Formosa Journal of Multidisciplinary Research (FJMR)
Vol. 4 No. 12, 2025: 5953-5970

Hybrid Analysis of Financial Distress in Professional Football Clubs Using Network Analysis and XGBoost: Evidence from the English Premier League

Deasy Aisyah Safitri Musyawaman
Universitas Negeri Surabaya, Indonesia
Corresponding Author: Deasy Aisyah Safitri Musyawaman
deasy.21094@mhs.unesa.ac.id

A R T I C L E I N F O
Keywords: Financial Distress, Network Analysis, XGBoost, Professional Football Clubs, English Premier League

Received : 08, November
Revised : 29, November
Accepted: 28, December

A B S T R A C T
This study investigates financial distress among English Premier League clubs using a hybrid approach combining financial indicators, non-financial performance signals, network-based features, and XGBoost. Using 224 observations from 32 clubs during the 2016/17–2022/23 seasons, the study constructs correlation-based financial networks to extract centrality and community features. Comparative prediction modeling shows that the hybrid model outperforms the baseline model, with significant differences in financial and network signals between distressed and healthy clubs. Results support signaling and agency theories, suggesting that both internal financial indicators and structural network positions reflect a club's financial vulnerability.

DOI: https://doi.org/10.55927/fjmr.v4i12.647
ISSN-E: 2829-8896
https://npaformosapublisher.org/index.php/fjmr

INTRODUCTION

Financial distress has become an increasingly visible issue in professional football, particularly in the English Premier League (EPL). Despite high revenue generation, many clubs face structural financial vulnerabilities driven by inflated spending, rising wage obligations, and persistent operating losses (Alaminos & Fernández, 2019; BBC, 2024). These conditions reflect deeper governance and incentive problems within modern football, where competitive pressure and financial inequality continue to widen risk exposure across clubs (Barajas & Rodriguez, 2010; Deloitte, 2024). Prior studies also demonstrate that distress prediction in football clubs remains challenging, as traditional models often yield varying levels of accuracy across contexts (Lutfiyyah & Bhilawa, 2021), emphasizing the need for more robust analytical approaches.

Traditional financial ratios have long been used to assess distress, yet growing evidence suggests they are insufficient on their own. Football clubs operate in an interconnected ecosystem shaped by transfer activities, financial exchanges, and competitive interactions, making network structures an important source of additional information (Bardoscia et al., 2021). Network indicators such as centrality and community structure offer insights into systemic influence and potential risk propagation, complementing financial metrics and aligning with signaling theory's view that observable patterns convey underlying conditions (Choudhury, 2024).

Alongside these theoretical advances, machine learning—particularly XGBoost—has proven effective for modeling complex financial distress patterns due to its scalability and strong predictive performance (Chen & Guestrin, 2016; Jabeur et al., 2023). Integrating network-derived variables into XGBoost models aligns with recent efforts to enhance risk forecasting through hybrid analytical approaches (Kadkhoda & Amiri, 2024).

This study contributes by combining financial ratios with network-based metrics to analyze distress among EPL clubs, offering empirical insight into how structural network positions may strengthen prediction beyond traditional indicators (Alaminos & Esteban, 2020). Methodologically, it introduces a hybrid XGBoost model that incorporates both financial and network signals, addressing existing gaps in football-focused machine learning research. The study aims to map the financial network structure among EPL clubs, develop an XGBoost-based distress prediction model, and evaluate whether integrating network features improves predictive performance.

LITERATURE REVIEW

Signaling Theory

Signaling Theory explains that organizations use observable indicators to reduce information asymmetry (Spence, 1973). Financial ratios such as profitability, liquidity, and leverage act as signals of stability or potential distress (Choudhury, 2024). In football, these financial signals—together with performance-related indicators—reflect the underlying health of clubs. Empirical evidence shows that financial ratios strongly predict distress among European clubs (Alaminos & Fernández, 2019).

Agency Theory

Agency Theory highlights conflicts between owners and managers arising from divergent interests (Jensen & Meckling, 1976). In football, complex ownership structures increase agency problems, especially when managers pursue short-term gains that jeopardize long-term financial stability (Panda & Leepsa, 2017). Inefficient decisions such as excessive player spending and high debt levels elevate distress risk (Scott & O'Brien, 2003).

Financial Distress

Financial distress refers to a condition in which firms struggle to meet short-term obligations due to liquidity pressures (Heryadi, 2023). In football, distress is often triggered by high wage bills, volatile revenues, and performance-based income risks (Barajas & Rodriguez, 2010; Plumley et al., 2021). Low liquidity and high leverage consistently emerge as key predictors of club distress (Alaminos & Fernández, 2019).

Network Analysis

Network Theory views organizations as interconnected entities whose structural positions shape systemic risk exposure (Newman, 2010; Bardoscia et al., 2021). In financial networks, relationships are often modeled using correlation-based structures (IMF, 2014). Centrality metrics—degree, closeness, betweenness—capture nodes' influence and vulnerability (Battiston et al., 2012). Studies show that adding network metrics enhances distress prediction accuracy beyond traditional financial indicators (Kadkhoda & Amiri, 2024).

Machine Learning and XGBoost

Machine learning captures nonlinear patterns relevant to financial distress prediction (Hastie et al., 2009; Mohri et al., 2018). XGBoost is particularly effective due to its boosting mechanism and regularization, improving accuracy while limiting overfitting (Chen & Guestrin, 2016; Zhang et al., 2021). Evidence shows XGBoost outperforms traditional models, and its performance increases further when network features are included (Jabeur et al., 2023; Kristanti et al., 2024; Kadkhoda & Amiri, 2024).

Football Industry Dynamics and Distress Risk

Professional football operates with volatile revenues and high fixed costs, making clubs vulnerable to financial shocks (Deloitte, 2023). In the English Premier League, high income does not eliminate distress risk due to transfer spending, wage inflation, and relegation exposure (GOV.UK, 2023). These dynamics highlight the need for predictive models that integrate financial, structural, and systemic indicators.

Hypothesis Development

Signaling Theory argues that observable financial indicators reduce information asymmetry by revealing the internal health of an organization (Spence, 1973). In professional football, financial ratios such as liquidity, leverage, and profitability function as key signals of stability or distress

(Choudhury, 2024). Empirical evidence shows that these indicators significantly differentiate distressed from healthy clubs, particularly through measures of liquidity pressure and excessive leverage (Alaminos & Fernández, 2019; Plumley et al., 2021). Agency Theory further explains that managerial decisions driven by misaligned incentives—such as excessive player spending or inefficient financial policies—can intensify distress (Jensen & Meckling, 1976; Scott & O'Brien, 2003). Based on this theoretical and empirical foundation, financial signals are expected to significantly influence distress outcomes.

H1: Financial signal variables have a significant effect on predicting financial distress among Premier League clubs.

Network Theory emphasizes that an entity's position within a network shapes its vulnerability to systemic disturbances (Newman, 2010; Bardoscia et al., 2021). Centrality measures such as degree, closeness, and betweenness capture the structural importance of an entity and its embeddedness in broader financial dynamics (Battiston et al., 2012). Prior studies indicate that integrating network features with machine learning models enhances predictive accuracy compared with financial variables alone (Kadkhoda & Amiri, 2024). XGBoost, in particular, performs better when enriched with relational features that capture complex interdependencies (Jabeur et al., 2023; Kristanti et al., 2024).

H2: Adding network-based features to the XGBoost model improves predictive performance compared to a model using only financial signals.

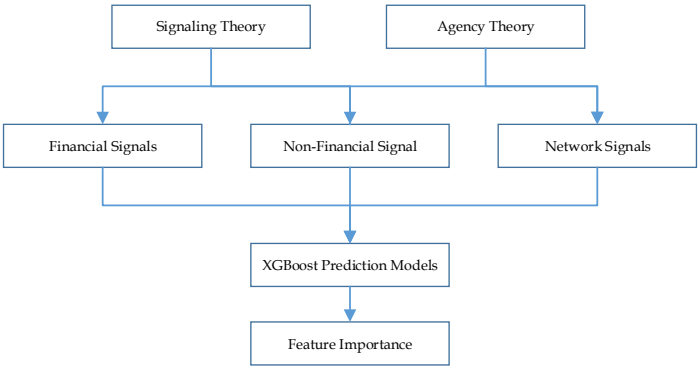

Figure 1. Conceptual Framework

METHODOLOGY

This study adopts a quantitative predictive design integrating financial statement analysis, network-theoretic modelling, and machine-learning classification to examine the determinants of financial distress among English Premier League clubs. The dataset contains 224 club-year observations (2016/17–2022/23), compiled from audited reports, and Transfermarkt. The explanatory variables comprise financial signals (liquidity, leverage, profitability, and expenditure ratios), a non-financial performance signal, and network-based signals derived from Spearman correlation networks. Financial distress is defined as a binary condition reflecting sustained negative earnings

accompanied by worsening short-term obligations. All data undergo systematic cleaning, normalization, and outlier diagnostics prior to analysis.

All analytical procedures are conducted in Python. Network structures are estimated using correlation matrices, with community detection performed via the Louvain algorithm and centrality metrics (degree, closeness, betweenness) computed to capture structural influence. Predictive modelling employs XGBoost with SMOTE resampling and stratified cross-validation, with model quality assessed using accuracy, AUC-ROC, recall, precision, and F1-score. Variable contributions are evaluated using XGBoost's Gain, Cover, and Weight metrics, enabling a concise assessment of the relative predictive power of financial and network signals.

Table 1. Variable

Signal Type	Variable Name	Formula	Explanation
Financial Signal	Debt Ratio (Alaminos & Fernández, 2019)	Total Liabilities / Total Assets	Measures the proportion of liabilities relative to assets. A higher ratio indicates a greater risk of financial distress.
	Current Ratio (Alaminos & Fernández, 2019)	Current Assets / Current Liabilities	Measures a club's ability to meet short-term obligations. Lower liquidity indicates a higher risk of distress.
	Return on Equity (ROE) (Barajaz & Rodríguez, 2010)	Net Income / Total Equity	Measures profitability generated from shareholders' equity. Low ROE may signal inefficient equity use and potential financial problems.
	Player Expenses Ratio (Barajaz & Rodríguez, 2010)	Player Expenses / Total Revenue	Large player expenses relative to revenue may create liquidity pressure and solvency risk.
	Return on Assets (ROA) (Barajaz & Rodríguez, 2010)	Net Income / Total Assets	Measures asset efficiency in generating revenue. Low ROA indicates poor financial performance.
Non-Financial Signal	Club Performance (Alaminos & Fernández, 2019)	Total points earned in one league season	Measures sporting performance. Poor performance may reduce broadcasting, sponsorship, and commercial income.
Network Signal	Betweenness Centrality (Kadhkoda et al., 2024)	Number of shortest paths passing through the club / total shortest paths	Indicates the club's role as an intermediary in the network. Low values suggest a loss of strategic position.
	Degree Centrality (Kadhkoda et al., 2024)	Number of significant correlation edges connected to the club	Indicates how financially connected a club is. Low degree may reflect isolation and potential distress.

	Closeness Centrality (Kadhkoda et al., 2024)	1 / (sum of shortest paths from the club to all others)	Measures financial proximity to other clubs. Low values indicate deviation and increased risk.
	Community Detection (Kadhkoda et al., 2024)	Actual edges within community / expected edges in a random network	Clubs sharing the same community label exhibit similar financial correlation patterns.

RESEARCH RESULT

Descriptive Statistics

Table 2. Descriptive Statistics (Summary)

Variable	Count	Mean	Std	Min	Median	Max
Debt Ratio	224	1.307408	0.817772	0.359909	1.061634	5.628806
Current Ratio	224	0.569833	0.462204	0.017837	0.494533	2.296346
ROE	224	0.290861	2.936628	-14.15630	0.102017	22.27143
Player Expenses Ratio	224	1.040891	0.432991	0.493124	0.923793	3.165353
ROA	224	-0.173980	0.672531	-4.969220	-0.067350	3.004285
Club Performance	224	33.17857	29.43297	0	39	98

Table 2. indicates that EPL clubs' financial variables are highly dispersed and asymmetric. Debt Ratio, Current Ratio, Player Expenses Ratio, and Club Performance show strong right-skewness, reflecting a few clubs with extreme leverage, liquidity, or spending. Conversely, ROE and ROA are left-skewed due to low or negative profitability, consistent with the industry's structural vulnerability to financial distress.

Table 3. Descriptive Statistics by Group (Distress vs Healthy)

Group	% of sample	Mean					
		Debt Ratio	Current Ratio	ROE	Player Exp. Ratio	ROA	Club Perf.
Distress (positive cases)	17.85%	1.622482	0.445114	0.530402	1.063392	-0.323307	32.22500
Healthy (negative cases)	82.15%	1.238913	0.596946	0.238787	1.035999	-0.141521	33.38587

Table 3. shows that distressed clubs (17.85% of the sample) exhibit higher leverage and lower liquidity than healthy clubs, alongside more volatile and often negative profitability (ROE, ROA)—a pattern consistent with evidence of financial fragility in professional football (Alaminos & Fernández, 2019; Deloitte, 2024). Healthy clubs (82.15%) display stronger liquidity, better overall

profitability, and slightly higher on-field performance, highlighting clear financial and operational gaps between the groups.

Network Analysis

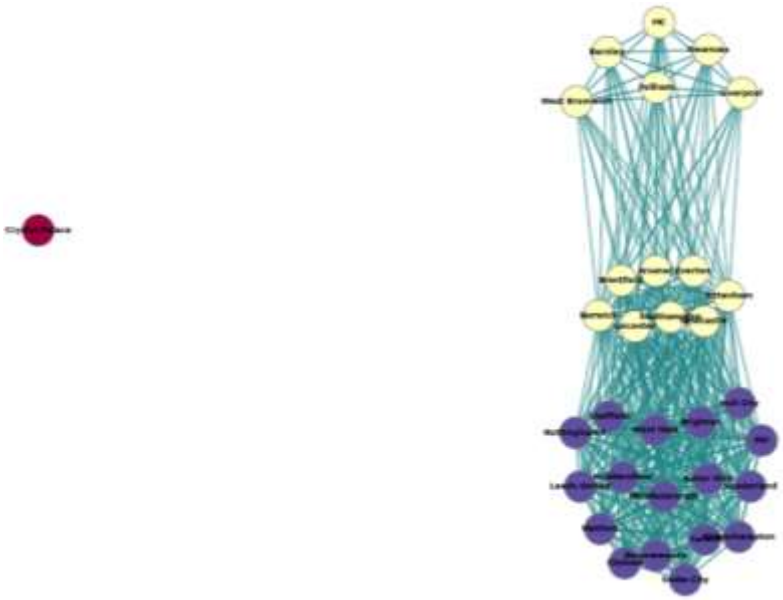

Figure 2. Financial Correlation-Based Network

Figure 2. shows that EPL clubs form financially correlated networks, with dense connections and clear community structures indicating shared financial patterns and interdependence (Bardoscia et al., 2021; Battiston et al., 2012). Several clubs—such as Arsenal, Brentford, Everton, Leicester City, Newcastle United, Norwich City, Southampton, and Tottenham Hotspur—visually exhibit higher degree, betweenness, and closeness centrality, while Crystal Palace appears more peripheral. These structural patterns highlight the systemic nature of financial distress and support the use of network-based features to improve distress prediction beyond traditional financial ratios (Alaminos & Fernández, 2019; Zhang & Luo, 2017; Song et al., 2021; Qiu et al., 2022; Alessi & Savona, 2021).

Table 4. Community Composition And Distress Incidence

Community ID	Distress	Healthy	Total Clubs	% Distress
0	0	1	1	0.00
1	6	8	14	42.86
2	12	5	17	70.59

Table 4. show that community structure aligns with financial distress patterns: Community 1 records a moderate distress rate (42.86%), while Community 2 exhibits the highest rate (70.6%), reflecting stronger financial interdependence and group-level vulnerability (Bardoscia et al., 2021; Song et al., 2021). Community 0—containing only Crystal Palace—shows no distress, indicating an isolated financial profile. Communities also differ in network roles, with high–degree centrality clubs such as Arsenal, Newcastle, and

Tottenham acting as connectivity hubs that may accelerate the transmission of financial risk across the network.

Signals Comparison – Distress vs Healthy Tests

Table 5. Shapiro–Wilk Normality Results (Distress vs Healthy)

Variable	Normality p (Distress)	Normality p (Healthy)	Distribution: Distress	Distribution: Healthy
Debt Ratio	0.0311	0.7117	Not Normal	Normal
Current Ratio	0.0060	0.0077	Not Normal	Not Normal
ROE	0.0000	0.0000	Not Normal	Not Normal
Player Expenses Ratio	0.3460	0.1836	Normal	Normal
ROA	0.0000	0.0025	Not Normal	Not Normal
Club Performance	0.0000	0.0000	Not Normal	Not Normal
Degree	0.0000	0.0000	Not Normal	Not Normal
Betweenness	0.0000	0.0000	Not Normal	Not Normal
Closeness	0.0000	0.0000	Not Normal	Not Normal
Community	0.0000	0.0000	Not Normal	Not Normal

Table 5. shows that the Shapiro–Wilk test indicates non-normal distributions for most financial variables in both distressed and healthy clubs. Debt Ratio is non-normal only in the distress group, while Player Expenses Ratio is the only financial variable that remains normally distributed in both groups. All network-signal variables (Degree, Betweenness, Closeness, Community) are also non-normal ($p < 0.05$). These results confirm that normality assumptions are largely unmet, justifying the use of non-parametric tests—specifically the Mann–Whitney U test—while Player Expenses Ratio proceeds to homogeneity testing and a t-test.

Table 6. Homogeneity Test (Levene) — Selected Variable

Variable	p-Value	Conclusion
Player Expenses Ratio	0.2233	Homogeneous

Table 6. shows that the homogeneity test for Player Expenses Ratio yields a p-value of 0.2233, which exceeds the 0.05 threshold. This indicates that the variances between groups are statistically homogeneous, allowing the variable to meet the assumptions required for further parametric analysis using the Independent Samples t-test.

Table 7. Group Difference Results

Variable	Test Used	p-Value	Conclusion
Debt Ratio	Mann–Whitney U	0.0006	Significant (Distress > Healthy)
Current Ratio	Mann–Whitney U	0.0048	Significant (Distress < Healthy)
ROE	Mann–Whitney U	0.2944	Not Significant
Player Expenses	Independent t-test	0.6916	Not Significant

Variable	Test Used	p-Value	Conclusion
Ratio ROA	Mann–Whitney U	0.2181	Not Significant
Club Performance	Mann–Whitney U	0.8207	Not Significant
Degree	Mann–Whitney U	0.1564	Not Significant
Betweenness	Mann–Whitney U	0.4224	Not Significant
Closeness	Mann–Whitney U	0.1564	Not Significant
Community	Mann–Whitney U	0.0095	Significant

Table 7. indicates that Debt Ratio (p = 0.0006) and Current Ratio (p = 0.0048) differ significantly between distressed and healthy clubs, while ROE, Player Expenses Ratio, ROA, and Club Performance show no significant differences. Among network-signal variables, only Community is significant (p = 0.0095), whereas Degree, Betweenness, and Closeness are not. Thus, H1 is partially supported, as significant differences appear in some financial and network signals but not in non-financial indicators.

Model Evaluation — Cross-Validation Results

Table 8. Per-fold evaluation: Base vs Hybrid Model (5-fold CV)

Base Model					
Fold	ROC_AUC	Recall	Precision	F1	Accuracy
1	0,631757	0,75	0,3	0,428571	0,644444
2	0,631757	0,875	0,259259	0,4	0,533333
3	0,472973	0,875	0,21875	0,35	0,422222
4	0,442568	0,25	0,5	0,333333	0,822222
5	0,694444	0,875	0,291667	0,4375	0,590909
Hybrid Model					
Fold	ROC_AUC	Recall	Precision	F1	Accuracy
1	0,668919	0,625	0,416667	0,5	0,777778
2	0,685811	1	0,275862	0,432432	0,533333
3	0,510135	1	0,216216	0,355556	0,355556
4	0,527027	1	0,228571	0,372093	0,4
5	0,746528	1	0,296296	0,457143	0,568182

Table 8. shows that the Hybrid Model consistently outperforms the Base Model, achieving higher ROC_AUC and F1 Scores and stronger recall, indicating better discrimination and improved detection of distress cases. However, both models still suffer from low precision, suggesting false positives remain an issue and threshold tuning may be needed. While the Base Model attains higher accuracy in some folds, the Hybrid Model provides a more effective early-warning signal overall.

Model Comparison – Base vs Hybrid Test

Table 9. Shapiro–Wilk normality for model metrics (Base vs Hybrid)

Metric	p-base	Dist Base	p-hybrid	Dist Hybrid
ROC_AUC	0.2659	Normal	0.3333	Normal
Recall	0.0053	Not Normal	0.0001	Not Normal

Metric	p-base	Dist Base	p-hybrid	Dist Hybrid
Precision	0.1116	Normal	0.3054	Normal
F1	0.4030	Normal	0.7235	Normal
Accuracy	0.9316	Normal	0.6431	Normal

Table 9. shows that most metrics in both the Base and Hybrid Models are normally distributed, except Recall, which is non-normal in both cases. This means ROC_AUC, Precision, F1, and Accuracy can be tested with parametric methods such as the Paired T-Test, while Recall requires a non-parametric test such as the Wilcoxon Signed-Rank Test.

Table 10. Paired comparison (Base vs Hybrid)

Metric	Test Used	p-Value	Conclusion
ROC_AUC	Paired t-test	0.0036	Significant (Hybrid > Base)
Recall	Wilcoxon signed-rank	0.3125	Not Significant
Precision	Paired t-test	0.6959	Not Significant
F1	Paired t-test	0.0385	Significant (Hybrid > Base)
Accuracy	Paired t-test	0.4609	Not Significant

Table 10. shows that only ROC_AUC and F1 Score differ significantly between the Base and Hybrid Models, with p-values of 0.0036 and 0.0385, respectively. Recall, tested with the Wilcoxon Signed-Rank Test, shows no significant difference (p = 0.3125), and the same applies to Precision (p = 0.6959) and Accuracy (p = 0.4609). These results indicate that H2 is partially accepted, as the Hybrid Model improves discrimination and achieves a better recall–precision balance, but does not produce significant gains across all metrics.

Feature importance (XGBoost)

Table 11. Feature importance (Gain, Weight, Cover)

Base Model				
Feature	**Gain**	**Weight**	**Cover**	**Rank**
Debt Ratio	1.249939	153	10.38125	1
Club Performance	1.129224	106	8.793637	2
ROA	0.877778	152.2	8.49746	3
Current Ratio	0.863688	133	7.150945	4
Player Expenses Ratio	0.668925	127.6	7.584859	5
ROE	0.660378	145.6	7.32357	6
Hybrid Model				
Feature	**Gain**	**Weight**	**Cover**	**Rank**
Degree	4.042232	18.6	27.13042	1
Community	1.68952	14.8	10.05573	2
Club Performance	1.122175	92.6	8.465103	3
Debt Ratio	0.935783	140.4	8.842959	4
ROA	0.89554	143.6	8.453078	5
Current Ratio	0.794071	130.8	6.742807	6
ROE	0.681406	132.6	7.841575	7
Player Expenses Ratio	0.621963	108.4	7.627553	8
Closeness	0	0	0	9
Betweenness	0	0	0	10

Table 11. shows a clear shift in feature importance when network variables are added. In the Base Model, financial ratios dominate, led by Debt_Ratio, Club_Performance, and ROA. In the Hybrid Model, network features—especially Degree and Community—become the top contributors, while financial variables move to lower ranks. This indicates that network signals substantially reshape the model's priorities and serve as the primary drivers of distress prediction.

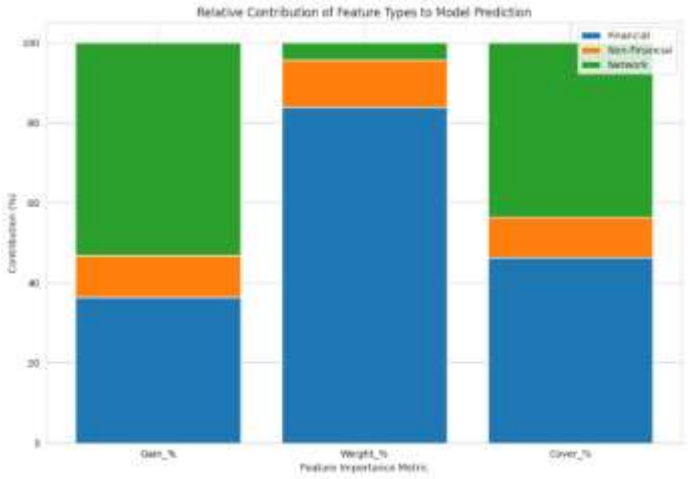

Figure 3. Relative Contribution of Feature Type to Model Prediction

Figure 3. shows distinct differences in the relative contribution of financial, non-financial, and network features across metrics. Network features dominate the Gain_% metric (≈50%), confirming their strong predictive value, while financial features contribute most to Weight_% (>80%), indicating they are used more frequently in tree splits. In terms of Cover_%, both financial and network features have substantial influence (≈46% and ≈43%), reflecting their impact on large portions of the data. Overall, the figure demonstrates that financial and network features play complementary roles: network features drive accuracy improvements, financial features guide more frequent splits, and both significantly shape model behavior, reinforcing the added value of network signals in the Hybrid Model.

DISCUSSION

Network Analysis

The visualization of the financial correlation network (Figure 2) shows the formation of communities and variations in clubs' positions within the network—findings that not only represent statistical associations but can also be interpreted as economic signals conveyed by each entity to the market and stakeholders. Within the framework of signaling theory, information is not always complete, and actors behave based on observable signals; relational positions (e.g., centrality) and community structures function as collective signals of a club's financial quality and stability (Spence, 1973; Choudhury, 2024). Accordingly, high centrality—as seen in clubs within Community 1 (e.g., Manchester City, Liverpool, Arsenal)—can be understood as a signal of

consistent performance and relatively strong financial capacity. These signals assist stakeholders (investors, sponsors, creditors) in assessing exposure and confidence toward the club (Alaminos & Fernández, 2019; Plumley et al., 2021).

Conversely, strong interconnectedness within a cluster that contains a high proportion of distress cases—such as Community 2 (e.g., Manchester United, Chelsea, Aston Villa)—produces collective signals of homogeneous risk patterns; in other words, negative signals from several clubs may reinforce perceptions of group-level vulnerability, prompting conservative risk evaluations from external stakeholders. This phenomenon is consistent with the concept of "herd signaling," in which similar financial patterns jointly signal the possibility of systemic distress within a group (Bardoscia et al., 2021; Song et al., 2021). Meanwhile, relatively isolated nodes—such as Crystal Palace in Community 0—send different signals: autonomy or unique financial profiles that may reflect resilience against systemic risk transmission, but also signal informational uncertainty that requires additional contextual interpretation (Alaminos et al., 2020).

From an agency and governance perspective, network position signals are also linked to managerial incentives and ownership structures: clubs with high centrality may be more motivated to maintain their reputation and long-term stability because reputational consequences are greater for highly visible entities (Jensen & Meckling, 1976). Conversely, clubs in high-risk clusters may face agency pressures that worsen fiscal behavior (e.g., transfer spending, high leverage), thereby reinforcing the negative signals detected through financial correlations (Barajas & Rodriguez, 2010; Plumley et al., 2021).

Comparison of Distress and Healthy Signals

Hypothesis 1 (H1) states that there are significant differences in financial signals and network signals between distressed clubs and financially healthy clubs in the English Premier League. Based on the Shapiro–Wilk normality test (Table 5) and difference tests including the Mann–Whitney U Test and the Independent Samples t-test, the decision for H1 is partially accepted. The group difference results in Table 7 show that not all variables differ significantly between distressed and healthy clubs. Among the financial signals, only Debt Ratio ($p = 0.0006$) and Current Ratio ($p = 0.0048$) consistently show significant differences. This indicates that debt structure and short-term liquidity serve as the primary differentiators of financial distress in football clubs. Conversely, profitability indicators such as ROE and ROA, as well as cost-efficiency metrics such as Player Expenses Ratio and Club Performance, do not differ significantly.

These findings align with Fan et al. (2023), who emphasize that leverage and liquidity are the most fundamental financial risk indicators in the global football industry. Fan et al. (2023) found that clubs with high leverage and limited liquid assets have a higher probability of financial distress regardless of sporting performance or short-term profits. This is also consistent with Alaminos & Fernández (2019), who state that financial failure in European

football clubs is more often triggered by imbalances in financial structure and debt pressure than by fluctuations in on-field performance.

The non-significance of the non-financial variable (Club Performance) reinforces this argument. Although Club Performance is included in the research objectives, the Mann–Whitney U Test indicates no significant difference between distressed and healthy clubs. This shows that sporting performance does not reliably reflect long-term or medium-term financial health. Agency theory helps explain this phenomenon: club managers may take risky decisions—such as aggressive player spending—to achieve short-term performance, even if such decisions increase future financial distress risk (Jensen & Meckling, 1976; Berle & Means, 1932). Barajas & Rodríguez (2010) similarly argue that high wage expenditure often fails to translate into financial sustainability, weakening performance as a signal of club viability.

In terms of network signals, only the Community variable ($p = 0.0095$) shows significant differences, whereas Degree, Betweenness, and Closeness do not. The significant Community result suggests that a club's structural position within network communities is relevant in distinguishing distress status. This supports Kadkhoda & Amiri (2024), who highlight that collective structural features—particularly community-level patterns—are more informative than individual centrality metrics. Community structure reflects systemic interdependence within the league, resonating with concepts of systemic risk in financial networks (Battiston et al., 2012; Bardoscia et al., 2021). The non-significance of individual centrality indicators indicates that a club's role within the network is not linearly associated with distress, especially in a highly homogeneous sample such as the EPL. This aligns with Qiu, Su, & Wang (2022), who argue that multilayer network structures and collective interactions are more effective in detecting crises than any single centrality metric.

From a signaling theory perspective, these results indicate that structural financial signals (debt and liquidity) are more credible and harder to manipulate than non-financial signals such as sporting performance (Spence, 1973; Choudhury, 2024). Distressed clubs issue strong negative signals through deteriorating financial ratios, while on-field performance may remain competitive due to aggressive financing strategies. These findings are consistent with Alaminos, Esteban, & Fernández-Gámez (2020), who assert that financial statements remain the primary signaling mechanism for stakeholders evaluating club sustainability.

Thus, H1 is partially supported because:

- Some financial signals (Debt Ratio and Current Ratio) differ significantly between distressed and healthy clubs.
- One network signal (Community) also differs significantly.
- Non-financial signals do not differ significantly.

These findings highlight that financial distress in the EPL is more strongly explained by financial structure and systemic network position than by sporting performance. The results reinforce the relevance of integrating financial analysis, signaling theory, agency theory, and network analysis.

Comparison of XGBoost Model Performance (H2)

Hypothesis 2 (H2) states that there is a significant difference in the predictive evaluation of financial distress between the Hybrid Model (network + XGBoost) and the Base Model (XGBoost without network features). Based on the Shapiro–Wilk test and paired comparison tests (Paired t-Test and Wilcoxon Signed-Rank Test), the decision for H2 is partially accepted. The difference results in Table 10 show that significant differences are found only in ROC_AUC ($p = 0.0036$) and F1 Score ($p = 0.0385$), whereas Recall, Precision, and Accuracy do not differ significantly. This indicates that adding network-signal features in the Hybrid Model improves class discrimination ability and the balance between recall and precision, but does not improve all evaluation metrics. Practically, the Hybrid Model better captures positive financial distress cases and produces higher F1 Scores, while overall accuracy remains similar.

These findings align with prior studies. Kadkhoda & Amiri (2024) show that integrating network features—such as Closeness Centrality—significantly enhances prediction accuracy. Fan et al. (2023) emphasize the importance of structural financial indicators such as leverage and liquidity, supporting the logical combination of financial and network information in Hybrid Models. Kristanti et al. (2023) highlight the strength of ensemble approaches such as XGBoost in balancing sensitivity and specificity, consistent with the improved F1 Score. Qiu, Su, & Wang (2022) emphasize the role of multilayer network structures in financial crisis detection. Alaminos & Fernández (2019) further reinforce that traditional financial indicators remain highly credible signals, validating the combined use of financial and network signals in predictive modeling.

From a signaling theory perspective, the improved performance of the Hybrid Model suggests that additional network information serves as a more credible signal of financial distress, particularly given the systemic interdependencies among clubs (Spence, 1973; Choudhury, 2024; Elwisam et al., 2024). Agency theory further explains that this information helps stakeholders monitor risk and reduce information asymmetry between managers and external parties (Jensen & Meckling, 1976; Panda & Leepsa, 2017).

Thus, H2 is partially supported because:

- ROC_AUC and F1 Score differ significantly between the Base and Hybrid Models.
- Recall, Precision, and Accuracy do not differ significantly.

These findings confirm that integrating network-signal features improves prediction performance—especially in class discrimination—although not across all metrics. They reinforce the relevance of combining financial analysis, signaling theory, agency theory, and network analysis.

Most Contributing Variables

The feature importance results for the Base Model and Hybrid Model show differences in variable priorities in predicting financial distress. In the Base Model, which uses only financial and non-financial features, Debt Ratio ranks highest, followed by Club Performance, ROA, Current Ratio, and Player Expenses Ratio, indicating that financial ratios are primary indicators of a club's

financial condition (Fan et al., 2023; Alaminos & Fernández, 2019). This aligns with signaling theory, where financial ratios communicate a club's financial health to stakeholders; managerial risk-taking behavior is reflected in these ratios, signaling early signs of distress. From an agency theory perspective, the dominance of financial ratios reflects potential information asymmetry between managers and shareholders, with Debt Ratio and Current Ratio acting as indirect monitoring mechanisms (Jensen & Meckling, 1976; Panda & Leepsa, 2017).

In the Hybrid Model, integrating network features significantly alters variable rankings: Degree and Community become the most dominant variables, while Debt Ratio and ROA drop in importance. This indicates that a club's network position and interconnections provide risk signals that are not captured by financial statements (Bardoscia et al., 2021; Kadkhoda & Amiri, 2024; Qiu et al., 2022). In signaling theory terms, network features serve as hidden signals that indicate systemic risk and interdependence among clubs, complementing financial ratios (Spence, 1973; Choudhury, 2024). From an agency perspective, the network structure highlights how managerial decisions in one club may affect the risk exposure of others, reinforcing the importance of monitoring systemic interactions (Jensen & Meckling, 1976).

Relative importance metrics (Gain_%, Weight_%, Cover_%) indicate that network features enhance prediction accuracy (Gain) and influence a large portion of observations (Cover), while financial features are used more frequently in tree splits (Weight). This suggests complementary roles between financial and network signals in predictive modeling (Alessi & Savona, 2021; Alaminos, Esteban, & Fernández-Gámez, 2020). These findings support H1 partially—some financial signals (Debt Ratio, Current Ratio) and network signals (Community) significantly differ between distressed and healthy clubs, while non-financial signals do not. They also support H2 partially, as the Hybrid Model improves prediction performance over the Base Model on ROC_AUC and F1, though not on all metrics.

Overall, integrating network features into financial distress prediction not only enhances model performance but also reinforces the relevance of signaling and agency theories by demonstrating that financial and network information provide complementary indicators of risk and managerial behavior in a systemically interconnected environment.

CONCLUSIONS AND RECOMMENDATIONS

The findings show that financial structure and network position function as the most reliable signals of financial distress among Premier League clubs. Distressed clubs are marked by high leverage, weak liquidity, volatile profitability, and elevated player-expense ratios, while healthy clubs maintain more stable capital structures. Debt Ratio and Current Ratio emerge as the strongest financial indicators, and Community is identified as the key network signal, indicating that structural embeddedness within the financial network conveys additional risk information. Significant differences appear only in leverage, liquidity, and network community structure, suggesting that

profitability, cost efficiency, and sporting performance do not distinguish distress status. Network patterns further reveal systemic risks not captured by individual ratios.

Predictive evaluation shows that adding network features to the XGBoost Hybrid Model improves ROC_AUC and F1-Score, with network variables becoming dominant in feature importance. Sporting performance indicators contribute little to prediction, reinforcing the value of combining financial and network signals. These results support signaling and agency theories by demonstrating that these indicators provide credible, difficult-to-manipulate signals that reduce information asymmetry and highlight systemic vulnerabilities.

Several practical implementations follow from these findings. Financial monitoring can prioritize leverage and liquidity; network-based surveillance can flag high-risk communities for early intervention; and hybrid predictive models can be adopted by regulators, investors, and auditors. Debt and liquidity policies can be refined through strengthened limits and buffers, while stakeholder decisions and risk communication frameworks can be improved by integrating standardized financial and network-based distress signals.

ADVANCED RESEARCH

This study has several limitations. The analysis is limited to English Premier League clubs, restricting generalizability to leagues with different financial structures and regulations. Non-financial variables are narrowly defined, as only Club Performance is included, leaving other relevant factors unexamined. The network model also remains limited, with only Community showing significant differences, indicating that systemic interactions may not be fully captured. The use of cross-sectional or season-level data prevents tracking long-term financial and network dynamics. Finally, although the Hybrid Model improves ROC_AUC and F1 Scores, gains are not consistent across other metrics, suggesting that predictive performance is still limited.

Future research may broaden the sample to other domestic and international leagues and incorporate additional non-financial indicators to improve generalizability and signal coverage. More advanced network models—such as temporal or multiplex networks—and longitudinal data are recommended to capture evolving interdependencies and financial dynamics. Further refinement of predictive models, including alternative algorithms or cost-sensitive methods, may also improve overall classification performance and strengthen distress detection.

REFERENCES

Alaminos, D., & Fernández, M. A. (2019). *Why do football clubs fail financially? A financial distress prediction model for European professional football industry*. PLoS ONE, 14(12), e0225989. https://doi.org/10.1371/journal.pone.0225989

Alaminos, D., Esteban, I., & Fernández-Gámez, M. A. (2020). Financial performance analysis in European football clubs. *Entropy, 22*(9), 1056. https://doi.org/10.3390/e22091056

Alessi, L., & Savona, R. (2021). Machine learning for financial stability. In S. Consoli, D. Reforgiato Recupero, & M. Saisana (Eds.), *Data science for economics and finance* (pp. 1–20). Springer. https://doi.org/10.1007/978-3-030-66891-4_4

Barajas, A., & Rodriguez, P. (2010). Spanish football clubs' finances: Crisis and player salaries. *International Journal of Sport Finance, 5*, 52–66.

Bardoscia, M., Barucca, P., Battiston, S., Caccioli, F., Cimini, G., Garlaschelli, D., Saracco, F., Squartini, T., & Caldarelli, G. (2021). The physics of financial networks. *Nature Reviews Physics, 3*(7), 490–507. https://doi.org/10.1038/s42254-021-00322-5

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., & Caldarelli, G. (2012). DebtRank: Too central to fail? Financial networks, the FED and systemic risk. *Scientific Reports, 2*(1), 541. https://doi.org/10.1038/srep00541

BBC. (2024, April 17). *Premier League PSR: Clubs total of £1bn of losses in 11 charts*. BBC. https://www.bbc.com/sport/football/68713522

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Choudhury, M. (2024). Signaling theory: An approach to organizational behavior research. *Journal of Accounting, Business and Management, 31*(2), 98–120. https://doi.org/10.31966/jabminternational.v31i2.1199

Deloitte. (2023, February 1). *Records tumble as Premier League clubs spend £815m in January transfer window*. https://www.deloitte.com/uk/en/about/press-room/records-tumble-as-premier-league-clubs-spend.html

Deloitte. (2024, June 25). *Annual review of football finance: Premier League clubs*. https://www.deloitte.com/uk/en/services/financial-advisory/research/annual-review-of-football-finance-premier-league-clubs.html

Fan, M., Chen, X., Liu, B., Zhou, F., Gong, B., & Tao, R. (2023). An analysis of financial risk assessment of globally listed football clubs. *Heliyon, 9*(12), e22886. https://doi.org/10.1016/j.heliyon.2023.e22886

GOV.UK. (2023, February 22). *Bold plan to protect long-term future of English football*. https://www.gov.uk/government/news/bold-plan-to-protect-long-term-future-of-english-football

Heryadi, A. D. (2023). Pengaruh financial distress, ukuran perusahaan dan leverage terhadap prudence pada perusahaan properti & real estate yang terdaftar di Bursa Efek Indonesia tahun 2017–2021. *Jurnal Ilmiah Akuntansi Rahmaniyah (JIAR), 6*(2). https://doi.org/10.51877/jiar.v6i2.286

IMF. (2014, May 28). *IMF survey: Financial networks key to understanding systemic risk*. https://www.imf.org/en/News/Articles/2015/09/28/04/53/sores052314a

Jabeur, S. B., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering.

Computational Economics, 61(2), 715–743. https://doi.org/10.1007/s10614-021-10227-1

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Kadkhoda, S. T., & Amiri, B. (2024). A hybrid network analysis and machine learning model for enhanced financial distress prediction. *IEEE Access, 12*, 52759–52777. https://doi.org/10.1109/ACCESS.2024.3387462

Kristanti, F. T., Febrianta, M. Y., Salim, D. F., et al. (2024). Advancing financial analytics: Integrating XGBoost, LSTM, and Random Forest algorithms for precision forecasting of corporate financial distress. *Journal of Infrastructure, Policy and Development, 8*(8), 4972. https://doi.org/10.24294/jipd.v8i8.4972

Lutfiyyah, I., & Bhilawa, L. (2021). Analisis akurasi model Altman modifikasi (Z"-Score), Zmijewski, Ohlson, Springate dan Grover untuk memprediksi financial distress klub sepak bola. *Jurnal Akuntansi, 13*(1), 46–60. https://doi.org/10.28932/jam.v13i1.2700

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. The MIT Press.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

Panda, B., & Leepsa, N. M. (2017). Agency theory: Review of theory and evidence on problems and perspectives. *Indian Journal of Corporate Governance, 10*(1), 74–95. https://doi.org/10.1177/0974686217701467

Plumley, D., Serbera, J.-P., & Wilson, R. (2021). Too big to fail? Accounting for predictions of financial distress in English professional football clubs. *Journal of Applied Accounting Research, 22*(1), 93–113. https://doi.org/10.1108/JAAR-05-2020-0095

Qiu, L., Su, R., & Wang, Z. (2022). Financial crisis prediction based on multilayer supervised network analysis. *Frontiers in Physics, 10*. https://doi.org/10.3389/fphy.2022.1048934

Scott, W. R., & O'Brien, P. C. (2003). *Financial accounting theory* (Vol. 3). Prentice Hall.

Song, J., Zhang, Z., & So, M. K. P. (2021). On the predictive power of network statistics for financial risk indicators. *Journal of International Financial Markets, Institutions and Money, 75*, 101420. https://doi.org/10.1016/j.intfin.2021.101420

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics, 87*(3), 355–374.

Zhang, J., & Luo, Y. (2017). Degree centrality, betweenness centrality, and closeness centrality in social network. In *Proceedings of the 2nd International Conference on Modelling, Simulation and Applied Mathematics* (pp. 300–303). Atlantis Press. https://doi.org/10.2991/msam-17.2017.68

Zhang, N., Qian, H., He, Y., Li, L., & Sun, C. (2021). A data-driven method for power system transient instability mode identification based on knowledge discovery and XGBoost algorithm. *IEEE Access, 9*, 154172–154182. https://doi.org/10.1109/ACCESS.2021.3124051