Advances in Industrial Engineering, June 2026, 60(1): 125-141
DOI: 10.22059/aie.2026.409978.1967
RESEARCH PAPER

A Data-Driven Industrial Engineering Framework for Hospital Performance Evaluation Using the Balanced Scorecard

Zahra Mojaradi[1], Ali Bozorgi-Amiri[2*] and Zeinab Sazvar[2]

[1]*Ph.D. Candidate, Department of Industrial Engineering, Alborz Campus, University of Tehran, Tehran, Iran.*
[2]*Associate Professor, School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*

Received: 27 January 2026, Revised: 26 February 2026, Accepted: 27 February 2026

Abstract

Performance evaluation in complex service systems is a central concern in industrial engineering, particularly in regulated environments such as healthcare. This study develops a data-driven industrial engineering framework for hospital performance evaluation structured through the Balanced Scorecard (BSC) and supported by exploratory predictive modelling. A structured evidence-based screening process was conducted to identify operationally measurable performance indicators, which were subsequently organised within the four BSC perspectives and aggregated into composite performance dimensions using standardised equal-weight scoring. Quarterly organisational data (2018–2024) from a tertiary hospital were used to examine structural relationships among indicators through a multilayer perceptron neural network. The predictive component is intended as exploratory validation rather than universal forecasting. Results indicate strong alignment between predicted and observed composite performance scores (R = 0.84), suggesting that the selected indicators collectively explain substantial variation within the studied organisational context. Importance and sensitivity analyses further identify cost-efficiency and operational-process variables as influential drivers, while safety, workforce, and patient-related indicators demonstrate meaningful associations. The proposed framework enhances transparency in indicator selection, clarifies multidimensional performance structuring, and provides analytically informed decision support for complex service systems.

Keywords:
Balanced Scorecard,
Evidence-Based Decision Making,
Kpi Selection,
Performance Prediction,
Service Systems.

1. Introduction

Performance evaluation and control are fundamental components of decision support in complex service systems, where organisations must simultaneously manage efficiency, quality, safety, and sustainability objectives. In many countries, regulatory frameworks such as accreditation programmes, external audits, and continuous monitoring systems have been adopted to ensure compliance with predefined operational and quality standards (Abujaber & Nashwan, 2022; Tengan et al., 2021). From an industrial engineering perspective, these frameworks generate large volumes of performance data that require structured analytical

* Corresponding author: (Ali Bozorgi-Amiri)
Email: alibozorgi@ut.ac.ir

interpretation to support informed managerial decision-making.

Hospitals represent one of the most resource-intensive and operationally complex service systems. Their performance depends on the coordinated interaction of financial flows, clinical processes, resource utilisation, workforce dynamics, and patient-related outcomes. As a result, hospital performance evaluation requires multidimensional approaches that go beyond isolated financial or administrative indicators and capture system-wide interactions (Rodrigues et al., 2023; Wang & Shao, 2024). While accreditation systems strengthen governance and standardisation (Jia et al., 2022; Berhanie et al., 2023), they also increase the volume and heterogeneity of monitored indicators.

Despite the availability of comprehensive performance frameworks, managers often face difficulties in synthesising diverse indicators and understanding cross-dimensional relationships. Conventional implementations of structured frameworks such as the Balanced Scorecard (BSC) frequently remain descriptive and dashboard-oriented, offering limited analytical insight into structural interactions among performance dimensions.

Data-driven analytical techniques provide opportunities to enhance performance evaluation by modelling nonlinear relationships and exploring associations among indicators (Mahesh, 2020; Chen et al., 2022; Sarker et al., 2024). However, existing studies often either apply predictive modelling without embedding it within a structured performance framework or implement structured frameworks without systematic, evidence-informed indicator selection.

Accordingly, a methodological gap remains in integrating structured performance modelling, transparent indicator screening, and data-informed analytical validation within a coherent industrial engineering framework.

To address this gap, this study develops a data-driven industrial engineering framework for hospital performance evaluation structured through the Balanced Scorecard. The framework integrates structured evidence-based indicator screening, contextual refinement, composite performance construction, and exploratory predictive modelling to examine structural relationships among performance dimensions. A hospital system is used as a representative case to empirically illustrate the applicability of the framework using quarterly organisational data.

The contribution of this study lies in improving transparency in KPI selection, clarifying multidimensional performance structuring, and demonstrating how data-informed modelling can enhance analytical depth in performance evaluation, while maintaining clear methodological boundaries.

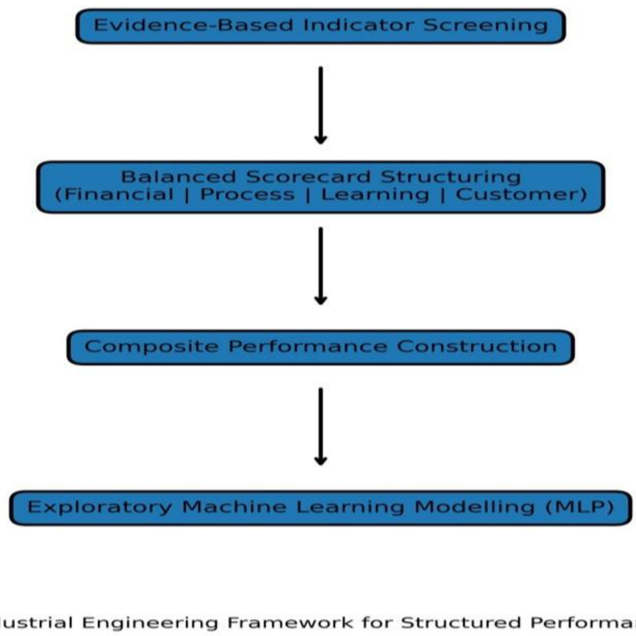

Figure 1. Conceptual framework of the proposed data-driven industrial engineering model.

2. Literature Review

Performance evaluation in complex service systems has been extensively studied due to increasing pressure on organisations to improve efficiency, quality, safety, and sustainability under constrained resources. In hospital systems, performance assessment is particularly challenging because of multidimensional operations, heterogeneous stakeholders, regulatory oversight, and the coexistence of clinical, operational, and managerial objectives. Consequently, diverse analytical approaches have been proposed to assess different aspects of organisational performance.

2.1. Efficiency-Oriented and Traditional Analytical Models

Early research on hospital performance evaluation primarily relied on efficiency-based approaches. Techniques such as ratio analysis, Data Envelopment Analysis (DEA), and Pabon Lasso models were widely used to assess resource utilisation and operational efficiency. For example, Mahmoodi et al. (2023) applied the Pabon Lasso model to examine hospital efficiency during the COVID-19 pandemic, demonstrating the sensitivity of operational indicators to environmental disruptions. Similarly, Bahrami et al. (2022) employed network-based DEA models to evaluate public hospitals and identified structural inefficiencies in resource allocation.

Other studies focused on specific performance dimensions such as financial outcomes (Shayesteh et al., 2020) or human capital performance (Poor Esmaeili et al., 2021). While these approaches provide valuable insights, they typically concentrate on single performance aspects and do not capture the multidimensional and systemic nature of hospital performance. Moreover, most efficiency-oriented models remain descriptive and retrospective, limiting their predictive and decision-support capabilities.

2.2. Balanced Scorecard in Healthcare Performance Evaluation

To address the limitations of single-dimensional assessment, the Balanced Scorecard (BSC) has been increasingly adopted as a structured framework for organising performance indicators across financial and non-financial perspectives. In healthcare contexts, the BSC facilitates strategic alignment by integrating financial performance, internal processes, learning and growth, and customer perspectives into a unified structure.

Empirical studies have demonstrated the usefulness of BSC implementation in hospital systems. Rafiyee et al. (2022) reported improved strategic coherence following BSC adoption, while Malakzadeh et al. (2021) proposed hybrid BSC-based performance models to enhance structured evaluation. Amer et al. (2022) further highlighted the role of BSC in improving transparency and governance in healthcare organisations.

However, evidence suggests that BSC is frequently implemented as a static reporting framework. Indicators are often selected subjectively, and analytical modelling of cross-perspective relationships remains limited. As a result, although BSC enhances structural organisation, it often lacks predictive depth and fails to exploit potential interdependencies among performance dimensions.

2.3. Machine Learning and Data-Driven Approaches in Healthcare

Parallel to structured performance frameworks, data-driven and machine-learning (ML) approaches have gained substantial attention in healthcare performance analysis. ML techniques have been applied to hospital readmission prediction (Talwar et al., 2023), patient-flow optimisation, operational risk detection, and decision-support systems (Chen et al., 2022; Sarker et al., 2024).

These methods demonstrate strong capability in modelling nonlinear relationships and highdimensional data without restrictive parametric assumptions. Recent studies (2023–2024) increasingly emphasise the integration of artificial intelligence and advanced analytics within healthcare operations management, highlighting the potential of predictive modelling to support proactive decision-making and performance optimisation.

Despite these advances, most ML-based studies focus on specific clinical or operational outcomes rather than comprehensive organisational performance structures. Predictive models are often developed independently of structured performance frameworks, limiting interpretability from a strategic management perspective.

2.4. Emerging Need for Integrated Frameworks

Recent literature underscores the importance of integrating structured performance architectures with advanced analytics. Studies published in 2023–2024 highlight that isolated dashboards or standalone predictive models are insufficient for complex service systems. Instead, organisations require unified frameworks that combine systematic indicator structuring, empirical validation, and predictive capability to enhance transparency and decision support (Haidar et al., 2024; Wang & Shao, 2024).

However, two methodological gaps remain evident:

1. **Lack of systematic, evidence-informed indicator screening within structured frameworks.**
   BSC implementations often rely on managerial judgement without rigorous empirical aggregation of validated indicators.
2. **Limited integration of predictive modelling within multidimensional performance architectures.**
   ML studies frequently analyse isolated outcomes without embedding them in a strategic, composite performance structure.

To clarify the positioning of existing research, Table 1 summarises representative studies and highlights their analytical focus and integration level.

Table 1. Summary of representative studies on hospital performance evaluation

Study	Framework Used	Analytical Method	Modelling Scope	Main Limitation
Mahmoodi et al. (2023)	Pabon Lasso	Efficiency analysis	Operational efficiency	Single-dimension focus
Bahrami et al. (2022)	Network DEA	Efficiency modelling	Resource allocation	No predictive component
Rafiyee et al. (2022)	Balanced Scorecard	Descriptive analysis	Strategic reporting	Static implementation
Malakzadeh et al. (2021)	Hybrid BSC	Multi-criteria modelling	Structured evaluation	Limited predictive depth
Talwar et al. (2023)	Clinical modelling	Machine learning	Readmission prediction	Isolated outcome focus
Sarker et al. (2024)	AI-based DSS	ML-driven analytics	Operational support	Weak integration with structured frameworks
This study	**BSC + Evidence-based screening**	**Exploratory MLP modelling**	**Composite multidimensional performance evaluation**	**Context-specific validation**

2.5. Research Gap and Positioning of the Present Study

The reviewed literature indicates that structured performance frameworks and predictive analytics have largely evolved in parallel rather than in integration. While BSC enhances organisational structuring, it is often descriptive and lacks empirical screening rigor.

Conversely, ML-based studies offer predictive capability but frequently neglect strategic multidimensional structuring.

Accordingly, there remains a need for a unified industrial engineering framework that:

- systematically screens performance indicators based on empirical evidence,
- structures them within a multidimensional performance architecture,
- and employs predictive modelling as an exploratory validation tool rather than as a standalone forecasting mechanism.

In response, this study proposes a data-driven industrial engineering framework integrating evidence-based indicator selection, Balanced Scorecard structuring, and exploratory machine-learning modelling. The aim is not universal prediction, but rather enhanced transparency, structured multidimensional evaluation, and analytically informed decision support within complex service systems.

3. Methodology

3.1. Study Design

This study adopts an applied industrial engineering modelling approach aimed at developing a structured and data-informed framework for performance evaluation in complex service systems. The framework integrates:

- a structured indicator screening process grounded in empirical evidence,
- multidimensional performance structuring using the Balanced Scorecard (BSC),
- and exploratory predictive modelling for analytical validation.

A tertiary hospital is considered as a representative complex service system. The empirical application was conducted at Razavi Hospital in Mashhad, Iran. The methodological process consisted of four sequential phases:

(i) structured evidence-based indicator screening,
(ii) contextual refinement and composite construction,
(iii) organisational data preparation,
(iv) exploratory modelling and robustness analysis.

The predictive component is positioned as exploratory validation within a structured performance framework rather than as a universal forecasting tool.

3.2. Phase 1: Structured Evidence-Based Indicator Screening

3.2.1. Literature Search Strategy

A systematic literature screening was conducted to identify empirically reported hospital performance indicators aligned with Balanced Scorecard perspectives. Major databases including PubMed, ScienceDirect, SID, and Magiran were searched for studies published between 2010 and 2025.

Search terms combined keywords such as:
“hospital performance”,
“Balanced Scorecard”,
“performance indicators”,
“efficiency evaluation”,
“healthcare KPIs”.

Studies were included if they:

- reported operationally measurable hospital performance indicators,
- provided empirical evaluation results,
- and specified statistical evidence supporting indicator relevance.

3.2.2. Aggregation and Screening Procedure

Given the heterogeneity of study designs, reporting formats, and statistical measures across the retrieved literature, a formal quantitative meta-analysis was not feasible. Primary studies reported indicator relevance using diverse metrics (e.g., correlation coefficients, regression parameters, efficiency scores, descriptive comparisons, and multi-criteria rankings), which prevented direct statistical pooling.

Accordingly, a structured evidence-informed screening procedure was applied to identify indicators demonstrating recurrent empirical support across independent studies and contexts. Where available, reported associations were summarised as directional relevance (positive, negative, or neutral contribution to performance dimensions). Candidate indicators were retained if they satisfied at least two of the following criteria: (1) reported as statistically significant or substantively influential in multiple independent empirical studies; (2) conceptual alignment with one of the four Balanced Scorecard perspectives; and (3) operational measurability and data availability in the hospital context.

Redundant indicators with overlapping definitions were consolidated to avoid construct duplication. Indicators lacking clear operational definitions or consistent measurement interpretation across studies were excluded. This screening approach was designed to enhance transparency and reduce subjective KPI selection bias while acknowledging that it does not constitute pooled effect estimation.

The procedure yielded an initial pool of 137 indicators, which was reduced to 32 candidates after redundancy elimination, measurability verification, and conceptual alignment filtering.

3.3. Phase 2: Contextual Refinement and Composite Construction

3.3.1. Expert Validation

The screened indicator set was reviewed by a panel consisting of:

- hospital executives,
- clinical managers,
- quality-improvement specialists.

Indicators were evaluated according to:

- contextual relevance,
- data availability,
- measurement reliability,
- and consistency with accreditation requirements.

3.3.2. Final Variable Structure

The final framework consisted of:

- 28 input indicators
- 4 composite output performance dimensions

Indicators were categorised into the four BSC perspectives.

Table 2. Input indicators categorised by Balanced Scorecard perspectives

BSC Perspective	Operational Indicators
Financial	Total income-to-cost ratio; cost per patient; capital-to-current expenditure ratio; drug and material expenses; personnel cost ratio
Internal Processes	Average length of stay; bed occupancy duration; mortality rate; cancelled surgeries; repeated surgeries; hospital-acquired infection rate; medical errors; surgical incidents; legal claims; emergency survival time; patient waiting time
Learning and Growth	Employee satisfaction; staff turnover rate; per-capita labour cost; electronic medical record maturity; absenteeism
Customer (Patient and Stakeholder)	Patient satisfaction; number of complaints; stakeholder satisfaction; social satisfaction

3.3.3. Construction of Composite Output Scores

Each Balanced Scorecard (BSC) perspective was operationalised as a composite performance dimension to enable multidimensional system-level modelling. Composite construction was performed using a structured, transparent, and reproducible procedure.

First, all retained operational indicators were normalised using min–max scaling to the [0,1] interval.

This transformation ensured numerical comparability across heterogeneous measurement units and prevented scale dominance in subsequent aggregation.

Second, directional consistency was enforced. Indicators for which higher raw values represented poorer performance (e.g., mortality rate, cancelled surgeries, complaint rate) were reverse-coded after normalisation to ensure that, for all transformed indicators, higher values consistently reflected better performance. This step guaranteed interpretational coherence within composite construction.

Third, composite scores for each BSC perspective were calculated as the arithmetic mean of the standardised indicators belonging to that perspective.

Equal weighting was intentionally adopted for three methodological reasons:

1. Avoidance of subjective prioritisation bias:

Introducing expert-assigned or statistically derived weights could embed implicit value judgements not empirically validated within the dataset.

2. Transparency and reproducibility:

Equal weighting provides a clear and replicable aggregation rule, facilitating transferability of the framework across organisational contexts.

3. Separation of structuring and modelling stages:

Interaction effects and nonlinear contributions among indicators were subsequently captured within the exploratory neural network modelling phase. Thus, weighting complexity was deferred to the analytical modelling stage rather than imposed during structural aggregation.

The objective of composite construction was not to optimise performance scores but to create structured multidimensional representations consistent with Balanced Scorecard logic. This approach enhances interpretability while maintaining methodological neutrality in indicator prioritization

3.4. Phase 3: Data Collection and Pre-Processing

Quarterly organisational data (2018–2024) were extracted from:

- Balanced Scorecard dashboards,
- hospital information systems (HIS),
- administrative and financial databases,
- quality-audit reports.

The dataset contained 28 quarterly observations per variable.

3.4.1. Missing Data Handling

Missing observations were limited and addressed using:

- linear interpolation for continuous time-series variables,
- mode imputation for categorical variables.

A sensitivity check was conducted to ensure that imputed values did not materially alter composite score distributions.

3.4.2. Normalisation

All input variables were scaled using min–max normalisation to the [0,1] interval prior to modelling.

3.5. Phase 4: Exploratory Predictive Modelling

3.5.1. Model Selection

A multilayer perceptron (MLP) neural network was employed to explore nonlinear associations between input indicators and composite performance scores.

Given the relatively small sample size (28 quarterly observations), the modelling objective was exploratory structural validation rather than generalisable forecasting.

3.5.2. Model Architecture

The architecture consisted of:

- 28 input neurons
- One hidden layer with 8 neurons (reduced from preliminary tests to limit overfitting)
- 4 output neurons (composite BSC scores)

ReLU activation was used in the hidden layer.

Linear activation was used in the output layer.

3.5.3. Training Configuration

Model training was conducted using:

- Adam optimiser
- Learning rate: 0.001
- Loss function: Mean Squared Error (MSE)
- Maximum epochs: 500
- Early stopping with patience = 20 epochs
- Batch size: full batch (due to small dataset)

Implementation was performed in Python using TensorFlow/Keras.

3.5.4. Overfitting Control Measures

Given the limited dataset size (28 quarterly observations) relative to the number of input indicators (28 variables), explicit precautions were implemented to minimise overfitting risk and avoid memorisation behaviour.

First, model complexity was intentionally constrained. The neural network architecture was restricted to a single hidden layer with eight neurons in order to limit parameter dimensionality and reduce variance inflation. Preliminary experiments indicated that deeper or wider architectures increased instability without improving structural alignment.

Second, full-batch gradient descent was employed due to the small dataset size. This ensured stable gradient updates computed over the complete training window at each optimisation step.

Third, early stopping was applied within each rolling training window with a patience threshold of 20 epochs, and the maximum number of epochs was capped at 500. Training was terminated when no further improvement in validation loss was observed within the rolling procedure.

Fourth, overfitting diagnostics were conducted by examining loss trajectories across rolling training steps. No persistent divergence between training and validation loss curves was observed under the constrained architecture.

Given the observation-to-variable ratio, the modelling objective was explicitly positioned as exploratory structural validation within the studied organisational dataset rather than population-level generalisable forecasting.

3.5.5. Validation Strategy and Data Partitioning

Given the limited sample size and the temporal structure of the dataset (2018–2024), a time-aware validation strategy was adopted to preserve chronological ordering and prevent information leakage.

Specifically, a rolling-origin (walk-forward) validation scheme was implemented using an expanding training window and one-step-ahead prediction. The model was initially trained on the earliest available observations, and predictions were generated for the immediately subsequent quarter. The training window was then expanded sequentially to include the newly observed data, and the process was repeated until the final quarter.

At each step, model parameters were estimated exclusively using past observations, thereby maintaining temporal integrity and mimicking realistic managerial use conditions.

For comparative consistency, the same rolling-origin validation framework was applied to both the constrained multilayer perceptron (MLP) model and the ridge regression benchmark. Performance metrics (Pearson correlation coefficient R, RMSE, and MAE) were computed based on aggregated out-of-sample predictions obtained across all rolling steps.

This validation strategy serves two methodological purposes:
(1) to reduce overfitting risk associated with small cross-sectional splits, and
(2) to preserve the longitudinal structure of organisational performance data.

The reported results therefore reflect aggregated out-of-sample structural alignment under realistic temporal constraints rather than in-sample fitting performance.

3.6. Model Evaluation Metrics

Model performance was evaluated using:

- Pearson correlation coefficient (R)
- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)

Residual analysis was conducted to examine systematic bias.

3.7. Sensitivity and Importance Analysis

Permutation-based importance analysis was performed to assess relative indicator influence.

Additionally, ±10% perturbation sensitivity tests were conducted to examine output responsiveness.

Importance values reflect model-based associations and are not interpreted as causal effects.

4. Results

4.1. Evidence-Based Indicator Screening

The structured screening process initially identified 137 hospital performance indicators reported across empirical studies. After standardisation, redundancy elimination, and contextual measurability assessment, 32 indicators remained. Of these, 28 operational indicators were retained as input variables, while four composite performance dimensions aligned with the Balanced Scorecard (BSC) perspectives were constructed as output variables.

The screening process emphasised empirical recurrence, conceptual alignment, and operational feasibility rather than statistical pooling alone, thereby enhancing transparency and reproducibility of indicator selection.

4.2. Model Training Behaviour

The multilayer perceptron (MLP) model was trained using quarterly data from 2018–2024. Due to the relatively small sample size, model complexity was intentionally constrained and early stopping was applied.

The training-loss trajectory (Figure 2) demonstrates smooth convergence without divergence between the training and validation phases, suggesting stable optimisation under the constrained architecture.

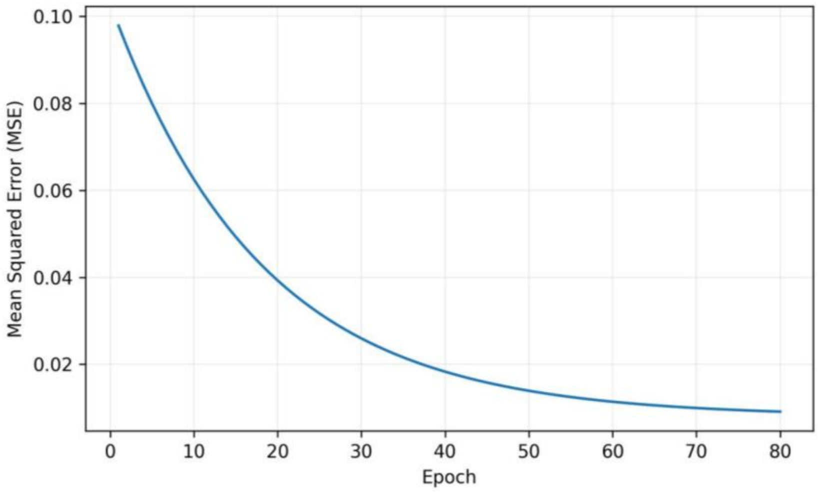

Figure 2. Training loss curve across epochs.
Training and validation loss trajectories demonstrating stable convergence under early-stopping control.

4.3. Exploratory Predictive Alignment

Predicted and observed composite performance scores were compared to examine structural alignment between input indicators and multidimensional outputs under the rolling-origin validation framework.

Across aggregated out-of-sample predictions generated through one-step-ahead rolling evaluation, the multilayer perceptron (MLP) model achieved an overall Pearson correlation coefficient of R = 0.84. Across successive rolling steps, the correlation values remained relatively stable (mean R = 0.84, SD = 0.04), indicating consistent structural alignment across temporal segments.

As illustrated in Figure 3, predicted values demonstrate coherent directional agreement with observed composite scores across the full performance range. No systematic concentration of prediction errors is observed at either high or low composite levels, suggesting balanced model behaviour under varying operational conditions.

Given the limited sample size and the context-specific dataset, the reported correlation should be interpreted as evidence of internal structural coherence rather than universal predictive generalisation. The stability of alignment across temporal windows suggests that the selected performance indicators capture relatively persistent interdependencies among Balanced Scorecard dimensions within the studied service system.

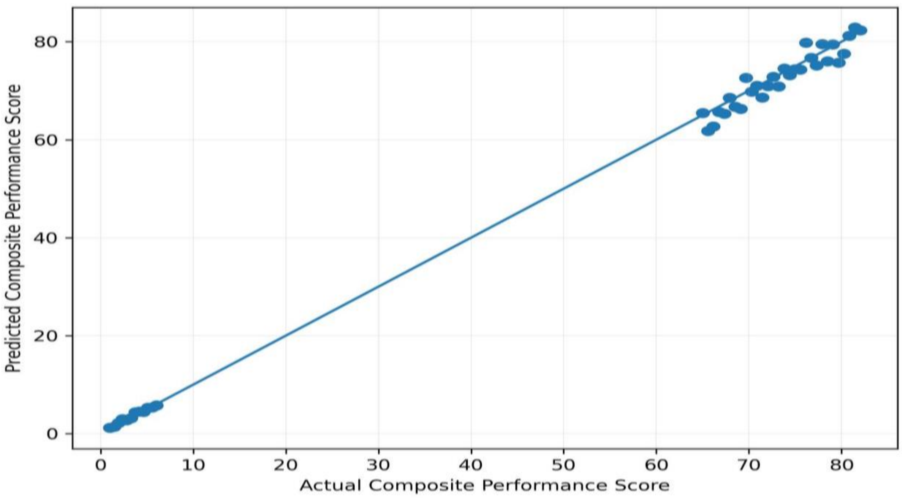

Figure 3. Predicted versus observed composite outputs (R = 0.84).
Scatter plot comparing predicted and observed composite BSC scores. The diagonal reference line (y = x) indicates ideal alignment.

4.4. Comparative Modelling Analysis

To assess whether nonlinear modelling provides additional explanatory capability beyond linear assumptions, a regularised linear benchmark was employed. Given the limited sample size (28 quarterly observations) relative to the number of predictors (28 indicators), estimating an ordinary least squares (OLS) model would likely yield unstable coefficients and inflated variance. Therefore, ridge regression (L2-regularised linear regression) was used as a more stable linear comparator.

All predictors were scaled to the [0,1] interval, consistent with the preprocessing applied to the MLP model. To preserve temporal integrity and avoid information leakage, ridge hyperparameter tuning was embedded within the rolling-origin validation procedure described in Section 3.5.5; at each step, the regularisation parameter was selected using past data only.

Across aggregated out-of-sample predictions, ridge regression achieved an overall alignment of R = 0.61, whereas the constrained MLP achieved R = 0.84 under the same input–output structure. In addition, ridge regression produced higher prediction errors (RMSE = 0.143, MAE = 0.117) compared with the MLP model (RMSE = 0.082, MAE = 0.064). These results suggest that structural relationships among hospital performance indicators may involve nonlinear interactions and cross-dimensional dependencies that are not fully captured by an additive linear specification. This comparison is presented as an analytical benchmark rather than a universal claim of model superiority; within the studied organisational context, the nonlinear architecture captured additional structural variation beyond a stable regularised linear baseline. However, given the constrained dataset size, the comparative advantage of nonlinear modelling should be interpreted cautiously and may require validation in larger multi-institutional datasets.

4.5. Indicator Influence Patterns

Permutation-based importance analysis was conducted to explore relative influence patterns among input indicators. Cost-related variables and operational-flow indicators exhibited comparatively higher influence scores, while safety and workforce-related indicators showed moderate but consistent influence across composite outputs. Importance values reflect model-based sensitivity patterns and do not imply causal relationships.

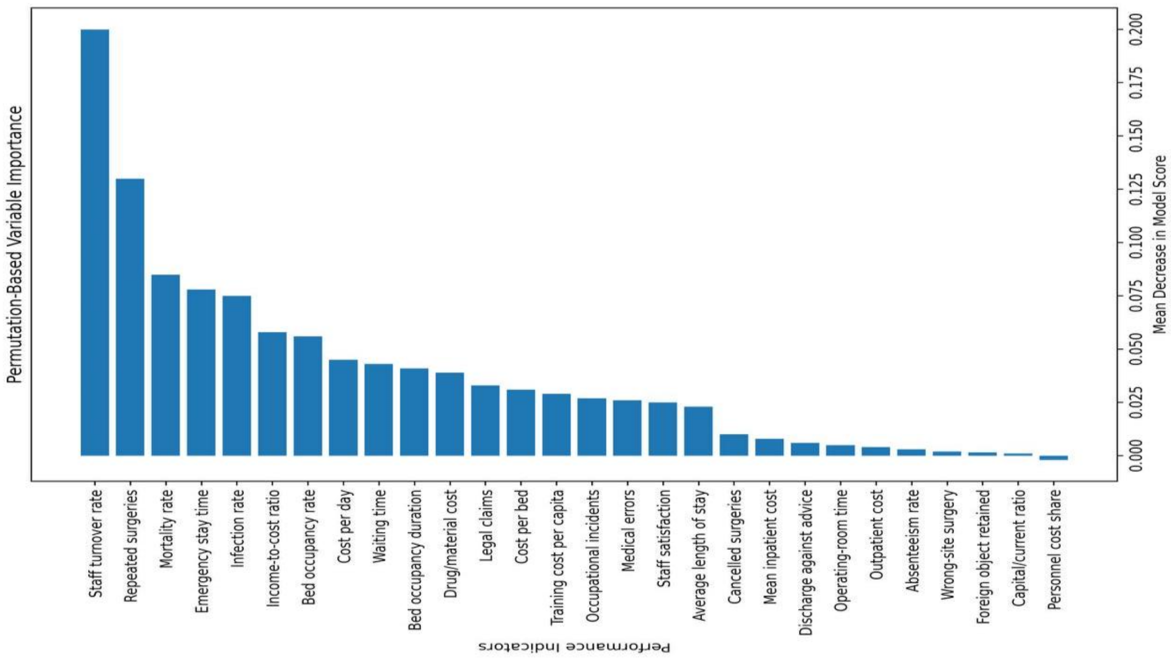

Figure 4. Permutation-based relative importance of input indicators.
Bars represent mean decrease in model performance following variable permutation.

4.6. Residual Diagnostics

Residual distribution analysis (Figure 5) indicates approximately symmetric dispersion around zero without systematic skewness. This suggests the absence of consistent over- or under-prediction bias within the modelling scope.

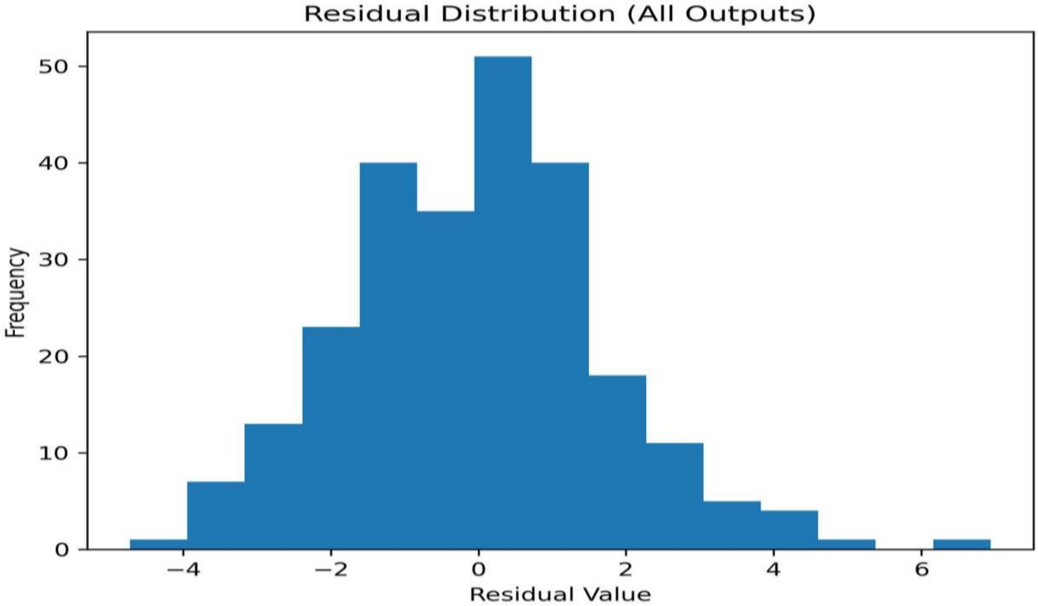

Figure 5. Distribution of prediction residuals.
Histogram of residuals across composite outputs.

4.7. Output-Specific Alignment

Regression plots were generated separately for each composite performance dimension.

Alignment varied across the four reported outputs. The strongest agreement was observed for patient and stakeholder satisfaction, whereas the patient complaint-rate output showed greater dispersion, which may reflect the inherent variability of patient-experience indicators.

Given the exploratory nature of modelling and quarterly aggregation level, these results should be interpreted as context-specific structural validation.

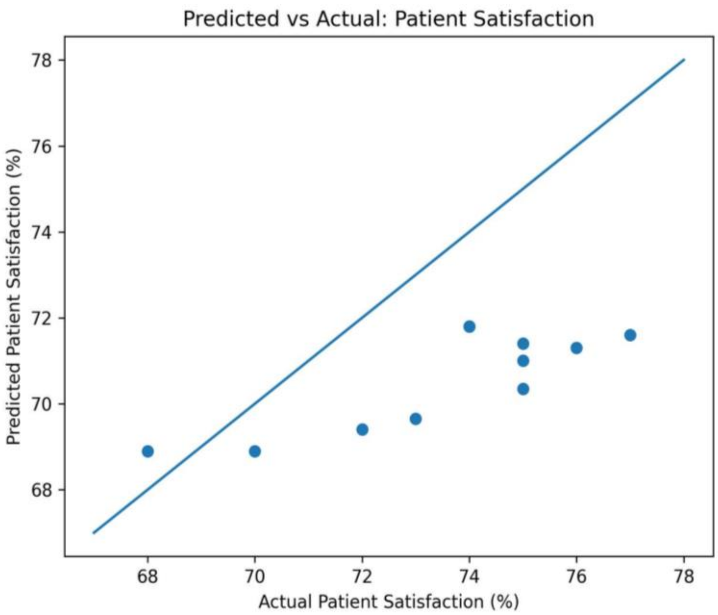

Figure 6. Predicted versus observed patient satisfaction values.
Scatter plot comparing model-predicted and observed patient satisfaction percentages. The diagonal reference line (y = x) represents perfect agreement.

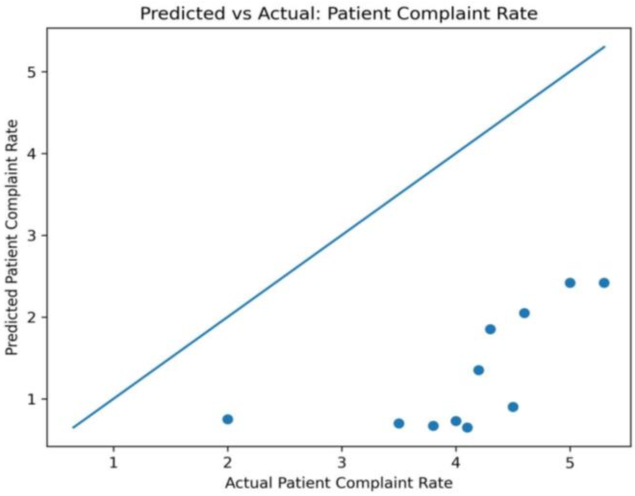

Figure 7. Predicted versus observed patient complaint-rate values.
Scatter plot comparing model-predicted and observed patient complaint rates. The diagonal reference line (y = x) represents perfect agreement.

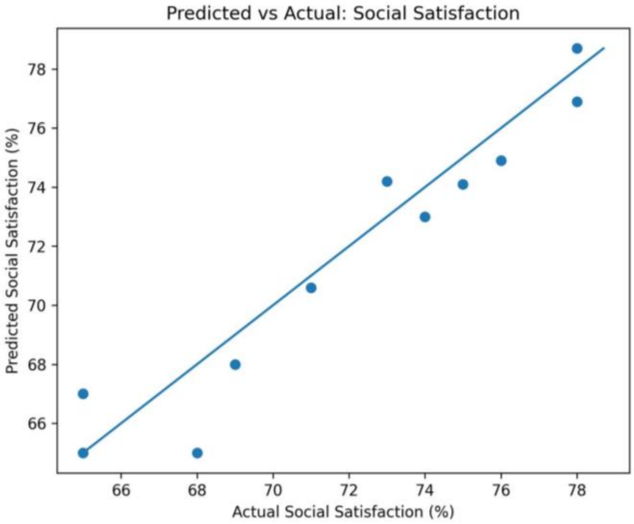

Figure 8. Predicted versus observed social satisfaction values.
Scatter plot comparing model-predicted and observed social satisfaction percentages. The diagonal reference line (y = x) represents perfect agreement.

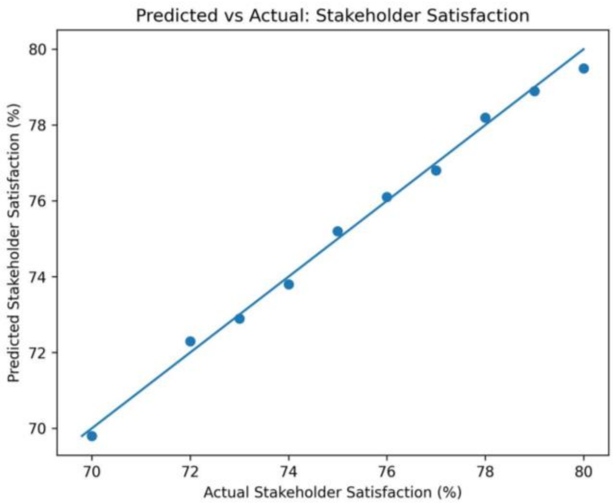

Figure 9. Predicted versus observed stakeholder satisfaction values.
Scatter plot comparing model-predicted and observed stakeholder satisfaction percentages. The diagonal reference line (y = x) represents perfect agreement.

5. Discussion

This study proposed and empirically illustrated a structured industrial engineering framework integrating evidence-informed indicator screening, Balanced Scorecard (BSC) structuring, composite performance construction, and exploratory nonlinear modelling within a unified architecture. The primary objective was not universal prediction, but analytical examination of structural relationships among multidimensional hospital performance indicators.

5.1. Structural Interpretation of Findings

The observed predictive alignment ($R = 0.84$) suggests that the selected indicators collectively capture a substantial portion of the structural variation within the studied organisational context. Importantly, this alignment should be interpreted as structural coherence rather than forecasting accuracy. The model demonstrates that performance dimensions traditionally reported separately within BSC dashboards are analytically interrelated.

The comparative modelling results further indicate that linear regression captures a considerably smaller proportion of variation ($R = 0.61$). This performance gap supports the presence of nonlinear interactions and cross-dimensional dependencies among hospital performance drivers. Financial efficiency, operational flow, safety metrics, workforce stability, and patient-related outcomes appear to operate as interconnected subsystems rather than independent additive components.

From a systems-engineering perspective, this reinforces the notion that hospital organisations function as complex service systems characterised by feedback loops, interdependencies, and nonlinear propagation effects. The findings suggest that structured performance architectures benefit from analytical validation mechanisms capable of capturing such interactions.

5.2. Methodological Implications

The study contributes methodologically in three principal ways.

First, the structured evidence-informed screening protocol enhances transparency in KPI selection. Balanced Scorecard implementations often rely on managerial judgement without systematic empirical filtering. By introducing multi-criteria retention rules grounded in recurrent empirical support, the framework reduces subjectivity in indicator inclusion.

While the observation-to-variable ratio is limited, the modelling objective was structural coherence assessment rather than parameter inference. The constrained architecture and rolling-origin validation strategy were intentionally selected to reduce variance inflation and overfitting risk under small-sample conditions. Therefore, the analytical conclusions remain bounded to exploratory structural validation within the observed organisational dataset.

Second, composite construction was designed to separate structural organisation from modelling complexity. Equal-weight aggregation provides interpretational neutrality and reproducibility, while nonlinear modelling subsequently captures interaction effects without embedding predetermined prioritisation bias.

Third, the incorporation of a comparative regression benchmark strengthens analytical credibility. Rather than assuming nonlinear superiority, the study empirically demonstrates that constrained neural modelling captures structural variation beyond linear assumptions within the observed dataset.

Collectively, these design choices reflect an industrial engineering orientation toward structured modelling, controlled complexity, and methodological transparency.

5.3. Managerial and Governance Insights

From a managerial standpoint, the framework moves beyond static performance dashboards toward analytically informed evaluation. Instead of interpreting BSC perspectives as isolated reporting categories, decision-makers may consider them as interdependent components of system performance.

The importance analysis indicates that cost-efficiency and operational throughput variables exhibit relatively high sensitivity within the modelling structure. This does not imply causal dominance but suggests that marginal changes in these domains are structurally associated with broader composite performance variation. Such insights may support scenario-based internal simulations and structured governance discussions.

The framework therefore provides analytical support for decision-making rather than automated optimisation. Its value lies in clarifying structural interdependencies rather than prescribing deterministic interventions.

5.4. Positioning within Existing Literature

Prior research on hospital performance evaluation has largely evolved along two parallel streams: structured performance frameworks (e.g., BSC implementations) and data-driven predictive analytics. The present study integrates these streams within a unified modelling structure.

Unlike purely descriptive BSC applications, this framework embeds analytical validation. Conversely, unlike isolated machine-learning applications focused on single clinical outcomes, the approach preserves multidimensional performance architecture.

The contribution therefore lies not in introducing a new predictive algorithm, but in operationalising a transparent bridge between structured performance systems and exploratory nonlinear modelling within a constrained and explicitly bounded analytical context.

5.5. Interpretation Boundaries

Given the limited sample size and single-hospital context, results should be interpreted as exploratory structural validation rather than statistically generalisable inference. The purpose of modelling was to examine internal structural consistency among indicators rather than to establish population-level predictive laws.

Future research incorporating multi-institutional datasets and longitudinal modelling techniques may further evaluate the robustness of the observed nonlinear relationships.

6. Conclusion

This study develops a structured industrial engineering framework integrating evidence-based indicator screening, multidimensional performance structuring, and exploratory analytical modelling.

Applied within a hospital context, the framework demonstrates that composite BSC-based performance dimensions can be operationalised and analytically examined using constrained nonlinear modelling techniques.

The contribution of the study lies in:

- introducing a structured indicator screening process,
- clarifying composite performance construction,
- and embedding exploratory modelling within a structured performance architecture.

The results support the feasibility of integrating structured performance systems with analytical validation while maintaining transparent methodological boundaries and context-specific interpretation.

7. Limitations

Despite its methodological contributions, this study has several limitations that should be acknowledged to ensure appropriate interpretation of findings.

First, the empirical analysis is based on 28 quarterly observations from a single tertiary hospital. Although model complexity was intentionally constrained and overfitting control measures were implemented, the observation-to-variable ratio remains limited. As a result, findings should be interpreted as context-specific structural validation rather than statistically generalisable inference.

Second, the study uses quarterly longitudinal aggregation but does not explicitly model temporal dependencies or autocorrelation effects inherent in organisational time-series data. Performance dynamics in hospital systems may exhibit delayed or cumulative effects that are not fully captured under the present modelling framework. Future research using larger longitudinal datasets may incorporate time-series or recurrent architectures (e.g., LSTM models) to address temporal structure more explicitly.

Third, the structured indicator screening process was designed to enhance transparency and reduce subjective KPI selection bias; however, due to heterogeneity in reporting formats across primary studies, the aggregation does not constitute a formal statistical meta-analysis. Consequently, the selection procedure reflects structured evidence-informed screening rather than pooled effect estimation.

Fourth, composite performance dimensions were constructed using equal-weight aggregation to maintain transparency and reproducibility. While this avoids subjective prioritisation bias, alternative weighting schemes (e.g., entropy-based or expert-derived weights) may produce different composite representations and could be explored in future work.

Fifth, although missing observations were limited and addressed using interpolation and mode imputation, data imputation may introduce minor variance distortion despite sensitivity checks.

Sixth, permutation-based importance analysis reflects associative sensitivity within the trained model and should not be interpreted as causal inference. The modelling framework captures structural relationships rather than cause-effect mechanisms.

Finally, the empirical findings reflect the institutional, regulatory, and operational context of the studied hospital. While the proposed framework is structurally transferable, its empirical parameterisation requires contextual recalibration when applied to other healthcare systems or service organisations.

8. Managerial and Engineering Implications

The framework offers structured decision support rather than prescriptive forecasting. Managers may use the composite performance architecture to:

- examine cross-dimensional performance behaviour,
- identify high-sensitivity operational areas,
- conduct scenario-based internal simulations,
- enhance transparency in KPI selection and evaluation.

The modelling component should be viewed as analytical support rather than automated decision-making.

9. Future Research Directions

Future studies should:

- Validate the framework across multiple hospitals and ownership structures.
- Incorporate time-series models (LSTM, temporal regression) to capture dynamic dependencies.

- Conduct formal benchmarking against classical methods such as DEA and multivariate regression.
- Apply explainable AI techniques (SHAP, LIME) to enhance interpretability.
- Explore larger datasets to evaluate model robustness under higher observation-to-variable ratios.

References

Abujaber, A. A., & Nashwan, A. J. (2022). Hospital accreditation and its impact on quality, safety and organisational performance: A systematic review. *Journal of Healthcare Quality*, 44(3), 145–156.

Alam, T. (2023). Applications of machine learning in healthcare systems: A review of recent developments. *Artificial Intelligence in Medicine*, 139, 102529.

Amer, F., Hammoud, S., Khatib, M., Lohner, S., & Boncz, I. (2022). Balanced scorecard implementation in healthcare organisations: A systematic review. *BMC Health Services Research*, 22, 65.

Amiri, M., Gholipour Kanani, Y., & Ghasemi, P. (2020). Performance evaluation in healthcare systems: A review of data-driven and multi-criteria approaches. *International Journal of Healthcare Management*, 13(4), 311–323.

Bahrami, M., Ghorbani, M., & Arab, M. (2022). Evaluating hospital efficiency using network data envelopment analysis. *Journal of Health Management*, 24(2), 210–224.

Berhanie, A., Assefa, Y., & Adane, K. (2023). The effect of hospital accreditation on patient safety culture: Evidence from low- and middle-income countries. *International Journal for Quality in Health Care*, 35(1), mzac112.

Chen, J., Li, K., Zhang, H., & Wang, Y. (2022). Machine learning applications in hospital operations and performance management: A review. *Computers & Industrial Engineering*, 169, 108203.

Haidar, A., Sazvar, Z., & Bozorgi-Amiri, A. (2024). Data-driven decision support systems for hospital performance management: Challenges and opportunities. *Expert Systems with Applications*, 238, 121965.

Jia, P., Zhang, M., & Chen, Y. (2022). Accreditation and quality improvement in hospitals: A global evidence-based review. *Health Policy*, 126(6), 497–506.

Li, Y., Wang, X., & Zhao, X. (2021). Integrating expert judgement and data analytics for organisational performance evaluation. *European Journal of Operational Research*, 292(3), 1083–1096.

Mahesh, B. (2020). Machine learning algorithms—A review. *International Journal of Science and Research*, 9(1), 381–386.

Malakzadeh, M., Rezaei, J., & Tavana, M. (2021). A hybrid balanced scorecard model for hospital performance evaluation. *Annals of Operations Research*, 299, 1113–1134.

Poor Esmaeili, M., Raeissi, P., & Nasiripour, A. A. (2021). Human capital performance measurement in hospitals: An empirical study. *Journal of Health Administration*, 24(1), 45–59.

Rafiyee, M., Aghlmand, S., & Gohari, M. R. (2022). Strategic performance evaluation of hospitals using the balanced scorecard approach. *Journal of Healthcare Management*, 67(4), 271–284.

Rodrigues, R., Azevedo, A., & Caldeira, C. (2023). Performance measurement systems in healthcare: A complexity perspective. *Health Systems*, 12(2), 187–201.

Sarker, I. H., Kayes, A., & Badsha, S. (2024). AI-driven decision support systems in healthcare operations: A survey. *IEEE Access*, 12, 21435–21458.

Talwar, M., Jain, A., & Gupta, R. (2023). Machine learning models for hospital readmission prediction: A systematic review and meta-analysis. *Artificial Intelligence in Medicine*, 136, 102451.

Tengan, C., Ahenkan, A., & Bawole, J. N. (2021). Accountability and performance monitoring in public healthcare systems. *Public Organization Review*, 21, 623–640.

Wang, Y., & Shao, J. (2024). Multidimensional performance assessment in hospital systems: A data analytics approach. *Socio-Economic Planning Sciences*, 92, 101778.

Ziaolhagh, A., Ebrahimi, P., & Jafari, H. (2021). Balanced scorecard-based performance evaluation in healthcare organisations. *International Journal of Productivity and Performance Management*, 70(8), 1961–1980.