: Variable Importance – TOBIN´S Q

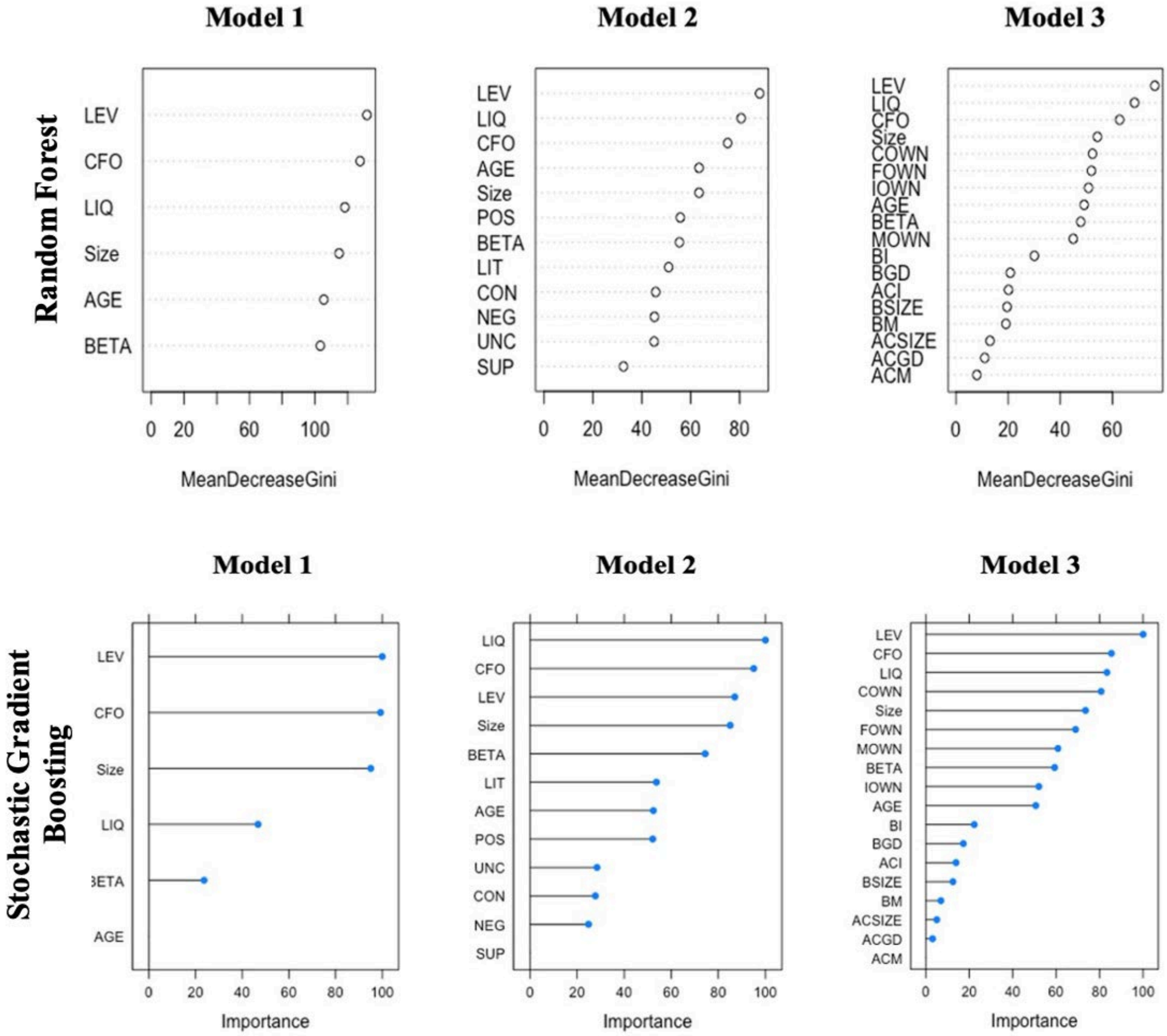

Figure 6: Variable Importance – MTB