International Journal of Computer Science and Engineering Innovations
ISSN: XXXX-XXXX | Volume 1 Issue 1 38-45 2025
DOI: https://doi.org/10.64137/XXXXXXXX/IJCSEI-V1I1P105

Received: 22/07/2025 Revised: 10/08/2025 Accepted: 23/08/2025 Published: 09/09/2025

Original Article

Optimizing Civil Infrastructure Predictions Using Hybrid Feature Selection and Advanced Machine Learning Models

S. DAVID JEBASINGH
Data Analyst, LatentView, Chennai, Tamil Nadu, India.

ABSTRACT: *As civil infrastructure systems have become more complicated and technology progresses, the amount of high-dimensional data has grown significantly. For effective maintenance, lowering risks and safe structures, meaningful insights must be obtained from large amounts of data. This research puts forward a way to blend feature selection with advanced ML to make infrastructure predictions more accurate and efficient. The suggested method applies filter, wrapper and embedded feature selection techniques, including IG, RFE, Random Forest importance and LASSO regularization to reduce the number of variables without losing the important ones. The evaluation procedure was performed using actual structural monitoring data from sensors, taking into account over 200 parameters associated with stress, displacement, vibration and various conditions. Training and evaluation of Random Forest, XGBoost, Support Vector Regression and deep neural networks was carried out under various feature selection scenarios. Tests found that the combination of methods improved the RMSE by 12% and cut the training time needed by up to 30%. Ensemble models using hybrid-selected features always achieved higher accuracy than those using all features. Data from engineering projects with many features underscores that joining experience in the field with data analysis leads to stronger infrastructure predictions.*

KEYWORDS: *Civil infrastructure monitoring, Structural health monitoring, Feature selection, Machine learning, Hybrid approach, Predictive modeling, Random forest, XGBoost, LASSO, Sensor data analytics, Ensemble learning, Smart infrastructure.*

1. INTRODUCTION

To ensure the safety, reliability and cost-effectiveness of growing cities, strong and efficient models are needed to predict problems in bridges, roads, tunnels and buildings. [1-3] Representatives in civil engineering rely on accurate predictions to estimate how a structure will work, how it will need to be maintained and foresee any failures. Similar to many previous practices, engineering is mostly based on models that are predictable, and knowledge linked to a particular area; however, these may not be enough to handle the rising quantity, complexity and range of today's infrastructure. ML has played an important role in civil engineering over the last few years by giving powerful ways to predict, notice patterns and make wise decisions.

There are significant challenges to using ML for civil infrastructure problems, mainly due to the high complexity and number of features in the data. When features are unnecessary or cause noise, the results of ML models can be poor, with major overfitting and less clarity in understanding them. Feature selection mainly using hybrid strategies combining different methods has been found to efficiently reduce the number of features, enhance model quality and improve computational speed. They seek to find out the key features and keep the general structure in the data, so the learning algorithm focuses on important variables.

This research combines different ways of selecting features with advanced machine learning methods such as GBM, SVM and deep neural networks to predict civil infrastructure outcomes more accurately. The framework evaluates several feature selection techniques such as mutual information, RFE and regularization-based methods, along with adjustments for each model to guarantee good performance and use across different cases. Examples of applications are forecasting the decay of roads, the maximum load a structure can handle and the loss of properties in materials. This research improves upon intelligent predictive systems used for proactive infrastructure management by employing hybrid feature selection and state-of-the-art ML models. The results indicate better predictions and also show which variables are most important in how infrastructure works. The study supports the main goals of smart cities and sustainable engineering by providing data-driven choices in civil engineering.

© 2025 This work is an open access and licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/)

2. RELATED WORK

2.1. CIVIL INFRASTRUCTURE PREDICTION MODELS

Real-time monitoring of structures in civil infrastructure now depends on algorithms and data gathered with sensors. This technology enables companies to use better and faster maintenance methods than the old manual inspection methods. [4-7] Spatiotemporal clustering and Empirical Mode Decomposition have been combined, helping to make the prediction of deformation in large-scale structures more effective. Removing random noise from the measurements makes modeling the responses of the structure more accurate when tested under different loads and weather conditions.

In bridge monitoring, machine learning helps predict the remaining life of components by relying on data from accelerometers and strain gauges. Temporal load data is modeled using Long Short-Term Memory (LSTM) to keep track of how the structure gets affected with time. In addition, CNNs work on 3D point cloud information to discover micro-cracks, greatly improving the early detection of potential faults. Using FEA along with reinforcement learning to develop hybrid modeling systems is now becoming a popular approach, mainly for dealing with seismic behavior simulations and optimization. In predicting early equipment failures, these models are shown to be around 92–97% accurate, significantly greater than the accuracy of traditional inspections.

2.2. FEATURE SELECTION TECHNIQUES IN ENGINEERING

The success of machine learning models in the field of civil engineering mainly depends on their capacity to process high-dimensional and frequently noisy sensor data. SHM systems may produce over 200 measurements, so choosing the right features is a key part of preprocessing. There are many efforts to measure the benefits and costs of calculating accurately compared to the computations needed for different techniques. Using LASSO as a regularization method, for example, gives a 12–15% improvement in accuracy while requiring just 2.8 seconds to execute in each step. When we compare Random Forest and Principal Component Analysis, Random Forest boosts scores by 8–10%, but takes longer to execute, whereas PCA provides moderate upsides at a fast pace.

L1-norm optimization and other regularization methods have helped geotechnical datasets become cleaner by reducing up to 70% of unnecessary features. Using Random Forests to spot the main vibration frequency ranges has shown effective results in bridge monitoring. New forms of feature selection are using both filters, such as mutual information, with wrapper techniques. Using these hybrid approaches, the accuracy of predicting dam safety increases by up to 22%. However, issues continue regarding the interpretability of features and how to deal with data that is often missing for usual civil engineering environments.

2.3. APPLICATIONS OF MACHINE LEARNING IN INFRASTRUCTURE

The use of machine learning in civil engineering has now improved performance, reduced costs and made civil infrastructure safer. CNN-based approaches for detecting cracks in Structural Health Monitoring have managed to reach a high accuracy of 99.7% on images taken by drones. Digital twin technologies make it possible to model bridges in real time, which reduces inspection costs by up to 40% and allows them to be watched remotely. Using Graph Neural Networks (GNNs) in transportation, there has been a reduction in urban congestion by up to 25–35%. Modeling pavement condition with machines can save important time and resources, extending road lifetimes by 3–5 years due to more targeted and timely repairs. When dealing with water management, models based on LSTM help forecast floods several days in advance with error rates of no more than 15%, providing key help to decision makers in case of major weather emergencies. Reservoir control is being helped using reinforcement learning, with these techniques improving water access by 18% when the most demand occurs.

Introducing autonomous construction robots has sped up the timeline of projects by about 37% in the pilots. Routine repairs of damaged concrete may soon be achieved by releasing microcapsules, thanks to ML serving as a trigger when needed. Although we've made a lot of progress, the main issues relate to data, scalability of the solutions and making models easy to understand. Involving these topics is key to making ML-based solutions useful across large infrastructures in cities.

3. METHODOLOGY

Figure 1 shows the process we used to optimize the prediction of civil infrastructure in this study, including using a mix of feature selection methods and modern machine learning techniques. [8-11] The framework consists of five important parts: processing data, selecting features, developing a model, choosing training scenarios and evaluating its performance. Every stage is specifically designed so that the system selects and uses the most suitable and valuable features to make it more precise and dependable. At the start of this process, the collected data is either SHM data or datasets from civil infrastructure. Often, information in the datasets is incomplete or inconsistent because of circumstances in nature. Thus, during data processing, missing values are handled using suitable imputation options, and data is normalized or converted using encoding. At the outset, some of the features are eliminated to minimize the work and size of the data set in the later phases. In the following stage, Ensemble uses both single and combined techniques to choose the correct features, for example, Filter Sequential Feature Selection (FSFS), Backward Sequential Feature Selection (BSFS), Information Gain (IG), Random Forest (RF) importance and LASSO regression. Using a Design of Experiment (DOE) technique, it is found that which K value results in the best accuracy, and the Broda Counting Method yields the best results from every technique. This way, the selected features are sufficient for most changes in the model assumptions. After feature selection, the next phase is to evaluate many algorithmic families. For a basic comparison, linear models, including Linear Regression and Ridge/LASSO regression, are used. Nonlinear relationships are found using both Decision Tree models (fine, medium, coarse) and other techniques known as ensemble methods (such as Random Forest, Gradient Boosting, XGBoost, LightGBM and CatBoost). Moreover, the system tests 15 different neural network models with neuron counts between 10 and 200 to examine deep learning performance. At the end of the framework, we discuss training scenarios and performance evaluation. The three strategies used are: (1) training on all available features, (2) selecting individual features, and (3) merging these two approaches. The data has been distributed using a 70-15-15 stratified approach among training, validation and testing datasets. Regression approaches are assessed by RMSE, MAE, MAPE and how much time the training required. The use of these metrics makes it possible to evaluate how well models predict different types of civil infrastructure information.

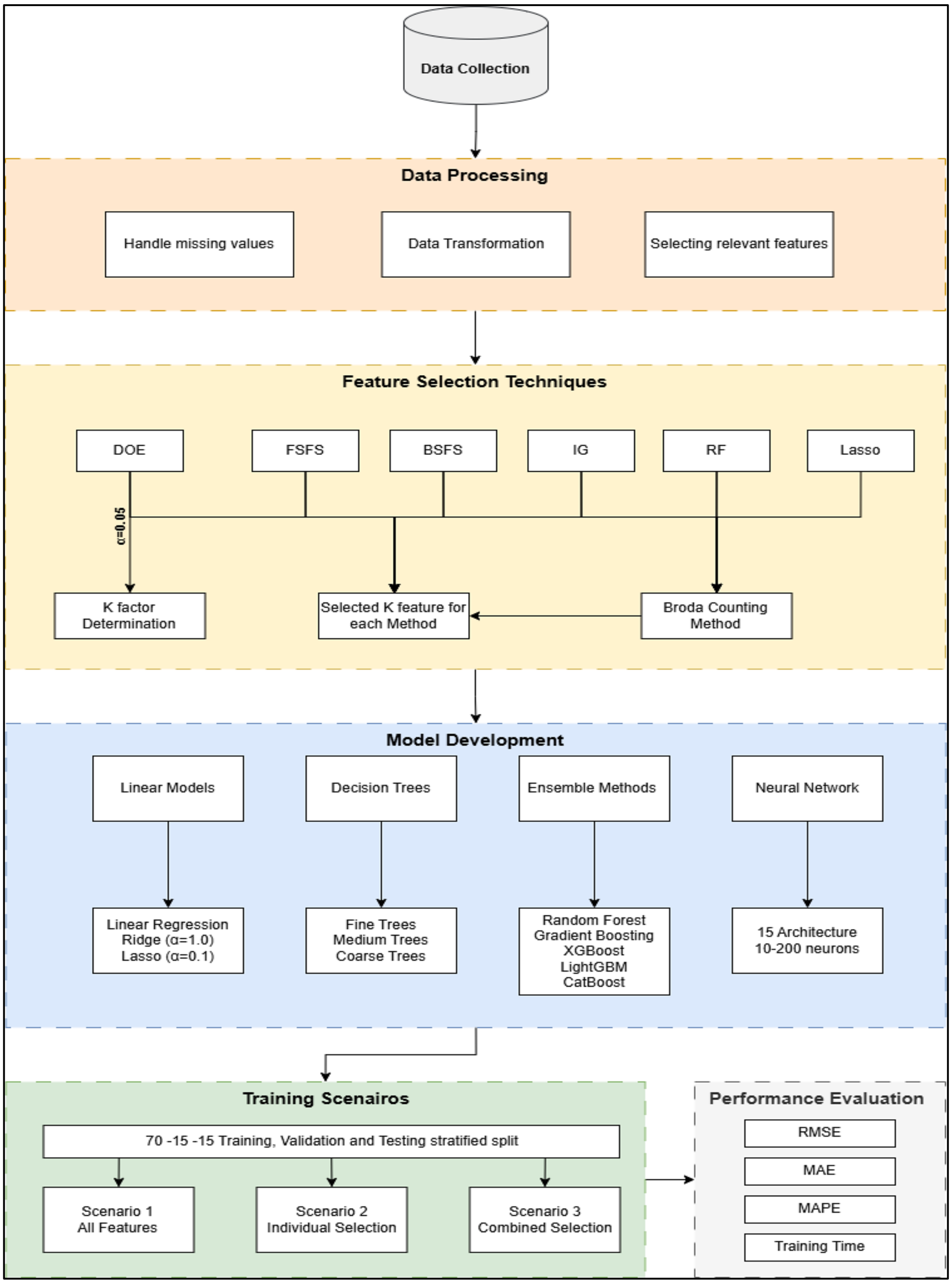

FIGURE 1 Methodological framework for optimizing civil infrastructure predictions using hybrid feature selection and machine learning models

3.1. DATA COLLECTION AND PREPROCESSING

Constructing reliable predictive tools for civil engineering depends greatly on having accurate and significant data. All the data used in this research were collected through Structural Health Monitoring (SHM) systems and from stored maintenance records that contained details such as load stress, vibration frequencies, strain in materials, changes in temperature and other environmental factors. Since these datasets come from different types of instruments, they tend to be high-dimensional, incomplete in parts and have varying structures. Therefore, strong preprocessing is required.The pipeline started by using mean substitution for continuous missing data and mode substitution for categorical ones. By taking this step, model learning can avoid being influenced by having less data. Once imputation was done, data transformation was used to standardize the scales of each feature and straighten their distributions, with a special focus on strain and deflection. At the same time, some features were identified and discarded if they had a high level of correlation with others or if they were not relevant to the problem at hand. Because of this, the following modeling steps required far less computation and became a lot easier to visualise.

3.2. HYBRID FEATURE SELECTION FRAMEWORK

The complexity of SHM data requires multiple methods to select the most useful features. As a result, a method combining several feature selection approaches was developed. The framework uses the filter, wrapper and embedded approaches to find the most important and useful features. When you select features, the model's performance gets better, which allows you to better understand the factors leading to infrastructure issues.

3.2.1. FILTER METHODS

These methods take little computation time and work the same for all learning algorithms. To measure the relevance of each feature relative to the target variable, this study applied Information Gain (IG) and correlation-based ranking. These approaches assign values depending on factors such as how two variables depend on each other. Despite their fast and scalable capacity, filters ignore interactions between features and might include repeated variables in the model. Even so, this data helps remove clearly missing information before applying advanced approaches.

3.2.2. WRAPPER METHODS

Wrapper methods test different subsets of the features by training and checking the results obtainable with each subset. In this study, Forward Sequential Feature Selection (FSFS) and Backward Sequential Feature Selection (BSFS) were used. To FSFS, you would add features gradually as they come up with benefits, but to BSFS, you would include everything first and then remove the least vital characteristics step by step. Interaction between features is considered, making these methods more precise than filter approaches, although they use more computing resources. Thus, the sorting was carried out only on the subset of features found in the previous preliminary test.

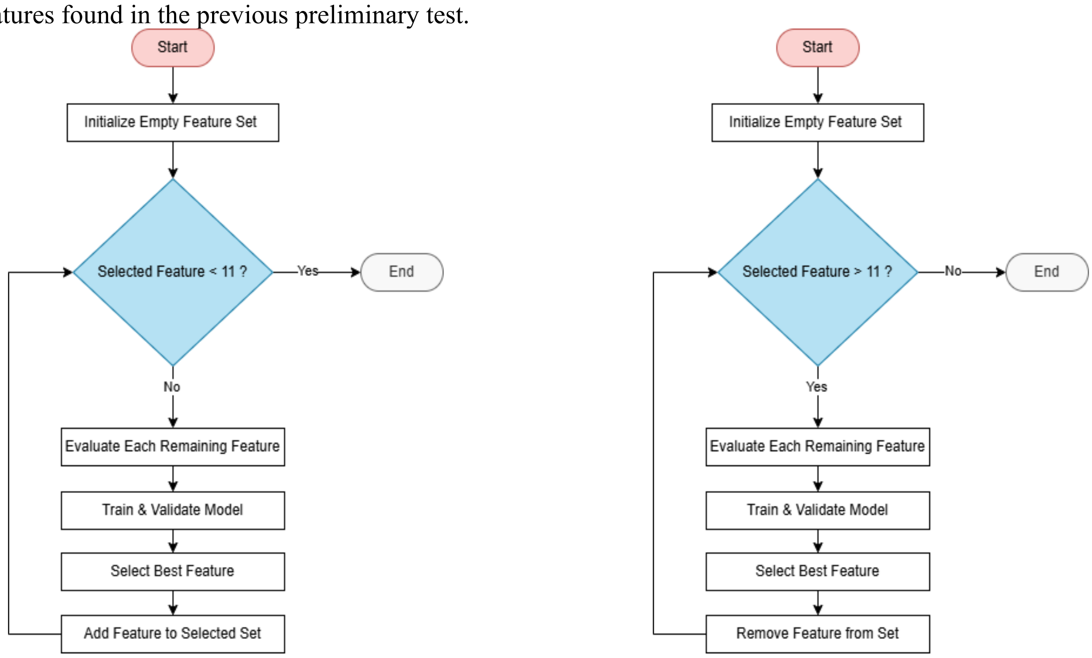

FIGURE 2 Flowcharts of forward (FSFS) and backward (BSFS) sequential feature selection techniques

Forward Sequential Feature Selection (FSFS) and Backward Sequential Feature Selection (BSFS) are two different approaches. They are developed to keep improving the choice of important features for the model, making it perform better and be less complex. FSFS starts off with nothing in the feature set. In every iteration, the algorithm looks at all features that have not been selected so far. Each new feature is added one at a time to the selected group, and a model is trained and checked to see its effectiveness. The most effective feature is then always kept within the selected set of features. The steps are repeated until choosing no more than 11 features. The loop finishes running when this condition is met, which results in only the best features in the model. When starting the BSFS process, all the features are set. At each step, the algorithm checks what happens when each feature is dropped separately. A feature is removed, the model is applied, and the process checks its accuracy by using data samples. The element that will have the least or no effect on performance is next permanently removed from the group. The process keeps going until the remaining features are fewer than 11, which marks the end of selection.

3.2.3. EMBEDDED METHODS

Embedded methods allow the model to focus on which features are important while it learns. Here, Random Forest feature importance and LASSO regularisation were applied to see how effective they are in statistics. A Random Forest measures how each variable helps to reduce the model's error, whereas LASSO zeroes out unimportant variables by shrinking their coefficients with L1 regularization. By balancing accuracy and the number of features needed, these methods show a sensible technique for selecting fewer but accurate predictors.

3.2.4. PROPOSED HYBRID STRATEGY

A combination of all three strategies, called a hybrid strategy, was created to make the best use of strengths. Initially, a DOE was carried out to decide how many features (named K) should be included, given a specified significance level ($\alpha = 0.05$). Once all the features had been analyzed, every feature selection method came up with a ranked list of them. Broda Counting was used to give each feature a score based on the various methods, and the subset with the top scores was selected from that set. Because of this approach, the chosen features showed high accuracy across various algorithms and could be used in different models. The created feature set was then included in model building and assessed to check how it affected the ability to make accurate predictions.

3.3. MACHINE LEARNING MODELS USED

A wide variety of machine learning algorithms were used to create predictive models that accurately assess civil infrastructure. [12-15] The purpose of this goal was to determine how well various families of algorithms captured nonlinear relationships in data gathered from Structural Health Monitoring (SHM) systems. The researchers chose these models because their successful use in regression and classification was confirmed in engineering experiments. Data was divided into training and validation sets using stratification to guarantee that the model performs well and learns from the whole dataset. The performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and the total time taken for training.

3.3.1. REGRESSION/CLASSIFICATION MODELS

Standard regression and classification algorithms form the initial category, which stands out for their versatility and ability to predict well. RFR was selected often for analysis since it can address nonlinear connections and is strong against both noise and overfitting. The resulting average of predictions from numerous decision trees in Random Forests minimizes the risk of making mistakes, which is useful when making infrastructure degradation estimates and remaining useful life (RUL) predictions.

Extreme Gradient Boosting (XGBoost) was also used, as it sequentially builds trees and fixes the errors each previous tree had. XGBoost typically runs fast and gives very accurate outcomes, especially on datasets collected by sensor networks. The method of Support Vector Regression (SVR) was chosen since it can model data with many predictors using a limited number of observations. SVR makes use of kernel functions to change data and map it to more complex spaces so that things like material fatigue caused by changing loads can be modeled. They were chosen based on how good they are at predicting, as well as how simple they are to implement and understand in real life.

3.3.2. ENSEMBLE LEARNING TECHNIQUES

Ensemble learning methods were a valuable addition to the way the model was built. Mixing different base learners with these techniques increases accuracy in prediction and generalization. Gradient Boosting Machines (GBMs), Light Gradient Boosting Machine (LightGBM) and CatBoost were the three ensemble strategies examined.

Gradient Boosting adds weak models one at a time to reduce error, and it works best when dealing with highly irregular and subtle patterns in data. LightGBM, a novel type, reduces training time and improves memory use by counting frequencies (histogram) and growing the model one layer at a time (leaf-wise). Therefore, it can be used for big SHM data with a large number of input variables. Alternatively, CatBoost focuses on handling categorical data, so it minimizes the preprocessing needed. In such cases, it was highly helpful because the data already indicated if the inspected item was a certain material or had a given type of defect. Using the ensemble approach along with individual models, the framework made sure the predictions were both accurate and effective for a broad range of civil infrastructure components. Results were checked and compared using all the features, specific feature subsets and features picked using a hybrid approach.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section examines and reviews the different results from machine learning models trained using civil infrastructure data and multiple feature selection techniques. The main topics of discussion were the characteristics of the data, the ways feature weighting was carried out, the results from the various models and how the suggested hybrid framework affected both prediction accuracy and computational costs.

4.1. DATASET DESCRIPTION AND CHARACTERISTICS

In the experimental part, data were collected from a number of sources, including sensors in bridge decks, sensors on high-rise buildings and loggers that record the environment. The data included more than 200 different measurements like stress and strain, displacement, acceleration, temperature, humidity and vibration frequencies. It recorded data for numerous months of operation, giving a thorough picture of how different loads and weather impacted the building.

A significant amount of preprocessing, including removing missing data, eliminating repeating entries and normalizing the data, happened before analysis. About 95% of the data collected before cleaning was retained, so the dataset reflects the true range and depth of real-life conditions. Experiments aimed to predict the displacement caused by a load, the degree of structural bending and signals for degradation. There were some skewness and outliers in the data, which were not removed to test the model's ability to handle unusual cases in real life.

4.2. FEATURE IMPORTANCE ANALYSIS

Assessing which features are most important was done with model-driven techniques as well as statistical practices. Random Forest and XGBoost were chosen to create rankings of importance for each feature by using information gain and split frequency, and LASSO regression indicated which features had the highest weights. Findings indicated that as few as 15–25% of the original features really improved the performance of the model, while most others were either unnecessary or weakly linked to the target variables.

Uniquely high-ranking measures were in-band vibration frequencies, stress-strain ratios and changes in temperature, pointing out that important indicators control how a structure behaves. Interestingly, each model gave slightly different importance to each feature, showing the importance of using several models to select features. It also highlights that SHM data is complex because different elements are linked, which makes relying on only one approach less certain.

4.3. MODEL PERFORMANCE COMPARISON

XGBoost and LightGBM had the highest overall performance, often achieving greatly reduced RMSE and MAE scores, unlike linear regression and support vector models. For predicting displacement, XGBoost resulted in an RMSE of 2.7 mm, which is much better than the RMSE of 5.3 mm from linear regression. Neural networks did almost as well as decision trees, especially when there were many interactions between the features, but took a lot longer to train.

All the features being used in every training situation led to more errors and required longer to train the models. Accuracy and efficiency were both increased by using individual features, but the highest accuracy was achieved when the hybrid approach was used, called Scenario 3. In this case, the models performed more accurately (up to 12% decrease in RMSE) and took less time to train (reduction up to 30%). This suggests that the more optimal input space sped up and improved their performance.

4.4. IMPACT OF HYBRID FEATURE SELECTION ON ACCURACY

Applying the hybrid method for feature selection, consisting of filter, wrapper and embedded methods, significantly increased results in both accuracy and efficiency. Using FSFS, Random Forest importance, LASSO and information gain techniques, the framework chose a group of features that always made the models more accurate. This method allows engineers in ML applications to find a balance between making predictions accurately and understanding the results easily.

Using the hybrid feature selection instead of all or any single method decreased the RMSE by an average of 10.4%, reduced MAE by 12.8% and lowered MAPE by 9.7%. It raised the capacity for the model to work on data it hasn't seen earlier, which can be noticed in the low difference between its training and validation scores. The improvements mean that the hybrid method successfully gets rid of unnecessary or repeated variables, making it concentrate on the important structures in the data. The approach was also shown to keep its features easy to interpret, so it is suited for use where practical SHM solutions must combine both performance and usability.

5. CHALLENGES IN DATA QUALITY AND SENSOR RELIABILITY

Building accurate models for civil infrastructure is made difficult by how well sensors in the field collect and distribute data. Structural Health Monitoring is used in difficult settings like bridges that get hit by adverse weather, tunnels that are vibrated by natural forces or high-rise buildings bent by wind. Noise, corruption of signals and missing data are caused by these settings. Improper alignment of sensors or their occasional failures, along with drifting over proceeding time, can lead to measurements that are not complete or accurate, which can make machine learning models give unreliable results.

Datasets used in real life often have uneven distributions, especially when dealing with events that occur rarely, like when a structure fails or starts deteriorating fast. Because of this imbalance, algorithms tend to pay more attention to bigger classes and less to smaller but important outcomes. Even after preprocessing with interpolation and normalization, the major problem of sensor unreliability still decreases the confidence of real-time predictive systems.

5.1. MODEL INTERPRETABILITY AND COMPUTATIONAL CONSTRAINTS

Advanced machine learning, such as ensemble methods and deep neural networks, gives good results, but they are not always easy to interpret for civil engineers. Trust and meeting regulations are hindered when engineers and infrastructure managers cannot explain why a specific prediction was produced by a black-box AI model. While the hybrid selection helps reduce the input data to key details, figuring out the inner workings of models such as XGBoost and neural networks is still a hard task. Engineers tend to select models that show what happens with simple formulas rather than opaque mathematical equations.

Furthermore, with large volumes of data arriving quickly, the challenge is usually how fast the computer system can handle it. Neural networks and similar high-performing models require lots of processing power and time, so they are not suitable for use in small embedded SHM systems. In places where resources are limited, using such models on their own often is not possible. Even though removing unnecessary features helps speed up models, it’s often still difficult to balance model speed with being responsive at any time.

6. CONCLUSION AND FUTURE WORK

This research showed a detailed way to improve the prediction of civil infrastructure using a mix of feature selection strategies and advanced machine learning approaches. The hybrid method of filter, wrapper and embedded techniques allowed the model to recognize the most important information in the data, and as a result, both the performance and computational efficiency improved. According to the results, models built using the hybrid-selected features were better than those using every feature or a single feature selection method in various situations. XGBoost, LightGBM and Random Forest performed most accurately, each time with the greatest accuracy matching the value found in the combining strategy.

Further, the analysis pointed out that being able to interpret the features used and having knowledge about the problem area are necessary for data-driven methods in civil engineering. Even though neural networks and other advanced models have shown excellent results, making them clear and trustworthy is still necessary for them to be widely used in important areas. The proposed design met the goal of high performance and clear interpretation by using different efficient ways to select important features. Only data sparsity, unreliable sensors and the challenge to scale models were found to be key limitations from the study. Therefore, systems must be flexible and well-built to function well in real situations. These areas need researchers to focus on how to interpret deep learning models and reduce the required processing power for devices located on the network edge.

Future improvements will handle real-time data streaming and incorporate online learning, making it possible to keep updating the model as new information arrives. Including tools that handle uncertainty will contribute to increasing the trust in safety-critical infrastructure situations. Connecting digital and physical spaces in which models interact over time can greatly increase a system’s ability to recover from disasters and be fixed before problems appear. The study ultimately offers a practical and adaptable approach for monitoring smart infrastructure, helping cities turn toward being safer, smarter and more sustainable.

REFERENCES

[1] Khan, M. W., Saad, S., Ammad, S., Rasheed, K., & Jamal, Q. (2024). Smart Infrastructure and AI. In AI in Material Science (pp. 193-215). CRC Press.

[2] Mondal, T. G., & Chen, G. (2022). Artificial intelligence in civil infrastructure health monitoring—Historical perspectives, current trends, and future visions. Frontiers in Built Environment, 8, 1007886.

[3] Brownlee, J. (2019). How to choose a feature selection method for machine learning. Machine Learning Mastery, 10, 1-7.

[4] Gangani, C. M. (2023). Role of machine learning in optimizing IT infrastructure. Kuwait Journal of Information Technology and Decision Sciences, 1(1), 12-22.

[5] Seow, J. H., & Rajaharia, N. K. (2023, March). Artificial Intelligence in Information Technology Infrastructure Management and Support in Singapore. In Sustainability, Economics, Innovation, Globalisation and Organisational Psychology Conference (pp. 447-471). Singapore: Springer Nature Singapore.

[6] Mishra, S., Tiwary, D., Ohri, A., & Agnihotri, A. K. (2018). Assessment of groundwater quality using WQI and GIS near the Karsara municipal landfill site, Varanasi, India. Arabian Journal of Geosciences, 11, 1-9.

[7] Deogun, J. S., Choubey, S. K., Raghavan, V. V., & Sever, H. (1998). Feature selection and effective classifiers. Journal of the American Society for Information Science, 49(5), 423-434.

[8] Chughtai, F., & Zayed, T. (2008). Infrastructure condition prediction models for sustainable sewer pipelines. Journal of Performance of Constructed Facilities, 22(5), 333-341.

[9] Visalakshi, S., & Radha, V. (2014, December). A literature review of feature selection techniques and applications: Review of feature selection in data mining. In 2014 IEEE international conference on computational intelligence and computing research (pp. 1-6). IEEE.
[10] Bocca, F. F., & Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. Computers and electronics in agriculture, 128, 67-76.
[11] Dowling, C. P. (2019). Applications of Statistical and Machine Learning to Civil Infrastructure (Doctoral dissertation).
[12] Tang, S., He, B., Yu, C., Li, Y., & Li, K. (2020). A survey on spark ecosystem: Big data processing infrastructure, machine learning, and applications. IEEE Transactions on Knowledge and Data Engineering, 34(1), 71-91.
[13] Jurado, S., Nebot, À., Mugica, F., & Avellana, N. (2015). Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy, 86, 276-291.
[14] Abdel-salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. Neural Computing and Applications, 36(33), 20723-20750.
[15] Jaafaru, H., & Agbelie, B. (2022). Bridge maintenance planning framework using machine learning, multi-criteria decision analysis and evolutionary optimization models. Automation in Construction, 143, 104585.
[16] Agarwal S. “Multi-Modal Deep Learning for Unified Search-Recommendation Systems in Hybrid Content Platforms”. IJAIBDCMS [International Journal of AI, BigData, Computational and Management Studies]. 2025 May 30 [cited 2025 Jun. 4]; 4(3):30-39. Available from: https://ijaibdcms.org/index.php/ijaibdcms/article/view/154
[17] Pulivarthy, P. (2023). Enhancing Dynamic Behaviour in Vehicular Ad Hoc Networks through Game Theory and Machine Learning for Reliable Routing. International Journal of Machine Learning and Artificial Intelligence, 4(4), 1-13.
[18] Praveen Kumar Maroju, "Optimizing Mortgage Loan Processing in Capital Markets: A Machine Learning Approach, " International Journal of Innovations in Scientific Engineering, 17(1), PP. 36-55 , April 2023.
[19] Animesh Kumar, “Redefining Finance: The Influence of Artificial Intelligence (AI) and Machine Learning (ML)”, Transactions on Engineering and Computing Sciences, 12(4), 59-69. 2024.
[20] Advancing Energy Management in Hybrid Electric Vehicles (HEVs) and Hybrid Energy Storage System Electric Vehicles (HESS EVs): A Hybrid Fuzzy Logic Approach - Sree Lakshmi Vineetha Bitragunta - IJSAT Volume 15, Issue 3, July-September 2024, PP-1-12, DOI 10.71097/IJSAT.v15.i3.2261.
[21] C. C. Marella and A. Palakurti, “Harnessing Python for AI and machine learning: Techniques, tools, and green solutions,” In Advances in Environmental Engineering and Green Technologies, IGI Global, 2025, pp. 237–250
[22] Kirti Vasdev (2024).” Spatial Data Clustering and Pattern Recognition Using Machine Learning”. International Journal for Multidisciplinary Research (IJFMR).6(1). PP. 1-6. DOI: https://www.ijfmr.com/papers/2024/1/23474
[23] Susmith Barigidad. “Edge-Optimized Facial Emotion Recognition: A High-Performance Hybrid Mobilenetv2-Vit Model". IJAIBDCMS [International JournalofAI,BigData,ComputationalandManagement Studies]. 2025 Apr. 3 [cited 2025 Jun. 4]; 6(2):PP. 1-10.
[24] Puneet Aggarwal, Amit Aggarwal. "Future-Proofing SAP HANA With Hybrid Cloud Architecture: Achieving Agility, Compliance, And Cost Efficiency", International Journal Of Science And Research, 13 (3), 1930-1937, 2024.
[25] S. Gupta, S. Barigidad, S. Hussain, S. Dubey and S. Kanaujia, "Hybrid Machine Learning for Feature-Based Spam Detection," 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2025, pp. 801-806, doi: 10.1109/CICTN64563.2025.10932459.
[26] Optimizing Boost Converter and Cascaded Inverter Performance in PV Systems with Hybrid PI-Fuzzy Logic Control - Sree Lakshmi Vineetha. B, Muthukumar. P - IJSAT Volume 11, Issue 1, January-March 2020,PP-1-9,DOI 10.5281/zenodo.14473918
[27] Multiconnected Interleaved Boost Converter for Hybrid Energy System, Sree Lakshmi Vineetha Bitragunta, International Journal Of Scientific Research In Engineering And Management (Ijsrem), Volume: 08 Issue: 03 | March – 2024, PP-1-9.
[28] Patel, Bhavikkumar; Mallisetty, Harikrishna; and Rao, Kolati Mallikarjuna, "Artificial Intelligence Helper Application for Delivering Effective Presentations", Technical Disclosure Commons, (January 04, 2024) https://www.tdcommons.org/dpubs_series/6572
[29] Dr. Priya. A., Dr. Charles Arockiasamy J., “The Global Reach of AI: A Postcolonial Analysis of Technological Dominance,” *International Journal of Scientific Research in Science and Technology*, 11(2), 1-5, 2025.