

Research paper

Image-based prediction of concrete mechanical properties through CNN segmentation and ANN modeling ^{*}Mati Ullah ^{id a}, Vagelis Plevris ^{id a}, Junaid Mir ^{id b}, Afaq Ahmad ^{id c,*}^a Department of Civil and Environmental Engineering, College of Engineering, Qatar University P.O. Box, 2713, Doha, Qatar^b Department of Electrical Engineering, University of Engineering & Technology Taxila, Pakistan^c Department of Built Environment, Oslomet University, Norway

ARTICLE INFO

Keywords:

Image processing

Deep learning

CNN

Aggregate segmentation

Concrete mechanical properties

Structural health monitoring

ABSTRACT

Accurate prediction of the mechanical properties of concrete is essential for assessing structural performance and ensuring safety. Traditional destructive and non-destructive testing methods are labor-intensive, costly, and often impractical for evaluating in-service structures due to specimen extraction requirements, calibration challenges, and limited field accuracy. This study presents an automated image-based framework that integrates Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) models to predict the compressive and tensile strengths of concrete from image-derived features. High-resolution images of concrete cylinders with different mix ratios and water-cement ratios were used to train a DeepLabV3+ segmentation model with EfficientNet-B5 as the encoder. The model achieved a mean Intersection over Union of 0.92 and an F1-score of 0.95 for coarse aggregate segmentation. Image analysis algorithms were then applied to extract aggregate characteristics, including proportion, size, roundness, and nearest-neighbor distance. Meanwhile, statistical descriptors of mortar pixels were derived to represent the matrix properties. These features served as inputs to multiple ANN models. The hybrid AggMortarNet model, combining aggregate and mortar characteristics, achieved the best predictive performance with $R^2 = 0.9680$ and RMSE = 0.0363 for compressive strength, and $R^2 = 0.9607$ and RMSE = 0.0368 for tensile strength, outperforming non-destructive methods. The results demonstrate an automated, data-driven approach with strong potential for Structural Health Monitoring (SHM) and quality assessment of both newly cast and existing concrete structures.

1. Introduction

Concrete plays a fundamental role in civil engineering structures due to its availability, strength, durability, and adaptability, making it indispensable for buildings, bridges, dams, and other infrastructure [1]. Accurate prediction of its mechanical properties is crucial for ensuring the safety, serviceability, and longevity of structures subjected to varying environmental conditions, loads, and stresses. Among these, compressive strength (f_c) and tensile strength (f_t) are the most widely used indicators in structural health assessment [2,3]. Traditionally, such properties have been determined through destructive and nondestructive testing. Destructive testing, conducted under controlled laboratory conditions, requires extracting core samples to measure compressive and split tensile strength, but this process is expensive, labor-intensive, and can compromise the integrity of the structure [4]. Nondestructive

methods such as ultrasonic pulse velocity (UPV) [5], rebound hammer (RH) [6], and SonReb [7] overcome some of these issues by estimating strength through empirical relationships without physical damage; however, they demand costly equipment, careful calibration and maintenance, and highly skilled operators [8]. The shortcomings of these conventional methods have paved the way for advanced approaches that provide accurate and non-destructive prediction of concrete properties, ensuring practical evaluation of existing infrastructure and real-time monitoring.

Over the past few decades, machine learning (ML) has gained prominence as a sophisticated alternative to conventional methods for evaluating the mechanical properties of concrete. Unlike traditional destructive or non-destructive techniques, ML-based approaches offer faster processing, improved precision, and the ability to capture complex, non-linear interactions between material parameters and strength outcomes

^{*} This work was supported by the project entitled “Forensic Analysis of Concrete Through Image Processing” FACIP, funded by the EU under the Grant agreement ID: 101153307, <https://cordis.europa.eu/project/id/101153307>.

^{*} Corresponding author.

E-mail addresses: matiullah@qu.edu.qa (M. Ullah), vplevris@qu.edu.qa (V. Plevris), junaid.mir@uettaxila.edu.pk (J. Mir), afaq.ahmad@oslomet.no, afaq.ahmad@oslomet.no (A. Ahmad).

<https://doi.org/10.1016/j.rineng.2025.108637>

Received 15 July 2025; Received in revised form 13 November 2025; Accepted 7 December 2025

Available online 21 December 2025

2590-1230/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

[9,10]. Moreover, these data-driven methods can be continuously refined with additional datasets, making them adaptable, cost-effective, and highly reliable tools for predicting concrete behavior [11,12]. Numerous studies have applied ML methods to assess the mechanical properties of concrete, using material composition as input variables and measured compressive strength as the output [13]. For instance, Timur et al. [14] employed ML regression models to estimate compressive strength and slump values using seven mix-related parameters. These parameters included factors such as aggregate type and cement fineness. In a study by Zeng et al. [15], three ML models, Support Vector Machine (SVM), ANN, and Adaptive Boosting (AdaBoost), are trained on images from 380 cube specimens with varying cement grades, water-binder ratios, and mineral admixtures, achieving about 95 % accuracy in compressive strength prediction. Abd et al. [16] employed SVM to predict the compressive strength of lightweight foamed concrete, using parameters such as density, sand size, and foam ratio, with an achieved accuracy of 98.7 %. Gradient Boosting, K-Nearest Neighbors, and Extreme Gradient Boosting (XGBoost) were applied in a study by Momshad et al. [17] to assess the engineering properties of banana leaf ash concretes. SHapley additive explanations and partial dependence plot analyses identified curing age, cement content, and ash proportion as dominant predictors. Sobuz et al. [18] evaluated the mechanical strength of recycled glass-based concretes with ensemble methods (XGBoost, Gradient Boosting, Random Forest, and AdaBoost), with XGB achieving the highest predictive accuracy of $R^2 = 0.95$. Ali et al. [19] conducted a comparative analysis of regression and ANN models for self-compacting green concrete, demonstrating that the ANN outperformed regression models with a performance of $R^2 = 0.94$ and $RMSE = 3.56$ MPa. Mohammed et al. [20] developed hybrid neuro-swarm (PSO-ANN) and neuro-imperialism (ICA-ANN) models for predicting strength of fly ash-modified concretes. The neuro-swarm model achieved superior accuracy with $R^2 = 0.91$. Zhao et al. [21] proposed optimized ML models for predicting the compressive strength of fiber-reinforced concretes. They found that the integrated Bald Eagle Search-XGBoost (BES-XGBoost) achieved the best performance, with $R^2 = 0.991$, $RMSE = 2.93$ MPa, and $MAPE = 3.8\%$, respectively. Gou et al. [22] applied gene expression programming (GEP), decision trees (DT), and SVR, to predict the compressive strength of self-compacting concrete incorporating recycled coarse aggregate and supplementary cementitious materials. GEP demonstrated superior accuracy with R^2 greater than 0.94, outperforming DT and SVR. Cement content, curing age, and water-to-binder ratio were identified as the most influential parameters.

The aforementioned methods have achieved high accuracy in predicting the strength of concrete in newly designed or laboratory-prepared mixes. Their application to predicting the mechanical properties of existing structures remains limited, as the original mix composition is often unknown or impractical to obtain. This limitation has driven the use of computer vision methods that predict the strength of existing concretes directly from image data, eliminating the need for prior mix information [23]. High-resolution images processed by deep learning (DL) models enable the extraction of critical features, and when trained on large labeled datasets, these models can generalize well to accurately estimate concrete properties [23]. Various studies have utilized image processing (IP) and DL models to assess concrete material properties based on image-derived data [24,25]. Bacsyiugit et al. [26] used linear, multi-linear, and non-linear regression analyses to predict the compressive strength of concrete specimens with varying W/C ratios (0.37 to 0.79) by combining image-derived percentages of aggregate, air voids, and cement matrix with UPV and HR test results. The multi-linear regression yielded the highest R^2 value of 0.9482. Dogan et al. [27] developed an IP and ANN-based method to predict the compressive strength, modulus of elasticity, and maximum deformation of concrete. Analyzing 144 images, the ANN achieved prediction accuracies between 97.18 % and 99.87 %, outperforming the traditional methods. In a study by Waris et al. [28], high-resolution images of 162 concrete slices with varying Cement Replacement Materials (CRM) ratios were analyzed to

extract statistical features of the image, ie, mean, median, standard deviation, skewness, third moment, and entropy. These features were used to train and validate ANN and Adaptive Neuro-Fuzzy Inference System (ANFIS) models, which were then compared with destructive and non-destructive test results. The method achieved high accuracy, with R^2 values of 99.7 % for ANN and 97.8 % for ANFIS, respectively. Similarly, Plevris et al. [29] proposed a nondestructive framework for predicting the mechanical properties of hybrid concretes incorporating silica fume and fly ash by integrating IP with ML. High-resolution concrete images were processed to extract statistical and texture-based features such as mean, skewness, entropy, and Local Binary Patterns (LBP), which were then used as input to the SVM, boosted ensemble regression (BRE), and Gaussian process regression (GPR) models. Among these, GPR achieved the highest predictive accuracy for both compressive and tensile strength, demonstrating its effectiveness in modeling complex nonlinear relationships.

Recent studies have demonstrated the potential of convolutional neural network (CNN) models to process concrete images directly, offering improved accuracy in predicting concrete strength. Lee et al. [30] used the AlexNet model to predict the concrete compressive strength from 4709 DSLR and 3846 microscope images. The DSLR images achieved a lower RMSE of 3.779 and a MAPE of 12.25 % compared to the microscope images with an RMSE of 4.875 and a MAPE of 18.31 %. The results indicated that DSLR images, with a wider range, were more effective in predicting accurate strength. Shiuly et al. [31] predicted compressive strength of cement mortar and concrete from digital images by using SVM and five CNN models (AlexNet, GoogleNet, VGG19, ResNet, and Inception-ResNet-V2). Inception-ResNet-V2 achieved the best results, with R^2 values of 0.9 for mortar and 0.85 for concrete, along with the lowest RMSE. Similarly, CNN models were utilized by Jang et al. [32] to estimate concrete compressive strength from images captured with a portable digital microscope. The DCNN models, especially the ResNet-based model, outperformed the ANN model, achieving the highest R^2 value of 0.96. In a study by Fang et al. [33], the effect of pore meso-characteristics on the tensile strength of cellular concrete was investigated with the UNet segmentation model. Surface porosity and pore size were identified as key determinants of splitting-tensile strength. A DL approach is developed by Yu et al. [34] to predict the compressive strength degradation of cement-based materials in marine environments. The deep learning model outperforms other commonly used models in predicting strength degradation caused by sulfate attack. Celik and Izdemir [24] proposed a hybrid model that integrates a CNN (MPaConvNet) with SVR to extract visual features from 1453 concrete surface images obtained from 203 core samples taken from real buildings. Their approach achieved an exceptional prediction accuracy with $R^2 = 0.9959$, outperforming traditional destructive and non-destructive methods, as well as other ML models.

1.1. Research significance

In most studies, the ML models for predicting concrete properties are trained on newly cast mixes with known compositions, making them practical for new designs but less applicable to existing structures where mix details are unknown. Although some studies have used surface images for strength prediction in existing concrete specimens, these methods still face notable limitations. For example, the study [26] relied on manual thresholding of grayscale images to separate aggregates, cement, and voids, which made the results highly dependent on material type and limited their generalizability. Further, it required combining image data with UPV and RH results, rendering the approach semi-empirical rather than fully image-based. Similarly, the studies [27–29] used hybrid concretes incorporating CRM and additives, which improve mortar color and texture, thereby simplifying pattern recognition IP and ANN models. Normal concrete without CRM lacks such visual enhancements, making aggregate-mortar boundaries less discernible and presenting a more complex challenge for image recognition models in pre-

dicting compressive strength. The studies [30–32] relied on CNN regression models. However, regression-based approaches may oversimplify complex relationships by focusing on general patterns rather than the detailed features of concrete components, such as aggregate size, shape, and distribution, which have a significant impact on concrete strength. If coarse aggregates in images appear gray, their similarity in color to fine aggregates and cement particles may cause regression models to overlook the aggregate characteristics, impacting prediction accuracy. Furthermore, in [24], improved predictive performance was reported by supplementing image-based features with manually measured aggregate diameter. While this hybrid approach enhanced the model's accuracy in estimating compressive strength, it rendered the method only partially automated. The reliance on manual measurements introduces additional effort, making the process cumbersome and less practical for large-scale or rapid field applications where fully automated image-based methods are more desirable.

This study introduces an automated, IP-inspired framework for predicting the mechanical properties of concrete using a DL-based coarse aggregate segmentation approach. The CNN segmentation model, DeepLabV3+, utilizes various backbone encoders, including Inception-ResNetV2, ResNet-101, Xception, and EfficientNet-B5, to isolate aggregates from mortar even in normal concrete with visually indistinct, gray-colored aggregates. From these segmented images, key aggregate properties, including proportion, size, shape, and spatial distribution, are automatically quantified using IP methods. Additionally, the impact of mortar on concrete strength is also considered by extracting statistical features of mortar pixels and incorporating them into an ANN model along with the aggregate properties. This dual-focus strategy allows the distinct contributions of aggregates and mortar to compressive strength to be captured, leading to a more robust and interpretable prediction model. By eliminating the need for manual measurements and offering an end-to-end automated pipeline, the proposed methodology enhances accuracy, efficiency, and generalizability, establishing a novel and practical pathway for evaluating the mechanical properties of both newly cast and in-service concretes. The current study is primarily based on laboratory-prepared concrete specimens under controlled imaging and testing conditions. However, the developed framework shows strong potential for field implementation, where it can be adapted for in-situ evaluation of existing structures using portable imaging systems and automated prediction models.

The key contributions of this study are:

1. A high-resolution image dataset of concrete cylinders with varying mix designs, W/C ratios, and compaction level was developed, featuring pixel-level ground truth masks. This dataset is the first to provide standard images focusing on gray aggregates specifically for deep learning-based concrete aggregate segmentation and strength prediction.
2. A customized DeepLabV3+ segmentation framework was developed and thoroughly evaluated across multiple DCNN backbone architectures to assess its segmentation performance.
3. Image processing algorithms were designed to quantify aggregate characteristics such as proportion, size, roundness, packing angle, and nearest-neighbor distance, enabling detailed evaluation of particle distribution.
4. An integrated hybrid prediction model (AggMortarNet) combined CNN-derived aggregate features with mortar statistical descriptors in an ANN framework, achieving high prediction accuracy and outperforming traditional UPV and RH methods.

The remainder of this paper is organized as follows: Section 2 presents the materials and experimental procedures. Section 3 details the segmentation of coarse aggregates. Section 4 represents image analysis techniques for the extraction of input features and evaluating the correlation between them. Section 5 describes the development and performance evaluation of the ANN models for strength prediction. Finally,

Table 1
Types and characteristics of materials.

Material	Description	Source / Origin
Fine Aggregates	Crushed Stone	Granite, Limestone
Coarse Aggregates	Crushed Stone	Granite, Limestone
Cement	OPC	Not Applicable

Section 6 summarizes the key findings, limitations, and recommendations for future research.

2. Materials and methods

The complete methodology starts with casting concrete cylinders using different mix ratios and W/C ratios, producing samples in compacted, non-compacted, and high W/C categories. DT and NDT testing are conducted to assess the compressive and tensile strengths. The cylinders are then sectioned, and high-resolution images are captured. These images are meticulously labeled to segment aggregates in the concrete images, creating a curated image dataset for further analysis. The flowchart depicted in Fig. 1 provides an overview of these processes. A CNN model is trained on the labeled images to isolate aggregates from the mortar. Subsequently, IP methods are applied to extract the essential features. The extracted numerical features are used as input, and experimental results from compressive and tensile strength tests are used as output for the ANN model training. Finally, the results of the predictions for DT, NDT, and ML are compared to evaluate the precision and reliability of the proposed framework.

2.1. Casting of cylinders

We prepared concrete samples using five distinct mix design ratios with varying proportions of cement, sand, and aggregates. The mix ratios employed were 1:4:8, 1:3:6, 1:2:4, 1:1.5:3, and 1:1:2. Table 1 presents the materials used in the concrete mixtures, along with their type and provenance. The aggregate used included gray-colored aggregates from granite and limestone rocks alongside colored aggregates from basalt, siliceous, and calcined materials, with a composition of approximately 70% gray aggregates and 30% colored aggregates. The higher proportion of gray-colored aggregates can help generalize the segmentation algorithms due to their color resemblance to cement paste. The distinct color of non-gray aggregates simplifies the segmentation task by creating clear boundaries for the object, making the segmentation task trivial. Additionally, fine and coarse aggregates with gray color are preferred options in the construction sector across various South Asian countries due to their abundance and suitability with regional construction materials [35].

As shown in Fig. 2, the coarse aggregate fraction primarily consists of particles retained between 4.75 mm and 25 mm sieves, while the fine aggregates pass through smaller sieve sizes down to 0.15 mm. The aspect ratio (length-to-width) of coarse aggregate generally varied between 1.2 and 1.8. The largest particle size recorded was 24.9 mm. The aggregate blend consisted of two main fractions: approximately half with particles passing the 25.6 mm sieve and the other half passing the 12.8 mm sieve. Prior to mixing, these fractions were carefully combined to achieve a uniform distribution, thereby enhancing the homogeneity of the concrete mix.

To investigate the effect of the W/C ratio on workability and aggregate distribution, each mix was made with three different W/C ratios: 0.4, 0.5, and 0.6, which resulted in fifteen unique mix designs as shown in Table 2. These procedures followed the standards outlined in ACI 318 [36] and ASTM C 94 [37]. A set of non-compacted cylinders and cylinders with very high W/C ratios was also cast to study the effect of unhealthy conditions, such as non-compaction and very high W/C ratios, on the mechanical properties. Overall, 350 cylindrical concrete samples were prepared according to the specified mix designs. For each

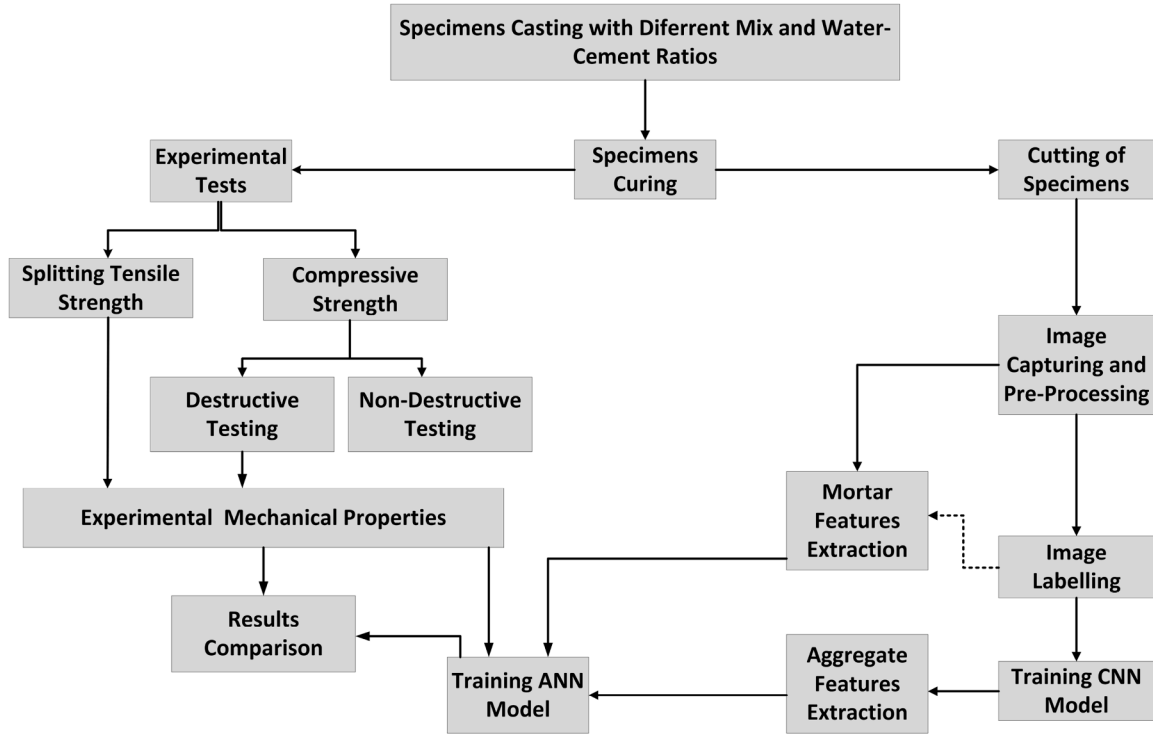


Fig. 1. Flowchart depicting stepwise procedure from dataset curation to ANN training.

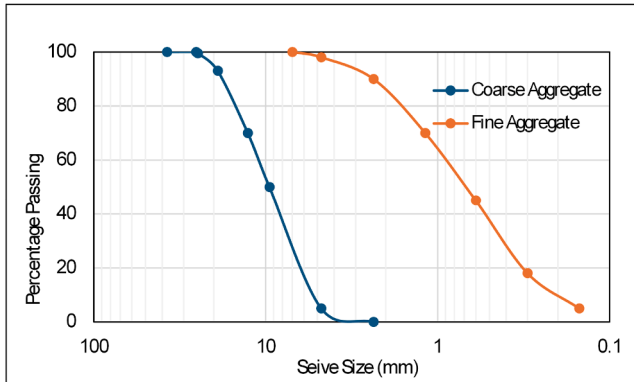


Fig. 2. Gradation curve of coarse and fine aggregates.

mix ratio and W/C ratio in each category, ten cylinders were cast. Two were used for IP, four for compressive strength tests, and four for tensile strength tests. Fig. 3 represents the stepwise procedure from cylinder casting and curing to dataset creation. All cylinders were demolded after 24 h and cured in water for 28 days before testing to ensure full strength development.

2.2. Destructive testing

The destructive testing results were employed as target values for the training and validation of the ANN model. To obtain these results, destructive tests were carried out to determine the mechanical properties of concrete using a compression testing machine (CTM). Compression and split-cylinder tests were conducted on concrete cylinders to assess their response to compressive and tensile stresses. The compressive strength and split tensile test of hardened concrete was performed according to standard test methods, ASTM C39 [38], and ASTM C496 [39], respectively, as depicted in Fig. 3.

Compressive strength: Compression tests were performed on right circular concrete cylinders using a compression testing machine equipped with a spherically seated upper platen to ensure uniform load distribution. The specimen ends were prepared to meet planeness and perpendicularity requirements. Each specimen was carefully centered on the lower platen using etched alignment marks, and the load was applied continuously and without shock at a controlled stress rate of 0.25 ± 0.05 MPa/s until failure, following ASTM C39 [38]. The compressive strength was calculated as

$$f_c = \frac{P_{\max}}{A} \quad (1)$$

where $P_{\max}$ is the maximum applied load and A is the average corrected cross-sectional area.

Splitting tensile strength: For the split-cylinder test, each cylinder was placed horizontally between the platens with two bearing strips, extending the full length of the specimen. A diametral line was drawn on each end to ensure the load passed through the cylinder's generatrix. Load was increased continuously to failure at a controlled splitting stress rate of 0.7–1.4 MPa/min, per ASTM C496 [39]. The splitting tensile strength was computed as

$$f_t = \frac{2P_{\max}}{\pi LD} \quad (2)$$

where L and D are the measured specimen length and diameter.

Table 2 displays the compressive and tensile strength experimental results. The variation in density among the investigated mixes showed a direct influence on both compressive and tensile strength outcomes. In general, specimens with higher density exhibited superior mechanical performance, while reductions in density corresponded to notable strength losses. For the compacted mixes, densities in the range of 2320–2549 kg/m³ resulted in compressive strengths up to 21 MPa and tensile strengths above 3 MPa. This reflects the denser microstructure achieved through proper compaction, which reduces entrapped air and improves the load transfer between cement paste and aggregates. In contrast, non-compacted mixes displayed comparatively lower densities, typically between 2270–2490 kg/m³, and this reduction was associated

Table 2

Overview of mix designs for sample casting, corresponding destructive testing results, and calculated density.

S.No.	Mix Design Notation	Mix Ratios	W/C	Category	Proportions in Kg/m ³				Density (kg/m ³)	Compressive Strength (MPa)	Tensile Strength (MPa)
					Cement	Fine Aggregate	Coarse Aggregate	Water Content			
1	CM1	1:4:8	0.4	Compacted	177.09	746.01	1544.04	70.84	2321.02	3.43	0.81
2	CM2	1:4:8	0.5	Compacted	177.09	746.01	1544.04	88.55	2341.85	4.06	0.97
3	CM3	1:4:8	0.6	Compacted	177.09	746.01	1544.04	106.25	2360.58	5.16	1.23
4	CM4	1:3:6	0.4	Compacted	230.22	727.36	1505.44	92.09	2357.79	5.84	1.32
5	CM5	1:3:6	0.5	Compacted	230.22	727.36	1505.44	115.11	2370.63	7.40	1.43
6	CM6	1:3:6	0.6	Compacted	230.22	727.36	1505.44	138.13	2412.73	8.78	2.15
7	CM7	1:2:4	0.4	Compacted	328.89	692.72	1433.75	131.56	2450.10	14.53	2.68
8	CM8	1:2:4	0.5	Compacted	328.89	692.72	1433.75	164.45	2426.95	11.66	2.63
9	CM9	1:2:4	0.6	Compacted	328.89	692.72	1433.75	197.33	2371.37	10.17	2.38
10	CM10	1:1.5:3	0.4	Compacted	418.59	661.23	1368.58	167.44	2468.56	18.07	3.20
11	CM11	1:1.5:3	0.5	Compacted	418.59	661.23	1368.58	209.30	2463.82	16.76	3.02
12	CM12	1:1.5:3	0.6	Compacted	418.59	661.23	1368.58	251.15	2418.52	14.37	2.97
13	CM13	1:1:2	0.4	Compacted	575.56	606.13	1254.53	230.22	2549.01	21.03	3.28
14	CM14	1:1:2	0.5	Compacted	575.56	606.13	1254.53	287.78	2500.54	14.78	3.07
15	CM15	1:1:2	0.6	Compacted	575.56	606.13	1254.53	345.34	2451.74	13.74	2.71
16	NC1	1:4:8	0.4	Non-Compacted	177.09	746.01	1544.04	70.84	2274.24	2.91	0.70
17	NC2	1:4:8	0.5	Non-Compacted	177.09	746.01	1544.04	88.55	2299.58	3.43	0.73
18	NC3	1:4:8	0.6	Non-Compacted	177.09	746.01	1544.04	106.25	2291.78	3.80	0.86
19	NC4	1:3:6	0.4	Non-Compacted	230.22	727.36	1505.44	92.09	2311.30	5.02	1.17
20	NC5	1:3:6	0.5	Non-Compacted	230.22	727.36	1505.44	115.11	2306.03	5.92	1.33
21	NC6	1:3:6	0.6	Non-Compacted	230.22	727.36	1505.44	138.13	2361.16	6.30	1.37
22	NC7	1:2:4	0.4	Non-Compacted	328.89	692.72	1433.75	131.56	2403.54	12.43	2.49
23	NC8	1:2:4	0.5	Non-Compacted	328.89	692.72	1433.75	164.45	2369.99	11.58	2.48
24	NC9	1:2:4	0.6	Non-Compacted	328.89	692.72	1433.75	197.33	2316.41	10.77	2.49
25	NC10	1:1.5:3	0.4	Non-Compacted	418.59	661.23	1368.58	167.44	2427.95	18.72	3.15
26	NC11	1:1.5:3	0.5	Non-Compacted	418.59	661.23	1368.58	209.30	2416.67	17.63	3.11
27	NC12	1:1.5:3	0.6	Non-Compacted	418.59	661.23	1368.58	251.15	2340.79	15.86	3.07
28	NC13	1:1:2	0.4	Non-Compacted	575.56	606.13	1254.53	230.22	2489.96	21.23	3.24
29	NC14	1:1:2	0.5	Non-Compacted	575.56	606.13	1254.53	287.78	2437.78	16.81	3.18
30	NC15	1:1:2	0.6	Non-Compacted	575.56	606.13	1254.53	345.34	2388.39	14.95	3.14
31	HWC1	1:4:8	0.9	Very High W/C	177.09	746.01	1544.04	159.38	2215.67	10.59	1.58
32	HWC2	1:3:6	0.9	Very High W/C	230.22	727.36	1505.44	207.20	2118.85	8.50	1.45
33	HWC3	1:2:4	0.9	Very High W/C	328.89	692.72	1433.75	296.00	2063.52	7.57	2.04
34	HWC4	1:1.5:3	0.9	Very High W/C	418.59	661.23	1368.58	376.73	2002.67	7.40	1.73
35	HWC5	1:1:2	0.9	Very High W/C	575.56	606.13	1254.53	518.00	1969.89	6.91	1.26

with decreases in both compressive and tensile strengths. Although the overall strength values remained within acceptable limits for structural concrete, the presence of additional voids and weaker particle packing diminished their mechanical capacity relative to compacted specimens of identical mix ratios. The effect became most pronounced in the very high W/C ratio category, where densities fell below 2200 kg/m³. These specimens demonstrated the lowest compressive and tensile strengths. The reduction can be attributed to excess mixing water, which leads to bleeding and the formation of interconnected voids that weaken the hardened matrix.

2.3. Non-destructive testing

NDT was carried out on the cylinders used for IP. Initially, the cylinders' surfaces were smoothed, and the RH test was conducted for each sample following ASTM C805 [6]. The resulting R-value was used to determine the compressive strength of the concrete samples. Additionally, the UPV test was performed according to ASTM C597 [5]. The travel time of ultrasonic waves at a frequency of 250 kHz through the concrete slice was measured, allowing for the calculation of wave velocity using the formula in Eq. (3).

$$v = \frac{T}{t} \quad (3)$$

where v represents the velocity, T denotes the slice thickness, and t is the time taken for the ultrasonic waves to pass through. The compressive strength was subsequently estimated based on the UPV values.

2.4. Concrete image dataset

A significant challenge in any computer vision task is ensuring access to appropriate and sufficient data for training, validation, and test-

ing. Currently, the only dataset available for segmentation of aggregate in concrete [40], consists of 1144 images with a resolution of 448 × 448 pixels depicting concrete mixtures having colored aggregates. This dataset is inadequate for studies involving gray aggregates. Therefore, a new image dataset was created and curated for this study.

2.4.1. Cutting of cylinders

Each cylindrical specimen, measuring 300 mm in height (H) and 152 mm in diameter (D), was sectioned into three parts at one-third and two-thirds of its height (roughly 100 mm from both the top and bottom ends) using a precision stone-cutting saw. The cutting process for obtaining the concrete slices is illustrated in Fig. 4. Due to the accumulation of fine material on the top surface of the first slice and the bottom surface of the third slice, these sections contained little to no aggregate detail and were therefore excluded from image acquisition. Consequently, only four faces were utilized for capturing high-resolution images to build the dataset.

2.4.2. Image acquisition

Image capturing was conducted in an enclosed environment to maintain uniform conditions and ensure high-quality data suitable for IP. All openings were covered with black sheets to block external light sources, thereby eliminating shadows and reflections. Controlled illumination was provided by two 30 W LED lamps positioned at fixed angles, producing an average light intensity of approximately 2000 lx. A Canon EOS 4000D digital camera with an 18-megapixel sensor and an ISO setting of 1000 was used to photograph a total of 280 slice images, each captured at a resolution of 5184 × 3456 pixels. A constant distance was maintained by mounting the camera on a stand at a distance of 550 mm from the sample's surface. The image acquisition setup for an example slice image is illustrated in Fig. 3.

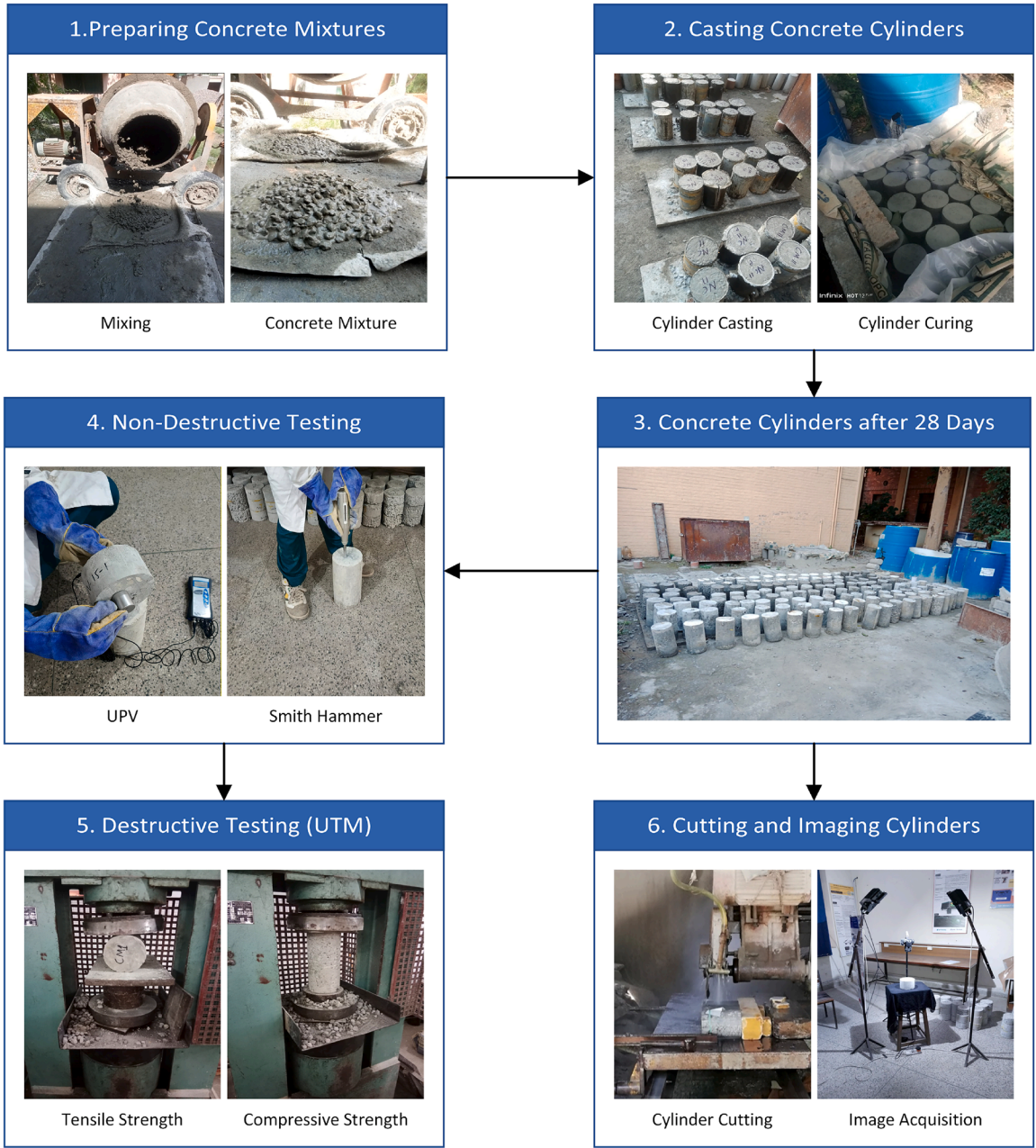


Fig. 3. Step-wise procedure from cylinder casting to dataset creation.

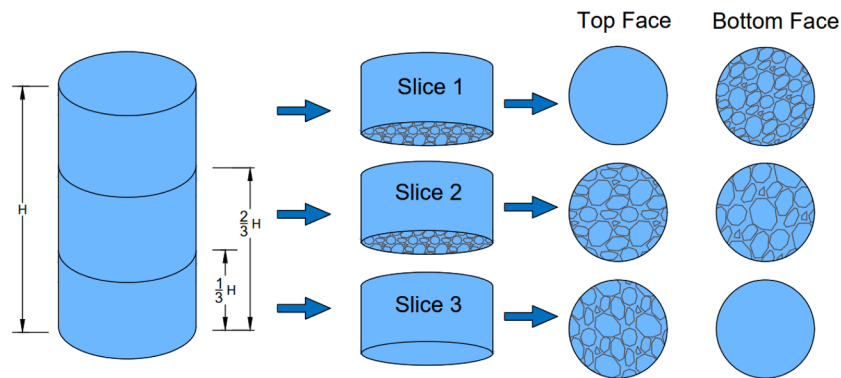


Fig. 4. Sectioning of concrete cylinders into three slices.

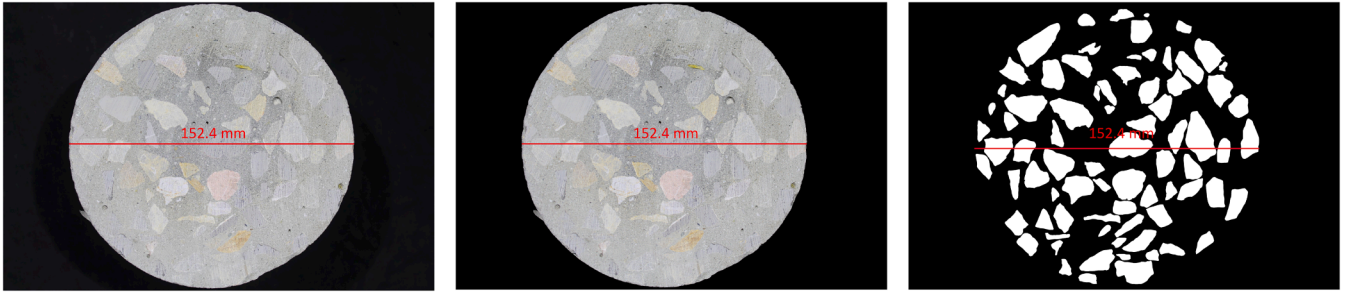


Fig. 5. Concrete slice images with varying processing stages: Original image (left), image with background removed (center), and annotated ground truth image (right).

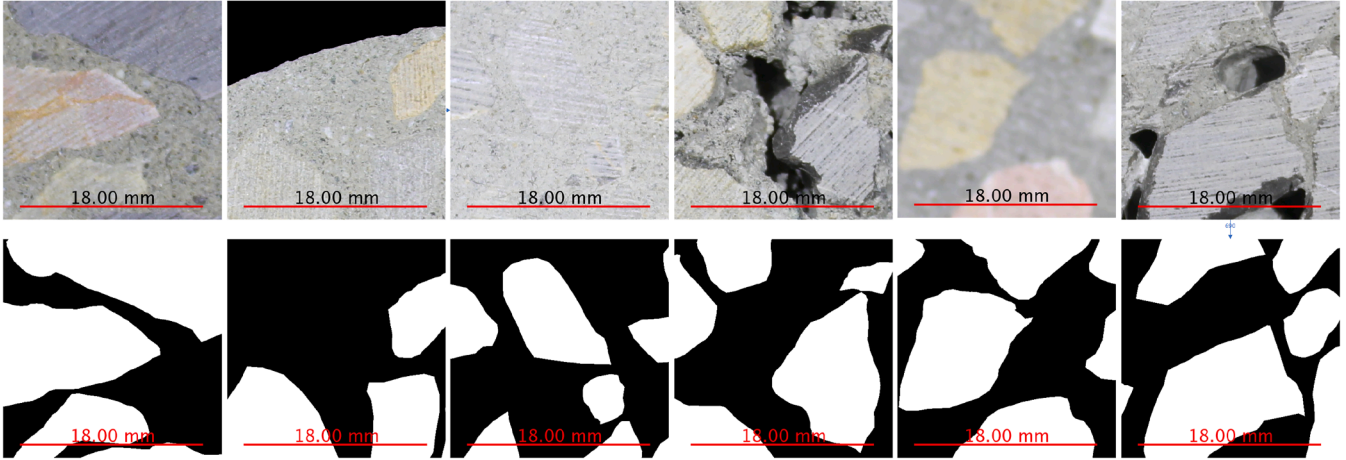


Fig. 6. Sample patched concrete images (top row) with their corresponding ground truth segmentation masks (bottom row).

2.4.3. Image annotation and quality control

Each captured slice image included both the surface of the concrete specimen and its surrounding background. To isolate the specimen surface from the surrounding background, several morphological pre-processing steps were applied. These included converting the image to grayscale, applying thresholding, filling internal gaps, performing closing operations, analyzing connected regions, and finally extracting the primary foreground object. This sequence of operations ensured that only the pixels corresponding to the concrete cylinder were retained for further analysis. Subsequently, the concrete slice images were annotated manually at the pixel level using the MATLAB Image Labeler tool [41], following a consistent labeling protocol. Two categories were defined: aggregate and background. Each image was annotated by two independent annotators with a background in civil and structural engineering image analysis. Both annotators used a digital pen and graphic tablets for precise tracing of aggregate boundaries, ensuring that fine edges and contact regions between aggregates and mortar were carefully delineated. During annotation, zoomed-in views were employed to enhance visibility of fine boundaries and ensure accurate edge placement, particularly in gray or low-contrast regions. Before commencing full annotation, a calibration session was conducted in which the annotators jointly labeled 20 representative images to harmonize labeling criteria such as aggregate inclusion threshold and edge treatment.

To quantify inter-annotator reliability after calibration, the Intersection-over-Union (IoU) and Cohen's Kappa (κ) metrics were calculated on a randomly selected subset of 80 annotated images. The resulting mean IoU between annotators was 0.94, and $\kappa = 0.95$, indicating excellent agreement and high labeling consistency. Discrepancies, mainly arising near partially visible aggregates or indistinct boundaries, were resolved through consensus discussion under the supervision of a senior structural engineer. Quality control was further enforced by visual inspection of all labeled images.

Annotations with visible boundary errors, missing aggregates, or mislabeled regions were rechecked and corrected. Binary masks representing the aggregate regions were subsequently created after the labels were reviewed and confirmed by the experts. Fig. 5 illustrates an example of a concrete slice, showing the original image with its background, the background-removed version, and the corresponding ground truth mask for aggregates. In the binary images, white pixels denote the aggregate regions identified in the reference labels. Each high-resolution 5184×3456 pixels image and its corresponding ground truth image were segmented into 74 smaller 512×512 pixels each, resulting in 20,720 images. Breaking down the high-resolution images into these standardized patches was essential for training the DCNN models. Images that included only non-cylindrical regions lacking information on aggregates or non-aggregate suspensions were excluded from the dataset due to their redundancy and lack of contribution to DCNN training. Consequently, the final dataset comprised 13,138 images. Fig. 6 illustrates sample images and their corresponding ground truths, where white pixels denote aggregates. Across the 280 annotated slice images, approximately 42.6% of the total pixels were identified as aggregates, while the remaining 57.4% corresponded to non-aggregate regions.

3. Segmentation and analysis of aggregate

Segmentation of aggregates plays a critical role in understanding the characteristics of aggregate properties. This process enables us to accurately quantify the aggregate distribution, size, and shape, which are essential for evaluating the mechanical properties of concrete. We can effectively differentiate between various components within concrete slice images by employing advanced segmentation techniques. This step not only improves the accuracy and reliability of the subsequent anal-

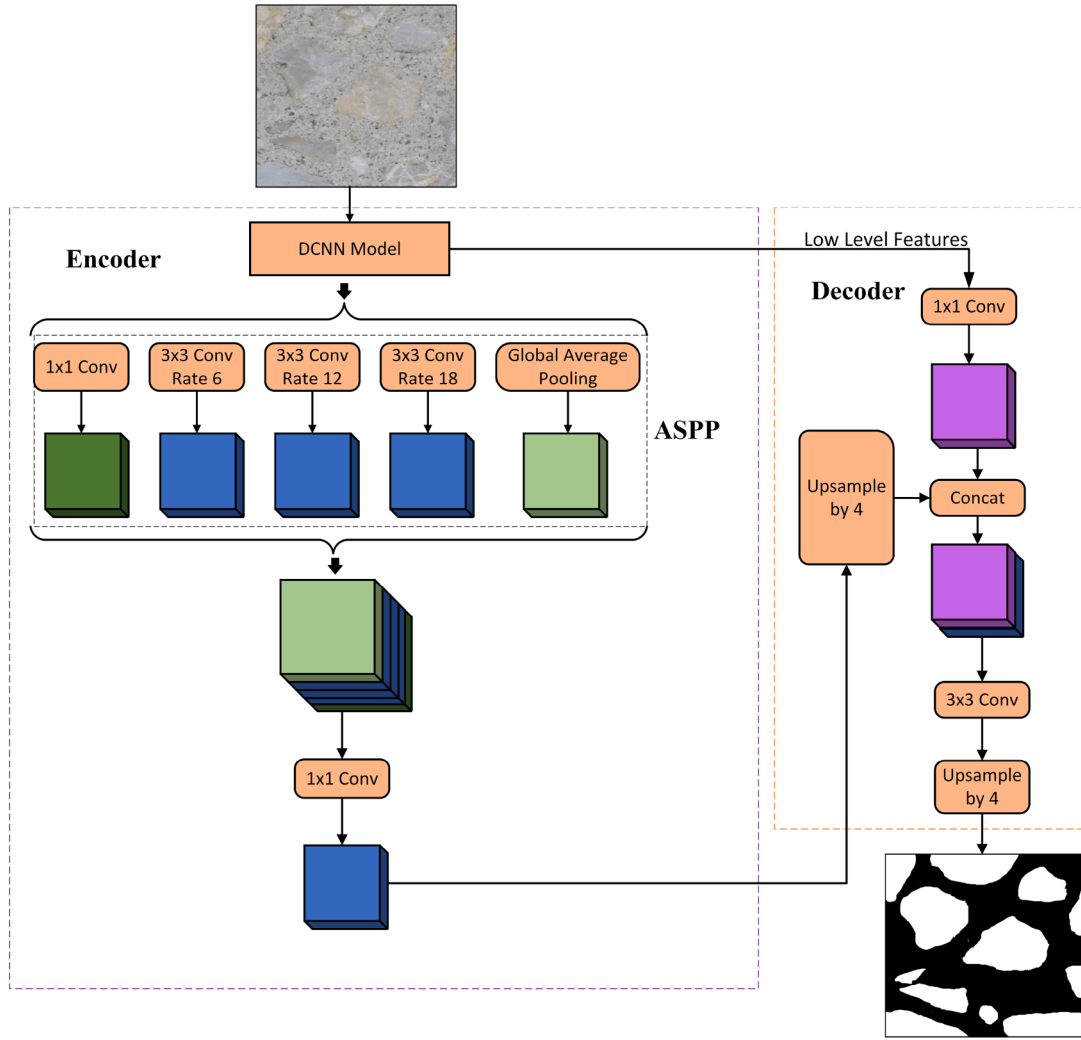


Fig. 7. Framework of the proposed DeepLabv3+ model, for aggregate segmentation.

ysis but also supports informed decision-making in optimizing concrete mixtures.

3.1. Segmentation architectures

Empowered by recent advancements in computational power and extensive databases, DCNNs have demonstrated exceptional performance in numerous computer vision tasks, particularly in image segmentation. Image segmentation involves partitioning an image into multiple segments or regions to simplify its representation or to facilitate more efficient analysis [42]. Numerous DL architectures tailored for semantic segmentation have been introduced, including fully convolutional network [43], U-Net [44], SegNet [45], and DeepLab [46,47], among others. These DCNN architectures are typically composed of an encoder-decoder structure, where the encoder utilizes several filters to capture image features, and the decoder generates the segmentation mask for the targeted region. Selecting an appropriate architecture plays a vital role in achieving precise segmentation. This study used DeepLabV3+ for concrete aggregate segmentation. DeepLabV3+ [48] is an improved deep learning model for image semantic segmentation. It enhances the capabilities of its predecessor, DeepLabV3 [46,47], by incorporating several advancements that boost performance, especially in capturing fine spatial details. Fig. 7 shows the complete architecture of DeepLabV3+ comprising three main components: the encoder (which features a DCNN backbone), the spatial pyramid pooling (ASPP), and the decoder.

3.1.1. Encoders block

The encoder incorporates a backbone DCNN to extract features from the input image. A DCNN consists of multiple layers with learnable weights and biases. It processes images through convolutional layers, activation functions, pooling operations, and dropout layers. Denoting the DCNN model by F , the feature extraction through the convolutional layers can be represented as:

$$a = F(I; \theta) \in \mathbb{R}^{W_a \times H_a \times D_a}, \quad (4)$$

where a represents feature map with width W_a and height H_a . D_a indicates the number of feature maps produced. The input image and parameter vector are denoted by I and θ , respectively. The output from the convolutional layers, followed by the activation function, undergoes batch normalization to assist training and enhance convergence. The activation value a_{ij}^{xy} of the i th feature map unit at the coordinates (x, y) and j th layer is computed by:

$$a_{ij}^{xy} = \rho \left(\sum_{r=1}^{W_j} \sum_{s=1}^{H_j} (v^{rs} w_{ij}^{rs}) + b_{ij} \right), \quad (5)$$

where v represents the input values, w denotes the kernel weights in the i th feature map unit of the j th layer, and b is the bias term. The dimensions of the kernel in the j th layer are given by W_j and H_j . The function ρ typically denotes the rectified linear unit (ReLU) or leaky

ReLU activation function and is defined as:

$$\rho(x) = \max(0, x) \text{ or } \max(\alpha x, x), \quad (6)$$

Where z represents the input to the function and α is a small constant for leaky ReLU, and Activation functions introduce non-linearity, essential for identifying complex patterns. Pooling layers downsample the spatial dimensions of feature maps, maintaining crucial features while reducing the computational burden. Additionally, dropout layers may be employed to combat overfitting by randomly deactivating some neurons during training. In addition to standard convolutions, DCNNs may incorporate specialized convolution techniques, such as dilated convolutions to capture features at multiple resolutions and scales or depthwise separable convolutions to decrease computational costs. The goal of a DCNN is to progressively reduce the spatial resolution of the input image while extracting more abstract and higher-level features.

3.1.2. Atrous spatial pyramid pooling (ASPP)

The ASPP module is designed to aggregate features at various scales, which helps handle objects of different sizes within semantic segmentation tasks. The receptive field size of a convolutional network defines the extent of the input that influences the resulting feature map. A receptive field that is too small may fail to capture larger contextual information, whereas one that is too large may overlook smaller objects. To address this, atrous convolution is employed, which increases the receptive field by introducing zeros between kernel elements, allowing the network to capture features at different scales without augmenting the number of parameters. ASPP implements multiple parallel atrous convolutions with varying dilation rates to capture multi-scale features effectively. It also integrates global average pooling to incorporate global contextual information. The features extracted at different scales are concatenated and then processed through a 1×1 convolution to produce the final output of the ASPP module.

3.1.3. Decoder

The DeepLabV3+ model's decoder is composed of two convolutional layers (both 1×1 and 3×3), two upsampling layers, and one skip connection. It merges high-level semantic data from the encoder with low-level spatial information. The initial step in the decoder involves upsampling the encoder output by a factor of 4 using bilinear interpolation, which is more computationally efficient than deconvolution methods. This upsampled feature map, now enriched with semantic information, is combined with a low-level feature map from the encoder, which has undergone 1×1 convolution processing. By integrating these high-level and low-level features, the model enhances the detail of the segmentation. The final steps involve applying a 3×3 convolution to the concatenated feature map and performing another upsampling operation to generate the ultimate segmentation map, which emphasizes the target areas within the image.

3.2. Training protocol

The training followed a systematic approach to ensure robust model development and accurate performance evaluation. The image dataset was split into sets of training, validation, and test, with 60 % allocated for training, 20 % for validation, and 20 % for testing. This division was performed randomly and was independent of the mix ratio and W/C. The segmentation models were developed and trained in MATLAB using the deep learning toolbox [41]. Experiments were conducted on a Windows-based workstation with an Intel Core i7-10700 CPU running at 2.90 GHz with 32 GB DDR4 and an NVIDIA GeForce RTX 3060 GPU. The convolutional layers' parameters were initialized with the Kaiming. He initialization method [49], and we used the Stochastic Gradient Descent (SGD) [50] with the momentum optimizer to update the model parameters after each iteration. The initial training parameters were: learning rate $1e-3$, momentum 0.9, learn rate drop period 10, learn rate drop factor 0.3, and L2 regularization 0.005. The network

Table 3

Specifications of backbone DCNNs used as Encoders in the Deeplabv3+.

Backbone DCNN	No of Layers	FLOPS Complexity	Trainable Parameters	Inference Time (ms)
InceptionResNetV2	572	59.44 B	55.86 M	41.9
Xception	126	8.07 B	22.87 M	4.8
ResNet101	101	7.80 B	44.6 M	3.9
EfficientNet-B5	342	8.8 B	22 M	10.3

was trained over 25 epochs and a mini-batch size of 8 while minimizing cross-entropy and Lovász-Softmax loss functions. The DL models employed in this study were pre-trained using the PASCAL VOC dataset [51]. However, due to the differences between PASCAL VOC and our dataset, CNN filter weights were fine-tuned using SGD on cross-entropy and Lovász-Softmax loss functions. After each training epoch, the data was shuffled, and checkpoints were saved in a temporary directory to evaluate the model's performance using loss and accuracy metrics.

The training data was shuffled after each epoch during training, and checkpoints were stored in a temporary directory to evaluate model performance based on loss and accuracy metrics. The loss factor underwent regular updates to reflect the discrepancy between predicted outputs and actual labels. We employed various data augmentation techniques to enhance the robustness of our deep learning model and mitigate overfitting. These techniques included vertical and horizontal flipping, random translations, variations in illumination, and the addition of Gaussian noise. Specifically, images were flipped vertically and horizontally with a probability of 0.35, applying mirror transformations that help the model learn invariant features. Random translations involved shifting images horizontally and vertically within a range of -30 to 30 pixels to promote spatial invariance. We added Gaussian noise with a zero mean and a standard deviation of 0.01, providing subtle yet effective perturbations to pixel values, which enhanced the model's generalization. Illumination variations were simulated utilizing the Jitter Color Function in MATLAB, allowing us to account for changes in lighting conditions. These augmentation strategies aimed to improve the robustness of our model and ensure that our findings are applicable across a broader range of conditions.

3.3. Metrics for evaluating segmentation performance

To quantitatively assess the performance of the segmentation model, several metrics are employed, including IoU, precision, recall, and F1 score. The model's accuracy and effectiveness in segmenting concrete aggregates are evaluated by comparing ground truth images to the segmented images. IoU measures the overlap between the ground truth and predicted segmentation, while precision represents the proportion of aggregate pixels classified correctly out of all pixels labeled as aggregates. Recall indicates the proportion of actual aggregate pixels that were correctly identified. The F1 Score combines precision and recall into a metric by calculating their harmonic mean. In this evaluation, aggregate pixels are considered as the positive class and non-aggregate pixels as the negative class. The definitions for these metrics are provided below:

$$IoU = \frac{TP}{TP + FN + FP} \quad (7)$$

$$\text{Precision (P)} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall (R)} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{F1 Score} = 2 \times \left(\frac{P \times R}{P + R} \right) \quad (10)$$

where TP denotes the number of aggregate pixels correctly identified as aggregates, FP represents the number of non-aggregate pixels incorrectly classified as aggregates, and FN signifies the number of aggregate pixels misclassified as non-aggregate.

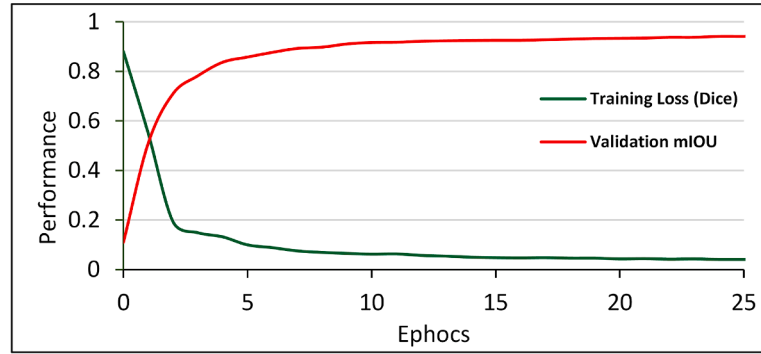


Fig. 8. Learning curves for EfficientNet-B5 DCNN.

Table 4

Performance metrics of different backbone DCNNs for aggregate segmentation.

Backbone DCNN	Performance Metrics			
	mIoU	F1-Score	Precision	Recall
InceptionResnetV2	0.83	0.93	0.89	0.95
ResNet101	0.90	0.94	0.93	0.95
Xception	0.88	0.93	0.91	0.96
EfficientNet-B5	0.92	0.95	0.94	0.96

3.4. Aggregate segmentation results

The performance of the DeepLabV3+ architecture in segmentation is significantly affected by the backbone DCNN model used for feature extraction. To explore this impact, we evaluated the DeepLabV3+ model with four distinct backbone architectures: InceptionResNetV2 [52], Xception [53], ResNet101 [54], and EfficientNet-B5 [55]. These backbones vary in complexity, layer configuration, and the types of layers utilized. Table 3 provides a comparative overview of these DCNNs. FLOPS Complexity refers to the number of Floating Point Operations Per Second a model requires to make predictions. Trainable Parameters are the model's adjustable values (weights and biases) that are learned during training to minimize error. Inference Time measured in milliseconds (ms) is the time it takes for a trained model to process an input and generate an output. Notably, the InceptionResNetV2 model, with its extensive layer count and substantial number of trainable parameters, is particularly complex and capable of extracting a wide range of features. However, increased network depth does not always translate to better segmentation performance; deeper networks can suffer from issues such as vanishing gradients and overfitting. Hence, we constructed the DeepLabV3+ framework utilizing various backbone DCNNs to assess and validate their effectiveness in the segmentation of aggregates in concrete images.

Table 4 summarizes the performance of the DeepLabV3+ model using various backbone DCNNs used as encoders for aggregate segmentation. The metrics quantify the segmentation accuracy and reliability of each architecture in distinguishing coarse aggregates from the mortar matrix. InceptionResNetV2 achieved the lowest performance, with an mIoU of 0.83 and an F1-score of 0.93. Although this architecture benefits from a deep residual-inception design that enhances representational richness, its large number of parameters increases training complexity and susceptibility to overfitting when dataset variability is limited. This led to occasional boundary misclassifications and excessive false positives, reflected in its comparatively low precision. The Xception network attained moderately higher accuracy with mIoU = 0.88 and F1 = 0.93, benefiting from depthwise separable convolutions that reduce computational cost while maintaining representational efficiency. Its high recall value of 0.96 suggests strong sensitivity in identifying

aggregate pixels, though at the expense of slightly lower precision, indicating that some non-aggregate regions were mistakenly classified as aggregates. The ResNet101 demonstrated a balanced and reliable performance, achieving mIoU = 0.90 and F1 = 0.94. Its deep residual connections enhanced gradient flow and feature reuse, enabling consistent learning across scales. The nearly equal precision of 0.93 and recall of 0.95 indicate stable generalization, making it a robust architecture for medium-to-high-resolution segmentation tasks. The EfficientNet-B5 backbone achieved the best overall results, with mIoU = 0.92 and F1 = 0.95. Its compound scaling strategy, which balances network depth, width, and input resolution, allowed for superior feature extraction with only 22M trainable parameters. The high precision = 0.94 and recall = 0.96 values indicate strong capability in delineating aggregate boundaries and minimizing both false positives and false negatives. Moreover, as shown in Fig. 8, the convergence behavior observed in the learning curves confirms stable optimization without signs of overfitting or underfitting, validating the network's efficiency and adaptability.

Fig. 9 displays examples of original slices alongside ground truth and output masks from various deep-learning models. In the visualizations, false positives are highlighted in pink, whereas false negatives are marked in green. The InceptionResNetV2 base model showed suboptimal performance in aggregate segmentation, as indicated by numerous pink pixels. In contrast, EfficientNet-B5 outperformed other DCNN models in terms of efficiency in parameters and processing speed, exceeding the performance of Xception and ResNet101.

4. Input feature extraction through image processing

4.1. Aggregate properties

The aggregate mask images created by the trained DeepLabV3+ model were analyzed using a range of IP methods to extract and quantify key aggregate characteristics. While the analysis is performed on two-dimensional (2D) images, it is essential to consider the three-dimensional representation of the aggregates. Based on Cavalieri's principle, the area proportion observed in a 2D cross-section can be considered equivalent to the corresponding volume proportion in three dimensions (3D) [56,57]. Previous studies have validated the reliability of 2D image analysis for inferring 3D geometric properties of materials and structures [58–60]. Consequently, this work assumes that the features of coarse aggregates identified in 2D images represent their spatial characteristics in 3D form.

4.1.1. Aggregate proportion

The proportion of aggregates plays a crucial role in determining the rheological behavior, mechanical performance, durability, and overall cost-effectiveness of concrete. These proportions were quantified using the binary masks of concrete slices. The total aggregate area within each cylinder image was computed by counting the white pixels, corresponding to aggregate regions, and multiplying the count by the unit pixel

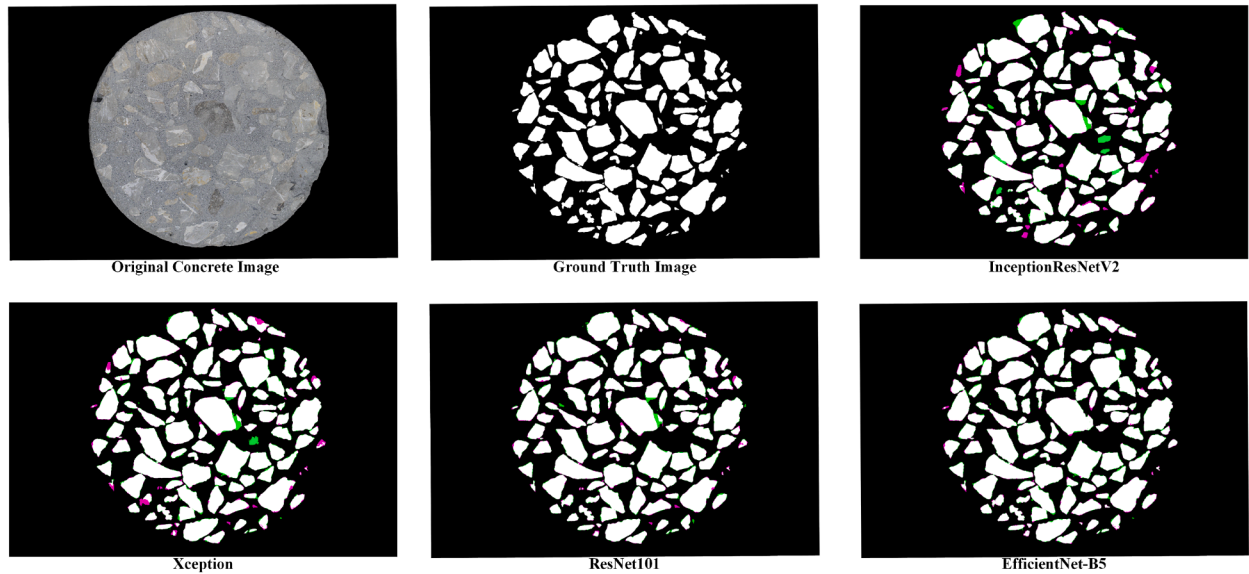


Fig. 9. Comparison of ground truth image and output segmentation masks generated by different backbone DCNN for an original size image from Concrete test dataset. False negatives and false positives pixels are marked in pink and green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

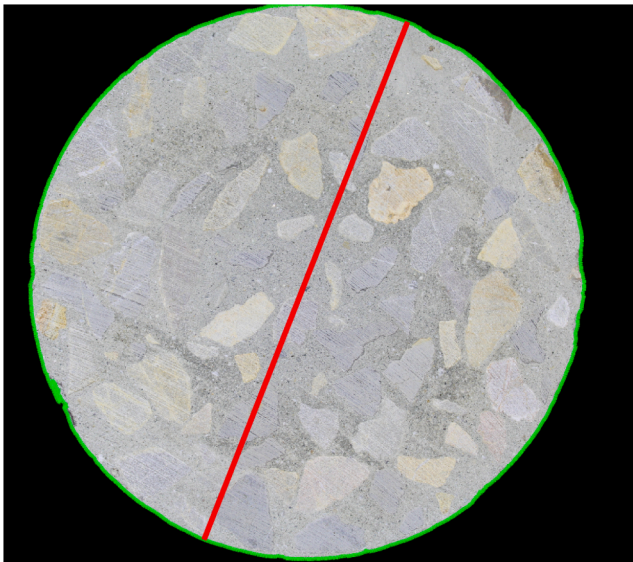


Fig. 10. Detection of the cylinder's boundary (green) and its diameter (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

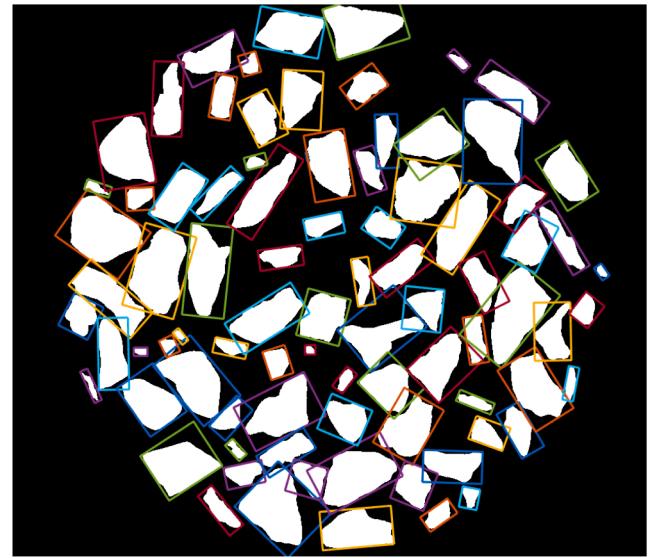


Fig. 11. Mask image showing minimum-area rectangles based on aggregate Feret diameters.

area in mm^2 . Since the imaging distance between the camera and the specimen varied slightly across images, the unit pixel area differed for each image. Accurate determination of an object's physical dimensions from digital images depends on the known imaging distance during image acquisition. However, maintaining a fixed imaging distance is often not feasible in real-time applications. To overcome this challenge, a procedure was developed to calculate the unit pixel area.

The boundary of the concrete cylinder, shown as a green contour in Fig. 10, was identified using the Moore-Neighbor tracing algorithm with 8-pixel connectivity [61,62]. The Moore-Neighbor tracing algorithm was chosen due to its efficiency in detecting connected components and its ability to delineate the boundary of concrete cylinders in the images accurately. Additionally, this algorithm provides robust performance in handling noise and small artifacts, common in concrete

images, thereby ensuring more reliable and precise measurements. The Moore-Neighbor tracing algorithm initiates by locating the first boundary pixel and then proceeds to follow the contour of the concrete cylinder in either a clockwise or counterclockwise direction. At each step, the algorithm selects the next boundary pixel based on predefined connectivity criteria, ensuring that all neighboring pixels contributing to the edge are included. This iterative process continues until the trace returns to the starting point, thereby completing an accurate outline of the cylinder's perimeter. The cylinder's diameter, represented by the red line in Fig. 10, is then measured between two diametrically opposite points on this boundary. The unit pixel area is subsequently determined by dividing the measured pixel diameter by the actual cylinder diameter of 152.4 mm (6 in.). This approach allows for precise conversion of pixel dimensions into real-world measurements, enabling accurate computation of aggregate characteristics in the following analyses.

4.1.2. Aggregate characteristics

The characteristics of aggregates, such as their size, orientation, and roundness, provide valuable insights into the composition of concrete. Measuring these features helps evaluate the uniformity of particle size and spatial distribution within the mixture, which facilitates the examination of the concrete mixture's consolidation, workability, compatibility, interlocking, and bonding [35,63]. The Feret rectangle technique [64] was utilized in this study to measure the characteristics of the coarse aggregate and voids from the image. This method involves enclosing each aggregate within the smallest possible rectangle based on its Feret diameter, as illustrated in Fig. 11. The Feret diameter refers to the maximum distance between any two points along an aggregate's boundary. To determine it, all boundary point pairs were evaluated by calculating their Euclidean distances, and the largest value was identified as the Feret diameter. A rectangle was then drawn around each aggregate, with its length equal to the Feret diameter and its width oriented perpendicularly. The rectangle's orientation was adjusted to achieve the minimum bounding area that fully enclosed the particle. The resulting width and length of each rectangle were recorded as the aggregate's respective dimensions. The longer side was taken as the aggregate's size, and the mean aggregate size was computed as the average of these values across all detected particles.

Aggregate roundness was quantified as the ratio of the longer side to the shorter side of each rectangle, where a value of one indicates a nearly square shape, and higher values denote greater elongation. The aggregate packing angle (APA) was also computed, representing the angular orientation of aggregates based on Feret rectangle analysis. For each Feret rectangle, principal orientation angle θ_i with respect to the horizontal axis was recorded. The APA was then calculated as the average orientation of all aggregates in the image. This metric reflects how consistently aggregates are aligned within the mixture. A narrow distribution of APA values indicates ordered packing, whereas a wide spread suggests random or irregular orientation.

The results were compared with manually measured data and AutoCAD-based evaluations [65], to verify the accuracy of this procedure. For manual verification, a transparent graph sheet with a 1 mm grid was placed over the cylinder surface, and the number of grid squares fully or partially covering aggregates was counted to estimate the total aggregate area, as illustrated in Fig. 12. The mean aggregate size and roundness were also measured and checked using Vernier caliper readings. In AutoCAD, the imported raster images were initially resized to match the actual cylinder diameter, after which the aggregate boundaries were traced by connecting key points along the edges to form polygons as shown in Fig. 12. The built-in area calculation command was then used to determine polygonal areas based on their boundary coordinates. This vector-based approach provided a useful comparison with the pixel-derived measurements. The total aggregate area was obtained by summing the individual areas of all the polygons.

A comparative analysis of aggregate parameters obtained from the manual measurements, proposed IP technique, and AutoCAD is summarized in Table 5. The aggregate masks used in this validation were

the predicted ones generated by the trained DeepLabV3+ segmentation model. The results demonstrate a strong agreement among all three approaches, confirming the reliability of the proposed IP technique. For aggregate area, the proposed method shows percentage errors ranging between 0.80 % and 2.83 %, indicating close correspondence with manual and AutoCAD measurements. Similarly, for aggregate mean size, the percentage error varies from 0.90 % to 4.35 %, while the aggregate mean roundness differs by less than 2.33 % across all mixes. These small deviations suggest that the proposed method can accurately quantify aggregate geometric characteristics from predicted masks.

4.1.3. Distance between aggregates

The distance between aggregates reflects their spatial arrangement and distribution within the concrete matrix. Maintaining adequate spacing among aggregates improves load transfer, structural stability, and crack resistance [66]. It also affects volumetric stability, risk of internal damage, and service life of concrete structures. To analyze the spatial arrangement of aggregates, the nearest-neighbor analysis method [67] was applied. As shown in Fig. 13, this approach measures the distance from each aggregate's centroid to those of surrounding aggregates to assess distribution characteristics. The nearest neighbor is defined as the aggregate whose centroid lies at the shortest Euclidean distance from a given particle. This measurement quantitatively describes the spacing and packing density of aggregates within the concrete matrix.

While nearest neighbor analysis focuses on the spacing between aggregates, another complementary descriptor is the number of direct contacts that each aggregate forms with its neighbors. This is expressed through the coordination number (CDN), which measures the degree of aggregate-aggregate contact. For each aggregate, neighboring particles were identified using centroid positions derived. If the centroid-to-centroid distance between two aggregates was less than a threshold defined as the mean equivalent diameter, the aggregates were considered in contact. The number of such contacts was counted for each aggregate, and the average across all aggregates in an image was reported as the coordination number (CDN). This descriptor captures the degree of aggregate interlock, with higher values indicating denser packing.

4.2. Mortar image analysis

The mechanical properties of concrete depend not only on the properties and distribution of aggregates but also on the properties of the mortar. However, predicting the mortar properties, i.e., the proportion of cement, fine aggregate, voids, and porosity, and W/C in mortar directly through IP, is challenging. Pixel distribution (PD) in image analysis can indirectly represent these properties. Fig. 14 shows aggregate-removed images of 1:2:4 mix ratio specimens with different W/C ratios. A noticeable change in surface appearance is observed in the histograms with varying W/C ratios and honeycombing. It was anticipated that the

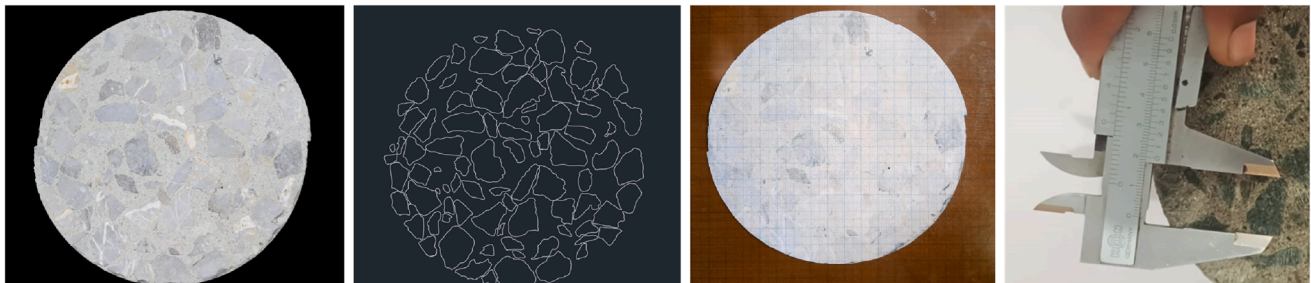


Fig. 12. Concrete specimen (first image), boundary tracing in AutoCAD (second image), area estimation using graph paper (third image), and dimension measurement with a Vernier caliper (fourth image).

Table 5
Comparison of aggregate parameters using proposed IP method, manual measurement and AutoCAD.

Parameters	Mix Ratio	Proposed Method	Manual Measurement	AutoCAD	Percentage Error
Aggregate Area (mm ²)	1:4:8	10550.34	10810.00	10720.00	2.40
	1:3:6	10058.42	10171.50	10083.88	1.11
	1:2:4	9335.35	9411.00	9297.12	0.80
	1:1.5:3	9100.79	8850.50	8775.00	2.83
	1:1:2	8210.45	8300.00	8225.00	1.08
Aggregate Mean Size (mm)	1:4:8	14.88	14.40	14.56	3.33
	1:3:6	15.36	15.50	15.64	0.90
	1:2:4	13.84	14.47	14.53	4.35
	1:1.5:3	16.34	15.80	15.85	3.42
	1:1:2	14.26	14.10	14.23	1.13
Aggregate Mean Roundness	1:4:8	1.64	1.61	1.58	1.86
	1:3:6	1.52	1.50	1.43	1.33
	1:2:4	1.73	1.75	1.81	1.14
	1:1.5:3	1.68	1.72	1.70	2.33
	1:1:2	1.45	1.44	1.42	0.69

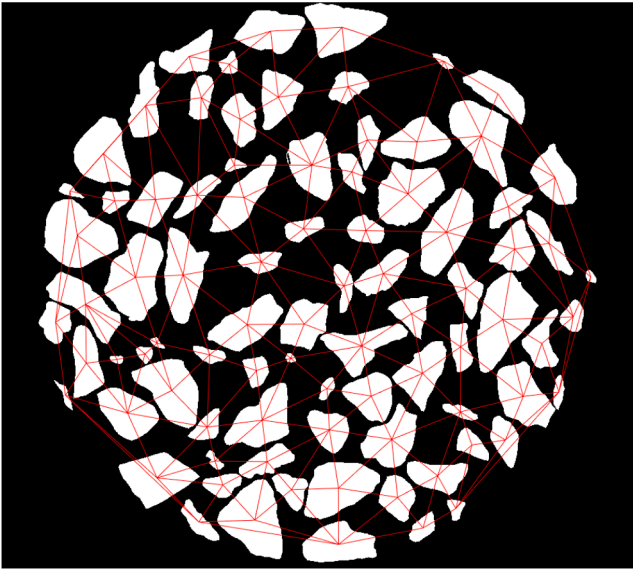


Fig. 13. Nearest-neighbor analysis for measuring distances between aggregates.

color of pixels of the mortar region would vary with changes in the cement dosage, texture, and presence of honeycombing. Thus, statistical features of pixel values were used to capture and represent these underlying mortar properties.

For mortar statistical analysis, the RGB images were first transformed into grayscale by applying a weighted sum of the red, green, and blue components, as given by the equation:

$$GS = 0.298 R + 0.589 G + 0.113 B \quad (11)$$

where R , G , and B denote the red, green, and blue color components of pixels, respectively, and GS represents the computed grayscale value. Subsequently, the binary mask images were superimposed onto the concrete images. Using a MATLAB-based image processing tool, the pixels corresponding to the aggregates in the ground truth images were set to black in the concrete images. This process excluded the aggregate pixels, leaving only the mortar pixels in the resulting images. Fig. 14 represents the aggregates-removed image. The non-zero pixel values in the aggregates-removed images represent only the mortar's appearance. Statistical features, such as the mean, standard deviation, range, skewness, median, entropy, variance, kurtosis, and third moment, were computed for these non-zero (mortar) pixels in the aggregate-removed images. These features help to represent the overall mortar pixels, aiming to optimize ANN processing time and enhance prediction accuracy

[68,69]. Mathematical formulations of the statistical features for mortar pixels are detailed in the following equations.

$$M = \frac{1}{n} \sum_{i=1}^n V_i \quad (12)$$

$$\text{Range} = \max(V_i) - \min(V_i) \quad (13)$$

$$\text{Std} = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_i - M)^2} \quad (14)$$

$$\text{Mo} = \frac{1}{n} \sum_{i=1}^n (V_i - M)^3 \quad (15)$$

$$E = \sum_i p_i \log_2(p_i) \quad (16)$$

$$\text{Sk} = \frac{1}{n} \sum_{i=1}^n \frac{(V_i - M)^3}{\text{Std}^3} \quad (17)$$

$$Me = \left(\frac{n+1}{2} \right)^{\text{th}} \quad (18)$$

$$\text{Var} = \frac{1}{n} \sum_{i=1}^n (V_i - M)^2 \quad (19)$$

$$\text{Kur} = \frac{1}{n} \sum_{i=1}^n \frac{(V_i - M)^4}{\text{Std}^4} \quad (20)$$

Where V indicates pixel values in the digitized matrix, M represents the mean value, n is the total number of data points, Std stands for the standard deviation, Mo and E denote the third moment and entropy, p is the normalized histogram counts, Sk is skewness, Median is the middle value of ordered pixel intensities, Var is variance, and Kur is kurtosis.

4.3. Correlation analysis of input features

Correlation analysis is a statistical tool used to quantify the strength and direction of linear relationships between variables. It helps identify which aggregate and mortar features are strongly correlated with one another and how they relate to the mechanical properties of concrete. Among the available measures of correlation, the Pearson correlation coefficient is the most widely used for assessing linear dependencies [70]. It ranges from -1 to $+1$, where positive values indicate direct relationships, negative values denote inverse relationships, and values near zero suggest weak or negligible associations. High absolute values greater than 0.8 typically indicate a strong interdependence [71], which raises multicollinearity concerns that can compromise the stability and interpretability of predictive models.

Fig. 15 presents the Pearson correlation coefficient heatmap, which provides insights into the relationships among aggregate features, mor-

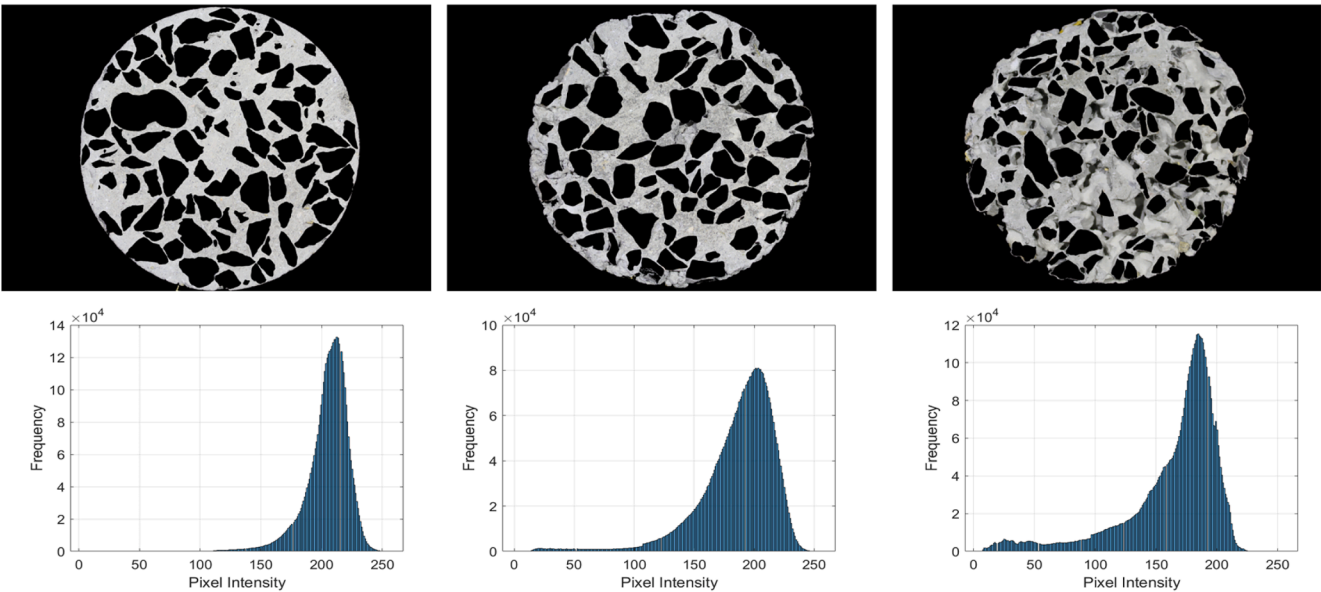


Fig. 14. Histogram of mortar pixels for 1:2:4 mix ratio with W/C ratio of 0.4 (left),0.6 (mid), and 0.9 (right).

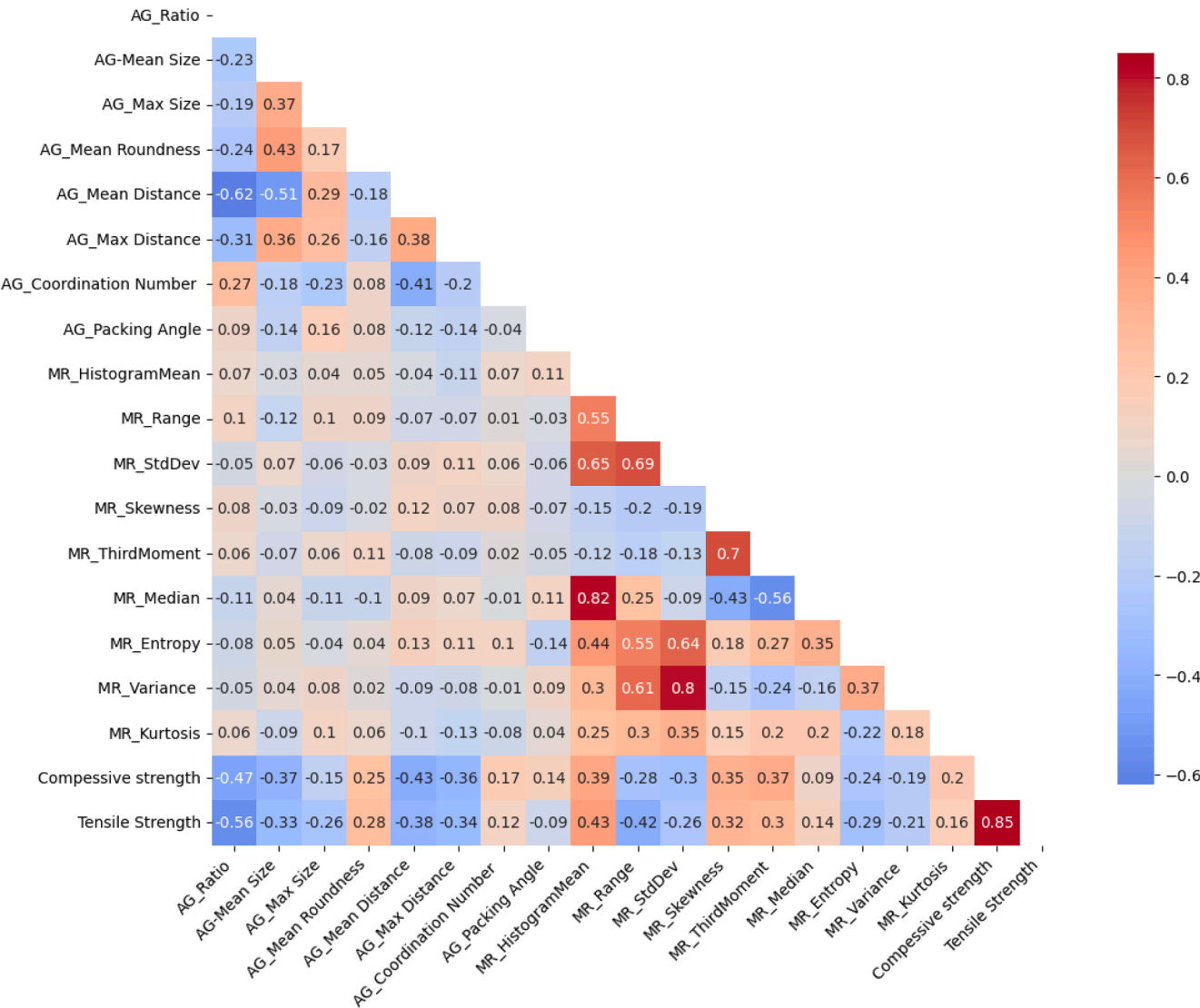


Fig. 15. Pearson correlation coefficient heatmap among aggregate descriptors, mortar statistical features, and target strengths.

tar statistical descriptors, and their association with compressive and tensile strengths. Among the aggregate features, aggregate ratio is strongly and negatively correlated with mean distance at -0.62, and mean size is inversely related to mean distance at -0.51. Together, these indicate that mixtures with higher aggregate content tend to exhibit tighter spacing, and larger particles also come with reduced inter-particle distances. Other pairings remain weak, and coordination number and packing angle contribute largely orthogonal information relative to the other aggregate descriptors. Within the mortar statistical features, strong multicollinearity is evident. The median correlates strongly with the mean at 0.82, while the standard deviation shows a nearly perfect relationship with variance at 0.80. Similarly, the third moment has notable correlations of 0.70 with skewness. The high correlations highlight redundancy among mortar descriptors, implying that some features may be removed without significant loss of information. These insights were used to guide the subsequent ablation study, where redundant variables were removed to create a leaner, more efficient set of input features without compromising predictive accuracy.

When comparing aggregates to mortar features, the correlations are generally weak, with most values falling between -0.13 and 0.12. This indicates that spatial and geometrical descriptors of aggregates are mainly independent of the intensity-based statistical measures of mortar. The relationship between aggregate features and mechanical strength is more pronounced. Compressive strength shows the strongest negative correlation with aggregate ratio at -0.47 and mean distance at -0.43, reflecting the detrimental effect of higher aggregate content and greater spacing on concrete strength. Tensile strength exhibits similar trends, with strong negative associations with the aggregate ratio at -0.56 and moderate negative correlations with the mean distance at -0.38. These results emphasize that aggregate characteristics are primary determinants of both compressive and tensile performance. In contrast, mortar features display relatively weak associations with strength. The third moment exhibits weak positive correlations with compressive strength at 0.37 and tensile strength at 0.30. Histogram mean and Range are slightly more influential, with histogram mean correlating at 0.39 with compressive strength and 0.43 with tensile strength, while Range correlates negatively with compressive strength at -0.28 and tensile strength at -0.42. The median shows modest negative correlation with compressive strength at 0.09 and a weak positive correlation with tensile strength at 0.14. Other descriptors such as kurtosis, variance, and skewness remain largely uncorrelated with strength, with $|r| \leq 0.2$. These results confirm that mortar's statistical descriptors contribute less explanatory power compared to aggregate properties, which dominate strength prediction.

5. ANN modelling and training

The features of mortar and aggregate extracted through digital IP were used to train ANN models to predict the mechanical properties of concrete. Compressive strength and tensile strength values derived from laboratory experiments were used as training outputs for the ANN. To systematically assess the contribution of mortar and aggregate features, four distinct ANN models were developed. Each model was trained on a specific subset of input features, allowing a direct comparison of the predictive capability of aggregate features, mortar features, and their combinations. The first model, termed *AggNet*, utilized CNN-derived features from the aggregates alone. The second model, *MortarStatNet*, relied exclusively on statistical descriptors computed from mortar pixels. The third model, *ConcreteStatNet*, incorporated statistical features representing the overall concrete microstructure, including both aggregates and mortar pixels. Finally, the fourth model, *AggMortarNet*, combined the CNN-derived aggregate features with the statistical descriptors of mortar, thereby integrating information from both phases of the composite material.

5.1. Architecture and training parameters

The ANN models used in this study consisted of one input layer, two hidden layers, and one output layer. The number of input neurons corresponded to the extracted feature dimensions, while the output layer contained two neurons representing compressive strength and tensile strength. Hyperparameter choices, including the number of layers, neurons, and training algorithm, were guided by the studies [28,72], where systematic trial-and-error searches on a comparable concrete dataset identified an optimal configuration with two hidden layers, each containing a number of neurons equal to the input features. Hidden layers employed nonlinear activations (tansig/hyperbolic tangent), and the output layer used a linear activation (purelin) to model continuous targets. Networks were trained in MATLAB with Levenberg–Marquardt backpropagation (trainlm) using mean squared error (MSE) as the loss function. Training limits were set to 500 epochs with a performance goal of 1×10^{-9} , a minimum gradient threshold of 1×10^{-7} , and early-stopping patience of `max_fail` = 100. The LM damping parameter μ was initialized at 10^{-3} and adapted during optimization with a decrease factor of 0.1, an increase factor of 10, and an upper bound of 10^{10} .

To ensure balanced representation, data division was carried out at the cylinder level. The dataset consisted of 70 concrete cylinders, with four images captured from each cylinder, resulting in a total of 280 images. To avoid potential data leakage, the splitting was performed strictly at the cylinder level, ensuring that all four images from a given cylinder remained within the same partition. The 70 cylinders were randomly divided in a 60/20/20 ratio: 42 cylinders (168 images) were assigned to training, 14 cylinders (56 images) to validation, and 14 cylinders (56 images) to testing. This strategy is stricter than image-level splitting, because images from the same cylinder can appear simultaneously in training and evaluation sets, and the model may artificially achieve higher accuracy due to memorization rather than true generalization. By keeping all images from each cylinder together, the evaluation results reflect the model's ability to generalize to entirely new specimens.

In ML, data preprocessing, such as autoscaling and standardization, is essential to ensure that all features contribute proportionally during model training. Without appropriate scaling, features with larger numerical magnitudes can unduly influence the learning process, causing the model to overestimate their importance and compromise prediction accuracy. Autoscaling techniques such as standardization (Z-score scaling) and normalization (min-max scaling) are widely used to mitigate this issue by rescaling data to a comparable range. Standardization transforms data to zero mean and unit variance, while normalization rescales features to a bounded range (0–1). In this study, min-max normalization was applied to all input and output variables because the extracted features had different numerical scales. Normalization was preferred over standardization because the ANN models employed activation functions that converge more effectively when input values are restricted to a bounded range. The normalization is defined as:

$$N = \frac{V - V_{\min}}{V_{\max} - V_{\min}} \quad (21)$$

where N represents the normalized value of the variable V , and $V_{\max}$ and $V_{\min}$ denote the maximum and minimum values, respectively. This preprocessing step eliminated scale-related biases, stabilized training dynamics, and ensured that the ANN learned based on the intrinsic relationships in the data rather than being skewed by feature magnitudes.

5.2. ANN results and discussion

5.2.1. Input features ablation study

Ablation study of input features in ML is an experimental technique used to quantify the importance and contribution of specific inputs to a model's performance. Systematically removing features and monitoring the resulting changes in fit and error metrics enables the identification

Table 6
Performance of models with aggregate and mortar features.

Model	Inputs	R ²	MAPE	RMSE	MSE
OLS-Aggregate Features	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance, CDN, APA	0.6402	51.2164	0.1906	0.0363
ANN-Aggregate Features	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance, CDN, APA	0.7781	38.4945	0.0942	0.0086
	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance, CDN	0.7743	39.3073	0.1096	0.0121
	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance	0.7768	38.0173	0.0954	0.0090
	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance	0.7613	39.8214	0.1135	0.0125
	Proportion, Mean Size, Max Size, Mean Roundness	0.7606	40.0396	0.1229	0.0156
	Proportion, Mean Size, Max Size	0.7525	40.9541	0.1355	0.0182
	Mean Roundness, CDN, APA	0.6743	52.5146	0.1423	0.0196
	Max Distance, Mean Roundness, CDN, APA	0.7182	40.1364	0.1451	0.0206
	Mean Distance, Max Distance, Mean Roundness, CDN, APA	0.7378	41.0271	0.1255	0.0164
OLS-Mortar only Statistical Features	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment, Median, Kurtosis, Variance	0.5144	60.5213	0.2139	0.0480
ANN-Mortar only Statistical Features	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment, Median, Kurtosis, Variance	0.7109	42.8177	0.1363	0.0188
	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment, Median, Kurtosis	0.7037	42.9275	0.1315	0.0168
	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment, Median	0.7058	43.0156	0.1326	0.0181
	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	0.7046	42.4828	0.1273	0.0165
	Mean, Standard Deviation, Range, Entropy, Skewness	0.6847	44.8889	0.1384	0.0184
	Mean, Standard Deviation, Range, Entropy	0.6694	44.0477	0.1475	0.0211
	Mean, Standard Deviation, Range, Entropy	0.6561	46.0876	0.1446	0.0207
	Entropy, Skewness, Third Moment	0.6123	45.6123	0.1493	0.0232
	Entropy, Skewness, Third Moment, Median	0.6514	43.9321	0.1502	0.0221
	Entropy, Skewness, Third Moment, Median, Kurtosis, Variance	0.67005	44.6942	0.1394	0.0188
OLS-Aggregate and Mortar both Statistical Features	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	0.7136	45.7856	0.1837	0.0351
ANN-Aggregate and Mortar both Statistical Features	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	0.8984	26.7804	0.0753	0.0056
OLS-Aggregate Properties + Mortar Statistical Features	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance, Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	0.7983	34.2334	0.1528	0.0234
ANN-Aggregate Properties + Mortar Statistical Features	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance, Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	0.9553	8.3140	0.0403	0.0016

of variables that are critical, complementary, or redundant in terms of predictive power. In this study, the ablation examines four different configurations: aggregates-only properties, mortar-only statistical features, whole-image statistics that summarize both phases, and a fused set that combines aggregate properties with mortar statistics. For each configuration, a linear regression baseline (ordinary least squares) is reported alongside an ANN trained on the same splits, allowing the effect of input choice to be separated from model capacity. This analysis examines how model performance changes when specific inputs are omitted. Table 6 reports the performance of the models with different aggregate and mortar features. The error metric values are calculated from the normalized predicted outputs. The compressive and tensile strengths exhibit a considerable difference in their numerical ranges, varying from 3.43–21.23 MPa and 0.8–3.4 MPa, respectively. Due to this significant difference in scale, directly comparing RMSE and MSE values computed from the original data could lead to misleading interpretations, as the same absolute error would represent very different relative deviations for compressive and tensile strengths. Therefore, normalized RMSE and MSE were used to enable scale-independent comparison across models. To minimize the impact of random initialization and training stochasticity, each ANN model for a given input configuration was trained ten times independently, and the reported values in Table 6 represent the mean performance across these runs. Repeated training ensures statistical robustness and guards against the influence of random weight initial-

ization, local minima, and minor batch-order variations. Consequently, any observed difference between input combinations can be confidently attributed to the quality and relevance of the features themselves rather than training stochasticity.

Within the aggregate features block, the ANN model trained on all eight variables (proportion, mean size, max size, mean roundness, mean distance, max distance, coordination number, and average packing angle) provided the lowest error profile with an RMSE of 0.0942, an MSE of 0.0086, a MAPE of 38.49%, and an R^2 of 0.7781. A compact six-variable variant (Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance) performs almost identically, showing a slightly lower MAPE of 38.02% and very similar $R^2 = 0.7768$ and RMSE = 0.0954. These results indicate that the coordination number and average packing angle add little beyond the six core descriptors. As more variables are removed, accuracy deteriorates. When only the mean size, the maximum size, and the proportion are retained, R^2 drops to 0.7525 and RMSE increases to 0.1355. When size and proportion are removed and only Mean Roundness, Coordination Number, and Average Packing Angle are kept, the model performs much worse, as R^2 drops to 0.6743 and RMSE rises to 0.1423. The linear regression baseline with the same aggregate inputs performs markedly worse than the ANN, with R^2 of 0.6402, RMSE of 0.1906, MSE of 0.03632, and MAPE of 51.2164%. The six geometric descriptors, i.e., Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, and Max Distance, were preserved as

Table 7
Adopted ANN architectures for different input feature blocks.

Model	Input Features	Architecture
AggNet	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance	6-6-6-2
MortarStatNet	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	6-6-6-2
ConcreteStatNet	Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment (whole image)	6-6-6-2
AggMortarNet	Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max Distance, Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment	12-12-12-2

they consistently produced the best balance of accuracy and parsimony, while indices such as Coordination Number and Average Packing Angle added little beyond noise.

For the mortar statistical block, increasing the number of inputs does not automatically improve accuracy. The ANN model, when trained on all nine features, achieves the highest R^2 value of 0.7109, but with an RMSE of 0.1363, MSE of 0.0188, and MAPE of 42.82 %. The input variant, including Mean, Standard Deviation, Range, Entropy, Skewness, and Third Moment, yielded the lowest errors in this block, with an RMSE of 0.1273, MSE of 0.0165, and MAPE of 42.48 % while maintaining a comparable R^2 of 0.7046. Adding Median, Kurtosis, and Variance resulted in only marginal or mixed changes, suggesting limited utility and some noise. Trimming further input features caused noticeable degradation. With three statistics, including Entropy, Skewness, and Third Moment, R^2 falls to 0.6123, and RMSE rises to 0.1493. Similarly, retaining only the Mean, Standard Deviation, Range, and Entropy, R^2 remains at 0.6561 with an RMSE of 0.1446. The corresponding linear regression baseline for mortar statistics is weaker overall, with an R^2 of 0.5144, RMSE of 0.2139, MSE of 0.0480, and MAPE of 60.5213 %. Consequently, the input variant, including Mean, Standard Deviation, Range, Entropy, Skewness, and Third Moment, is found to be the most effective configuration for mortar texture. Removing any of these statistics leads to clear error inflation and a sudden drop in R^2 when the set becomes too small. These six indices were retained because they capture both central tendency and variability (Mean, Standard Deviation, Range) as well as higher-order distributional characteristics (Entropy, Skewness, Third Moment), thereby reflecting mortar texture and porosity. Redundant variables, such as Median, Kurtosis, and Variance, were excluded due to strong multicollinearity and weak correlation with the dependent variable, ensuring a concise yet informative input space.

To further test whether aggregate-mortar interactions could be adequately represented through global statistics, the ANN model was evaluated using whole-image features that summarized both phases together. Extending the best predicted mortar inputs to the whole image confirmed that global texture is informative. Using Mean, Standard Deviation, Range, Entropy, Skewness, and Third Moment computed over both aggregate and mortar pixels, the linear regression baseline reported an R^2 of 0.7136, MAPE of 45.7856 %, RMSE of 0.1837, and MSE of 0.0351, while the corresponding ANN model reached an R^2 of 0.8984, MAPE of 26.7804 %, RMSE of 0.0753, and MSE of 0.0056.

Fusing the most effective combination of aggregate features with the most effective mortar statistical features yielded the best overall results. The linear regression model with combined inputs achieved an R^2 of 0.7983, a MAPE of 34.23 %, an RMSE of 0.1528, and an MSE of 0.0234. In contrast, the ANN model achieved an R^2 of 0.9553, a MAPE of 8.3140 %, an RMSE of 0.0403, and an MSE of 0.0016. Relative to the whole-image statistical features, the RMSE decreased from 0.0753 to 0.0403 and the MAPE from 26.7804 % to 8.3140 %, with a clear improvement in R^2 from 0.8984 to 0.9553. The difference in performance between linear regression and the ANN model in all configurations confirms that the ANN is indispensable for capturing nonlinear feature-strength relationships.

Based on the ablation results, four ANN configurations were finalized, each reflecting the optimal input set for its respective feature block as shown in Table 7. AggNet was built on aggregate descriptors (Proportion, Mean Size, Max Size, Mean Roundness, Mean Distance, Max

Table 8
Performance of models for compressive and tensile strength (single validation).

Target	Model	R^2	MAPE	RMSE	MSE
Compressive Strength	AggNet	0.7996	37.0250	0.0912	0.0088
	MortarStatNet	0.7309	42.6500	0.1422	0.0194
	ConcreteStatNet	0.9032	25.2530	0.0700	0.0053
	AggMortarNet	0.9680	7.8542	0.0363	0.0012
Tensile Strength	AggNet	0.7785	40.5401	0.1108	0.0123
	MortarStatNet	0.7125	43.5189	0.1322	0.0166
	ConcreteStatNet	0.8771	26.5505	0.0710	0.0047
	AggMortarNet	0.9607	8.8914	0.0368	0.0013

Distance) with a 6-6-6-2 structure comprising two hidden layers of six neurons each and a two-neuron output layer. MortarStatNet employed mortar statistics (Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment) in the same 6-6-6-2 layout, ensuring a direct comparison with AggNet while focusing exclusively on mortar features. ConcreteStatNet used whole-image descriptors (Mean, Standard Deviation, Range, Entropy, Skewness, Third Moment) that summarize both phases globally, also adopting the 6-6-6-2 design to maintain comparability with the single-phase models. The most comprehensive model, AggMortarNet, fused twelve inputs into a 12-12-12-2 configuration with two hidden layers of twelve neurons each and a two-neuron output layer, achieving the highest predictive accuracy by explicitly combining aggregate geometry and mortar texture. Fig. 16 illustrates the architecture of the AggmortarNet Model, which consists of an input layer with twelve neurons, two hidden layers with nonlinear activations, and a two-neuron output layer for compressive and tensile strength.

5.2.2. Model performance analysis

To ensure accuracy, predictive models must be validated using appropriate techniques, including single validation, multiple runs performance variability, and k-fold cross-validation. In single validation, the dataset is partitioned once into training, validation, and testing subsets, providing a straightforward measure of model performance. The optimized input sets identified through the ablation study were consistently applied, allowing each model to be evaluated under conditions that highlight its predictive strengths while minimizing redundancy from highly correlated variables. For each model configuration, the results of single validation on the cylinder-level split data are reported in Table 8 using the best determined input set. The statistical parameters, including R^2 , MAPE, RMSE, and MSE, demonstrate the predictive capabilities of each model under optimized input sets identified through the ablation study. The results indicate that the models maintained strong accuracy and adaptability when applied to unseen datasets.

The AggMortarNet model demonstrated the highest performance for both compressive and tensile strength predictions as shown in Figs. 17 and 18. For compressive strength, it achieved an R^2 of 0.9680 with very low errors, MAPE = 7.8542 %, RMSE = 0.0363, and MSE = 0.0012. For tensile strength, the accuracy remained equally high with R^2 = 0.9607, MAPE = 8.89 %, and RMSE = 0.0368. AggMortarNet exhibited the tightest clustering around the $y = x$ line, with most data points falling within the ± 25 % error margins, reflecting its high accuracy. AggMortarNet combines the detailed properties of aggregates (Proportion, size, shape, and distribution) and the statistical values of the mortar pix-

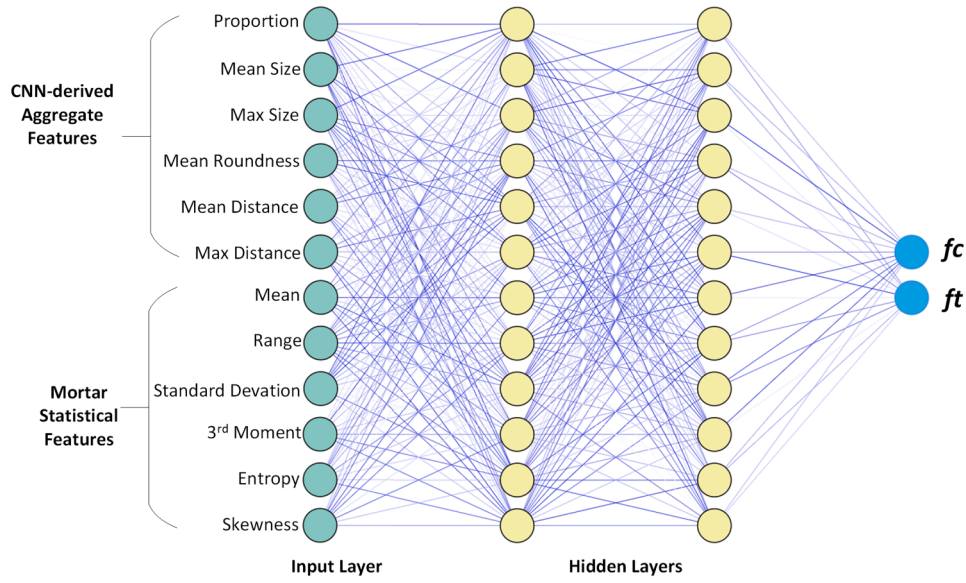
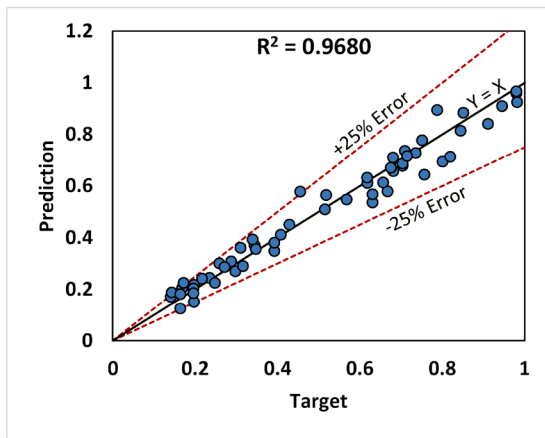
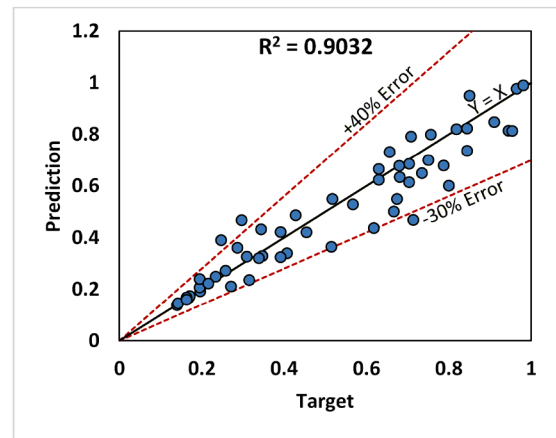


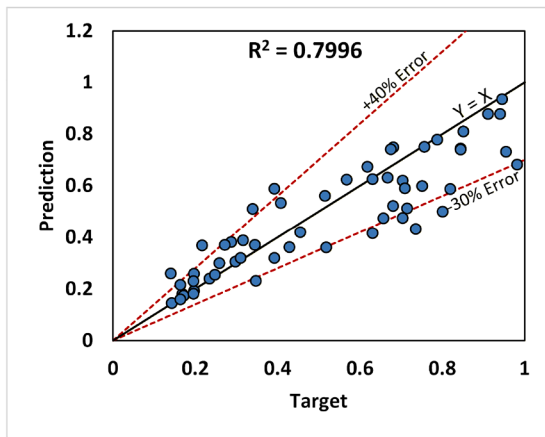
Fig. 16. Architecture of the optimal ANN (12-12-12-2) for AggMortarNet.



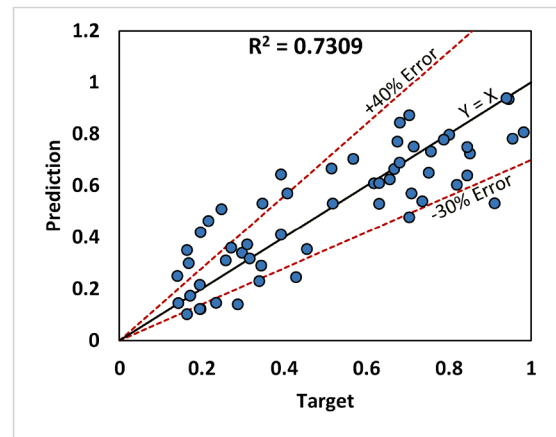
(a) AggMortarNet



(b) ConcreteStatNet



(c) AggNet



(d) MortarStatNet

Fig. 17. Comparison of predicted versus target compressive strength values for four ANN models: (a) AggMortarNet, (b) ConcreteStatNet, (c) AggNet, and (d) MortarStatNet.

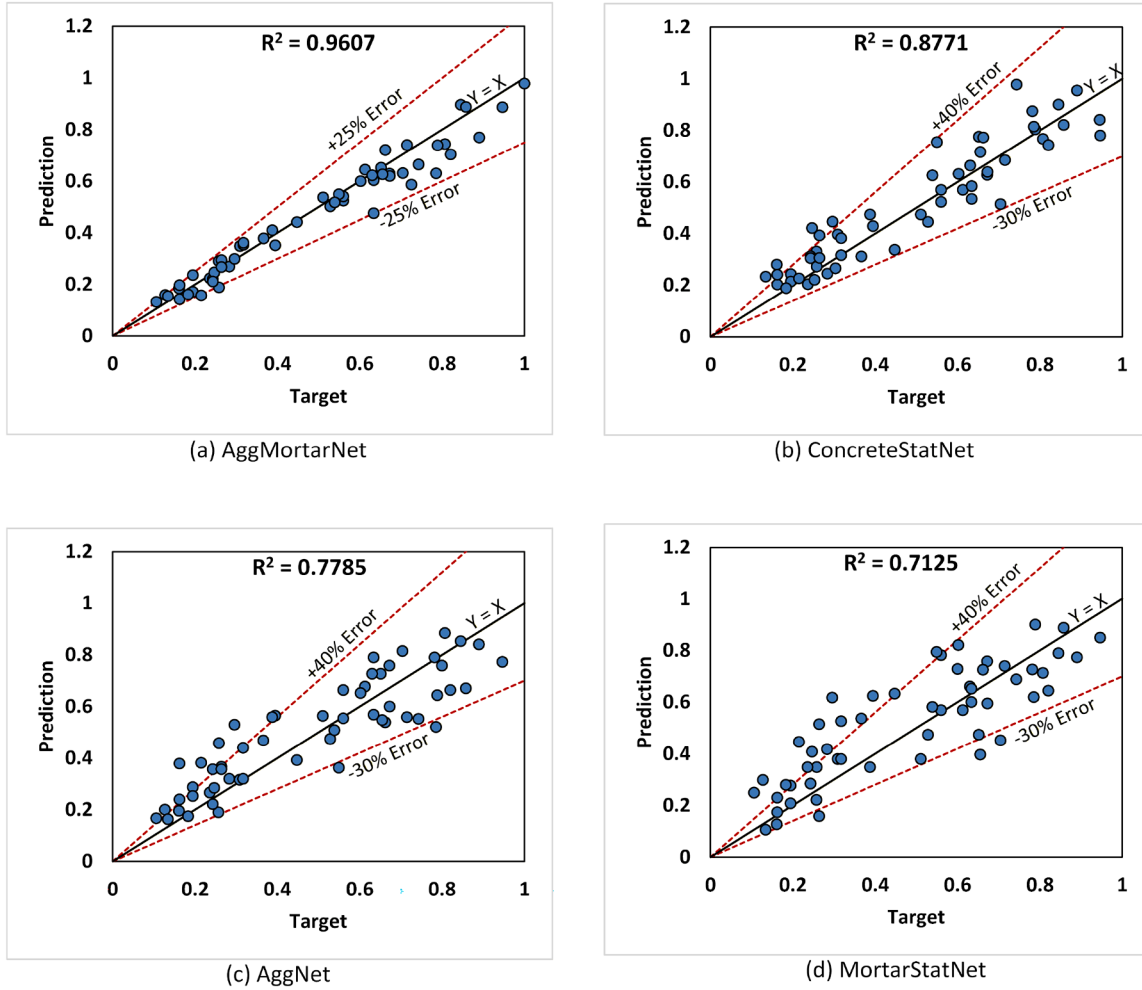


Fig. 18. Comparison of predicted versus target tensile strength values for four ANN models: (a) AggMortarNet, (b) ConcreteStatNet, (c) AggNet, and (d) MortarStatNet.

els. These results confirm that by incorporating both the morphology of aggregates and the statistical characteristics of the mortar, this model captures more detailed and relevant features that affect the mechanical properties of concrete.

The AggNet model showed moderate predictive power by using only aggregate descriptors. For compressive strength, it attained $R^2 = 0.7996$ with MAPE = 37.03% and RMSE = 0.0912, while for tensile strength it achieved $R^2 = 0.7785$, MAPE = 40.54%, and RMSE = 0.1108. AggNet exhibited larger deviations, with several points falling outside the +40% and -30% error bounds, which corresponds to their lower R^2 values. These results highlight the importance of aggregates in determining concrete strength, though ignoring mortar reduces accuracy. Similarly, MortarStatNet performed the weakest among all models. For compressive strength, it yielded $R^2 = 0.7309$ with high error levels, MAPE = 42.65%, RMSE = 0.1422. For tensile strength, it obtained $R^2 = 0.7125$, MAPE = 43.52%, and RMSE = 0.1322. MortarNet also showed wider dispersion, with several points exceeding the +40% and -30% error bands, indicating reduced predictive capability. The comparatively poor performance confirms that mortar statistical features alone provide limited explanatory power relative to aggregate morphology.

The ConcreteStatNet model also produced strong results, though slightly lower than AggMortarNet. For compressive strength, the model achieved $R^2 = 0.9032$, with MAPE = 25.25% and RMSE = 0.0700. For tensile strength, it obtained $R^2 = 0.8771$, MAPE = 26.55%, and RMSE = 0.0710. ConcreteStatNet's predictions are contained mostly within $\pm(30-40\%)$ error margins, indicating moderate performance.

This shows that statistical descriptors of both aggregates and mortar pixels capture essential trends, but without explicit aggregate morphology, the predictions are less precise. Aggregates and mortar both play crucial roles in concrete's strength and behavior. Focusing solely on aggregates or mortar neglects the contributions of the other, limiting the model's ability to predict mechanical properties accurately. Models that consider both components perform better as they capture a more complete picture of the concrete's structure. However, ConcreteStatNet uses only the statistical values of pixels from aggregates and mortar as input parameters. While this provides useful information, it does not capture the specific morphological properties of the aggregates, which are critical in determining the mechanical properties of concrete. The additional aggregate morphology data enhances the prediction accuracy, making AggMortarNet more precise than ConcreteStatNet.

Multi-run evaluation

To ensure that a single random initialization does not bias results, the models were further assessed under multiple independent runs. For each run, the networks were retrained with different weight initializations, and the distributions of R^2 , MAPE, RMSE, and MSE were recorded, as illustrated in Fig. 19. This multi-run analysis provides a more robust estimate of model reliability and reduces the influence of lucky chance in training outcomes. The averaged performance metrics across runs confirm that the reported improvements are systematic rather than dependent on a specific initialization. The boxplots demonstrate that AggMortarNet consistently outperformed all other models across repeated

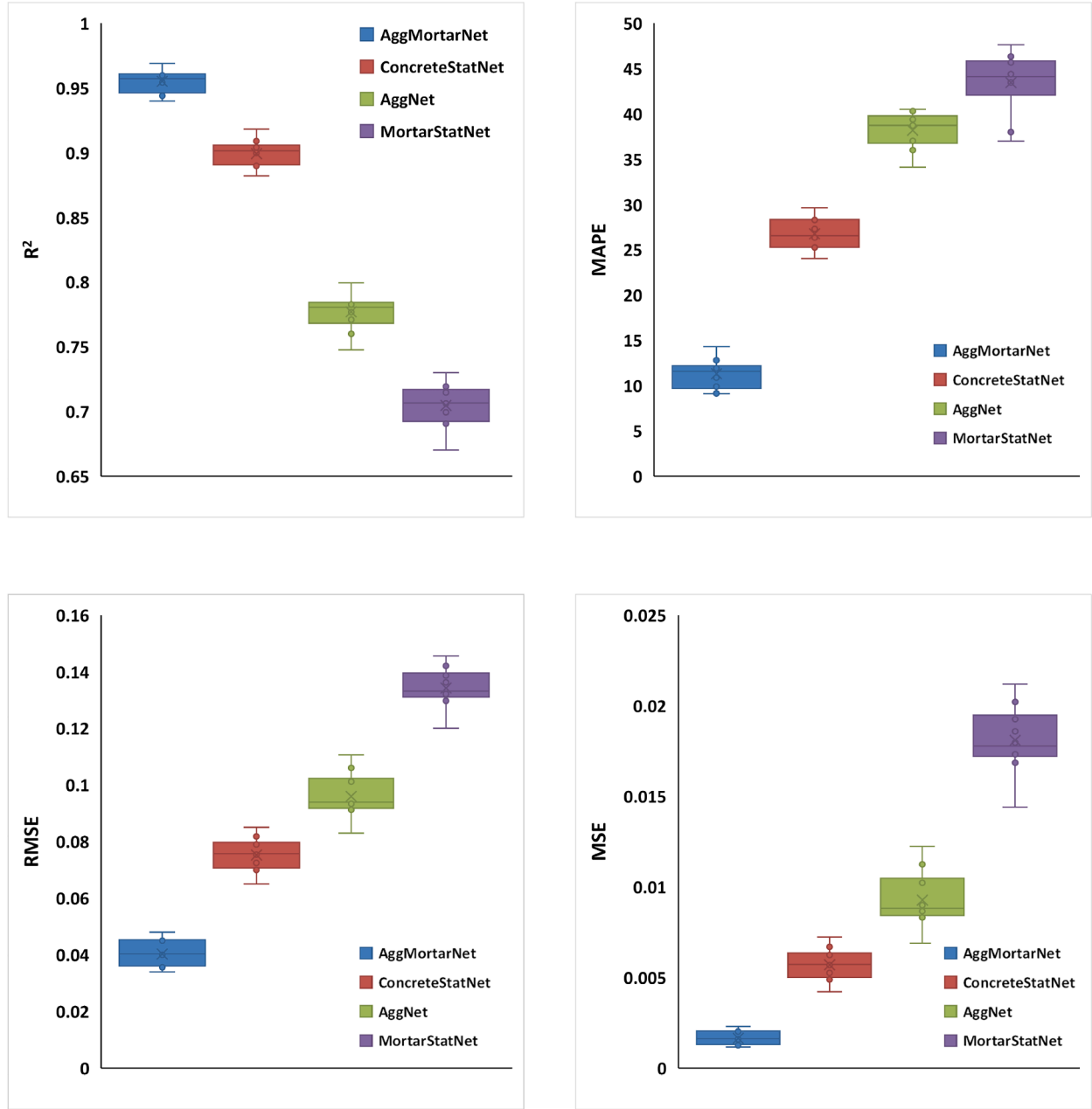


Fig. 19. Performance evaluation of AggNet, MortarStatNet, ConcreteStatNet, and AggMortarNet, across multiple runs.

trials. It achieved the highest R^2 values clustered around 0.96, with the narrowest spread, indicating stable and accurate predictions. Its error metrics also remained lowest, with MAPE values centered close to 10%, RMSE around 0.04, and MSE below 0.002, confirming both accuracy and robustness. ConcreteStatNet exhibited median R^2 values near 0.90, with MAPE around 25% and RMSE of 0.07, which were higher than those of AggMortarNet but still demonstrated consistent performance with relatively small variance across runs. AggNet followed with moderate predictive power. Its R^2 values averaged around 0.78 to 0.80, while the error distributions were broader, suggesting greater sensitivity to initialization compared to AggMortarNet and ConcreteStatNet. MortarStatNet showed the weakest and least stable performance. The R^2 values centered near 0.70, with wider dispersion, and the error metrics were the highest, with MAPE above 44%, RMSE around 0.13, and MSE close to 0.02, confirming its limited capacity to generalize across multiple training runs. The multi-run analysis confirms that AggMortarNet

not only provides the highest predictive accuracy but also demonstrates the greatest robustness and stability, making it the most reliable model among the tested architectures.

Cross-validation results

A simple train-test split can lead to an optimistic or pessimistic performance estimate if the data is not representative, and it can result in a model that is overfit to a particular subset of the training data. The k-fold cross-validation technique mitigates these risks by systematically partitioning the dataset and providing a more comprehensive assessment of the model's ability to generalize to unseen data. To ensure a robust and reliable evaluation of the ANN models, a 5-fold cross-validation using the cylinder-level splitting procedure was implemented. The 70 cylinders were randomly partitioned into five folds of 14 cylinders each. In each run, four folds (56 cylinders) were used for training, while the remaining fold (14 cylinders) served as the independent test set. This

Table 9

Performance of models for compressive and tensile strength (5-fold cross-validation).

Target	Model	R ²	MAPE	RMSE	MSE
Compressive Strength	AggNet	0.7826	37.0250	0.0912	0.0088
	MortarStatNet	0.7147	43.6500	0.1326	0.0169
	ConcreteStatNet	0.9091	26.3521	0.0770	0.0064
	AggMortarNet	0.9502	9.1453	0.0450	0.0019
Tensile Strength	AggNet	0.7475	40.3210	0.1134	0.0129
	MortarStatNet	0.6930	46.3527	0.1456	0.0201
	ConcreteStatNet	0.8823	26.5537	0.0710	0.0047
	AggMortarNet	0.9447	9.8604	0.0410	0.0016

procedure ensured that no cylinder contributed images to more than one partition within a fold. The final performance metric for each ANN model was calculated as the average of the metrics obtained from all five iterations. This approach ensures that every data point is used for both training and validation, which is particularly beneficial for maximizing data utilization.

The results of the 5-fold cross-validation are presented in Table 9. The fluctuations between single validation and 5-fold cross-validation models were minimal, generally within the range of 0.01 to 0.03 for R^2 values and corresponding small variations in error metrics, confirming the robustness and stability of the ANN architectures. For compressive strength, AggMortarNet consistently achieved the highest predictive accuracy in both strategies, with $R^2 = 0.9680$ in single validation and 0.9502 under 5-fold cross-validation. It also maintained the lowest error levels, with MAPE $\approx 9.14\%$ and RMSE ≈ 0.0450 , confirming its superior generalization ability. ConcreteStatNet achieved an R^2 of 0.9032 in single validation and 0.9091 in 5-fold cross-validation, with moderate error values, MAPE around 26.35%, and RMSE about 0.077. AggNet showed weaker performance, with R^2 falling from 0.7996 in single validation to 0.7826 in cross-validation, while MortarStatNet performed the lowest overall, with $R^2 = 0.7309$ in single validation and 0.7147 in cross-validation, and MAPE exceeding 42% in both cases.

For tensile strength, the ranking pattern was consistent. AggMortarNet delivered the best results, with $R^2 = 0.9607$ in single validation and 0.9447 in 5-fold cross-validation, together with the lowest error metrics. ConcreteStatNet followed with R^2 values of 0.8771 in single validation and 0.8823 in cross-validation, coupled with stable error levels having MAPE $\approx 26.55\%$, and RMSE ≈ 0.071 . MortarStatNet remained the weakest, with $R^2 = 0.7125$ in single validation and 0.6930 under cross-validation, and the highest errors, MAPE = 46.35%, RMSE = 0.146.

The 5-fold cross-validation results demonstrated that the models, particularly AggMortarNet, generalize well across independent folds without signs of overfitting. The close agreement between training and validation outcomes, together with the consistently high R^2 values and low error metrics across both strategies, indicates that the predictive capability of the models is not dependent on a specific data split. This stability confirms that the models capture meaningful relationships between aggregate and mortar features rather than memorizing noise, ensuring reliable performance on unseen data.

Taylor diagram

The performance of the developed ANN models for predicting the compressive and tensile strengths of concrete was evaluated using Taylor diagrams, which provide a concise statistical representation of model performance by simultaneously comparing three key parameters: the correlation coefficient (R), the standard deviation (σ), and the centered root mean square deviation (RMSD). In these diagrams, the radial distance from the origin represents the normalized standard deviation with respect to the observed data, the azimuthal position corresponds to the correlation coefficient, and the distance from the reference point (observed data point) indicates the centered RMSD. Hence, models plotted

closer to the reference arc and having higher correlation with lower centered RMSD are considered superior in predictive performance.

As shown in Fig. 20, the AggMortarNet model, which integrates both aggregate properties and mortar statistical features, achieved the highest correlation coefficient of $R = 0.98$ and the lowest normalized standard deviation, which indicates excellent agreement with experimental data and minimal prediction dispersion. ConcreteStatNet also performs well, having $R = 0.95$, but shows slightly higher variance, while AggNet and MortarStatNet are positioned farther from the reference, reflecting larger deviations and weaker correlations due to their reliance on single-phase features.

For tensile strength, all models show slightly lower correlations compared to the compressive case. AggMortarNet remains the top-performing model, with $R = 0.97$, but exhibits a modest increase in normalized standard deviation, suggesting that tensile behavior is more sensitive to micro-defects and interfacial bonding effects. ConcreteStatNet shows a correlation of $R = 0.94$, whereas AggNet and MortarStatNet exhibit higher centered RMSD and lower correlations, reflecting reduced accuracy and weaker predictive reliability for tensile strength. The Taylor diagram analysis confirms that integrating both aggregate morphology and mortar texture in a single hybrid model, AggMortarNet, yields the most accurate, stable, and generalizable prediction of concrete's mechanical properties.

5.2.3. Comparison with existing traditional methods

The normalized ANN prediction values were denormalized to retrieve their actual compressive strength values in megapascal (Mpa) for comparison with traditional methods. This process employed the inverse of the min-max normalization function, as defined in Eq. (22). This ensures that the predicted strengths correspond directly to experimentally measured results.

$$V = N \times (V_{\max} - V_{\min}) + V_{\min} \quad (22)$$

Here, V denotes the denormalized (actual) compressive strength in MPa, N is the normalized ANN output, and $V_{\max}$ and $V_{\min}$ are the maximum and minimum compressive strength values in the dataset, respectively. After denormalization, the predicted compressive strengths from AggMortarNet, RH, and UPV were compared cylinder-by-cylinder against the experimental results. Fig. 21 presents this comparison in terms of predicted and target strengths along with the corresponding percentage error distributions.

For compacted cylinders, the AggMortarNet model achieved predictions that remained within a narrow error range of $\pm 10\%$ for most specimens, indicating a strong agreement with actual compressive strength. By contrast, UPV and RH exhibited wider deviations, particularly in medium to high-strength mixes. In CM6, the ANN error was 6.46%, whereas UPV underestimated strength by -23.23% and RH by -38.35% . A similar pattern was observed in CM9, where ANN was off by -3.87% , while UPV and RH were underestimated by -19.01% and -34.64% , respectively. These results show that ANN is more capable of generalizing across different mix ratios, while UPV and RH are highly sensitive to material density and surface conditions.

The non-compacted cylinders represent a more challenging case due to weaker particle packing, higher void content, and uneven aggregate distribution. In this category, errors increased slightly for ANN, yet it remained consistently better than UPV and RH. For example, NC13 showed an ANN error of only 4.59%, while UPV and RH underestimated strength by -22.36% and -25.24% . In NC10, the ANN slightly overestimated compressive strength by 20.67%, but this was still closer to the target compared with RH (-25.28%). The relatively higher ANN error in some non-compacted mixes can be attributed to several factors. Compaction improves the uniformity of aggregate distribution in concrete, whereas non-compacted cylinders may have uneven aggregate distribution and varying densities, affecting the ability of the proposed model to accurately capture and represent these features. Additionally, non-compacted cylinders tend to have more air voids and segregation issues,

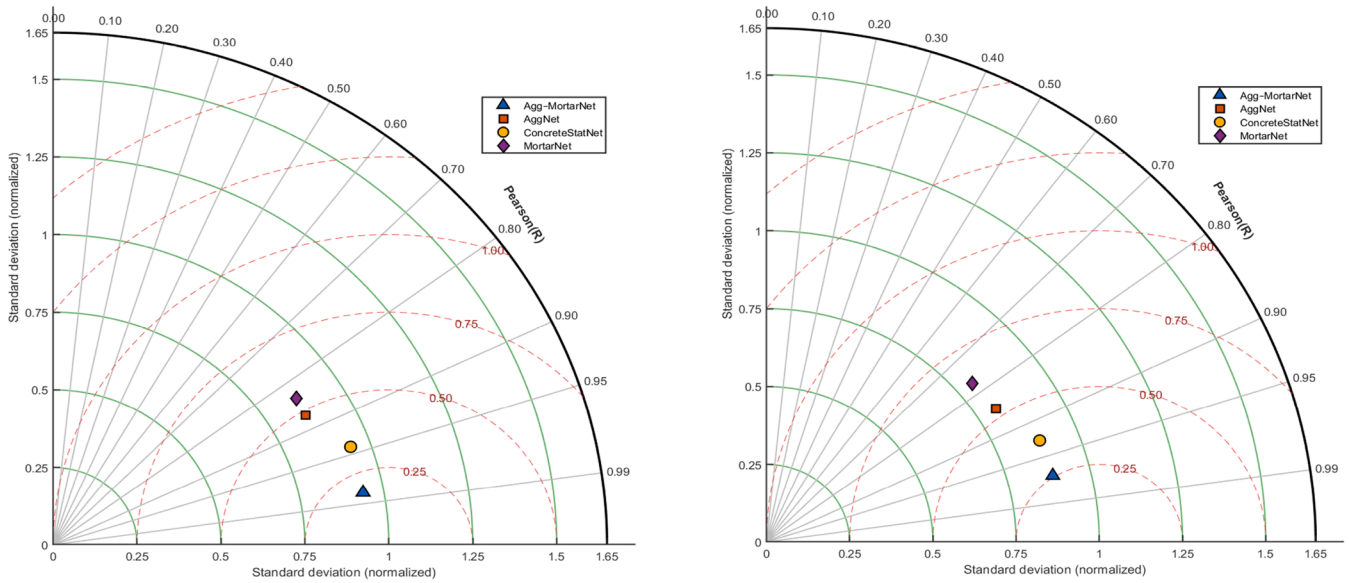


Fig. 20. Taylor diagrams showing the performance of ANN models; compressive strength (left), tensile strength prediction (right).

which can alter the concrete's structural integrity and visual appearance, impacting feature extraction and ultimately leading to relatively less accurate ANN predictions. Despite these challenges, the ANN maintained an overall trend of keeping errors within $\pm 15\%$, which proves its reliability compared with the large underestimations consistently produced by UPV and RH, often exceeding -30% to -40% .

Cylinders with very high W/C ratios presented the weakest overall performance due to excessive porosity, segregation, and bleeding, resulting in a more porous and inconsistent concrete structure. These defects not only compromise the material's integrity but also impact its surface hardness and texture, rendering traditional NDT methods particularly unreliable. In this category, UPV and RH performed poorly, with errors ranging from -23.33% to -40.53% . WC1 recorded a UPV error of -23.33% and a RH error of -36.14% . In comparison, the ANN model limited its error to 7.17% , which represents a significant improvement. In WC4, where ANN underestimated strength by -17.07% , UPV error was still higher at -16.67% . This confirms that while ANN accuracy is slightly reduced in highly porous mixes, it still outperforms traditional methods by capturing deeper structural features rather than relying on surface hardness or wave propagation speed, both of which are heavily affected by porosity.

The AggMortarNet model demonstrated the highest agreement with experimental results, showing minimal deviations and stable error trends across all cylinders. Conversely, RH and UPV methods exhibited wider fluctuations and frequent underestimations in both low- and high-strength regions. The ANN approach integrates multiple image-derived descriptors, making it more resilient to variations in mix design, compaction, and curing. This positions AggMortarNet as a reliable, non-destructive, and cost-effective alternative to conventional NDT techniques for strength evaluation in concrete structures.

Table 10 presents a comparative evaluation of the proposed model AggMortarNet against conventional DT and NDT methods in terms of coefficient of determination (R^2), practical applicability, calibration requirements, cost, and the degree of specimen destructiveness. The conventional compression test yields the most accurate results with $R^2 = 1.000$, as it measures the actual load-bearing capacity; however, it is destructive, time-consuming, and requires both specialized equipment and specimen extraction. This approach involves extracting a core sample from the structure, which leads to damage and is both time-consuming and expensive. NDT techniques offer fast alternatives for the prediction of compressive strength of concrete. Among these methods, the

UPV test achieved a reasonably high correlation of $R^2 = 0.913$, but it depends heavily on surface conditions, requires frequent calibration, and involves moderately high operational costs. The RH test, although faster and cheaper, exhibited slightly lower accuracy of $R^2 = 0.869$ due to its sensitivity to surface hardness, moisture, and aggregate type, thereby limiting its reliability for diverse concrete mixes. The proposed AggMortarNet model demonstrates superior predictive performance among all non-destructive approaches, achieving an R^2 of 0.968 . It can operate rapidly on image data without the need for calibration or specialized equipment, and its cost is minimal compared to UPV and RH methods. Moreover, because it relies solely on image-based analysis, it is entirely non-destructive, making it highly suitable for large-scale and repeated structural assessments.

The results of the AggMortarNet model for predicting the tensile strength of concrete for compacted, non-compacted, and high W/C cylinders are shown in Fig. 22. For compacted cylinders, the percentage error varied between -20.36% and $+11.40\%$, indicating a generally consistent prediction performance with moderate deviations across samples. For non-compacted cylinders, the prediction errors ranged from -21.77% to $+14.15\%$, showing higher variability compared to compacted specimens. The significant deviation in this group suggests that non-compacted concrete poses greater challenges for accurate prediction, likely due to the presence of voids and non-uniform aggregate distribution, which affect the ANN's feature extraction capability. In the case of high W/C ratio cylinders, the errors ranged from -24.81% to $+4.60\%$, reflecting the effect of higher water content on surface texture and feature uniformity. Overall, despite a few outliers, the proposed AggMortarNet model demonstrated strong generalization ability and reliable performance in predicting tensile strength across different compaction and mix conditions.

5.2.4. Comparison with existing studies

Table 11 summarizes the comparative performance of the best predicted model, AggMortarNet, against existing image-based studies for predicting the compressive strength of concrete. The RMSE and MSE were calculated using denormalized data to enable comparison with existing studies in the common engineering unit of strength (MPa). Among the earlier studies, Shiuly et al. [31], Jang et al. [32], and Lee et al. [30] achieved R^2 values ranging between 0.764 and 0.85 , with MAPE as high as 17.77% . These moderate results highlight the limited predictive capacity of CNN regression models trained directly on microscopic or



Fig. 21. Comparison of percentage error in concrete compressive strength predicted by ANN (top), UPV (middle), and RH (bottom).

Table 10

Comparison of compressive strength results of the developed model with traditional methods.

Methods	R ²	Application Practice	Calibration	Cost	Destructive Situation
Ultrasonic Pulse Velocity	0.913	Fast	Required	Moderate	Non-destructive
Rebound Hammer	0.869	Fast	Required	Low	Non-destructive
AggMortarNet	0.968	Fast	Not required	Very low	Non-destructive
Compression Test	1.000	Slow	Required	High	Destructive

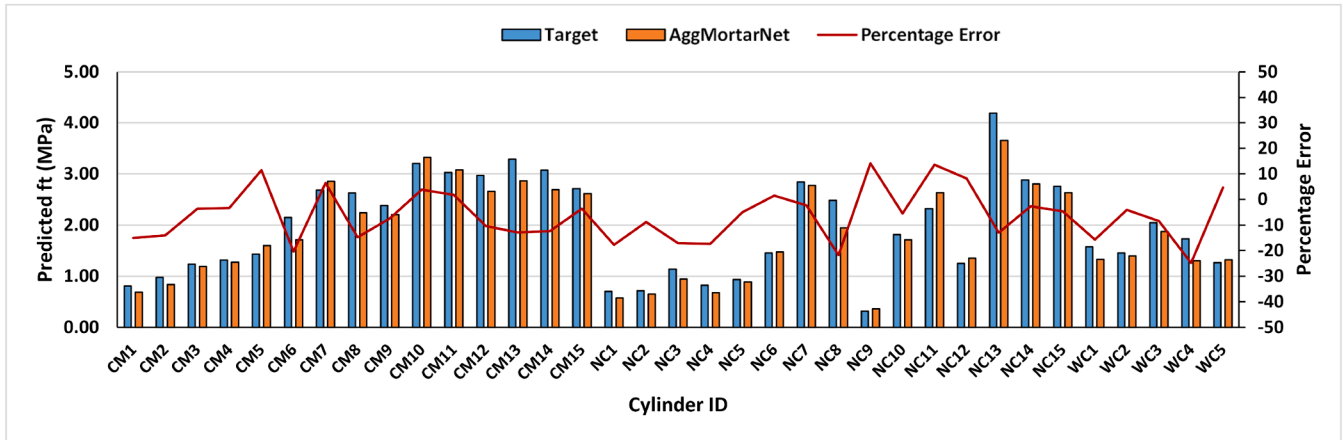


Fig. 22. Percentage error of concrete tensile strength predicted by ANN.

Table 11

Comparison of proposed work with existing studies on predicting compressive strength of concrete using image data.

Models	Number of Images	R^2	MAPE	RMSE (MPa)	MSE
Shiuly et al. (2022) [31]	2580 microscope images from 306 mortar and concrete cylinders	0.85	–	3.200	–
Başıyigit et al. (2012) [26]	42 images from 28 cubic specimens	0.9482	2.279	2.424	–
Çelik & Özdemir (2024) [24]	1453 images from 203 cores	0.9959	2.090	0.6836	0.4673
Waris et al. (2022) [28]	324 images from 54 hybrid cylinders	0.978	–	–	–
Lee et al. (2020) [30]	4709 DSLR & 3846 microscopic images	–	12.25	3.779	–
Jang et al. (2019) [32]	4000–5400 microscopic surface images from 27 cylinders	0.764	17.765	4.463	–
Plevris et al. (2025) [29]	324 images from 54 hybrid cylinders	0.970	0.70	0.1170	0.01000
AggMortarNet	280 images from 70 cylinders	0.9680	7.8542	1.314	1.7318

DSLR images without explicit feature extraction. The absence of image segmentation meant that these models primarily learned global appearance patterns rather than physical characteristics of aggregates such as size, shape, or distribution, thereby restricting their interpretability and robustness across different concrete types. Başıyigit et al. [26] achieved a comparatively higher accuracy with $R^2 = 0.9482$, RMSE = 2.424 MPa by applying manual thresholding on grayscale images to distinguish aggregates, cement paste, and voids. However, the results were highly dependent on material type, lighting, and operator judgment. Moreover, the combination of image data with UPV and RH tests made the approach semi-empirical rather than fully image-based, limiting its automation and generalizability.

Çelik & Özdemir [24] achieved the highest predictive accuracy with $R^2 = 0.9959$ and RMSE = 0.6836 MPa, by combining CNN-extracted image features with regression-based modeling to estimate compressive strength. While this hybrid strategy demonstrated remarkable precision, it required extensive manual data preprocessing and the inclusion of experimentally measured aggregate diameter. Consequently, the workflow was not fully automated and remained labor-intensive, relying on auxiliary equipment for obtaining physical measurements. Waris et al. [28] and Plevris et al. [29] also reported high predictive performance of $R^2 = 0.978$ and $R^2 = 0.970$, respectively, by employing hybrid concretes containing CRM. Such materials improved the surface color and texture of mortar, facilitating image pattern recognition. Their strong performance can therefore be attributed in part to the enhanced visual contrast introduced by CRM rather than the intrinsic capability of the prediction models.

The proposed AggMortarNet model integrates CNN-based aggregate segmentation with advanced image analysis and ANN-based prediction, achieving a robust performance of $R^2 = 0.9680$ and RMSE = 1.314 MPa. The model simultaneously quantifies aggregate morphology (proportion, size, shape, and spatial distribution) and the statistical features of mortar pixels, thereby capturing both geometric and textural attributes directly from images. Notably, it was developed and validated using normal concrete with gray-colored aggregates, which exhibit minimal

contrast between aggregate and mortar phases, a condition that significantly increases the complexity of image-based segmentation and analysis. Despite this visual challenge, AggMortarNet attained predictive accuracy comparable to, and in some cases approaching, that of CRM-based or semi-manual hybrid approaches. Additionally, the entire process is fully automated, eliminating the need for manual feature extraction or supplementary laboratory measurements.

6. Conclusion

This study developed an automated framework for predicting the mechanical properties of concrete through the integration of DL segmentation and image-derived feature analysis. The approach linked visual information of concrete constituents to their strength behavior, offering a non-destructive and data-driven alternative to traditional testing. Following are the key findings

- The DeepLabV3+ model with the EfficientNet-B5 encoder achieved highly accurate segmentation of coarse aggregates with mIoU = 0.92, and F1 = 0.95 even in gray, low-contrast concrete images, demonstrating robustness to limited color differentiation between mortar and aggregates.
- Among the extracted aggregate descriptors, aggregate proportion, mean size, and nearest-neighbor distance showed the strongest correlations with compressive and tensile strengths, confirming that aggregate geometry and packing density dominate the mechanical response.
- From the mortar region, statistical descriptors such as mean gray value, entropy, and third moment were found to be the most informative for capturing variations in the cement matrix due to changes in W/C ratio, compaction, and presence of air voids.
- The hybrid AggMortarNet model, combining aggregate geometry and mortar statistical features, achieved the highest prediction accuracy with $R^2 = 0.9680$ and RMSE = 0.0363 for compressive strength, and $R^2 = 0.9607$ and RMSE = 0.0368 for tensile strength.

- The superior performance of AggMortarNet stems from the complementary nature of the input features, where aggregate descriptors provided structural and spatial cues, while mortar features captured texture and density variations influencing strength.
- The proposed framework surpassed ultrasonic pulse velocity and rebound hammer methods by more than 6–10% in prediction accuracy, demonstrating its potential as a fast, non-destructive, and calibration-free alternative for strength estimation.

This study deepens the theoretical understanding of how concrete microstructure governs its mechanical behavior. By establishing quantitative relationships between image-derived descriptors of aggregates and mortar and their corresponding compressive and tensile strengths, it reveals that spatial distribution, morphology, and textural heterogeneity are dominant predictors of material performance. The integration of deep segmentation and statistical learning demonstrates that visual information encoded in digital images can be transformed into interpretable, physics-related parameters, thereby linking computer vision with the fundamental mechanics of cementitious materials. This approach contributes to the broader field of material informatics by providing a scalable framework for translating microstructural data into predictive models of engineering properties. Moreover, the developed methodology enables engineers and practitioners to estimate concrete strength directly from high-resolution images without destructive sampling or specialized equipment. It can be incorporated into quality-control routines, forensic assessments, and SHM systems for existing infrastructure, ensuring faster and more objective evaluations.

This present study was limited to cylindrical specimens analyzed under controlled laboratory lighting. Variations in surface finish, illumination, and image noise in field conditions may affect segmentation accuracy. Moreover, the dataset did not include concretes with admixtures or fibers, which could influence the learned feature relationships. Future work will expand the dataset to include a wider variety of concrete mixtures and aggregate types, enhancing model generalization. Furthermore, we aim to use vision transformers to enhance segmentation accuracy. We also aim to apply the proposed method to real structures by removing the concrete cover up to a depth of 10–30mm, taking images of the exposed concrete, and applying the proposed method to assess the mechanical properties of in-situ concrete. This approach will validate the method's applicability in real-world scenarios.

CRedit authorship contribution statement

Mati Ullah: Writing – original draft, Formal analysis, Data curation; **Vagelis Plevris:** Writing – original draft, Project administration, Methodology, Funding acquisition; **Junaid Mir:** Writing – original draft, Supervision, Resources, Conceptualization; **Afaq Ahmad:** Writing – original draft, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization.

Data availability

Data will be made available on request.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of using Generative AI in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) to improve the clarity, grammar, and readability of the manuscript. The authors reviewed and edited all content generated by the tool and take full responsibility for the final version of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] C.R. Gagg, Cement and concrete as an engineering material: an historic appraisal and case study analysis, *Eng. Fail. Anal.* 40 (2014) 114–140. <https://doi.org/10.1016/j.engfailanal.2014.02.004>
- [2] K. Yan, H. Xu, G. Shen, P. Liu, Prediction of splitting tensile strength from cylinder compressive strength of concrete by support vector machine, *Adv. Mater. Sci. Eng.* 2013 (1) (2013) 597257. <https://doi.org/10.1155/2013/597257>
- [3] B. Badarloo, A. Kari, F. Jafari, Experimental and numerical study to determine the relationship between tensile strength and compressive strength of concrete, *Civil Eng. J.* 4 (11) (2018) 2787–2800. <https://doi.org/10.28991/cej-03091198>
- [4] J. Helal, M. Sofi, P. Mendis, Non-destructive testing of concrete: a review of methods, *Electron. J. Struct. Eng.* 14 (1) (2015) 97–105. <https://doi.org/10.56748/ejse.141931>
- [5] Standard test method for ultrasonic pulse velocity through concrete, (2024), (ASTM C597, American Society for Testing and Materials (ASTM)). [accessed August 11, 2024]. <https://doi.org/https://store.astm.org/c0597-22.html>
- [6] Standard test method for rebound number of hardened concrete (2024) (ASTM C805, American Society for Testing and Materials (ASTM)). [accessed August 11]. <https://doi.org/https://www.astm.org/Standards/C805.html>
- [7] G. Uva, F. Porco, A. Fiore, The SonReb method: critical review and practical aspects, *Proc. Ital. Concr. Days 2016* (2018) 161–171. https://doi.org/10.1007/978-3-319-78936-1_12
- [8] B. Bolborea, C. Baera, S. Dan, A. Gruin, D.-D. Burduhos-Nergis, V. Vasile, Concrete compressive strength by means of ultrasonic pulse velocity and moduli of elasticity, *Materials* 14 (22) (2021) 7018. <https://doi.org/10.3390/ma14227018>
- [9] Z. Babović, B. Bajat, V. Đokić, F. Đorđević, D. Drašković, N. Filipović, B. Furht, N. Gačić, I. Ikodinović, M. Ilić, et al., Research in computing-intensive simulations for nature-oriented civil-engineering and related scientific fields, using machine learning and big data: an overview of open problems, *J. Big Data* 10 (1) (2023) 73. <https://doi.org/10.1186/s40537-023-00731-6>
- [10] M. Noman, M. Yaqub, M. Salman, M. Faizan, S. Mahmoudabadi, A. Ahmad, Predicting axial load capacity of CFRP fire-damaged RC columns through DANN, *Innovative Infrastruct. Solutions* 10 (12) (2025) 571. <https://doi.org/10.1007/s41062-025-02371-6>
- [11] W. Qayyum, R. Ehtisham, A. Bahrami, C. Camp, J. Mir, A. Ahmad, Assessment of convolutional neural network pre-trained models for detection and orientation of cracks, *Materials* 16 (2) (2023) 826. <https://doi.org/10.3390/ma16020826>
- [12] M. Noman, M. Salman, M. Yaqub, Artificial neural network modelling approach to predict the effect of CFRP composites on the axial strength of rectangular and square columns, in: *Proceedings of the 3rd International Conference on Advances in Civil and Environmental Engineering (ICACEE-2024)*, 2024, pp. 400–408. Conference dates: 21–22 February 2024. <https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/2105/286>
- [13] R. Kumar, E. Althaqafi, S.G.K. Patro, V. Simic, A. Babbar, D. Pamucar, S.K. Singh, A. Verma, Machine and deep learning methods for concrete strength prediction: a bibliometric and content analysis review of research trends and future directions, *Appl. Soft Comput.* 164 (2024) 111956. <https://doi.org/10.1016/j.asoc.2024.111956>
- [14] M. Timur Cihan, Prediction of concrete compressive strength and slump by machine learning methods, *Adv. Civil Eng.* 2019 (1) (2019) 3069046. <https://doi.org/10.1155/2019/3069046>
- [15] Z. Zeng, Z. Zhu, W. Yao, Z. Wang, C. Wang, Y. Wei, Z. Wei, X. Guan, Accurate prediction of concrete compressive strength based on explainable features using deep learning, *Constr. Build. Mater.* 329 (2022) 127082. <https://doi.org/10.1016/j.conbuildmat.2022.127082>
- [16] A.M. Abd, S.M. Abd, Modelling the strength of lightweight foamed concrete using support vector machine (SVM), *Case Stud. Constr. Mater.* 6 (2017) 8–15. <https://doi.org/10.1016/j.cscm.2016.11.002>
- [17] A.M. Momshad, M.H. Islam, S.D. Datta, M.H.R. Sobuz, R. Aayaz, M.K.I. Kabbo, M.M.H. Khan, Assessing the engineering properties and environmental impact with explainable machine learning analysis of sustainable concrete utilizing waste banana leaf ash as a partial cement replacement, *Cleaner Eng. Technol.* 24 (2025) 100886. <https://doi.org/10.1016/j.clet.2025.100886>
- [18] M.H.R. Sobuz, M.K.I. Kabbo, T.S. Alahmari, J. Ashraf, E. Gorgun, M.M.H. Khan, Microstructural behavior and explainable machine learning aided mechanical strength prediction and optimization of recycled glass-based solid waste concrete, *Case Stud. Constr. Mater.* 22 (2025) e04305. <https://doi.org/10.1016/j.cscm.2025.e04305>
- [19] B.A. Ali, A.S. Mohammed, M.A. Salih, S.H. Salih, M. Omer, R. Kurda, W. Emad, P.G. Asteris, P. Sihag, Artificial neural network and soft computing models to predict the compressive strength in self-compacting green concrete, *Innovative Infrastruct. Solutions* 9 (11) (2024) 412. <https://doi.org/10.1007/s41062-024-01723-y>
- [20] A.S. Mohammed, R. Kurda, D.J. Armaghani, M. Hasanipanah, Prediction of compressive strength of concrete modified with fly ash: applications of neuro-swarm and neuro-imperialism models, *Comput. Concrete* 27 (5) (2021) 489–499. <https://doi.org/10.12989/cac.2021.27.5.489>

- [21] N. Zhao, H. Zhang, P. Xie, X. Chen, X. Wang, Prediction of compressive strength of multiple types of fiber-reinforced concrete based on optimized machine learning models, *Eng. Appl. Artif. Intell.* 152 (2025) 110714. <https://doi.org/10.1016/j.engappai.2025.110714>
- [22] J. Gou, A. Zaman, F. Farooq, Machine learning-based prediction of compressive strength in sustainable self-compacting concrete, *Eng. Appl. Artif. Intell.* 161 (2025) 112190. <https://doi.org/10.1016/j.engappai.2025.112190>
- [23] F. Alsakka, S. Assaf, I. El-Chami, M. Al-Husseini, Computer vision applications in offsite construction, *Autom. Constr.* 154 (2023) 104980. <https://doi.org/10.1016/j.autcon.2023.104980>
- [24] G. Celik, M. Ozdemir, Determination of concrete compressive strength from surface images with the integration of convolutional neural networks and support vector regression methods, *Measurement* (2024) 115331. <https://doi.org/10.1016/j.measurement.2024.115331>
- [25] M. Ullah, J. Mir, S.S. Husain, M.L. U.R. Shahid, A. Ahmad, Concrete forensic analysis using deep learning-based coarse aggregate segmentation, *Autom. Constr.* 162 (2024) 105372. <https://doi.org/10.1016/j.autcon.2024.105372>
- [26] C. Başyigit, B. Comak, Ş. Kılıncarslan, İ.S. Üncü, Assessment of concrete compressive strength by image processing technique, *Constr. Build. Mater.* 37 (2012) 526–532. <https://doi.org/10.1016/j.conbuildmat.2012.07.055>
- [27] G. Dogan, M.H. Arslan, M. Ceylan, Concrete compressive strength detection using image processing based new test method, *Measurement* 109 (2017) 137–148. <https://doi.org/10.1016/j.measurement.2017.05.051>
- [28] M.I. Waris, V. Plevris, J. Mir, N. Chairman, A. Ahmad, An alternative approach for measuring the mechanical properties of hybrid concrete through image processing and machine learning, *Constr. Build. Mater.* 328 (2022) 126899. <https://doi.org/10.1016/j.conbuildmat.2022.126899>
- [29] V. Plevris, A.T. Al-Sayegh, J. Mir, A. Ahmad, Nondestructive evaluation of hybrid concrete properties using image processing and machine learning, in: *Structures*, 79, Elsevier, 2025, p. 109423. <https://doi.org/10.1016/j.istruc.2025.109423>
- [30] S. Lee, Y. Ahn, H.Y. Kim, Predicting concrete compressive strength using deep convolutional neural network based on image characteristics, *Comput. Mater. Continua* 65 (1) (2020). <https://doi.org/10.32604/cmc.2020.011104>
- [31] A. Shiuly, D. Dutta, A. Mondal, Assessing compressive strengths of mortar and concrete from digital images by machine learning techniques, *Front. Struct. Civil Eng.* 16 (3) (2022) 347–358. <https://doi.org/10.1007/s11709-022-0819-z>
- [32] Y. Jang, Y. Ahn, H.Y. Kim, Estimating compressive strength of concrete using deep convolutional neural networks with digital microscope images, *J. Comput. Civil Eng.* 33 (3) (2019) 04019018. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000837](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000837)
- [33] X. Fang, C. Wang, H. Li, X. Wang, S. Zhang, X. Luo, H. Jia, Influence of mesoscopic pore characteristics on the splitting-tensile strength of cellular concrete through deep-learning based image segmentation, *Constr. Build. Mater.* 315 (2022) 125335. <https://doi.org/10.1016/j.conbuildmat.2021.125335>
- [34] Y. Yu, C. Zhang, X. Xie, A.M. Yousefi, G. Zhang, J. Li, B. Samali, Compressive strength evaluation of cement-based materials in sulphate environment using optimized deep learning technology, *Dev. Built Environ.* 16 (2023) 100298. <https://doi.org/10.1016/j.dibe.2023.100298>
- [35] M. Ullah, J. Mir, A. Ahmad, Assessing the influence of cement replacement materials on concrete performance through image processing, *Measurement* (2025) 118014. <https://doi.org/10.1016/j.measurement.2025.118014>
- [36] A. Committee, Building Code Requirements for Structural Concrete (ACI 318-08) and Commentary, 2008. https://www.concrete.org/Portals/0/Files/PDF/Previews/318-08_preview.pdf
- [37] User's guide to ASTM specification C 94 on ready-mixed concrete, (2024), (ASTM C805, American Society for Testing and Materials (ASTM)). [accessed August 11, 2024]. <https://store.astm.org/mnl49-2nd-eb.html>
- [38] Standard test method for compressive strength of cylindrical concrete specimens, (2024), (ASTM C39/C39M 14, American Society for Testing and Materials (ASTM)). [accessed August 11]. https://store.astm.org/c0039_c0039m-14.html
- [39] Standard test method for splitting tensile strength of cylindrical concrete specimens, (2024), ASTM C496/C496M-04e1, American Society for Testing and Materials (ASTM). [accessed August 11]. <https://store.astm.org/c0496-96.html>
- [40] M. Coenen, T. Schack, D. Beyer, C. Heipke, M. Haist, Semi-supervised segmentation of concrete aggregate using consensus regularisation and prior guidance, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* 2 (2021) 83–91. <https://doi.org/10.5194/isprs-annals-V-2-2021-83-2021>
- [41] MathWorks, MATLAB, MathWorks, 2022. Available online: <https://www.mathworks.com/products/matlab.html>
- [42] M. Ullah, V. Plevris, A. Ahmad, Deep learning-based concrete aggregate segmentation and analysis for forensic applications, in: *Proceedings of the 10th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering (COMPdyn)*, 2025. <https://2025.compdyn.org/proceedings/pdf/26186.pdf>
- [43] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440. <https://doi.org/10.48550/arXiv.1411.4038>
- [44] M. Krithika Alias AnbuDevi, K. Suganthi, Review of semantic segmentation of medical images using modified architectures of UNET, *Diagnostics* 12 (12) (2022) 3064. <https://doi.org/10.3390/diagnostics12123064>
- [45] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12) (2017) 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [46] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4) (2017) 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [47] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation. *arXiv, arXiv preprint arXiv:1706.05587* 5 (2017). <https://doi.org/10.48550/arXiv.1706.05587>
- [48] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818. <https://doi.org/10.48550/arXiv.1802.02611>
- [49] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034. <https://doi.org/10.48550/arXiv.1502.01852>
- [50] Y. Liu, Y. Gao, W. Yin, An improved analysis of stochastic gradient descent with momentum, *Adv. Neural Inf. Process. Syst.* 33 (2020) 18261–18271. <https://doi.org/10.48550/arXiv.2007.07989>
- [51] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, *Int. J. Comput. Vis.* 88 (2010) 303–338. <https://doi.org/10.1007/s11263-009-0275-4>
- [52] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 31, 2017. <https://doi.org/10.1609/aaai.v31i1.11231>
- [53] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258. <https://doi.org/10.48550/arXiv.1610.02357>
- [54] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. <https://doi.org/10.48550/arXiv.1512.03385>
- [55] M. Tan, EfficientNet: rethinking model scaling for convolutional neural networks, *arXiv preprint arXiv:1905.11946* (2019). <https://doi.org/10.48550/arXiv.1905.11946>
- [56] J. Han, K. Wang, X. Wang, P.J.M. Monteiro, 2D Image analysis method for evaluating coarse aggregate characteristic and distribution in concrete, *Constr. Build. Mater.* 127 (2016) 30–42. <https://doi.org/10.1016/j.conbuildmat.2016.09.120>
- [57] A. Baddeley, E.B.V. Jensen, *Stereology for Statisticians*, CRC Press, 2004.
- [58] M.B. Haha, E. Gallucci, A. Guidoni, K.L. Scrivener, Relation of expansion due to alkali silica reaction to the degree of reaction measured by SEM image analysis, *Cem. Concr. Res.* 37 (8) (2007) 1206–1214. <https://doi.org/10.1016/j.cemconres.2007.04.016>
- [59] E. Masad, B. Muhunthan, N. Shashidhar, T. Harman, Internal structure characterization of asphalt concrete using image analysis, *J. Comput. Civil Eng.* 13 (2) (1999) 88–95. [https://doi.org/10.1061/\(ASCE\)0887-3801\(1999\)13:2\(88\)](https://doi.org/10.1061/(ASCE)0887-3801(1999)13:2(88))
- [60] X. Li, X. Lv, Y. Zhou, Z. You, Y. Chen, Z. Cui, A. Diab, Homogeneity evaluation of hot in-place recycling asphalt mixture using digital image processing technique, *J. Clean. Prod.* 258 (2020) 120524. <https://doi.org/10.1016/j.jclepro.2020.120524>
- [61] T.H. Mandeel, M.I. Ahmad, M.N. Md Isa, S.A. Anwar, R. Ngadiran, Palmprint region of interest cropping based on Moore-neighbor tracing algorithm, *Sens. Imaging* 19 (2018) 1–14. <https://doi.org/10.1007/s11220-018-0199-6>
- [62] S. Biswas, R. Hazra, Robust edge detection based on modified Moore-Neighbor, *Optik* 168 (2018) 931–943. <https://doi.org/10.1016/j.jijleo.2018.05.011>
- [63] J.F. Dong, Y. Xu, Z.W. Guan, Q.Y. Wang, Freeze-thaw behaviour of basalt fibre reinforced recycled aggregate concrete filled CFRP tube specimens, *Eng. Struct.* 273 (2022) 115088. <https://doi.org/10.1016/j.engstruct.2022.115088>
- [64] W. Wang, Image analysis of particles by modified Ferret method–best-fit rectangle, *Powder Technol.* 165 (1) (2006) 1–10. <https://doi.org/10.1016/j.powtec.2006.03.017>
- [65] AutoCAD software, (2023). Version, <https://www.autodesk.com/products/autocad/overview>
- [66] J.F. Dong, Q.Y. Wang, Z.W. Guan, H.K. Chai, High-temperature behaviour of basalt fibre reinforced concrete made with recycled aggregates from earthquake waste, *J. Build. Eng.* 48 (2022) 103895. <https://doi.org/10.1016/j.jobe.2021.103895>
- [67] P. Barbini, G. Cevenini, M.R. Massai, Nearest-neighbor analysis of spatial point patterns: application to biomedical image interpretation, *Comput. Biomed. Res.* 29 (6) (1996) 482–493. <https://doi.org/10.1006/cbmr.1996.0035>
- [68] S. Sergyan, Color histogram features based image classification in content-based image retrieval systems, in: *2008 6th International Symposium on Applied Machine Intelligence and Informatics, IEEE*, 2008, pp. 221–224. <https://doi.org/10.1109/SAMI.2008.4469170>
- [69] U. Petronas, Mean and standard deviation features of color histogram using laplacian filter for content-based image retrieval, *J. Theor. Appl. Inf. Technol.* 34 (1) (2011) 1–7. <https://www.jatit.org/volumes/Vol34No1/Vol34No1.pdf>
- [70] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient, in: *Noise Reduction in Speech Processing*, Springer, 2009, pp. 1–4. https://doi.org/10.1007/978-3-642-00296-0_5.pdf
- [71] L. Zhang, F. Huettmann, X. Zhang, S. Liu, P. Sun, Z. Yu, C. Mi, The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution, *MethodsX* 6 (2019) 2281–2292. <https://doi.org/10.1016/j.mex.2019.09.035>
- [72] A. Ahmad, D.M. Cotsosovs, N.D. Lagaros, Framework for the development of artificial neural networks for predicting the load carrying capacity of RC members, *SN Appl. Sci.* 2 (2020) 1–21. <https://doi.org/10.1007/s42452-020-2353-8>