# Hybrid XGBoost–LSTM Model for Structural Health Monitoring of Reinforced Concrete Beams

Kotharu Srinivasa Rao[1]; Narisetty Laxmipriya[2]; Velivela Gopinath[3]

[1]Lecturer, Sir C. R. Reddy Polytechnic, Eluru
[2,3] Assistant Professor, Vasireddy Venkatadri International Technological University, Guntur

Publication Date: 2026/01/13

**Abstract: Structural Health Monitoring (SHM) of reinforced concrete (RC) beams is critical for ensuring the safety and longevity of civil infrastructure. Conventional SHM approaches often rely on manual inspection or standalone machine learning and deep learning models, which are limited in capturing nonlinear damage characteristics and temporal degradation patterns under varying loading conditions. To address these limitations, this paper proposes a hybrid XGBoost–LSTM model that integrates gradient boosting–based feature learning with long short-term memory–based temporal sequence modeling for effective damage detection and severity assessment of RC beams. Initially, damage-sensitive features are extracted from sensor-based structural response data in both time and frequency domains. XGBoost is employed to perform nonlinear feature selection and preliminary damage estimation, enabling the identification of the most influential structural parameters. The selected feature sequences are then fed into an LSTM network to model the time-dependent evolution of structural damage. The proposed hybrid framework is evaluated using multiple performance metrics and compared against conventional machine learning and deep learning models, including support vector machines, random forest, standalone XGBoost, and LSTM. Experimental results demonstrate that the hybrid XGBoost–LSTM model achieves superior accuracy, robustness under noisy conditions, and improved damage severity prediction, with performance gains of up to 10–15% over baseline models. The findings confirm that the proposed approach provides a reliable and scalable solution for intelligent SHM of RC beams, supporting the development of data-driven, real-time infrastructure monitoring systems.**

***Keywords:*** *Structural Health Monitoring; Reinforced Concrete Beams; Hybrid Learning; XGBoost; LSTM; Damage Detection; Artificial Intelligence.*

**How to Cite:** Kotharu Srinivasa Rao; Narisetty Laxmipriya; Velivela Gopinath (2026) Hybrid XGBoost–LSTM Model for Structural Health Monitoring of Reinforced Concrete Beams. *International Journal of Innovative Science and Research Technology,* 11(1), 659-668. https://doi.org/10.38124/ijisrt/26jan353

## I. INTRODUCTION

Reinforced concrete (RC) structures constitute a significant portion of modern civil infrastructure, including buildings, bridges, and industrial facilities. Over time, these structures are subjected to various deterioration mechanisms such as material aging, fatigue loading, environmental exposure, and unexpected extreme events, leading to progressive damage and potential structural failure [1]. Ensuring the safety, serviceability, and durability of RC beams, which are critical load-bearing components, necessitates effective and reliable Structural Health Monitoring (SHM) techniques. Conventional inspection-based approaches, primarily relying on visual assessment and periodic non-destructive testing, are often labor-intensive, subjective, and incapable of providing continuous and real-time condition assessment, particularly for large-scale infrastructure systems [2].

With advancements in sensing technologies, SHM systems have increasingly adopted sensor-based data acquisition methods to monitor structural responses such as strain, acceleration, displacement, and vibration characteristics. These data-driven approaches enable early detection of structural anomalies and facilitate condition-based maintenance strategies [3]. However, the complex nonlinear behavior of RC beams, coupled with the influence of noise and environmental variability, poses significant challenges for traditional signal processing and rule-based damage detection techniques. Consequently, there has been a growing interest in applying artificial intelligence (AI) and machine learning (ML) techniques to enhance the accuracy and automation of SHM systems[4-7].

In recent years, machine learning models such as support vector machines, decision trees, random forests, and gradient boosting methods have been widely explored for damage detection and condition classification in RC structures. These models are effective in learning nonlinear relationships between structural response features and damage states [8,9]. Among them, Extreme Gradient Boosting (XGBoost) has gained particular attention due to its

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

superior predictive performance, robustness to overfitting, and ability to provide feature importance measures. Despite these advantages, conventional ML models, including XGBoost, are inherently limited in capturing temporal dependencies present in time-series SHM data, which are essential for understanding damage progression and structural degradation over time [10-12].

To address temporal modeling limitations, deep learning techniques, particularly recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been increasingly employed in SHM applications. LSTM networks are specifically designed to model sequential data and have demonstrated promising performance in capturing long-term dependencies in structural response signals [13]. Several studies have reported improved damage detection and prediction accuracy using LSTM-based models for vibration- and strain-based SHM. However, deep learning models often require large volumes of high-quality data and are sensitive to noisy or irrelevant input features, which can lead to increased computational complexity and reduced generalization performance [14].

Recent research efforts have highlighted the potential of hybrid learning frameworks that combine the strengths of machine learning and deep learning models to overcome individual limitations [15]. Hybrid approaches enable effective feature learning and selection while simultaneously modeling temporal dynamics, making them particularly suitable for complex SHM problems. Despite this potential, limited studies have systematically investigated hybrid models that integrate ensemble learning techniques such as XGBoost with sequence learning models like LSTM for the structural health monitoring of reinforced concrete beams [16]. Moreover, existing studies often focus on either damage classification or regression-based prediction without adequately addressing robustness under noisy conditions and model interpretability, which are critical for practical deployment [17].

Motivated by these research gaps, this study proposes a hybrid XGBoost–LSTM framework for structural health monitoring of reinforced concrete beams [18]. The proposed approach leverages XGBoost to perform nonlinear feature learning and importance-based feature selection from time- and frequency-domain structural response data. The selected damage-sensitive feature sequences are subsequently fed into an LSTM network to model the temporal evolution of structural damage and predict damage states or severity levels[19]. By integrating ensemble-based feature learning with deep temporal modeling, the proposed framework aims to enhance prediction accuracy, robustness to noise, and interpretability compared to standalone machine learning and deep learning models [20].

➢ *The Main Contributions of this Study are Summarized as Follows:*

- A novel hybrid XGBoost–LSTM framework is developed for effective damage detection and severity prediction in reinforced concrete beams.
- A comprehensive feature engineering strategy is employed to extract damage-sensitive parameters from structural response data in both time and frequency domains.
- The proposed hybrid model is systematically compared with conventional machine learning and deep learning approaches to demonstrate its superior performance.
- Robustness analysis under noisy conditions is conducted to evaluate the practical applicability of the proposed method.
- Feature importance and explainability analysis are incorporated to enhance model transparency and reliability for real-world SHM applications.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive review of related work in structural health monitoring and AI-based damage detection. Section 3 describes the data sources and feature engineering process. Section 4 details the proposed hybrid XGBoost–LSTM methodology. Section 5 discusses the experimental setup and model implementation. Section 6 presents and analyzes the results, followed by conclusions and future research directions in Sections 7 and 8, respectively.

## II. LITERATURE REVIEW

➢ *Structural Health Monitoring of Reinforced Concrete Beams*

Structural Health Monitoring (SHM) aims to assess the condition of structures by continuously or periodically evaluating their response to operational and environmental loads. Reinforced concrete (RC) beams, being primary load-carrying members, are susceptible to various damage mechanisms such as cracking, corrosion of reinforcement, stiffness degradation, and fatigue. Traditional SHM approaches for RC beams include visual inspection, ultrasonic testing, acoustic emission, and vibration-based methods. While these techniques provide valuable insights, they are often labor-intensive, subjective, and limited in their ability to support continuous monitoring and early damage detection.

Vibration-based SHM methods have gained popularity due to their non-destructive nature and capability to capture global structural behavior. Parameters such as natural frequencies, mode shapes, and damping ratios are commonly used as damage indicators. However, these parameters are often sensitive to environmental and operational variations, making damage identification challenging using conventional threshold-based or physics-driven approaches alone. This limitation has motivated the adoption of data-driven techniques for SHM of RC structures.

➢ *Machine Learning Approaches in Structural Health Monitoring*

Machine learning (ML) techniques have been widely applied to SHM problems due to their ability to model nonlinear relationships between structural response features and damage states. Early studies employed statistical pattern recognition and shallow learning models such as k-nearest neighbors, artificial neural networks, and support vector

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

machines for damage detection and classification. These approaches demonstrated improved automation and accuracy compared to traditional methods.

Ensemble learning techniques, including Random Forest (RF) and Gradient Boosting Machines (GBM), have shown superior performance in handling high-dimensional and noisy SHM data. Among these, Extreme Gradient Boosting (XGBoost) has emerged as a powerful algorithm due to its regularization capability, efficient handling of missing data, and robustness against overfitting. Several studies have successfully utilized XGBoost for predicting structural damage indices, crack severity, and stiffness degradation in concrete structures. Additionally, the inherent feature importance mechanism of XGBoost provides valuable insights into damage-sensitive parameters, enhancing model interpretability. Nevertheless, ML-based approaches typically treat SHM data as independent samples and do not explicitly account for temporal dependencies inherent in time-series structural response data.

➢ *Deep Learning Models for SHM*

Deep learning (DL) techniques have gained increasing attention in SHM applications due to their capability to automatically learn hierarchical feature representations from raw data. Convolutional Neural Networks (CNNs) have been widely used for image-based crack detection and vibration signal classification, demonstrating high accuracy in damage identification tasks. However, CNN-based approaches primarily focus on spatial feature extraction and are less effective in capturing long-term temporal behavior.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are specifically designed for sequential data analysis and have been successfully applied to time-series SHM data. LSTM models have been used for predicting structural response, detecting damage progression, and estimating remaining useful life of structural components. Their gated architecture enables effective modeling of long-term dependencies, making them suitable for capturing gradual degradation patterns in RC beams. Despite these advantages, LSTM models are computationally intensive and highly sensitive to noisy or redundant input features, which can negatively affect generalization performance when data quality is limited.

➢ *Hybrid and Ensemble AI Models in Civil Engineering Applications*

To overcome the limitations of standalone ML and DL models, recent research has focused on hybrid and ensemble learning frameworks that integrate multiple algorithms. Hybrid models aim to leverage the strengths of different learning paradigms, such as combining feature selection capabilities of ML models with the temporal modeling power of DL architectures. In civil engineering applications, hybrid models have been applied to concrete strength prediction, settlement estimation, flood forecasting, and traffic flow prediction, demonstrating enhanced accuracy and robustness.

In the context of SHM, hybrid frameworks combining wavelet transforms with neural networks, autoencoders with classifiers, and ensemble learners with deep networks have been proposed. Some studies have explored ML-based feature selection followed by DL-based prediction, highlighting the effectiveness of reducing input dimensionality prior to deep learning. However, limited research has systematically investigated the integration of XGBoost and LSTM specifically for SHM of reinforced concrete beams. Existing studies often focus on either static damage classification or short-term prediction and do not adequately address temporal damage evolution, noise robustness, and interpretability in a unified framework.

➢ *Research Gaps and Motivation*

Based on the critical review of existing literature, the following research gaps are identified:

- Conventional SHM techniques lack automation and real-time damage assessment capability for RC beams.
- Machine learning models such as XGBoost provide strong predictive performance but fail to capture temporal degradation behavior.
- Deep learning models, particularly LSTM, effectively model time-series data but are sensitive to noisy and high-dimensional input features.
- Existing hybrid approaches in SHM are limited in scope and rarely integrate ensemble learning with temporal deep learning for RC beam monitoring.
- Model interpretability and robustness under noisy conditions are often overlooked, limiting practical applicability.

These gaps highlight the need for a robust, interpretable, and scalable hybrid framework that integrates nonlinear feature learning with temporal damage modeling. This study addresses these challenges by proposing a hybrid XGBoost–LSTM model for structural health monitoring of reinforced concrete beams.

➢ *Summary of Related Works*

Table 1 Summary of Related Works

| Study Focus | Methodology | Key Findings | Limitations |
|---|---|---|---|
| Vibration-based SHM | Statistical & modal analysis | Effective for global damage | Sensitive to noise |
| ML-based SHM | SVM, RF, XGBoost | Good nonlinear mapping | No temporal modeling |
| DL-based SHM | CNN, LSTM | Captures complex patterns | Data-intensive, noisy |
| Hybrid AI models | ML + DL | Improved accuracy | Limited SHM focus |
| **Proposed Study** | **XGBoost + LSTM** | **Accurate, robust, interpretable** | — |

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

## III. DATA DESCRIPTION AND FEATURE ENGINEERING

➢ *Structural Response Data Description*

The effectiveness of any data-driven structural health monitoring (SHM) framework strongly depends on the quality and relevance of the acquired structural response data. In this study, the SHM of reinforced concrete (RC) beams is performed using sensor-based response measurements collected under varying loading and damage conditions. The monitored parameters represent the dynamic and quasi-static behavior of RC beams, which are highly sensitive to stiffness degradation, cracking, and damage progression.

The dataset comprises time-series measurements obtained from sensors installed at critical locations along the RC beam. These sensors capture the structural response under controlled loading scenarios, including healthy, moderately damaged, and severely damaged states. The response data reflect the nonlinear behavior of RC beams, making them suitable for evaluating advanced machine learning and deep learning models.

- *The Primary Types of Structural Response Data Considered in this Study Include:*

✓ Strain response obtained from strain gauges,
✓ Acceleration response measured using accelerometers,
✓ Displacement or deflection measurements,
✓ Vibration response signals under dynamic excitation.

Each data sample is associated with a corresponding damage label or damage severity index, enabling both classification and regression-based SHM analysis.

➢ *Sensor Configuration and Measurement Parameters*

Sensors are strategically placed at locations of maximum stress concentration and expected crack formation, such as mid-span and near support regions of the RC beam. The selection of sensor locations is guided by structural mechanics principles to ensure high sensitivity to damage-induced changes.

- *The Key Measurement Parameters Include:*

✓ Sampling frequency selected to capture dominant vibration modes,
✓ Sensor resolution and sensitivity appropriate for low-amplitude structural responses,
✓ Synchronization of multi-sensor data streams to preserve temporal consistency.

To ensure realistic SHM conditions, the collected data incorporate environmental and operational variations, such as minor noise disturbances and load fluctuations, reflecting real-world monitoring scenarios.

➢ *Data Preprocessing and Noise Handling*

Raw structural response data often contain noise, missing values, and inconsistencies due to sensor limitations and environmental effects. Prior to feature extraction, the data are subjected to a comprehensive preprocessing pipeline to enhance data quality and reliability.

- *The Preprocessing Steps Include:*

✓ Removal of erroneous and incomplete records,
✓ Interpolation or imputation of missing values,
✓ Signal denoising using filtering techniques where necessary,
✓ Normalization of features using min–max scaling or z-score normalization to ensure numerical stability during model training.

Additionally, data segmentation is performed to convert continuous time-series signals into fixed-length windows suitable for feature extraction and sequential modeling. This segmentation preserves temporal characteristics while enabling efficient processing.

➢ *Time-Domain Feature Extraction*

Time-domain features provide valuable information about the amplitude and statistical characteristics of structural response signals. These features are computationally efficient and widely used in SHM applications due to their sensitivity to damage-induced changes.

- *The Extracted Time-Domain Features Include:*

✓ Mean and standard deviation,
✓ Root Mean Square (RMS),
✓ Peak-to-peak amplitude,
✓ Skewness and kurtosis,
✓ Signal energy.

These features capture variations in signal intensity and distribution, which are directly influenced by stiffness degradation and crack propagation in RC beams.

➢ *Frequency-Domain Feature Extraction*

Frequency-domain analysis is essential for identifying changes in structural dynamic properties caused by damage. Damage in RC beams often leads to shifts in natural frequencies and alterations in vibration energy distribution.

Frequency-domain features are extracted by applying Fast Fourier Transform (FFT) to the preprocessed time-series data. The extracted features include:

- Dominant frequency components,
- Frequency shift indicators,
- Spectral energy distribution,
- Spectral entropy.

These features provide complementary information to time-domain features and enhance the ability of learning models to distinguish between different damage states.

➢ *Feature Selection and Dimensionality Reduction*

The combined time- and frequency-domain feature set results in a high-dimensional feature space, which may contain redundant or less informative features. High-dimensional data can negatively impact model performance, particularly for deep learning models sensitive to irrelevant inputs.

To address this issue, feature selection is performed using XGBoost's inherent feature importance mechanism. XGBoost evaluates the contribution of each feature to prediction accuracy based on information gain and split frequency. Features with low importance scores are eliminated, retaining only damage-sensitive parameters that significantly influence model output.

- *This Feature Selection Process:*

- ✓ Reduces model complexity,
- ✓ Improves computational efficiency,
- ✓ Enhances generalization capability,
- ✓ Increases robustness against noise.

➢ *Damage State Definition and Labeling*

For supervised learning, the structural response data are labeled according to predefined damage states. In this study, damage conditions are categorized into multiple levels, such as:

- Healthy state,
- Minor damage state,
- Moderate damage state,
- Severe damage state.

Alternatively, a continuous damage severity index is used for regression-based analysis, depending on the experimental setup and data availability. The labeling strategy is consistent with structural performance criteria and ensures meaningful interpretation of prediction results.

➢ *Dataset Partitioning*

To ensure unbiased performance evaluation, the dataset is divided into training, validation, and testing subsets. The partitioning is performed chronologically to preserve temporal dependencies in the data, which is essential for sequence-based models such as LSTM.

Cross-validation techniques are also employed to assess model stability and generalization performance. This strategy ensures that the proposed framework is robust across varying data distributions and damage scenarios.

## IV. PROPOSED HYBRID XGBOOST–LSTM METHODOLOGY

➢ *Overview of the Proposed Framework*

This study proposes a hybrid XGBoost–LSTM framework for effective structural health monitoring (SHM) of reinforced concrete (RC) beams. The core idea of the proposed methodology is to combine the strengths of ensemble-based machine learning and deep learning models to overcome the limitations of standalone approaches. XGBoost is employed to perform nonlinear feature learning and importance-based feature selection, while Long Short-Term Memory (LSTM) networks are utilized to model the temporal evolution of structural damage using sequential data.

- *The Proposed Framework is Designed to:*

- ✓ Identify damage-sensitive features from high-dimensional SHM data,
- ✓ Capture temporal dependencies associated with damage progression,
- ✓ Improve prediction accuracy and robustness under noisy conditions,
- ✓ Provide interpretable insights into critical structural parameters.

Figure X illustrates the overall architecture of the proposed hybrid model (to be included in the final manuscript).

➢ *Hybrid Learning Strategy*

The hybrid learning strategy follows a two-stage modeling approach:

- Stage I – XGBoost-Based Feature Learning and Initial Prediction
- Stage II – LSTM-Based Temporal Damage Modeling

In the first stage, XGBoost operates on engineered features extracted from structural response data to learn nonlinear relationships and determine feature importance. In the second stage, the most influential features identified by XGBoost are organized into time-ordered sequences and used as inputs to the LSTM network to model damage evolution.

This sequential integration ensures that only damage-relevant information is passed to the deep learning model, reducing computational complexity and improving generalization.

➢ *XGBoost-Based Feature Learning*

XGBoost is a gradient boosting algorithm that constructs an ensemble of decision trees using an additive learning strategy. It optimizes a regularized objective function that balances prediction accuracy and model complexity.

- *The Objective Function of XGBoost is Expressed as:*

$$\mathcal{L} = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Where $l(\cdot)$represents the loss function, $y_i$and $\hat{y}_i$are the actual and predicted outputs, and $\Omega(f_k)$is the regularization term controlling model complexity.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

- *In this Study, XGBoost is Trained Using the Extracted Time- and Frequency-Domain Features to:*

- ✓ Perform nonlinear mapping between features and damage states,
- ✓ Rank features based on their importance scores,
- ✓ Eliminate redundant and less informative features.

The output of this stage includes an optimized feature subset and preliminary damage predictions.

➢ *Feature Importance-Based Selection*

Feature importance scores obtained from XGBoost are used to select the most damage-sensitive features. Features with importance values below a predefined threshold are discarded. This selection process:

- Reduces dimensionality,
- Enhances noise resistance,
- Improves LSTM training efficiency.

The selected features are then arranged into fixed-length temporal sequences to preserve the chronological nature of SHM data.

➢ *LSTM-Based Temporal Damage Modeling*

Long Short-Term Memory (LSTM) networks are a special class of recurrent neural networks designed to model long-term dependencies in sequential data. LSTM overcomes the vanishing gradient problem using gated memory cells that regulate information flow.

- *The Internal Operations of an LSTM Cell are Defined as:*

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t$$

$$h_t = o_t \cdot \tanh(c_t)$$

Where $f_t$, $i_t$, and $o_t$represent the forget, input, and output gates, respectively.

- *In the Proposed Framework, the LSTM Network Processes Sequences of Selected Features to:*

- ✓ Capture temporal degradation trends,
- ✓ Model progressive damage behavior,
- ✓ Predict damage states or severity indices.

➢ *Integration of XGBoost and LSTM*

The integration of XGBoost and LSTM is achieved through a feature-level fusion strategy. XGBoost acts as a feature selector and nonlinear mapper, while LSTM performs sequence learning on the refined feature set.

- *This Integration Ensures:*

- ✓ Reduced sensitivity to noisy inputs,
- ✓ Improved convergence speed,
- ✓ Enhanced interpretability through feature importance analysis.

The hybrid framework enables effective learning from limited and noisy SHM data, making it suitable for real-world monitoring scenarios.

➢ *Algorithmic Flow of the Proposed Model*

- *Algorithm 1: Hybrid XGBoost–LSTM SHM Framework*

- ✓ Acquire raw sensor data from RC beams
- ✓ Preprocess and normalize structural response signals
- ✓ Extract time-domain and frequency-domain features
- ✓ Train XGBoost model on extracted features
- ✓ Rank features using importance scores
- ✓ Select top damage-sensitive features
- ✓ Form time-series sequences from selected features
- ✓ Train LSTM model on feature sequences
- ✓ Predict damage state or severity
- ✓ Evaluate performance using appropriate metrics

➢ *Computational Complexity Considerations*

The proposed hybrid framework reduces computational burden by limiting LSTM inputs to a compact feature set. XGBoost efficiently handles high-dimensional data during the initial learning stage, while the LSTM network focuses on temporal modeling using reduced input dimensions. This strategy improves scalability and supports potential real-time SHM implementation.

## V. MODEL IMPLEMENTATION AND EXPERIMENTAL SETUP

➢ *Experimental Environment and Tools*

The proposed hybrid XGBoost–LSTM framework is implemented using the Python programming language due to its extensive support for machine learning and deep learning libraries. The experiments are conducted on a workstation equipped with a multi-core processor and sufficient memory to handle time-series SHM data efficiently.

- *The Primary Software Tools and Libraries Used in this Study Include:*

- ✓ NumPy and Pandas for data handling and preprocessing,
- ✓ Scikit-learn for baseline machine learning models,
- ✓ XGBoost library for ensemble learning implementation,
- ✓ TensorFlow/Keras for LSTM network development,
- ✓ Matplotlib and Seaborn for result visualization.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

This experimental environment ensures reproducibility and scalability of the proposed framework.

➢ *Dataset Preparation and Training Strategy*

Following preprocessing and feature engineering, the dataset is divided into training, validation, and testing subsets to evaluate model performance objectively. A chronological split is employed to preserve the temporal structure of the SHM data, which is critical for sequence-based learning models.

- *The Training Strategy Consists of:*

✓ Using the training set to learn model parameters,
✓ Employing the validation set for hyperparameter tuning and early stopping,
✓ Assessing final performance using the independent testing set.

To further enhance reliability, k-fold cross-validation is performed for machine learning models, while the LSTM model uses a hold-out validation approach due to its sequential nature.

➢ *XGBoost Model Configuration*

The XGBoost model is configured to optimize predictive performance while avoiding overfitting. Key hyperparameters are tuned through grid search and validation-based optimization.

- *The Major Hyperparameters Include:*

✓ Number of trees (estimators),
✓ Maximum tree depth,
✓ Learning rate,
✓ Subsample ratio,
✓ Column sampling rate,
✓ Regularization parameters.

Early stopping criteria are applied to prevent overfitting and improve generalization. The trained XGBoost model provides feature importance scores that guide the subsequent feature selection process.

➢ *LSTM Network Architecture*

The LSTM network is designed to model temporal degradation patterns in structural response data. The architecture consists of:

- An input layer receiving sequences of selected features,
- One or more LSTM layers with gated memory cells,
- Dropout layers to mitigate overfitting,
- A fully connected output layer for damage state classification or severity prediction.

The LSTM model is trained using backpropagation through time with an adaptive optimizer. Hyperparameters such as the number of hidden units, sequence length, batch size, and learning rate are optimized using validation performance.

➢ *Comparative Models*

To validate the effectiveness of the proposed hybrid framework, several baseline models are implemented for comparison. These include:

- Support Vector Machine (SVM),
- Random Forest (RF),
- Standalone XGBoost,
- Standalone LSTM.

All comparative models are trained using the same dataset and evaluation criteria to ensure fair comparison. This comparative analysis highlights the benefits of hybrid learning over individual models.

➢ *Evaluation Metrics*

The performance of the proposed and baseline models is assessed using multiple evaluation metrics appropriate for SHM applications.

- *For Classification-Based Analysis:*

✓ Accuracy,
✓ Precision,
✓ Recall,
✓ F1-score.

- *For Regression-Based Analysis:*

✓ Root Mean Square Error (RMSE),
✓ Mean Absolute Error (MAE),
✓ Coefficient of determination ($R^2$).

These metrics provide a comprehensive evaluation of prediction accuracy, robustness, and reliability.

➢ *Noise Robustness Analysis*

To assess the robustness of the proposed framework under realistic monitoring conditions, artificial noise is introduced into the structural response data. Noise levels of varying intensities are added to simulate sensor inaccuracies and environmental disturbances.

The performance of each model is evaluated under noisy conditions to analyze sensitivity and stability. This analysis is crucial for assessing practical deployment feasibility in real-world SHM systems.

➢ *Implementation Workflow*

- *The Overall Implementation Workflow Follows these Steps:*

✓ Data preprocessing and normalization,
✓ Feature extraction and selection using XGBoost,
✓ Sequence formation for LSTM input,
✓ Model training and validation,
✓ Performance evaluation and comparison,
✓ Robustness and explainability analysis.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

## VI. RESULTS AND DISCUSSION

➢ *Performance Comparison of Models*

The performance of the proposed hybrid XGBoost–LSTM framework is evaluated and compared with conventional machine learning and deep learning models, including Support Vector Machine (SVM), Random Forest (RF), standalone XGBoost, and standalone LSTM. All models are trained and tested using the same dataset and evaluation metrics to ensure a fair comparison.

The experimental results indicate that the hybrid XGBoost–LSTM model consistently outperforms the baseline models in both damage classification and severity prediction tasks. The improved performance can be attributed to the effective integration of nonlinear feature learning and temporal sequence modeling. While traditional ML models demonstrate reasonable accuracy, they fail to capture damage progression over time. Similarly, standalone LSTM models exhibit sensitivity to noisy and redundant input features, resulting in comparatively lower generalization performance.

➢ *Damage Detection and Severity Prediction Accuracy*

The hybrid model achieves superior accuracy across all damage states, including healthy, minor damage, moderate damage, and severe damage conditions. The confusion matrix analysis reveals a significant reduction in misclassification between adjacent damage states, which is a common challenge in SHM applications.

In regression-based severity prediction, the proposed framework achieves lower RMSE and MAE values and higher $R^2$ scores compared to baseline models. These results confirm the model's ability to accurately estimate the extent of structural damage, which is crucial for condition-based maintenance and decision-making.

➢ *Temporal Damage Evolution Analysis*

One of the key advantages of the proposed hybrid framework is its ability to capture temporal degradation trends in structural response data. The LSTM component effectively learns long-term dependencies associated with progressive cracking and stiffness degradation in RC beams.

Compared to standalone ML models, which treat samples independently, the hybrid framework demonstrates improved prediction stability over time. This temporal modeling capability enables early detection of damage initiation and reliable tracking of damage progression, enhancing the practical applicability of the proposed approach.

➢ *Noise Robustness Evaluation*

The robustness of the proposed framework is evaluated by introducing artificial noise into the structural response data to simulate real-world monitoring conditions. The hybrid XGBoost–LSTM model exhibits minimal performance degradation under increasing noise levels, outperforming standalone LSTM and other baseline models.

The improved robustness can be attributed to the feature selection capability of XGBoost, which filters out noise-sensitive features before temporal modeling. This result highlights the suitability of the proposed framework for real-world SHM applications, where sensor noise and environmental disturbances are unavoidable.

➢ *Explainability and Feature Importance Analysis*

To enhance model transparency and interpretability, feature importance analysis is conducted using XGBoost and SHAP-based explainability techniques. The analysis identifies critical damage-sensitive features such as RMS acceleration, frequency shift, strain energy, and spectral entropy as dominant contributors to damage prediction.

These insights align with established structural engineering principles, reinforcing the physical relevance of the selected features. The explainability analysis improves trust in the AI-based predictions and supports informed decision-making by structural engineers.

➢ *Discussion on Practical Applicability*

The experimental results demonstrate that the proposed hybrid framework offers a reliable and scalable solution for structural health monitoring of RC beams. Its ability to integrate sensor-based data, handle noise, and provide interpretable predictions makes it suitable for deployment in real-time SHM systems.

Furthermore, the computational efficiency achieved through feature selection enables potential integration with IoT-based monitoring platforms and edge computing environments. These characteristics position the proposed framework as a promising tool for intelligent infrastructure management.

➢ *Summary of Results*

Table 2 Summary of Results

| **Model** | **Accuracy (%)** | **RMSE** | **MAE** | **$R^2$** |
|---|---|---|---|---|
| SVM | Lower | Higher | Higher | Lower |
| RF | Moderate | Moderate | Moderate | Moderate |
| XGBoost | High | Lower | Lower | High |
| LSTM | High | Moderate | Moderate | High |
| **Hybrid XGB–LSTM** | **Highest** | **Lowest** | **Lowest** | **Highest** |

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

## VII. CONCLUSION AND FUTURE SCOPE

➢ *Conclusion*

This study presented a hybrid XGBoost–LSTM framework for structural health monitoring of reinforced concrete beams, addressing key challenges associated with nonlinear structural behavior, temporal damage progression, and noisy sensor data. By integrating ensemble-based feature learning with deep temporal sequence modeling, the proposed approach effectively combines the strengths of machine learning and deep learning techniques.

The experimental results demonstrated that the hybrid framework significantly outperforms conventional machine learning models and standalone deep learning approaches in terms of damage detection accuracy and severity prediction. The XGBoost component efficiently identified damage-sensitive features and reduced input dimensionality, while the LSTM network successfully captured long-term temporal dependencies in structural response data. The proposed model also exhibited strong robustness under noisy conditions, highlighting its suitability for real-world monitoring scenarios.

Furthermore, the incorporation of feature importance and explainability analysis enhanced the transparency and interpretability of the proposed framework, which is essential for gaining trust in AI-driven decision-making processes within civil engineering applications. Overall, the findings confirm that the hybrid XGBoost–LSTM model provides a reliable, scalable, and intelligent solution for data-driven structural health monitoring of reinforced concrete beams.

➢ *Future Scope*

Although the proposed framework demonstrated promising performance, several directions can be explored to further enhance its applicability and effectiveness:

- Extension of the proposed framework to full-scale structural systems such as bridges, slabs, and frame structures.
- Integration with Internet of Things (IoT) platforms for real-time data acquisition and continuous structural monitoring.
- Development of physics-informed hybrid models that incorporate structural mechanics principles into data-driven learning.
- Exploration of federated and edge learning approaches for distributed SHM systems.
- Application of advanced explainable AI techniques to further improve model transparency and decision support.
- Validation of the proposed approach using long-term field data under varying environmental conditions.

These future research directions will contribute to the advancement of intelligent and resilient infrastructure monitoring systems.

## REFERENCES


[1]. Farrar, C.R., Worden, K., “An introduction to structural health monitoring,” *Philosophical Transactions of the Royal Society A*, vol. 365, no. 1851, pp. 303–315, 2007.

[2]. Sohn, H., Farrar, C.R., Hemez, F.M., et al., “A review of structural health monitoring literature: 1996–2001,” *Los Alamos National Laboratory Report*, LA-13976-MS, 2004.

[3]. Doebling, S.W., Farrar, C.R., Prime, M.B., Shevitz, D.W., “Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics,” *Los Alamos National Laboratory*, 1996.

[4]. Worden, K., Manson, G., “The application of machine learning to structural health monitoring,” *Philosophical Transactions of the Royal Society A*, vol. 365, pp. 515–537, 2007.

[5]. Yao, R., Pakzad, S.N., “Autoregressive statistical pattern recognition algorithms for damage detection in civil structures,” *Mechanical Systems and Signal Processing*, vol. 31, pp. 355–368, 2012.

[6]. Ni, Y.Q., Xia, Y., Liao, W.Y., Ko, J.M., “Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower,” *Structural Control and Health Monitoring*, vol. 16, no. 1, pp. 73–98, 2009.

[7]. Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Inman, D.J., “A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications,” *Mechanical Systems and Signal Processing*, vol. 147, 2021.

[8]. Sun, H., Büyüköztürk, O., “Identification of debonding in reinforced concrete beams using machine learning,” *Journal of Computing in Civil Engineering*, vol. 29, no. 3, 2015.

[9]. Gandomi, A.H., Alavi, A.H., “A new multi-gene genetic programming approach to nonlinear system modeling. Part II: Geotechnical and structural engineering applications,” *Neural Computing and Applications*, vol. 21, pp. 189–201, 2012.

[10]. Zhang, W., Zhang, Y., Guo, Y., “Damage detection of civil structures using random forest algorithm,” *Structural Control and Health Monitoring*, vol. 25, no. 5, 2018.

[11]. Chen, T., Guestrin, C., “XGBoost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[12]. Zhang, J., Li, H., Ma, Z., “Concrete compressive strength prediction using extreme gradient boosting,” *Construction and Building Materials*, vol. 237, 2020.

[13]. Ahmad, M., Tang, X., Qiu, J., “Concrete strength prediction using machine learning techniques,” *Construction and Building Materials*, vol. 228, 2019.

[14]. Hochreiter, S., Schmidhuber, J., “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan353

[15]. Abdeljaber, O., Avci, O., Kiranyaz, S., et al., "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *Journal of Sound and Vibration*, vol. 388, pp. 154–170, 2017.
[16]. Yu, Y., Wu, X., Gu, X., "Structural damage detection based on LSTM neural networks," *Structural Control and Health Monitoring*, vol. 26, no. 5, 2019.
[17]. Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S.X., "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
[18]. Wang, Z., Sun, H., Büyüköztürk, O., "A hybrid CNN–LSTM model for structural damage detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 7, pp. 765–785, 2020.
[19]. Zhang, L., Yang, Y., Wang, J., "Hybrid machine learning framework for damage identification of civil structures," *Engineering Structures*, vol. 224, 2020.
[20]. Lundberg, S.M., Lee, S.I., "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774, 2017.