SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Integrated AI and Image Processing Frameworks for Structural Monitoring, Mechanical Performance, and Engineering Optimization

[1]Dr. V. Senthilkumar, [2]Binod Kumar Pattanayak, [3]Anly Antony M, [4]Dr Brahm Prakash, [5]Mridini M Gawas, [6]Patel Dixitkumar Bharatbhai

[1]Professor, Department of Mechanical Engineering, SRM TRP Engineering College, Trichy, Tamilnadu – 621105, E-mail: trpvsk12@gmail.com
[2]Professor, Department of Computer Science and Engineering, Institute of Technical Education and Research, Siksha O Anusandhan Deemed to be University, Bhubaneswar, Odisha, India, E-mail: binodpattanayak@soa.ac.in
[3]Assistant Professor, Department of Computer Science and Engineering, Sahrdaya college of Engineering and Technology(Autonomous), APJ Abdul Kalam Technological University, E-mail: anlyantony1991@gmail.com
[4]Associate Professor & HOD CSE(Cyber Security) MITs Deemed to be University brahmprakashdahiya@gmail.com Orchid ID:- https://orcid.org/0000-0002-1755-5251
[5]Assistant Professor, School of Physical and Applied Science, Goa University, E-mail: mridinigawas@unigoa.ac.in
[6]Associate Professor, Department of Civil Engineering, Gokul Global University, Siddhpur, Gujarat, 384151.
E-mail: dixitpatel18795@gmail.com

Abstract
India requires frameworks for condition assessment and material design that are both infrastructure-scaled and budget-conscious while retaining engineering accountability. This study presents an artificial-intelligence and image-processing decision framework that connects road-damage monitoring, concrete compressive-strength modelling, and observed-space engineering optimisation at an analytical level. A field-scale proof-of-concept is not attempted; instead, three connected modules with demonstrated value are presented using Indian data. For the monitoring task, the India subset of the RDD2022 database is used, which consists of 9,665 road images; the training set contains 6,831 damage annotations. Of these annotations, potholes dominate the frequency at 46.7% compared with 1.0% for transverse cracks, a 46.9:1 ratio. Bounded region classification was piloted on the 1,221-image India data subset allocated with N-RDD2024. Histogram-of-oriented gradients, local texture, intensity, colour, and bounding-box descriptors yielded 1,826 features. An RBF support-vector machine, selected using five-fold stratified group validation, achieved 79.7% accuracy, 0.682 balanced accuracy, and 0.727 macro-F1 on the supplied held-out validation split. The recall for transverse cracking was low at 0.375. For mechanical performance, 188 experimental recycled-aggregate concrete records from IIT Bhubaneswar were grouped into 60 composition groups. With five-fold group-wise validation, Extra Trees were chosen as the regressor, with an out-of-fold R^2 of 0.893, an RMSE of 5.85 MPa and an MAE of 4.23 MPa. Curing age, fly ash, total cementitious material, and water-to-total-cementitious ratio were the most influential variables. For the optimisation task, a retrospective search was conducted within the subset of measured 28-day compressive strengths to screen conservative minimum-cement concrete recipes for 35, 45, and 55 MPa targets. It is concluded that a viable Indian framework would combine vision-based low-cost triage, material models with reduced leakage, observed-space optimisation with a transparent safety margin, and mandatory engineer acceptance, rather than relying on black-box AI to replace established inspection, testing, and design practices.

Keyword: artificial intelligence; structural health monitoring; image processing; road damage; recycled aggregate concrete; compressive strength; explainable machine learning; engineering optimisation; India

Introduction

Structural monitoring has shifted from ad-hoc visual inspection towards evidence-based diagnosis, but the underlying engineering imperative remains the same: damage has to be discoverable at a stage when responsible intervention is feasible, and the supporting evidence has to be compelling enough for an engineer to take action. Farrar and Worden have captured the essence of structural health monitoring as a statistical pattern-recognition problem in which "operational and environmental variability are inextricably linked to response features related to damage" [1]. In subsequent reviews, the focus has shifted towards wireless, mobile and vision-based sensors, an indication of the importance of contact-less measurement, but also the awareness of the need for calibration and physical interpretation [2-4]. Within this context, artificial intelligence appears most valuable not as an overarching classifier, but rather as a means to accelerate the formation of evidence amenable to prioritized engineering action.

The need is particularly pressing in the Indian infrastructure sector, in which the growth in demand for infrastructure has outpaced the evolution of inspection norms. According to the Ministry of Road Transport and Highways, National Highway length has grown from 91,287 km in 2014 to 146,195 km in 2024 [28]. This highway network spans vast metropolitan regions, urban clusters, high-temperature plains, coastal areas, the Himalayas, and areas of heavy monsoon rainfall. Fully manual inspection is prohibitively expensive and inconsistent, while fully automated computer vision solutions are not viable, given the frequent changes in road appearance due to weather, roadwork, markings, shadows, water-logging, camera angle, and local traffic composition. An Indian-specific infrastructure monitoring solution needs to be both high-frequency (re-deployable at the level of National Highways Authority of India, State Public Works Departments and municipal services) and conservative (fit for regulatory approval).

A similar consideration applies to material engineering. India generated 340.11 million tonnes and utilised 332.63 million tonnes of fly ash in 2024-25 [29]. According to one government report, "road and flyover construction activity; cement industry; and production of bricks and tiles account for 32%, 27% and 14%, respectively, of total fly ash generated" [30]. Construction and demolition recycling, as well as the use of fly ash and other industrial by-products in concrete, can reduce pressure on natural aggregate supplies. However, recycled aggregate concrete needs to be carefully engineered since its performance characteristics depend on gradation, water content, recycled-to-primary aggregate ratio, and degree of curing among other factors. Indian and global researchers have therefore used a combination of experimental and machine learning approaches to explore the performance characteristics of recycled aggregate concrete [16-20]. However, we see that the practical value of these insights is limited because the machine-learning models must be interpreted to arrive at what we may term conservative specifications. What is required is an evidence chain that goes from field reports through laboratory work, uses traceable validation rules, and applies predictive models conservatively within physically realistic parameters. This paper reports on three related issues. First is the use of road-damage images as a basis for large-scale evidence in infrastructure monitoring, which is a major requirement in a country like India with a large and diverse road network. Second, we present a study that predicts the compressive strength of concrete mixes using Indian laboratory data and uses group-wise validation so that the same mix family is not included in both training and test sets.

Third, it uses the measured 28-day compressive-strength subset to screen the observed mix space, identifying potential alternatives within the experimental domain and applying a one-OOF-RMSE safety margin rather than defining calibrated prediction intervals. Together, these contributions form an analytical decision-support architecture in which computer vision provides field evidence, mechanical prediction informs laboratory trial mixes, and the two evidence streams meet at a conservative decision layer.

The contribution is based on field evidence from road images, laboratory records that report mechanical performance, prediction-leakage avoidance, and conservative optimisation of the observed space. We also see that the approach's value is brought to light by the negative results, which in this case are of particular note because of the limited recall performance in transverse cracking. This, in turn, brings into question the deployability of what may appear to be a very effective vision-based infrastructure monitoring system.

2. Related Research and Indian Engineering Context

2.1 Image intelligence for structural and pavement monitoring

Computer vision has potential in both local response and global analysis. The works of Dong and Catbas [3], and Spencer and others [4] illustrate the range from crack, spalling and corrosion detection to displacement, modal response, and vibration measurement. Region-based convolutional networks have been used to identify multiple structural damage types [5]. Comparative studies of concrete cracks show that learning-based approaches can often be preferable to intensity-based edge detection algorithms for certain appearances [6]. Nevertheless, although all these works are relevant to India, their reported accuracy is not directly comparable because the input data distributions differ across countries and depend on the height of the camera, road conditions, maintenance technology, traffic composition, and weather. A detector that has been trained on well-prepared close-range images may not work reliably in the monsoon or on images captured through the windscreen of a moving car.

Conversely, we see that smartphone-based apps are a useful way to move out of the lab and into real-world application. Maeda et al. [7] report on the use of deep neural networks for road-damage detection from smartphone images. The RDD program has grown from its start in several countries to the RDD2020 project [8], which provides pooled data [9], and to the present RDD2022 dataset, which has 47,420 images from six countries [10]. We see that between-country variation is large and that local and ensemble approaches are key strategies

[11]. As for India, what is important is not only that an app can run on a smartphone, but that the target dataset represents defect types, road conditions, and operating environments relevant to the country.

2.2 AI for mechanical-performance prediction

Concrete strength prediction can be addressed within the paradigm of nonlinear machine learning, because the binder chemistry, water-to-binder ratios, aggregate gradation, recycled-aggregate qualities, admixture dosages, and curing age are interrelated. The IIT Bhubaneswar experimental program includes the use of recycled coarse aggregate in combination with fly ash, ground granulated blast furnace slag, and metakaolin [16,17]. An earlier study conducted in India evaluated the re-use potential of construction and demolition aggregates for concrete production [18]. Poon et al. demonstrate that the moisture condition of recycled aggregate influences both fresh and hardened properties [19]. These findings rationalize the need for predictive modelling while underscoring the value of laboratory certification.

Tree ensembles can be a preferred choice of model for the analysis of structured concrete data due to their ability to capture interactions without postulating any particular functional form. The crucial difference between Random Forests [21], Extremely Randomized Trees [22], and Gradient Boosting [23] lies in the balance between their variance and bias. Nevertheless, the reliability of prediction always depends on the validation design: if, for instance, records from mixes with the same proportions at different ages were divided between the training and validation sets, the result would mainly reflect relationships within the same mix group. In this study, mix proportions were used as groups, not cube-age observations.

2.3 Engineering optimisation and circular material use

Concrete optimisation is a multi-criteria problem involving strength, workability, durability, cost, embodied carbon and material availability. Yeh proposed a computer-aided framework that utilised predictive function and constrained mix optimisation [20], and non-dominated sorting provides a general language for identifying solutions in which improvement in one direction must be offset by deterioration in another [25]. In Indian practice, such approaches should be considered as adjuncts to trial mixes, constituent testing and code provisions. IS 10262:2019 provides guidance on concrete mix-proportioning [26], and IS 456:2000 specifies the general design and construction requirements for plain and reinforced concrete [27]. Hence, the present study is limited to a screening within an observed laboratory space, and the concrete mix design does not override the requirements of IS 456:2000.

3. Materials and Methods

3.1 Integrated research design

The proposed architecture comprises two evidence streams. The field stream is generated using georeferenced road images, inspection records and, if available, response or environmental sensors. The image processing, quality control and damage classification steps yield local evidence related to the asset chainage, each one associated with its confidence and uncertainty levels. The laboratory stream is built from the concrete proportions and measured cube strength, and passes through group-validated mechanical models and evidence-optimisation layers. Both streams meet at a decision layer that is designed to support condition scores, repair priorities and candidate materials. Inspection and laboratory activities then update the data rather than models being fixed once and for all. Figure 1 shows the architecture.

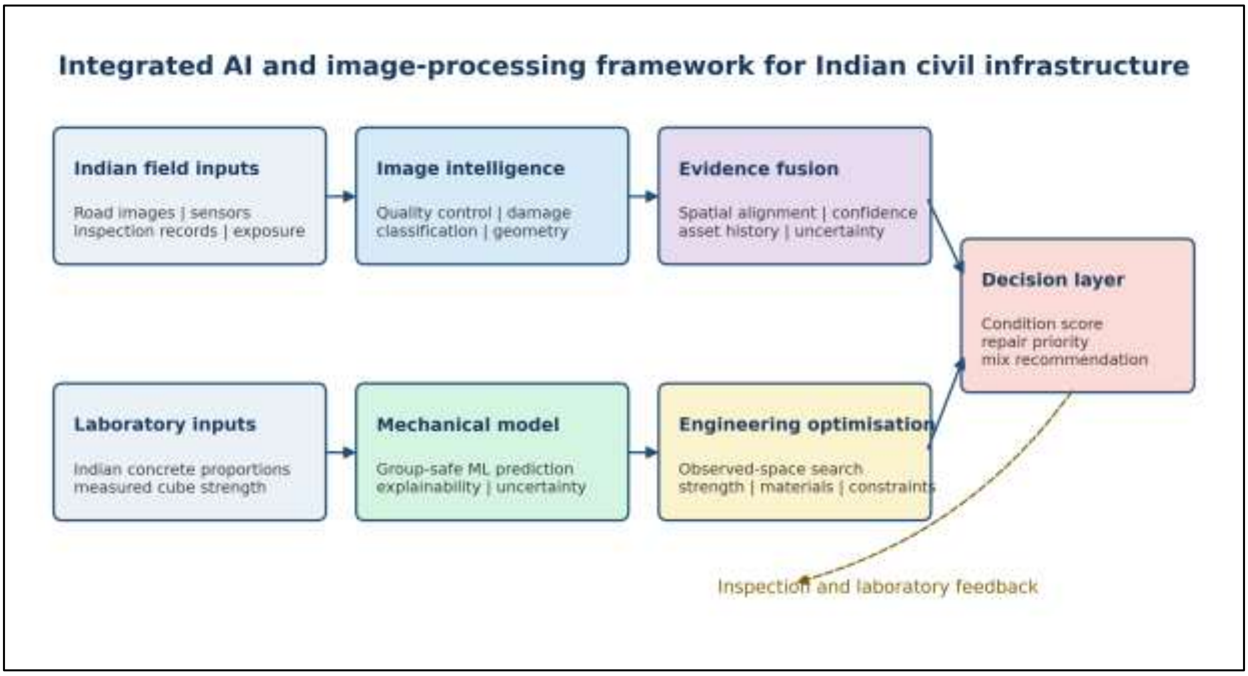

Figure 1. Integrated AI and image-processing framework for Indian civil infrastructure. The two empirical branches are connected at the decision layer; synchronized field deployment is proposed rather than claimed in this study.

3.2 Data sources and analytical roles

Three public datasets and two Indian context-specific government statistics sources were used. The RDD2022 release provided the deployment-scale class distribution for India [10,31]. The India VOC subset in version 3 of N-RDD2024 provided the images with Pascal VOC annotations for a reproducible region-classification pilot study [12]. The IIT Bhubaneswar concrete dataset provided the mechanical performance data [16,17]. Finally, the government road network and fly-ash statistics were used for providing the context for India's construction sector [28-30], and are not used as predictors. Table 1 makes the distinction for clarity, so that national context, training data, and study results are not conflated.

Table 1. Traceable Indian evidence used in the study.

Source	Indian coverage	Analytical role	Licence / authority
RDD2022 [10,31]	7,706 train + 1,959 test images; Delhi, Gurugram and mainly Haryana	Dataset-scale class and imbalance audit	CC BY 4.0 dataset
N-RDD2024 India VOC v3 [12]	1,221 images; 2,800 annotated objects in supplied train/validation splits	Four-class region-classification pilot	CC BY 4.0 dataset
IIT Bhubaneswar concrete [16,17]	188 laboratory records; 60 unique mix-composition groups	Strength prediction, interpretation and observed-space search	CC BY 4.0 dataset
PIB / MoRTH and environmental releases [28–30]	National highways and 2024–25 fly-ash statistics	Indian infrastructure and circularity context only	Government of India

3.3 Road-image audit and bounded classification pilot

RDD2022 describes four classes of pavement damage: longitudinal cracks (D00), transverse cracks (D10), alligator cracks (D20) and potholes (D40). The numbers of occurrences were taken from the published India training set, which was cross-verified with the stated value of 6831 annotations. The distribution was calculated as a percentage of each class from the total, while balance was calculated as a ratio of the largest to the smallest class. This is an analysis of readiness, rather than a performance evaluation

For image processing, all the 1221 images and their corresponding VOC XML annotations from the provided N-RDD2024 India split were parsed. Of these, 976 were training images, containing 2244 objects, while 245 were validation images, containing 556 objects. To maintain consistency with the core RDD2022 class set, only the D00, D10, D20 and D40 classes were retained for analysis. This reduced the number of training regions to 1792 across 835 images and the number of validation regions to 449 across 212 images. The original India image code was used to identify image identities, which were found to be disjoint between the two splits. Each ground-truth box was expanded by 10 percent on either side, clipped to the image borders and resized to 64 × 64 pixels. As such, this experiment evaluates the ability to classify a region of interest rather than perform localisation, object detection or image-level tasks.

The descriptor consists of 1826 variables, including 1764 HoG features, calculated with 8 × 8 cells, 2 × 2 blocks and nine unsigned orientations [13], 32-bin LBP histogram, 16-bin greyscale histogram, RGB mean and standard deviation, four gradient statistics and four normalised box geometry measurements. Three class-weighted models were trained: linear SVM, RBF SVM [14] and Extra Trees [22]. The models were evaluated using five-fold stratified group validation, performed on the training split, with the original image identities as groups and macro-F1 score as the selection criterion. The best model was retrained on all the training regions and evaluated once on the provided validation split. Accuracy, balanced accuracy, macro-precision, macro-recall, macro-F1 and weighted-F1 scores were computed. The scores are reported with 95 percent bootstrap confidence intervals, computed with 2000 iterations sampling image identities from the validation split, with replacement. Note that this procedure respects the original split and only samples whole images, thus leaving multiple regions from the same image in the sample.

3.4 Concrete database preparation

The 188 records from the IIT Bhubaneswar source PDF file were parsed into 17 columns: cement, fly ash, GGBS, metakaolin, TCM (total cementitious materials), water, water/TCM ratio, superplasticiser, viscosity-modifying admixture, natural and recycled coarse aggregates (20 mm and 10 mm), sand, age, measured compressive strength, and serial number. All 188 records were used, with the check that cement + fly ash + GGBS + metakaolin = TCM in all rows having passed without issues. The 188 records were grouped into 60 unique mix-composition groups having the same mix variables, excluding age, measured strength, and serial number. For each group, the four regressors described below were compared using five-fold shuffled GroupKFold cross-validation with random state 42. Since all the entries for a given composition were in one group, the question was whether a model would be able to generalise to an unseen mix group, rather than interpolate by using records from the same composition at different ages. Ridge regression was used as a linear regularized baseline, plus Random Forest [21], Extra Trees [22] and Gradient Boosting [23]. All models were implemented in scikit-learn [24].

3.5 Performance metrics and model selection

For measured strength y_i and prediction $\hat{y}_i$, mean absolute error, root-mean-square error and coefficient of determination were calculated for each fold and after concatenating all out-of-fold predictions. For selection, the Extra Trees model was chosen based on the lowest mean fold RMSE, with R^2 and MAE serving as secondary metrics.

$$\mathrm{MAE} = (1/n)\, \Sigma_i\, |y_i - \hat{y}_i| \tag{1}$$

$$\mathrm{RMSE} = \sqrt{[(1/n)\, \Sigma_i\, (y_i - \hat{y}_i)^2]} \tag{2}$$

$$R^2 = 1 - [\Sigma_i\, (y_i - \hat{y}_i)^2 \,/\, \Sigma_i\, (y_i - \bar{y})^2] \tag{3}$$

3.6 Predictive interpretation

Grouped permutation importance, referring to aggregation across group-wise validation folds rather than joint permutation of correlated variables, was evaluated on each validation fold by permuting one variable at a time in the held-out fold and measuring the increase in root mean squared error (RMSE). The value reported for each variable is the average increase in RMSE across the 5-fold validation. The interpretation of these numbers should be that the fitted predictors depend on these variables—it is a measure of variable importance, not a material effect. Thus, cement, TCM, water and water/TCM ratio should be interpreted jointly since they are related during the design process and their individual contribution cannot be separated or interpreted physically by this analysis.

3.7 Observed-space engineering optimisation

Optimisation was carried out on unique compositions that had an observed 28-day result. The Extra Trees model was refitted to all 188 cases and used to predict the 28-day strength for these mixes. The supplementary cementitious material (SCM) proportion was defined as (fly ash + GGBS + metakaolin)/TCM, and the recycled coarse aggregate fraction was determined by dividing the recycled by the total coarse aggregate. A mix was defined as non-dominated if there was no other observed mix that used no more cement, had no lower predicted strength, and no lower SCM proportion, with at least one strict improvement [25].

For target strength f_t, a candidate was screened as feasible only when its model prediction minus one global out-of-fold RMSE met the target, as shown in Equation (4). Among passable options, the one with the least cement was chosen; higher proportion of supplementary materials and higher conservative strength estimates were secondary criteria in case of a tie. The one-RMSE reduction is a transparent screening margin that reflects normal engineering caution, not a calibrated 95% lower confidence limit.

$$\hat{\mathbf{f}}_{28}(\mathbf{x}) - \mathbf{RMSE}_{oo}\mathbf{f} \geq \mathbf{f}_t \qquad (4)$$

The screened candidate mixes are decision-support examples. They necessitate characterisation of constituents, determination of fresh properties, evaluation of durability, trial batching and conformity to IS 10262 and IS 456 prior to their structural use [26,27].

4. Results

4.1 Readiness of the Indian road-image evidence

The RDD2022 India subset comprises of 7,706 training images and 1,959 test images, making up a total of 9,665 images. The labeling distribution for the training set was heavily skewed towards potholes and alligator cracking. Out of 6,831 labels, 3,187 of them were potholes, 2,021—alligator cracks, 1,555—longitudinal cracks, and only 68—transverse cracks. This makes the largest-to-smallest ratio as high as 46.9 to 1. As Table 2 and Figure 2 demonstrate, high overall accuracy could be misleading when the last category is underrepresented.

Table 2. RDD2022 India training-label distribution.

Code	Damage type	Instances	Share (%)	Relative to D10
D00	Longitudinal crack	1,555	22.8	22.9×
D10	Transverse crack	68	1.0	1.0×
D20	Alligator crack	2,021	29.6	29.7×
D40	Pothole	3,187	46.7	46.9×
Total	All four classes	6,831	100.0	—

Note. Shares are computed from the published India training counts [10,31]. Rounding may cause the displayed percentages to sum to 100.1%.

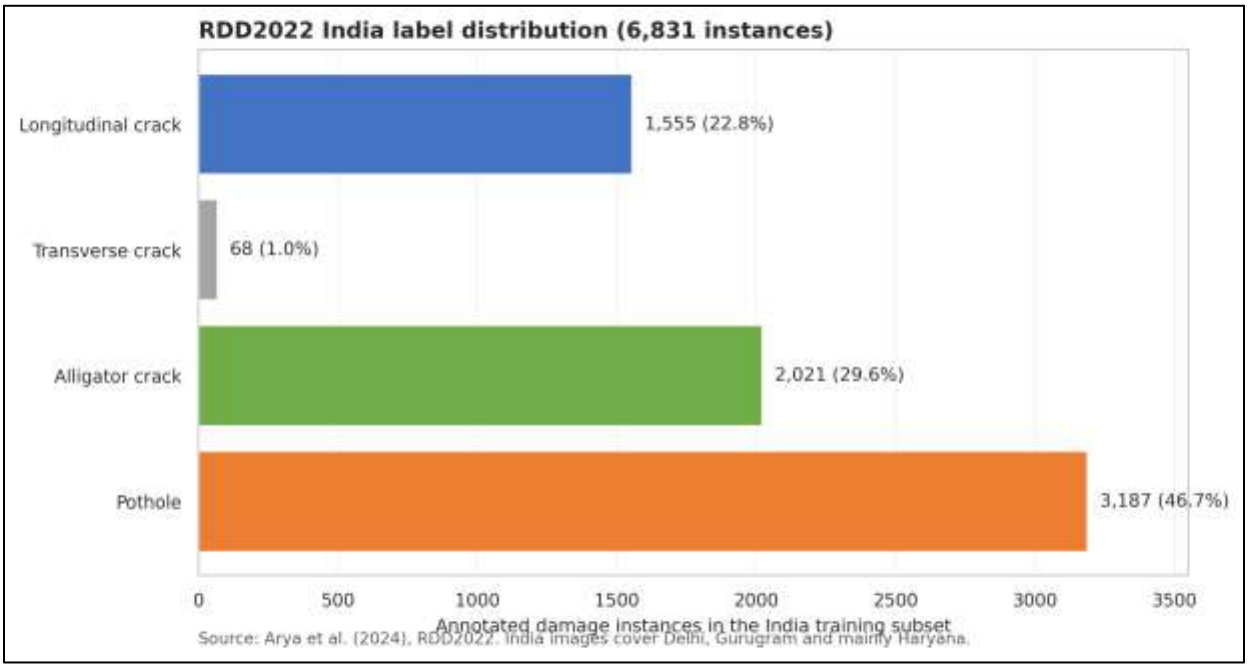

Figure 2. RDD2022 India label distribution. The visual imbalance is a deployment concern because transverse cracks form only 1.0% of annotated instances. Source data: Arya et al. [10,31].

4.2 Road-damage region-classification pilot

The RBF SVM achieved the best mean macro-F1 in grouped inner validation (0.708 ± 0.049) compared to Extra Trees (0.607) and linear SVM (0.584). On the held-out validation split, it achieved 79.7% accuracy, 0.682

balanced accuracy, 0.727 macro-F1, and 0.794 weighted-F1. The image-group bootstrap interval was 75.4 – 83.7% for accuracy and 0.594 – 0.810 for macro-F1.

Table 3. Training-group model selection and held-out validation performance for the road-region pilot.

Model	Inner macro-F1	Inner balanced acc.	Inner accuracy	Held-out macro-F1
RBF SVM	0.708 ± 0.049	0.678	0.794	0.727
Extra Trees	0.607 ± 0.051	0.584	0.754	0.625
Linear SVM	0.584 ± 0.036	0.580	0.722	0.599

Note. Model selection used five-fold stratified group validation on the training split. Held-out metrics are shown for transparency; only the preselected RBF SVM is interpreted as the pilot result.

Class-level analysis thus highlights the performance differences of the model. Potholes have the highest recall and F1-score, reaching 0.882 and 0.842, respectively. The F1-score for alligator cracks is 0.773, while it is 0.748 for longitudinal cracks. Transverse cracks have a precision of 1.000 because the model rarely predicts D10. However, their recall is only 0.375 because five out of eight D10 validation regions were predicted as potholes. The model has only 27 training regions and 8 validation regions for this class, so the precision estimate is unstable and should not be considered a reliable metric.

Table 4. Per-class performance of the selected RBF SVM on 449 held-out annotated regions.

Code	Damage type	Train n	Valid. n	Precision	Recall	F1
D00	Longitudinal crack	415	104	0.787	0.712	0.747
D10	Transverse crack	27	8	1.000	0.375	0.545
D20	Alligator crack	559	134	0.785	0.761	0.773
D40	Pothole	791	203	0.806	0.882	0.842

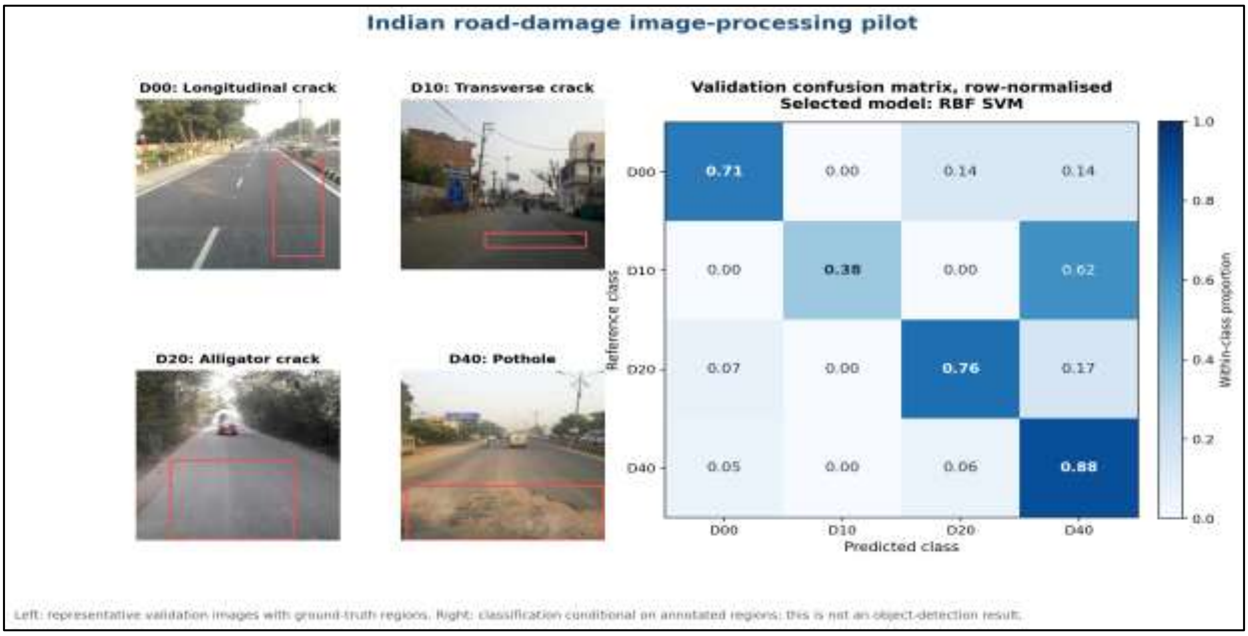

Figure 3. Indian road-damage image-processing pilot. Representative validation images are reproduced under CC BY 4.0 from the N-RDD2024 India subset [12]. The confusion matrix is conditional on ground-truth regions and is not an object-detection result.

4.3 Concrete database characteristics

The concrete database included 3, 7, 14, 28, 56, and 90-day strengths with 32, 49, 17, 49, 17, and 24 records, respectively. Measured compressive strength varied between 6.00 and 73.76 MPa, with an average of 34.42 MPa. The water-to-TCM ratio varied between 0.25 and 0.75, while cement content varied between 113 and 680 kg/m^3. The coarse recycled aggregates were used as the only coarse aggregates in 176 out of 188 records, which is consistent with a report that indicated recycled concrete aggregate was utilised as the only coarse aggregate in most samples [17]. Table 5 below provides an overview of variables relevant to interpretation and optimisation.

Table 5. Selected descriptive statistics for the IIT Bhubaneswar concrete records (n = 188).

Variable	Mean ± SD	Median	Range
Cement (kg/m^3)	338.57 ± 146.91	309.00	113–680
Fly ash (kg/m^3)	50.00 ± 101.27	0.00	0–385
GGBS (kg/m^3)	60.02 ± 113.02	0.00	0–385
Metakaolin (kg/m^3)	11.45 ± 26.73	0.00	0–102
TCM (kg/m^3)	460.04 ± 145.28	486.00	226–680
Water (kg/m^3)	170.45 ± 14.17	168.00	146–221
Water/TCM	0.418 ± 0.164	0.350	0.25–0.75
Curing age (days)	27.45 ± 28.37	14.00	3–90
Strength (MPa)	34.42 ± 17.94	—	6.00–73.76

Note. TCM = total cementitious material; GGBS = ground-granulated blast-furnace slag.

4.4 Leakage-safe strength prediction and interpretation

Extra Trees produced the lowest mean group-wise fold RMSE, which was 5.59 MPa, and hence was chosen. The concatenated out-of-fold RMSE for Extra Trees is 5.85 MPa. Additionally, its MAE is 4.23 MPa, and its R^2 equals 0.893. Gradient Boosting was the second best, whereas ridge regression was the worst. This result indicates a performance gap between nonlinear ensemble decision trees and ridge regression in this laboratory experiment; however, no statistical significance test was performed. The difference between Extra Trees and Gradient Boosting was smaller and does not establish statistical superiority.

Table 6. Five-fold composition-group validation for concrete compressive-strength prediction.

Model	MAE (MPa)	RMSE (MPa)	R^2	OOF RMSE	OOF R^2
Extra Trees	4.15 ± 1.16	5.59 ± 1.49	0.894 ± 0.057	5.85	0.893
Gradient Boosting	4.49 ± 1.01	6.01 ± 1.22	0.881 ± 0.042	6.18	0.881
Random Forest	4.90 ± 1.10	6.54 ± 1.35	0.858 ± 0.054	6.75	0.858
Ridge	5.72 ± 0.70	7.03 ± 1.00	0.836 ± 0.046	7.15	0.840

Note. Mean ± sample SD across five folds. OOF = metrics after concatenating all out-of-fold predictions. Every age record for a composition remained in the same fold.

Grouped permutation importance, aggregated across the group-wise validation folds, increased the validation-set RMSE by 6.05 MPa for curing age, 3.97 MPa for fly ash, 3.44 MPa for TCM, 2.80 MPa for water-to-TCM ratio, and 1.34 MPa for cement. Superplasticiser was a secondary predictor, with its permutation increasing validation RMSE by 0.59 MPa. For all other individual predictors, the increase was below 0.20 MPa (see Figure 4), showing that the fitted model was less reliant on each individual predictor within this validation, not that these materials are unimportant. The plot also shows the out-of-fold agreement, with greater scatter at high strengths, likely due to a lack of data points, and predictions showing regression toward the centre.

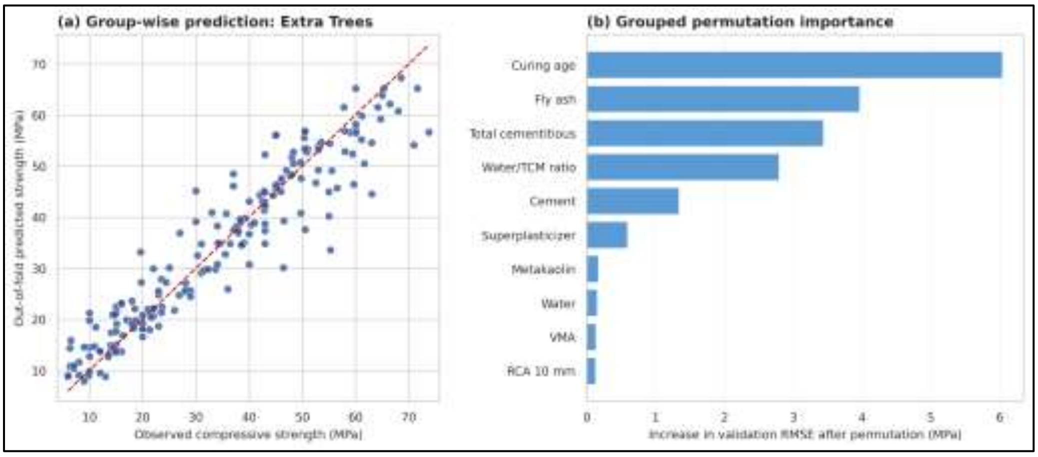

Figure 4. Composition-group Extra Trees predictions and grouped permutation importance. Variables were permuted individually; importance is predictive, not causal, and correlated mix variables should be interpreted jointly.

4.5 Conservative optimisation within measured mix space

Within the retrospective 28-day observed candidate space there were 47 unique composition groups, of which 9 were non-dominated according to the cement–strength–supplementary-material criteria. The use of the one-RMSE strength deduction produced distinct minimum-cement candidates for the 35, 45, and 55 MPa targets. The 35 MPa candidate used 165 kg/m^3 of cement and 385 kg/m^3 of GGBS; its modelled strength was 47.62 MPa, compared to its margin-adjusted value of 41.77 MPa. The 45 MPa candidate used 275 kg/m^3 of cement and an equal proportion of GGBS, with its conservative value being 53.56 MPa. In turn, the 55 MPa candidate used 468 kg/m^3 of cement and 82 kg/m^3 of metakaolin, with a conservative value of 61.13 MPa. Finally, all three candidates used recycled aggregate as the coarse-aggregate fraction in the recorded mix.

Table 7. Minimum-cement candidates meeting target strength after a one-OOF-RMSE deduction.

Target	Cement	SCM	w/TCM	RCA 20/10	Pred.	Margin-adj.	Observed	SCM (%)
35	165	385 GGBS	0.31	322/373	47.62	41.77	47.69	70.0
45	275	275 GGBS	0.29	349/404	59.41	53.56	59.00	50.0
55	468	82 MK	0.35	337/389	66.97	61.13	70.92	14.9

Note. All mass values are kg/m^3 and all strength values are MPa. SCM = supplementary cementitious material; MK = metakaolin; RCA 20/10 = 20 mm/10 mm recycled coarse aggregate; w/TCM = water-to-total-cementitious ratio. Superplasticiser/VMA/sand were 4.40/1.10/729, 6.05/1.10/789 and 6.05/0.28/760 kg/m^3 for the 35, 45 and 55 MPa rows, respectively. These are screened laboratory candidates, not approved structural mix designs.

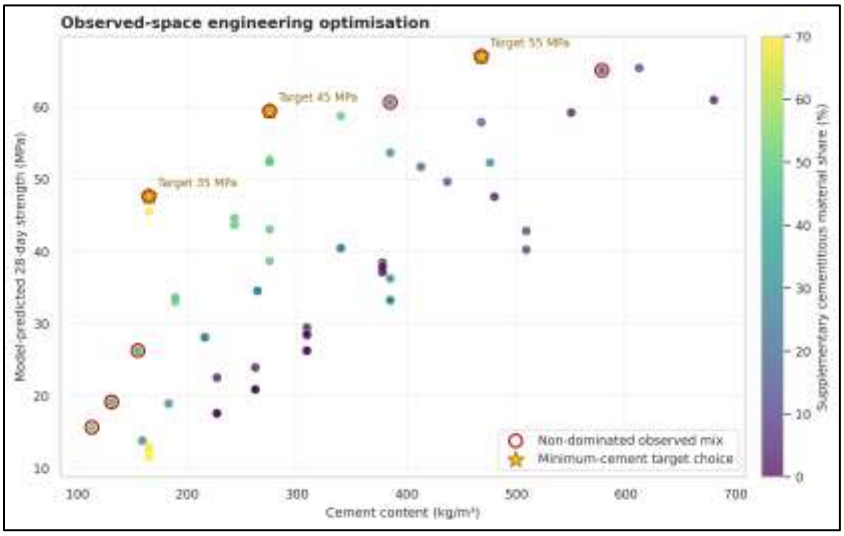

Figure 5. Observed-space engineering optimisation. The search does not extrapolate beyond measured composition families; gold stars denote minimum-cement choices after subtracting one out-of-fold RMSE from predicted strength.

5. Discussion

5.1 Meaning of the integrated framework

The two empirical branches address different engineering tasks, but both involve the same discipline. In the road branch, an image model transforms an unstructured observation into a class and confidence. In the concrete branch, a regression model converts proportions and curing age into a strength estimate. In both cases, the result is not an end in itself but must be combined with other evidence about location, exposure, history, uncertainty and an acceptance rule. With this in mind, it is best to think of the integrated framework as an evidence-management system: visual information may be used to identify promising locations for further examination, while mechanical prediction and optimisation can structure laboratory trials. The road pilot serves as an important reminder that distinguishing between detection and classification is crucial: the pilot's 79.7% accuracy applies only to cases where a ground-truth region has been identified. In practice, a field system must also locate damaged areas, discard false positives, accumulate relevant statistics and associate detected geometries with maintenance quantities. Deep detectors and segmentation networks offer one possible solution [5,7-11], but these methods must be evaluated using Indian sampling and acceptance criteria. The current conventional descriptor provides a useful baseline against which to evaluate these alternatives. It demonstrates that the D10 class is still challenging, even when location information is available.

The concrete results presented above demonstrate the benefit of a leakage-safe benchmark. For this small and probably heterogeneous laboratory database, the out-of-fold performance is good—0.893 R^2 and 5.85 MPa OOF RMSE—but no theoretical maximum was used for comparison. Grouping-based partitioning provides a more credible assessment than completely randomised record-level partitioning, but the error is still significant and may be unacceptable for some applications. However, subtracting one RMSE from the predicted value before selecting candidates provides a transparent safety margin during screening and may reduce overly optimistic choices. This effectively means that the error budget has been considered at the decision-making stage, but it is not a calibrated reliability score. It acts more conservatively than selecting candidates by the point estimate alone.

5.2 Practical pathway for Indian agencies and laboratories

A practical road-monitoring implementation can begin with a calibrated smartphone or dash-mounted camera on an inspection vehicle. Images should be time-stamped and geo-referenced to route, chainage and carriageway; poorly focused images and images of water-covered or shadowed road surfaces should be tagged for exclusion from inference. The model result should be logged to an asset register as a potential defect observation with image, location, class confidence and model version. High-consequence or low-confidence detections should be sent for visual inspection by a junior engineer and formal acceptance by an authorised engineer. Repeated surveys enable the categorisation of persistent or progressive defects. This workflow represents a reasonable allocation of responsibility, suitable for Indian State PWDs and ULBs, as it preserves professional accountability without eliminating domain knowledge from the process.

Class balance should be managed as an acquisition priority. The 68 transverse-crack labels in RDD2022 would not be sufficient to achieve reliable state-level generalisation, and class weighting will not compensate for the absence of positive examples from different road surfaces, repair regimes, lighting conditions, seasons and so on. Focal loss or resampling can be applied to this task [15], but they cannot compensate for the absence of D10 appearance diversity. It would be preferable to collect D10 samples with a focus on surfaces, weather and roadworks. Equally, a bridge deck crack detector should not be expected to generalise to pier spalling, bearing corrosion or other defect classes without site-specific evidence from each potential deployment region.

For concrete laboratories, the most promising application is the development of a model-driven database for trial-mix records. For each trial batch that uses the laboratory's materials, a record should be made of cement, fly ash, GGBS, metakaolin, recycled aggregate, chemical admixture, moisture corrections, fresh properties, curing regimes and specimen details. The model can be used to prioritise trial mixes, and the actual results can be filed in the database. This is more valuable in a context with many local variations in fly ash and recycled aggregate, as is the case in India [29,30], rather than a single national model.

5.3 Reliability, governance and standards

An engineering AI system must have a documented operating domain. For the road branch, this should include the states and road types covered, camera hardware, capture speed, weather, minimum defect size and classes. For the concrete branch, this should include constituent ranges, mix ages, test methods, and the laboratory population. Incoming data outside of these ranges must be flagged as out-of-domain rather than returning apparently reasonable predictions. Model cards, version control, audit logs, and periodic blind testing are therefore technical requirements, not luxuries.

Human review should be prioritized based on risk. A false negative road surface crack is not the same as an erroneous recommendation for structural concrete. The required assurance level for a road-surface image should depend on asset criticality and the consequence of error. A concrete recommendation requires independent trial batches, mean strength calculations, workability and durability tests, and a review of IS 10262 and IS 456 before any recommendation can be finalized [26,27]. The model must retain the original row of data and the explanation for why a recommendation was made so that an engineer can understand why a candidate was chosen.

6. Limitations

The road-image study has four caveats, the first being that the empirical pilot is a classifier conditional on annotated boxes, not an end-to-end detector/segmentation model, and therefore cannot estimate whole-image false positives, negatives, crack width, area of damage, or Pavement-Condition Index. Second, the Indian images are concentrated heavily in Delhi, Gurugram, and Haryana and are not representative of India’s wide range of climates, geologies, and road-building practices. Third, D10 is underrepresented, and its precision and recall are therefore poorly estimated. Fourth, N-RDD2024 was created using modified copies of the original RDD2022 images. The images were divided into training and validation sets while preserving image IDs, but the validation set was not independently surveyed by Indian authorities. Finally, the concrete study has limited scope: only 188 records from one lab's 60 composition groups are used in this pilot. The mix variables are not independent, so individual-variable permutation importance does not support causal interpretation, while models using recycled aggregates, fly ash, GGBS, metakaolin, or admixtures may need more testing. Compressive strength was the only performance outcome modelled; slump, density, modulus of elasticity, splitting tensile strength, shrinkage, permeability, durability, cost, and embodied carbon were not modelled. The one-RMSE rule of thumb is a conservative way to screen candidates, but is not a design factor itself—an optimisation searching solely across 28-day strengths could miss a better-performing untested mix.

Finally, the integration was demonstrated at an analytical and decision-architecture level. Road images and concrete laboratory records were not obtained from the same asset, and there was also no response sensor stream available. A future prospective study would entail associating repeated field images, inspection records, and load/response measurements with material batches, repairs, and subsequent performance observations under one asset identifier.

7. Conclusions

This study highlights the potential value of an integrated AI and image-processing system for India, provided that such technology embraces traceable data and recognizes jurisdictional confines. The audit of RDD2022 India revealed a 46.9:1 class imbalance, while the bounded N-RDD2024 region pilot achieved only 79.7% validation accuracy and 0.727 macro-F1 with a 0.375 transverse-crack recall. Therefore, the main lesson that can be drawn from this analysis is that the use of bounded region classification as an analytical tool has its limitations, and a targeted data-gathering approach is required before triaging road damage in India at the national level.

As for compressive-strength performance, the composition-group validated 188 IIT Bhubaneswar records allowed Extra Trees to achieve an out-of-fold R^2 of 0.893, an RMSE of 5.85 MPa, and an MAE of 4.23 MPa. The retrospective observed-space search consequently identified minimum-cement candidates for 35, 45, and 55 MPa targets after applying one OOF-RMSE deduction. This workflow screens measured candidate mixes and can be used to prioritise trial mixes that include recycled materials and supplementary cementitious materials, but it does not bypass Indian standards.

The recommended implementation, accordingly, is a human-in-the-loop framework that employs low-cost imaging for inspections, group-safe mechanical models for laboratory trials, conservative reductions for optimisation, and engineering judgement for approvals. This approach is suggested for further consideration in the Indian context because it focuses on both scale and affordability while also being compatible with jurisdiction-specific due-diligence requirements.

References

1. Farrar CR, Worden K. An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A. 2007;365:303–315. https://doi.org/10.1098/rsta.2006.1928
2. Sony S, Laventure S, Sadhu A. A literature review of next-generation smart sensing technology in structural health monitoring. Structural Control and Health Monitoring. 2019;26:e2321. https://doi.org/10.1002/stc.2321
3. Dong C-Z, Catbas FN. A review of computer vision-based structural health monitoring at local and global levels. Structural Health Monitoring. 2021;20(2):692–743. https://doi.org/10.1177/1475921720935585
4. Spencer BF Jr, Hoskere V, Narazaki Y. Advances in computer vision-based civil infrastructure inspection and monitoring. Engineering. 2019;5(2):199–222. https://doi.org/10.1016/j.eng.2018.11.030
5. Cha Y-J, Choi W, Suh G, Mahmoudkhani S, Büyüköztürk O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Computer-Aided Civil and Infrastructure Engineering. 2018;33(9):731–747. https://doi.org/10.1111/mice.12334
6. Dorafshan S, Thomas RJ, Maguire M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Construction and Building Materials. 2018;186:1031–1045. https://doi.org/10.1016/j.conbuildmat.2018.08.011
7. Maeda H, Sekimoto Y, Seto T, Kashiyama T, Omata H. Road damage detection and classification using deep neural networks with smartphone images. Computer-Aided Civil and Infrastructure Engineering. 2018;33(12):1127–1141. https://doi.org/10.1111/mice.12387

8. Arya D, Maeda H, Ghosh SK, Toshniwal D, Mraz A, Kashiyama T, Sekimoto Y. Deep learning-based road damage detection and classification for multiple countries. Automation in Construction. 2021;132:103935. https://doi.org/10.1016/j.autcon.2021.103935
9. Arya D, Maeda H, Ghosh SK, Toshniwal D, Sekimoto Y. RDD2020: An annotated image dataset for automatic road damage detection using deep learning. Data in Brief. 2021;36:107133. https://doi.org/10.1016/j.dib.2021.107133
10. Arya D, Maeda H, Ghosh SK, Toshniwal D, Sekimoto Y. RDD2022: A multi-national image dataset for automatic road damage detection. Geoscience Data Journal. 2024;11(4):846–862. https://doi.org/10.1002/gdj3.260
11. Arya D, Maeda H, Sekimoto Y. From global challenges to local solutions: A review of cross-country collaborations and winning strategies in road damage detection. Advanced Engineering Informatics. 2024;60:102388. https://doi.org/10.1016/j.aei.2024.102388
12. Kaya Ö, Çodur MY. N-RDD2024: Road damage and defects, version 3. Mendeley Data. 2024. https://doi.org/10.17632/27c8pwsd6v.3
13. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005;1:886–893. https://doi.org/10.1109/CVPR.2005.177
14. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20:273–297. https://doi.org/10.1007/BF00994018
15. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. 2017:2980–2988. https://doi.org/10.1109/ICCV.2017.324
16. Biswal US, Mishra M, Singh MK, Pasla D. Experimental investigation and comparative machine-learning prediction of the compressive strength of recycled aggregate concrete incorporated with fly ash, GGBS and metakaolin. Innovative Infrastructure Solutions. 2022;7:242. https://doi.org/10.1007/s41062-022-00844-6
17. Biswal US, Pasla D, Mishra M. Experimental data set for concrete compressive strength prediction from IIT Bhubaneswar concrete laboratory, version 2. Mendeley Data. 2022. https://doi.org/10.17632/5wkxzmzwnz.2
18. Rao A, Jha KN, Misra S. Use of aggregates from recycled construction and demolition waste in concrete. Resources, Conservation and Recycling. 2007;50(1):71–81. https://doi.org/10.1016/j.resconrec.2006.05.010
19. Poon CS, Shui ZH, Lam L, Fok H, Kou SC. Influence of moisture states of natural and recycled aggregates on the slump and compressive strength of concrete. Cement and Concrete Research. 2004;34(1):31–36. https://doi.org/10.1016/S0008-8846(03)00186-8
20. Yeh I-C. Computer-aided design for optimum concrete mixtures. Cement and Concrete Composites. 2007;29(3):193–202. https://doi.org/10.1016/j.cemconcomp.2006.11.001
21. Breiman L. Random forests. Machine Learning. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324
22. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine Learning. 2006;63:3–42. https://doi.org/10.1007/s10994-006-6226-1
23. Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics. 2001;29(5):1189–1232. https://doi.org/10.1214/aos/1013203451
24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html
25. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 2002;6(2):182–197. https://doi.org/10.1109/4235.996017
26. Bureau of Indian Standards. IS 10262:2019, Concrete Mix Proportioning—Guidelines. Second revision. New Delhi: BIS; 2019. https://www.services.bis.gov.in/
27. Bureau of Indian Standards. IS 456:2000, Plain and Reinforced Concrete—Code of Practice. Fourth revision. New Delhi: BIS; 2000. https://www.services.bis.gov.in/
28. Press Information Bureau, Government of India. Year End Review 2024: Ministry of Road Transport and Highways. 2024. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2091508
29. Press Information Bureau, Government of India. Fly Ash Management in Focus as Railways & NTPC Bring Stakeholders Together. 2025. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2160714
30. Press Information Bureau, Government of India. India's Green Pathway. 2026. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2249751