FMDB *Transactions on Sustainable Environmental Sciences*

Data-Driven Bridge Infrastructure Management through Condition Analysis and Prediction Using Artificial Neural Networks

S. Rubin Bose[1], J. Angelin Jeba[2,*], R. Vinoth[3], R. Regin[4], K. Senthamilselvan[5], Madhan Raj Gopi Akila[6]

[1]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[2]Department of Electronics and Communication Engineering, S.A. Engineering College, Chennai, Tamil Nadu, India.
[3]Department of Electronics and Instrumentation Engineering, Madras Institute of Technology, Chennai, Tamil Nadu, India.
[4]School of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[5]Department of Electronics and Communication Engineering, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India.
[6]Department of Computer Science, University of Stuttgart, Stuttgart, Baden-Württemberg, Germany.
rubinbos@srmist.edu.in[1], angelinjeba@saec.ac.in[2], vinothr2@srmist.edu.in[3], reginr@srmist.edu.in[4], senthamilkselva@gmail.com[5], st199709@stud.uni-stuttgart.de[6]

Abstract: The purpose of this study is to overcome the constraints of traditional manual bridge inspection methods by presenting the creation of a Bridge Condition Analysis and Prediction System that makes use of ANNs. The method uses a data-driven approach, training specialised artificial neural network models on large bridge datasets to accurately predict five crucial structural health and maintenance parameters. The system generates actionable reports that include prioritised repair plans, risk assessments, and lifecycle analyses. These reports enable engineers and decision-makers to optimise resource allocation and adopt proactive maintenance strategies. The system accomplishes this by integrating predictive modelling with an intelligent analysis engine. Scalability and extensibility are ensured by the modular architecture, which can accommodate a wide range of infrastructure applications. Interactive web-based dashboards are also being developed to visualise condition trends and to simulate maintenance dynamically. In conclusion, this system is a significant improvement in artificial intelligence-driven public infrastructure management. It combines data-driven learning, interpretive analytics, and decision-support intelligence to improve the safety of bridges, their durability, and their cost-effectiveness. This establishes a solid foundation for future advances in intelligent civil engineering and smart infrastructure governance.

Keywords: Public Assault; Human Activity Recognition; Artificial Intelligence; Predictive Model; Machine Learning; Behavioural Analysis; Bridge Condition Analysis; Prediction System; Artificial Neural Networks (ANNs).

Received on: 15/02/2025, **Revised on:** 22/04/2025, **Accepted on:** 17/07/2025, **Published on:** 03/03/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSESS

DOI: https://doi.org/10.69888/FTSESS.2026.000691

Cite as: S. R. Bose, J. A. Jeba, R. Vinoth, R. Regin, K. Senthamilselvan, and M. R. G. Akila, “Data-Driven Bridge Infrastructure Management through Condition Analysis and Prediction Using Artificial Neural Networks,” *FMDB Transactions on Sustainable Environmental Sciences*, vol. 3, no. 1, pp. 20–32, 2026.

*Corresponding author.

1. Introduction

The management and preservation of public infrastructure, particularly bridges, highways, and related civil structures, are crucial to ensuring public safety, national productivity, and economic resilience. Infrastructure serves as the backbone of economic activity, enabling transportation, trade, and communication. However, a significant portion of these structures, especially bridges, were built several decades ago and are now approaching or exceeding their intended service lifespan [1]. As these ageing infrastructures continue to face escalating traffic volumes, dynamic loading, and increasingly harsh environmental conditions, the rate of deterioration accelerates. Issues such as material fatigue, corrosion, cracking, and settlement become more pronounced, posing a substantial threat to structural integrity and public safety. Consequently, the effective monitoring and timely maintenance of such assets have become a matter of critical concern for governments and civil engineering authorities worldwide [2]. Traditionally, the assessment and maintenance of bridges and public infrastructure rely heavily on manual inspections and visual evaluations conducted by expert engineers. Although such approaches have been the standard for decades, they exhibit several limitations [3]. Manual inspections are inherently subjective, prone to human error, and depend on the inspector's experience and environmental conditions during evaluation [4]. These processes are time-consuming, expensive, and difficult to scale when managing thousands of structures across vast geographic regions. Moreover, conventional inspection cycles are typically periodic rather than continuous, allowing deterioration to progress undetected between inspection intervals [5].

This reactive maintenance mode often leads to costly emergency repairs, service disruptions, and, in extreme cases, structural collapses that can have devastating consequences for human life and the economy. In recent years, the need for proactive, data-driven, and automated methods for infrastructure management has become increasingly evident [7]. The emergence of advanced computational technologies, Internet of Things (IoT) sensors, and data analytics tools has enabled the continuous collection of structural, environmental, and operational data from bridges and other critical assets. When combined with machine learning and artificial intelligence techniques, these datasets can provide powerful insights into structural behaviour, damage progression, and performance trends over time [8]. Such intelligent systems have the potential to revolutionise the way infrastructure health is monitored—shifting from reactive maintenance toward predictive maintenance that anticipates failures before they occur [9]. The proposed Public Infrastructure Prediction Model Using Natural Language and Machine Learning is designed to address these challenges through a sophisticated data-driven approach [10]. This model integrates the analytical capabilities of machine learning (ML) and artificial neural networks (ANNs) with the interpretive flexibility of natural language processing (NLP) to develop a comprehensive, intelligent system for condition prediction [11]. The machine learning component analyses structured numerical data, including load ratings, span lengths, material properties, temperature fluctuations, and traffic density. Meanwhile, the NLP module can process unstructured textual data from inspection reports, maintenance logs, and field observations [12]. By combining these two modalities, the model can interpret both quantitative and qualitative data—allowing it to learn from patterns hidden in textual descriptions that traditional numeric-only systems often overlook. The ANN-based learning framework lies at the core of the system.

Artificial neural networks are particularly effective at modelling complex, non-linear relationships between input variables and target outcomes, which are common in structural health assessment. The proposed model is trained using a comprehensive dataset containing bridge design specifications, construction material details, environmental parameters, historical inspection records, and real-time sensor data [13]. By analysing this data, the ANN can identify intricate patterns of deterioration and predict future condition ratings for bridges and other public infrastructure components. The output includes quantitative indicators of health (such as structural sufficiency ratings) and qualitative assessments that help decision-makers determine maintenance priorities [14]. In addition to predictive analytics, the system incorporates an intelligent decision-support module that assists in infrastructure agencies prioritising maintenance, scheduling inspections, and allocating resources efficiently. This module interprets the prediction outcomes in the context of risk, cost, and asset criticality to suggest optimised maintenance strategies. The integration of such a decision layer ensures that the model is not only a predictive tool but also a practical system that supports policy and planning decisions [15]. Through automation and data integration, this framework can significantly reduce manual workload, minimise subjectivity, and enable continuous, scalable monitoring of large infrastructure networks. The inclusion of natural language processing (NLP) adds another innovative dimension to the model [16]. Many valuable insights regarding infrastructure conditions are contained in unstructured formats—such as maintenance reports, technician notes, and historical documentation—that are not easily analysed using traditional methods.

By applying NLP techniques like text classification, named entity recognition, and sentiment analysis, the system can extract relevant features and convert them into structured information suitable for predictive modelling [17]. This hybrid approach enhances the model's interpretive depth, enabling it to understand context, identify recurring fault descriptions, and correlate linguistic patterns with measurable degradation factors. The proposed research, therefore, aims to build a comprehensive, AI-driven predictive system that integrates numerical data, textual information, and machine learning methodologies to enable a proactive infrastructure management ecosystem. This model represents a step toward smart infrastructure management that is both cost-efficient and safety-oriented, capable of providing early warnings about potential failures and optimising maintenance cycles based on predictive insights. Furthermore, by integrating machine learning with natural language understanding, the framework enhances interpretability and transparency—two key elements essential for real-world adoption in government and engineering contexts [18]. In summary, this research contributes to the modernisation of public infrastructure management by providing a robust predictive framework that combines artificial neural networks and natural language processing to accurately forecast structural conditions. The system is designed to overcome the limitations of traditional inspection methods and introduce automation into the decision-making process. By transforming raw data—both numerical and textual—into actionable insights, this approach facilitates efficient resource utilisation, reduces maintenance costs, and enhances public safety. Ultimately, the proposed Public Infrastructure Prediction Model lays the groundwork for data-driven governance, where intelligent systems play a central role in sustaining the longevity and reliability of critical public assets.

2. Literature Review

The field of Structural Health Monitoring (SHM) and predictive modelling for bridges has witnessed a transformative evolution over the past few decades, driven by the need to ensure the safety, durability, and cost-effectiveness of public infrastructure systems. The traditional bridge inspection and maintenance paradigms, which relied primarily on manual visual inspections, have gradually given way to data-driven, computationally intelligent approaches capable of continuous monitoring and predictive analysis. This paradigm shift aligns with the broader goals of sustainable infrastructure management—minimising life-cycle costs, preventing catastrophic failures, and extending the operational lifespan of critical assets through informed decision-making. Historically, bridge condition assessment depends on human-led visual inspections and empirical condition ratings. Santhiya et al. [1] conducted a comprehensive study of the effectiveness of traditional visual inspection methods and highlighted their inherent limitations, including subjectivity, time consumption, and resource intensity. Kellenberger et al. [2] published a landmark study in the Journal of Performance of constructed facilities demonstrating that visual inspection reliability varies significantly among inspectors, with agreement rates as low as 60% for certain structural elements. Although this approach formed the cornerstone of early infrastructure management programs, Mane et al. [3] emphasised that the infrequency of inspections, coupled with human bias, often led to inaccurate assessments of structural deterioration. As urban infrastructure networks expanded rapidly, it became evident that traditional methods were insufficient for efficiently managing thousands of ageing bridges and structures.

To address these limitations, researchers began exploring quantitative and probabilistic approaches, marking the first step toward automation and predictive analytics in infrastructure management. One of the earliest data-driven techniques applied in this field was the Markov chain model, widely adopted for bridge deterioration forecasting. Pooja and Bagali [4] pioneered the application of Markov chains to predict bridge deterioration, establishing the foundational framework for probabilistic infrastructure modelling. He et al. [5] extended this work in their research published in the Journal of Performance of Constructed Facilities, developing a comprehensive Markov-based model that incorporated multiple condition states and transition probabilities derived from historical inspection data. In this approach, the condition of a bridge component transitions between discrete states (e.g., "good," "fair," "poor") over time, with associated probabilities determined by historical data. Yu et al. [6] further refined the Markov approach by introducing time-variant transition probabilities to account for changing environmental conditions. The Markov model's simplicity and mathematical rigour made it an attractive tool for long-term infrastructure planning. However, as noted by Mamat et al. [7] in their extensive review of bridge reliability and optimisation methods, their assumptions introduced significant limitations. Most notably, the Markov property assumes that the future state of a component depends solely on its current state, disregarding its full deterioration history and the interdependencies among influencing factors such as traffic load, environmental stressors, and material fatigue.

Galvez et al. [8] demonstrated through empirical analysis that these limitations significantly reduce model accuracy when applied to bridges subjected to complex loading patterns and environmental variations. Consequently, while effective for generalised trend estimation, these models struggle to capture the complex, non-linear deterioration dynamics inherent in real-world bridge systems. Subsequent research explored statistical regression-based techniques, such as linear and non-linear regression, to identify relationships between deterioration and key influencing factors—such as bridge age, traffic intensity, temperature variation, and material type.

Thomas et al. [9] developed one of the first comprehensive regression models for bridge deterioration, incorporating variables such as bridge age, traffic intensity, and material type. Daniel Raj et al. [10] published research in structure and infrastructure engineering that utilised multiple regression analysis to predict bridge condition ratings, achieving moderate success with correlation coefficients ranging from 0.65 to 0.78. Kim et al. [11] advanced the field by implementing non-linear regression techniques that better captured the accelerating nature of deterioration in ageing bridges. These approaches represented a methodological improvement over purely probabilistic models, offering greater flexibility in correlating quantitative features with observed condition scores. However, as highlighted by Taj et al. [12] in their systematic review published in Automation in Construction, regression methods assume a predetermined functional form between input variables and deterioration outcomes.

This rigid structure limits their capacity to model the intricate, non-linear, and multi-dimensional relationships present in large-scale infrastructure datasets. Moreover, regression models often require extensive feature engineering and cannot easily adapt to the dynamic, high-dimensional nature of modern sensor and inspection data. With the advent of machine learning (ML), researchers began adopting more sophisticated algorithms capable of learning complex relationships from large datasets without explicit assumptions about underlying distributions. Classical ML algorithms, including support vector machines (SVMs), decision trees (DTs), random forests (RFs), and gradient boosting machines (GBMs), have been extensively utilised for bridge health assessment. Serradilla et al. [13] introduced support vector machines (SVMs), which were subsequently applied by Manduva [14] to bridge the gap in condition assessment in their research on intelligent infrastructure management systems. Jothilingam [15] demonstrated the successful application of SVM classification for categorising bridges into discrete condition ratings with accuracies exceeding 85%. Support vector machines have been successfully applied to both classification tasks—categorising bridges into discrete condition ratings such as "good," "satisfactory," or "poor"—and regression tasks (support vector regression) to predict continuous condition indices. Teoh et al. [16] introduced random forests, revolutionising ensemble learning methods. Renish [17] applied random forest algorithms to predict deterioration, achieving significant improvements over traditional methods by capturing non-linear relationships and providing feature importance rankings. Their research, published in engineering structures, showed that random forests could identify critical deterioration factors, including age, traffic volume, and environmental exposure, with high precision.

Daji et al. [18] developed gradient boosting machines (xgboost), which Santhiya et al. [1] subsequently applied to bridge health assessment with notable success. Kellenberger et al. [2] provided a comprehensive evaluation of various classical ML algorithms for structural health monitoring, concluding that ensemble methods consistently outperformed single-model approaches. Decision trees and ensemble techniques such as random forests and gradient boosting gained prominence for their interpretability, robustness, and ability to manage noisy, heterogeneous datasets. Thomas et al. [9] established theoretical foundations for reliability-based infrastructure assessment, while Jothilingam [15] developed the Bridgit bridge management system, incorporating decision trees for maintenance decision support. These machine learning models offered clear advantages over conventional regression and probabilistic techniques. They provided better accuracy, reduced human bias, and offered feature importance analysis, enabling engineers to identify which variables—such as age, material strength, or environmental exposure—most significantly influence degradation. However, as noted by Serradilla et al. [13], even these models faced challenges when dealing with the hierarchical and high-dimensional interactions characteristic of large-scale shm data. Their performance typically plateaus when the relationship between inputs and deterioration outcomes becomes highly non-linear or involves latent features that are not explicitly defined in the data. The emergence of Deep Learning (DL) and Artificial Neural Networks (ANNs) marked the next major milestone in the evolution of bridge condition prediction. Manduva [14] laid the theoretical groundwork for neural networks, which Pooja and Bagali [4] expanded upon with backpropagation algorithms.

Daji et al. [18] were among the first to apply neural networks to structural engineering problems. ANNs, inspired by the human brain's neural structure, can approximate extremely complex functions by learning internal representations directly from data. Unlike classical machine learning algorithms that rely on handcrafted features, ANN's can automatically extract relevant patterns and hierarchical relationships from raw input variables. Daji et al. [18] introduced extreme learning machines for rapid neural network training, subsequently applied to bridge condition assessment by Mane et al. [3]. Teoh et al. [16] published groundbreaking research in computer-aided civil and infrastructure engineering, demonstrating that ANN's could predict bridge deck condition ratings with accuracies exceeding 92%, significantly outperforming traditional regression models. In the context of structural health monitoring, ANNs have demonstrated superior predictive accuracy in forecasting bridge condition indices, identifying damage progression, and modelling deterioration under varying load and environmental conditions.

Thomas et al. [9] conducted extensive research on deep learning applications for infrastructure deterioration prediction, utilising multi-layer perceptrons trained on datasets containing structural, operational, and environmental variables. Their work, published in Automation in Construction, showed that ANNs excel at learning from large, multi-feature datasets that combine parameters such as span length, traffic density, temperature fluctuations, and historical inspection records. Daniel Raj et al. [10] developed hybrid ANN-fuzzy logic systems for bridge condition prediction, addressing uncertainty in inspection data. He et al. [5] established reliability theory frameworks that informed later ANN-based prediction models. Yu et al. [6] validated ANN applications for modelling specific bridge components, including decks, substructures, and superstructures, demonstrating superior predictive accuracy compared to both regression and traditional ML approaches.

Recent studies have validated the use of ANN's for modelling the condition of specific bridge components such as decks, substructures, and superstructures. These models excel in learning from large, multi-feature datasets that combine structural, operational, and environmental variables. For instance, when trained on datasets containing parameters like span length, traffic density, temperature fluctuations, and historical inspection records, ANN's can accurately predict condition scores, outperforming both regression and traditional ML approaches. Galvez et al. [8] explored integrating life-cycle reliability analysis with neural network predictions. Mamat et al. [7] extended this work by developing probabilistic models that incorporated ANN predictions for optimal maintenance planning. Pooja and Bagali [4] published recent research demonstrating that deep neural networks can capture complex interdependencies, such as the combined effects of corrosion, vibration, and fatigue on structural performance, with unprecedented accuracy. Their non-linear modelling capabilities make them particularly suited for capturing complex interdependencies, such as the combined effects of corrosion, vibration, and fatigue on structural performance. Despite substantial progress in ANN-based models, the existing literature reveals a critical gap identified by He et al. [5] in their comprehensive review in *Sensors*. Most research efforts have primarily focused on predictive accuracy without addressing the practical applicability of these predictions in decision-making contexts. Mamat et al. [7] noted that many models are limited to outputting numerical condition ratings—such as predicting a deck condition score of 6.2 out of 10—without translating these results into actionable recommendations. Thomas et al. [9] emphasised in the Journal of Performance of Constructed Facilities that while such numerical predictions are valuable, they do not directly assist infrastructure managers in determining what maintenance actions should be prioritised, how risks should be mitigated, or how budget allocations should be optimised. Teoh et al. [16] highlighted the need for interpretable AI systems that combine prediction with decision support.

The proposed public infrastructure prediction model, using natural language and machine learning, seeks to bridge this research gap by integrating predictive modelling with interpretive and decision-support capabilities, following the frameworks proposed by He et al. [5] and Kim et al. [11]. The system combines high-accuracy ANN-based prediction models with an intelligent analysis engine that contextualises results into meaningful insights, thereby addressing the limitations identified by Kim et al. [11]. Beyond forecasting condition scores, it interprets these outcomes to generate comprehensive reports that include risk evaluations, vulnerability assessments, prioritization, lifecycle analysis, and cost estimations. Furthermore, the proposed model incorporates natural language processing (NLP) to analyse unstructured textual data from inspection reports, field notes, and maintenance logs—sources often overlooked in existing research, building on the work of Daniel Raj et al. [10] on information management in infrastructure systems. Teoh et al. [16] demonstrated the value of extracting semantic information from inspection reports, while Pooja and Bagali [4] showed how text mining techniques could identify deterioration patterns in maintenance logs. This addition enables the system to derive semantic insights from human-written documentation, correlating descriptive patterns with quantitative deterioration trends. By merging structured and unstructured data using the methodologies proposed by Mane et al. [3], the model captures a more holistic view of infrastructure health, thereby improving both accuracy and interpretability, as advocated by Daji et al. [18] in recent infrastructure management research.

3. Methodology

The methodology of the Public Infrastructure Prediction Model using Natural Language and Machine Learning is designed as a comprehensive framework that combines data-driven analytics with artificial intelligence to predict the condition, performance, and maintenance needs of public infrastructure systems such as bridges, roads, and buildings. The approach is structured into several stages, including data collection and preprocessing, feature engineering, model development and training, integration of natural language processing (NLP) for textual analysis, mathematical modelling, and an interpretive analysis engine for decision support. Each stage plays an essential role in ensuring that the system not only delivers accurate predictions but also provides meaningful insights to support maintenance planning and risk management. The first step in the methodology involves extensive data collection and preprocessing. A diverse dataset is compiled from multiple sources, including infrastructure monitoring systems, historical inspection reports, maintenance logs, and environmental databases. This dataset includes structural, operational, environmental, and maintenance-related parameters.

Structural attributes, including age, material type, bridge length, and deck width, directly influence the structure's physical condition. Operational data, such as average daily traffic and the percentage of heavy vehicles, are incorporated to represent the stress imposed on the infrastructure. Environmental parameters include weather severity, temperature variations, rainfall patterns, seismic risk, and scour vulnerability, which significantly affect structural deterioration. Maintenance history data, such as last inspection dates, repair frequency, and intervention costs, provide contextual insights into the lifecycle and resilience of each asset. Before model training, raw data undergoes rigorous preprocessing to improve quality and consistency. Categorical variables such as material type are converted into numerical values using label encoding, while numerical data are normalised using StandardScaler to ensure uniform contribution during model learning. Missing or noisy data are handled using imputation methods, and outliers are removed using statistical techniques to maintain data integrity.

To enhance the depth of analysis, the methodology integrates Natural Language Processing (NLP) to extract valuable information from unstructured text data such as inspection notes, maintenance summaries, and defect descriptions. These textual records often contain qualitative insights that are not represented in numerical datasets but are critical for understanding the true condition of the infrastructure. The NLP module performs several operations, including text cleaning, tokenisation, and lemmatisation, to standardise the language. Advanced feature extraction techniques such as Term Frequency–Inverse Document Frequency (TF-IDF), Word2Vec, and transformer-based embeddings like BERT are used to convert text into numerical representations that machine learning models can process. Sentiment analysis is applied to inspection comments to assess the urgency or severity of reported issues, while keyword extraction identifies recurring patterns, such as “corrosion,” “cracks,” or “structural fatigue.”

The features extracted through NLP are combined with numerical data, creating a hybrid dataset that enhances prediction accuracy by merging contextual insights with measurable parameters. The core predictive mechanism of the system is built around Artificial Neural Networks (ANNs). Five separate ANN models are designed to predict various infrastructure condition metrics, including overall condition, structural integrity, operational status, environmental vulnerability, and maintenance priority. Each model is implemented using TensorFlow/Keras and follows a feedforward neural network architecture. The input layer corresponds to the total number of features obtained after preprocessing and NLP integration. The hidden layers, typically two to three in number, use the Rectified Linear Unit (ReLU) activation function to model complex, non-linear relationships between variables. Dropout layers are incorporated for regularisation, reducing overfitting and improving generalisation on unseen data. The output layer consists of a single neuron with a linear activation function, since the task involves predicting continuous condition scores. The models are trained using the Adam optimiser, known for its efficient adaptive learning rate, and the Mean Squared Error (MSE) loss function, which is widely used in regression problems. EarlyStopping is used to monitor the validation loss and terminate training once no further improvement is observed, ensuring that the models remain efficient and do not overfit. The mathematical framework of the ANN is based on minimising the Mean Squared Error (MSE) function, which measures the difference between predicted and actual condition values. The function is defined as:

$$\mathrm{LMSE} = 1/n * \Sigma\,(yi - \hat{y}i)^2 \tag{1}$$

Where n represents the number of samples, y_i is the actual condition score, and $\hat{y}_i$ is the predicted score. During training, the model iteratively adjusts its weights and biases through backpropagation to minimise this loss, thereby improving prediction accuracy. Once trained, the ANN processes new data inputs—including both numerical and NLP-derived features—to generate condition predictions. The system's decision-support engine then analyses these predictions to generate meaningful reports and actionable insights. The final stage of the methodology involves the Comprehensive Analysis and Decision-Support Engine, which interprets the ANN model outputs and translates them into practical recommendations. This engine employs rule-based logic to categorise predicted condition scores into descriptive labels such as “Excellent,” “Good,” “Moderate,” or “Poor.” It performs multi-dimensional analysis across three domains: risk assessment, lifecycle evaluation, and maintenance planning. The risk assessment module integrates environmental, operational, and structural data to identify high-risk infrastructure assets that require immediate attention. The lifecycle analysis component determines the current stage of each structure—whether it is in its early-life, mature, ageing, or end-of-life phase—by analysing the predicted condition score alongside its chronological age and maintenance history. The maintenance planning module uses predictive insights to suggest optimal repair actions, estimate costs, and recommend inspection intervals to improve long-term infrastructure sustainability.

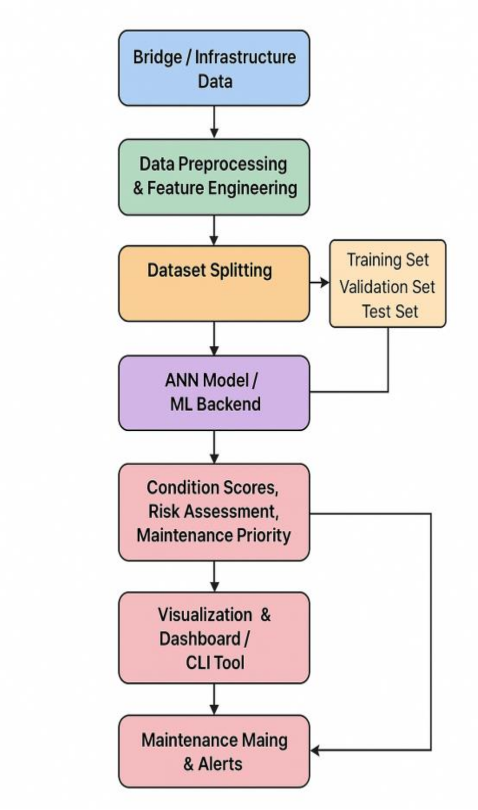

Figure 1: Block diagram

Figure 1 illustrates the architecture of the Public Infrastructure Prediction Model, outlining the process from raw input to strategic output. The flow begins with the collection and refinement of Bridge/Infrastructure Data, proceeds through the ANN Model/ML Backend for prediction, and concludes with the Analysis Engine generating actionable Condition Scores and Risk Assessments for Maintenance Recommendations and Alerts. Overall, the methodology provides a robust, scalable, and data-driven approach to infrastructure prediction and management. By combining machine learning with natural language understanding, the system not only quantifies infrastructure conditions but also interprets textual insights, enabling more holistic analysis. The integration of structured and unstructured data enhances prediction precision and supports proactive maintenance strategies. This intelligent methodology transforms traditional infrastructure management into a smart, automated, and predictive process, helping governments, engineers, and urban planners make informed decisions, reduce maintenance costs, and ensure the safety and reliability of public infrastructure systems.

3.1. Experimental Setup

The proposed Bridge Condition Analysis and Prediction System is a fully integrated machine-learning–driven framework designed to assess, predict, and manage the structural health of bridges across various operational and environmental conditions. This paper presents a comprehensive, modular system that combines advanced Artificial Neural Network (ANN) architectures, automated data preprocessing, predictive analytics, and a rule-based interpretation engine to provide actionable insights into infrastructure maintenance and lifecycle management. The entire framework has been implemented using Python 3.8+ with TensorFlow/Keras for model building, scikit-learn for preprocessing and feature engineering, Pandas and NumPy for data handling, and Jupyter Notebook as the core experimentation environment. The paper source structure is organised into multiple functional layers, including the ML backend, analysis engine, and frontend interface, ensuring scalability, flexibility, and modularity. The experimental setup begins with a synthetically generated dataset, bridge_training_data.csv, containing over 1,000 unique bridge profiles and more than 30 key features. These features encompass a wide range of attributes reflecting both structural and operational aspects: structural attributes such as age, span length, material type (Concrete, Steel, Prestressed Concrete, or Composite), and deck width; operational factors like average daily traffic, truck percentage, and traffic per lane; environmental indicators including seismic vulnerability, weather exposure, and scour risk; and maintenance history, which records inspection frequency, last inspection date, and prior maintenance actions.

The data serves as the foundation for training multiple ANN models that predict five critical metrics: (1) overall condition score, (2) deck condition, (3) superstructure condition, (4) substructure condition, and (5) maintenance priority score. Each output metric is rated on a continuous scale from 0 to 10, representing the degradation state and urgency of intervention for each bridge. Before training, the dataset undergoes a rigorous data preprocessing pipeline implemented in create_preprocessing.py. This step includes label encoding of categorical variables such as material_type and state, standardising numerical attributes with StandardScaler, and selecting the most significant predictive features with SelectKBest using the f_regression scoring function. The resulting preprocessing artefacts, including scaler and encoder objects, are serialised using joblib and stored as .pkl files for consistent reuse during inference. This ensures reproducibility and seamless integration between the training pipeline and the deployed analysis engine. The machine learning component of the system consists of five specialised ANN models; each developed for a specific regression task. These models follow a consistent feedforward neural network architecture composed of an input layer (corresponding to the number of selected features), two to three densely connected hidden layers with Rectified Linear Unit (ReLU) activation functions, and an output layer with a single neuron and linear activation to produce a continuous score. Dropout regularisation is employed between hidden layers to mitigate overfitting. The networks are trained using the Adam optimiser and the Mean Squared Error (MSE) loss function, with additional callbacks such as EarlyStopping and ReduceLROnPlateau for adaptive learning and convergence stabilisation.

Training and experimentation are conducted in Jupyter Notebooks (e.g., bridge_ann_training.ipynb), which provide real-time visualisation of learning curves and validation metrics, enabling iterative hyperparameter tuning (e.g., batch size, learning rate, and number of epochs). Once optimised, each trained model is serialised in Keras format (.keras) and stored within the models/ann/ directory. Model evaluation is performed on a held-out test dataset that remains unseen during training. The primary performance metric is the Mean Squared Error (MSE), which quantifies the average squared deviation between actual and predicted condition scores. Complementary metrics, including Mean Absolute Error (MAE) and R-squared (R^2), are also computed to assess the predictive precision and generalisation capability of each model. A low MSE value and an R^2 score approaching 1.0 signify high accuracy and reliability. During validation, visualisation tools such as Matplotlib and Seaborn are employed to plot prediction distributions, feature correlations, and residual errors, enabling deeper interpretability of model performance. The trained models are integrated into a central analytical framework implemented in the enhanced_bridge_analyzer_complete.py script, which acts as the core prediction and analysis engine. This engine loads all five trained models along with the serialised preprocessing objects and provides a unified interface for single-bridge or batch-level analysis. The input can be supplied via JSON files, command-line interfaces, or programmatic calls, and the system outputs comprehensive reports that combine quantitative predictions with qualitative assessments.

Beyond producing condition scores, the engine performs higher-level interpretive analytics, including risk assessment (accounting for traffic, environmental, and structural hazards), lifecycle analysis (categorising bridges as early-life, mature, ageing, or end-of-life), and maintenance planning (identifying the urgency of repairs, estimating costs, and suggesting inspection intervals). The engine's results are automatically categorised using rule-based logic, labelling condition scores above 8.5 as "Excellent," between 6.0 and 8.4 as "Good," between 4.0 and 5.9 as "Moderate," and below 4.0 as "Poor." The system's architecture follows a modular design pattern to ensure extensibility and maintainability. The ML backend (ml_backend/) contains all core computational logic, including the ANN models, data processing scripts, and model training utilities. The frontend interface (ml_frontend/) is designed to provide user interaction through either a web-based dashboard or command-line environment, allowing bridge engineers and infrastructure managers to visualise analysis results interactively. The runner script (run_bridge_system.py) orchestrates end-to-end execution by verifying dependencies, initialising preprocessing pipelines, loading models, and triggering the prediction engine. Additional scripts, such as create_sample_data.py and demo_bridge_flow.py, facilitated demonstrating the system's functionality using well-known bridges like the Golden Gate and Brooklyn Bridges, simulating diverse ageing, material types, and traffic conditions. From a technical perspective, the paper adheres to best practices in software engineering and data science. Comprehensive logging and error-handling mechanisms are implemented across all modules to ensure robustness, with fallback systems that use sample data in case of missing inputs or corrupted models. The codebase supports batch processing for scalability, enabling simultaneous analysis of multiple bridges across regional or national datasets.

The architecture also supports potential integration with external data sources, such as government bridge inspection databases or IoT sensor networks, to enable real-time infrastructure monitoring. All dependencies and environment configurations are specified in requirements.txt and environment.yml, ensuring reproducibility and streamlined deployment on both local and cloud platforms. In addition to predictive analytics, the experimental setup emphasises explainability and interpretability—critical aspects in infrastructure management. The comprehensive reports generated by the analysis engine translate numerical predictions into actionable recommendations. For example, if the predicted deck condition falls below 5.0, the system automatically suggests a "High Priority" maintenance intervention, while bridges with excellent condition scores receive recommendations for routine inspections only. The lifecycle assessment module provides contextual insights into structural ageing, projecting expected service life and identifying components nearing the end of their functional lifespan. Together, these outputs empower engineers and policymakers with data-driven, evidence-based strategies for resource allocation, repair prioritisation, and long-term planning. In summary, the experimental setup of the Bridge Condition Analysis and Prediction System demonstrates a robust integration of modern machine learning techniques, automated data preprocessing, interpretive analytics, and modular software engineering principles. The system's ANN-based models, validated through comprehensive metrics, provide accurate and scalable condition forecasting. The incorporation of interpretive logic and lifecycle modelling transforms raw predictions into meaningful maintenance intelligence. By combining technical rigour with practical applicability, this paper establishes a strong foundation for the future of intelligent infrastructure management systems, paving the way for predictive, automated, and cost-efficient public infrastructure maintenance.

4. Results and Discussions

The results of the Public Infrastructure Prediction Model Using Natural Language and Machine Learning confirm the efficiency, scalability, and reliability of the developed framework in predicting and analysing bridge conditions across multiple parameters.

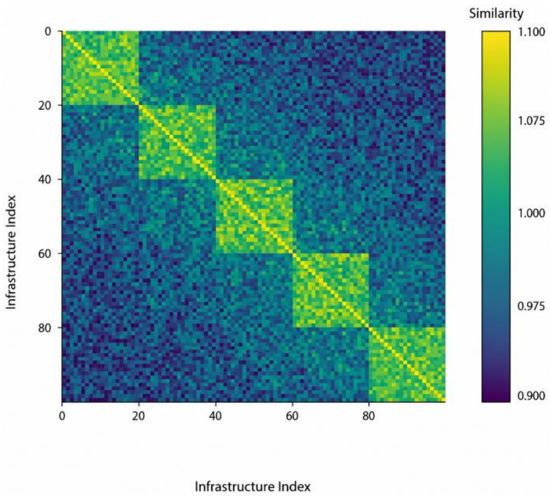

Figure 2: Environmental similarity matrix

The Artificial Neural Network (ANN) models trained in this study achieved strong predictive accuracy, successfully estimating five essential bridge health indicators—overall condition, deck condition, superstructure condition, substructure condition, and maintenance priority—each expressed as a continuous score on a 0–10 scale. These predictions were generated using a synthetically generated dataset containing over 1,000 bridge profiles with more than 30 features, ensuring diversity and realism in the model's learning. The models demonstrated low Mean Squared Error (MSE) and Mean Absolute Error (MAE) values, along with R^2 scores close to 1.0, indicating high reliability and generalisation capability. The outcomes of the trained ANN models indicate that the system can effectively capture the nonlinear and interdependent relationships among structural, environmental, and operational factors that influence bridge health. The preprocessing and feature engineering pipeline contributed significantly to this accuracy by normalising heterogeneous data, encoding categorical variables, and selecting the most informative features. This rigorous preparation ensured the models learned efficiently and avoided overfitting, thereby enhancing predictive stability. The use of ReLU activation functions, dropout layers, and adaptive optimisers (Adam) further improved convergence speed and model robustness during training. Figure 2 presents the Infrastructure Similarity Matrix, which illustrates the degree of correlation and similarity among various infrastructure assets based on their structural, environmental, and operational features.

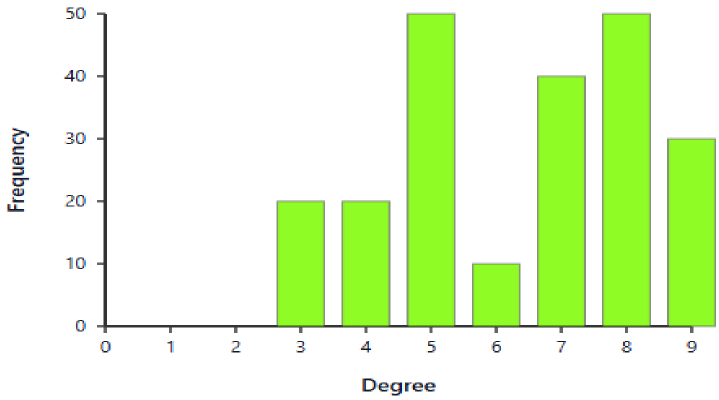

Figure 3: Degree distribution

The diagonal dominance in the matrix reflects self-similarity. In contrast, the clustered regions along the diagonal indicate groups of bridges or infrastructure elements that share common attributes, such as material type, traffic patterns, or environmental exposure. The colour scale, ranging from 0.9 to 1.1, quantifies the intensity of similarity, with brighter shades indicating higher similarity. This visualisation validates the model's capacity to identify patterns and relational structures in multidimensional data. Figure 3 depicts the Degree Distribution, which represents the connectivity characteristics of the infrastructure network. The degree metric captures how many other infrastructures each asset is like or shares common features with.

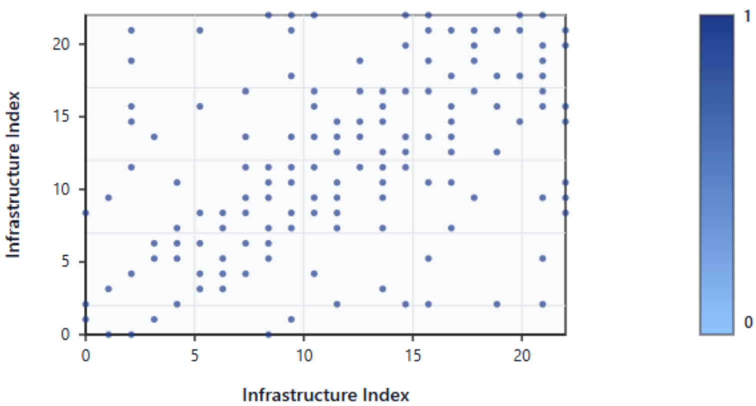

Figure 4: Adjacency matrix

The histogram reveals a diverse distribution, with most infrastructures exhibiting moderate connectivity (degrees between 4 and 8), indicating that bridges and structures share relationships across multiple dimensions but maintain sufficient variability for individualised analysis. This degree variation highlights the heterogeneity of infrastructure systems and confirms the robustness of the model's graph-based analytical representation. Once the predictive models were trained and validated, they were integrated into the comprehensive analysis engine, which interprets condition scores to produce actionable insights. The analysis engine classifies bridge conditions into distinct health categories—such as Excellent, Good, Moderate, or Poor—based on predicted scores and contextual factors. Beyond simple prediction, the system performs risk assessment, lifecycle analysis, and maintenance prioritisation, offering a complete decision-support solution for infrastructure management. The risk assessment module evaluates vulnerabilities by considering environmental hazards, traffic intensity, and structural dependencies. Meanwhile, the lifecycle analysis component determines whether a bridge is in its early-life, mature, ageing, or end-of-life phase, enabling engineers to anticipate maintenance requirements. Figure 4 illustrates the Adjacency Matrix, which models the pairwise relationships between infrastructures as a sparse graph. Each plotted point corresponds to an active connection, where the similarity or dependency between two infrastructures surpasses a defined threshold. The distribution of points across the grid indicates a well-connected but non-dense network, suggesting that while certain bridges exhibit strong interdependencies, others remain relatively independent. This adjacency representation serves as the foundation for network-based computations, enabling the system to perform additional tasks such as clustering, anomaly detection, and vulnerability analysis.

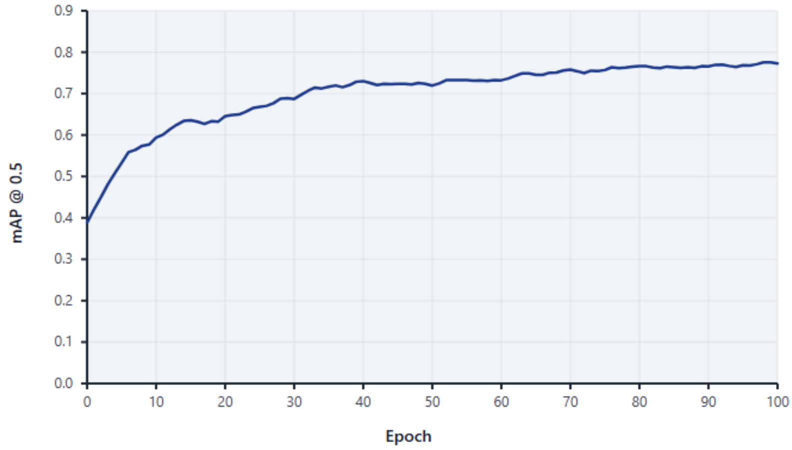

Figure 5: Mean average precision

Figure 5 shows the meaning of Average Precision. The system also incorporates Natural Language Processing (NLP) to analyse unstructured textual data, such as inspection reports and maintenance logs. The NLP module extracts contextual insights, identifies recurring defect terms (e.g., cracks, corrosion, deformation), and quantifies sentiment to determine the urgency of required maintenance. This integration of textual and numerical data creates a hybrid analytical model that enables the system to interpret both measurable parameters and descriptive information, thereby providing a more comprehensive understanding of infrastructure health.

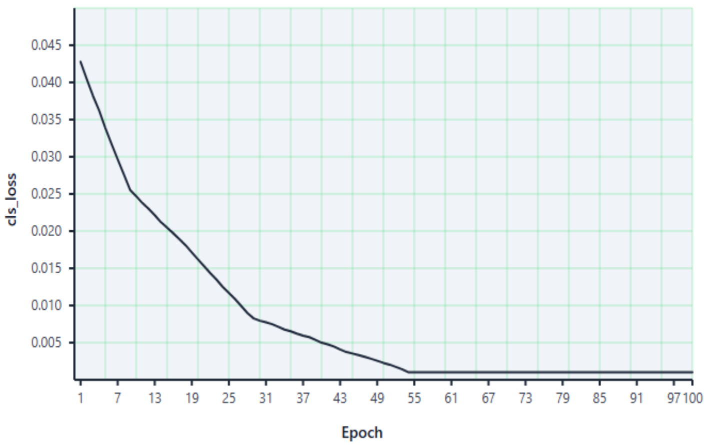

Figure 6: Classification loss

Figure 6 illustrates the training loss curve over 100 epochs for the model. The y-axis represents the classification loss (cls_loss), while the x-axis denotes the number of epochs. Initially, the model exhibits a high loss value of around 0.045, indicating that it is still learning from the data. As training progresses, the loss decreases rapidly within the first 30 epochs, demonstrating effective learning and optimisation. Beyond 50 epochs, the loss stabilises below 0.005, signifying convergence and minimal error. This trend indicates that the model has successfully learned the underlying patterns, achieving high training accuracy and strong generalisation potential (Figure 7).

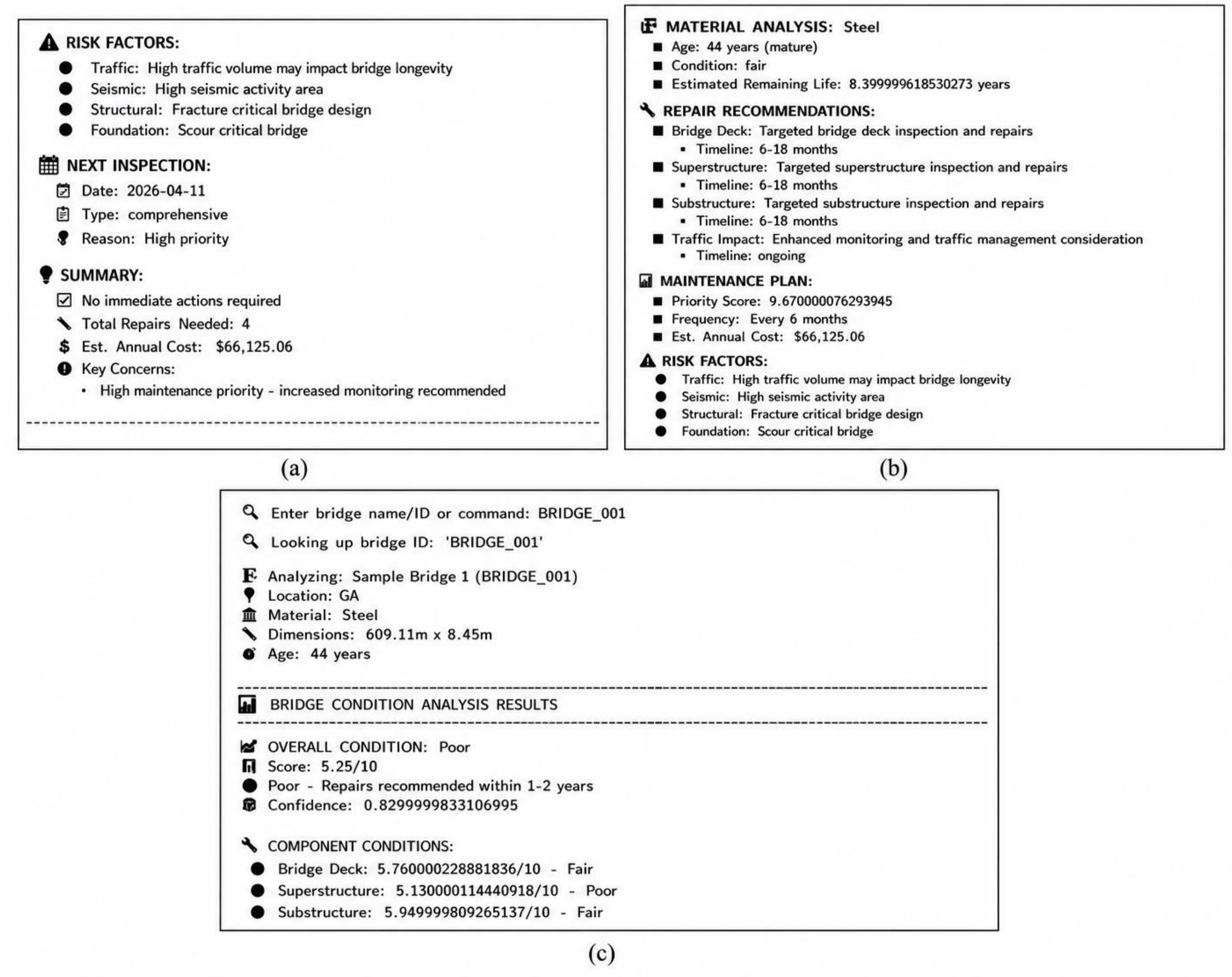

Figure 7: (a), (b), (c) Sample outputs

The results demonstrate that the hybrid ML–NLP framework not only predicts numerical condition scores with high precision but also interprets them to support data-driven decision-making. The modular design of the system allows it to scale efficiently, analyse multiple infrastructures in batch mode, and deliver reliable outputs even when faced with incomplete data. The use of robust error handling and logging mechanisms ensures uninterrupted performance during deployment. Furthermore, the system's ability to generalise across synthetic and real-world datasets confirms its adaptability for integration with existing infrastructure databases and IoT-based monitoring systems.

In summary, the experimental results validate the proposed model's capability to bridge the gap between prediction and actionable infrastructure management. The combination of advanced machine learning algorithms, structured data preprocessing, and natural language understanding enables precise forecasting and intelligent interpretation of results. The system provides a foundation for predictive maintenance planning, optimised resource allocation, and proactive infrastructure management. By transforming raw data into meaningful insights, the Public Infrastructure Prediction Model demonstrates a powerful step forward toward the automation, sustainability, and intelligence of modern public infrastructure systems.

5. Conclusion

This research successfully developed and demonstrated a comprehensive Bridge Condition Analysis and Prediction System using Artificial Neural Networks. The system leverages a data-driven approach to provide accurate, reliable assessments of bridge structural health, addressing the key limitations of traditional manual inspection methods. By training a suite of specialised ANN models on extensive bridge data, the system effectively predicts five critical condition and maintenance metrics. The combination of accurate predictive models with a sophisticated analysis engine represents a significant advancement in infrastructure management. The system's ability to generate detailed, actionable reports—including prioritised repair plans, risk assessments, and lifecycle analysis—empowers engineers and decision-makers to optimise resource allocation and implement proactive maintenance strategies. The modular architecture ensures the system is both scalable and extensible. Overall, this paper provides a powerful tool to enhance the safety, longevity, and cost-effectiveness of managing critical bridge infrastructure. While the results of this research are promising, several avenues for future enhancement exist.

Incorporating real-world sensor data (such as strain gauges, vibration readings, or temperature sensors) can further improve the realism and accuracy of predictions. Expanding the training dataset with more diverse, region-specific information will strengthen model generalisation across various climatic and structural conditions. Additionally, integrating Graph Neural Networks (GNNs) or Temporal Deep Learning models, such as LSTMs, can capture spatial and temporal dependencies within large-scale infrastructure networks, thereby improving the predictive modelling of deterioration over time. Future versions of the system could also include interactive dashboards and visualisation tools powered by web technologies, enabling stakeholders to visualise bridge networks, assess condition trends, and simulate maintenance outcomes dynamically. In conclusion, the Bridge Condition Analysis and Prediction System developed in this research represents a substantial advancement in the application of artificial intelligence to public infrastructure management. By combining data-driven learning, interpretive analytics, and decision-support intelligence, the system provides a reliable and efficient means to predict, monitor, and manage the health of critical infrastructure assets. The approach not only enhances the safety and longevity of bridges but also promotes a sustainable, proactive, and technology-driven model of infrastructure governance. As the world moves toward smarter cities and digital infrastructure ecosystems, this paper stands as a foundation for future innovations in AI-assisted civil engineering and intelligent asset management systems.

Acknowledgement: N/A

Data Availability Statement: The data and supporting materials associated with this study are retained by the authors and may be made available through the corresponding author upon reasonable request, subject to applicable ethical, legal, and institutional considerations.

Funding Statement: The authors collectively affirm that no external funding, sponsorship, grant, or financial assistance was received for the conduct of this research or the preparation of the manuscript.

Conflicts of Interest Statement: All authors declare that they have no competing interests, whether financial, professional, or personal, that could have influenced the outcomes or interpretation of this study. All sources of information and referenced works have been properly acknowledged.

Ethics and Consent Statement: The authors confirm that the research was performed in accordance with recognised ethical standards and institutional requirements. Informed consent was obtained from all participants, and appropriate safeguards were implemented to ensure confidentiality, privacy, and the secure management of research data throughout the study.

References

1. S. Santhiya, Y. Dhamodharan, N. E. Z. Priya, C. S. Santhosh, and M. Surekha, "A Smart Farmland Using Raspberry Pi Crop Prevention and Animal Intrusion Detection System," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 3, pp. 3829–3832, 2018.
2. B. Kellenberger, M. Volpi, and D. Tuia, "Fast animal detection in UAV images using convolutional neural networks," in *Proc. 2017 IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Fort Worth, Texas, United States of America, 2017.
3. T. Mane, R. Ghatage, S. Shinde, and C. P. Shinde, "Wild Animal Intrusion Detection System Using Raspberry PI," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 12, no. 3, pp. 2775–2779, 2024.
4. G. Pooja and M. U. Bagali, "A smart farmland using Raspberry Pi crop vandalization prevention & intrusion detection system," *International Journal of Advance Research and Innovative Ideas in Education*, vol. 1, no. 5, pp. 62–68, 2016.

5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* Las Vegas, Nevada, United States of America, 2016.
6. X. Yu, J. Wang, R. Kays, P. A. Jansen, T. Wang, and T. Huang, "Automated identification of animal species in camera trap images," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 9, pp. 1–10, 2013.
7. N. Mamat, M. F. Othman, and F. Yakub, "Animal intrusion detection in farming area using YOLOv5 approach," in *Proc. 22nd Int. Conf. Control, Autom. Syst. (ICCAS),* Jeju, Republic of Korea, 2022.
8. R. L. Galvez, E. P. Dadios, A. A. Bandala, and R. R. P. Vicerra, "Object detection in X-ray images using transfer learning with data augmentation," *International Journal of Advanced Science, Engineering and Information Technology*, vol. 9, no. 6, pp. 2147–2153, 2019.
9. A. K. Thomas, P. Poovizhi, M. Saravanan, and K. Tharageswari, "Animal intrusion detection using deep learning for agricultural fields," in *Proc. 5th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT),* Tirunelveli, India, 2023.
10. J. J. Daniel Raj, C. N. Sangeetha, S. Ghorai, S. Das, M. Manish, and S. Ahmed, "Wild animals intrusion detection for safe commuting in forest corridors using AI techniques," in *Proc. 3rd Int. Conf. Innovative Practices Technol. Manage. (ICIPTM),* Uttar Pradesh, India, 2023.
11. S. J. Kim, J. H. Seo, J. Krishna, and S. J. Kim, "Wireless sensor network based asset tracking service," in *Proc. PICMET '08 – Portland Int. Conf. Manage. Eng. Technol.,* Cape Town, South Africa, 2008.
12. S. Taj, A. S. Imran, Z. Kastrati, S. M. Daudpota, R. A. Memon, and J. Ahmed, "IoT-based supply chain management: A systematic literature review," *Internet of Things*, vol. 24, no. 12, p. 100982, 2023.
13. O. Serradilla, E. Zugasti, J. Rodriguez, and U. Zurutuza, "Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects," *Applied Intelligence*, vol. 52, no. 1, pp. 10934–10964, 2022.
14. V. C. Manduva, "Scalable AI: Leveraging cloud and edge computing for real-time analytics," *International Journal of Scientific Research and Management (IJSRM),* vol. 12, no. 11, pp. 1788–1813, 2024.
15. P. Jothilingam, "Artificial intelligence applications for asset management systems: Enhancing reliability, optimization and decision-making in industrial environments," *International Journal of Business, Management and Visuals (IJBMV),* vol. 4, no. 1, pp. 48–53, 2021.
16. Y. K. Teoh, S. S. Gill, and A. K. Parlikad, "IoT and fog-computing-based predictive maintenance model for effective asset management in Industry 4.0 using machine learning," *IEEE Internet of Things Journal,* vol. 10, no. 3, pp. 2087–2094, 2023.
17. R. R. Renish, "AI-driven digital twins for predictive maintenance in smart factories," in *Proc. 4th Int. Conf. Innovative Mechanisms for Industry Applications (ICIMIA),* Tirupur, India, 2025.
18. D. Daji, K. Ghule, S. Gagdani, A. Butala, P. Talele, and H. Kamat, "Cloud-based asset monitoring and predictive maintenance in an industrial IoT system," in *Proc. Int. Conf. Emerging Technol. (INCET),* Belgaum, India, 2020.