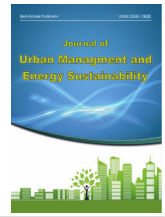
International Journal of Urban Management and Energy Sustainability (IJUMES)

Homepage: http://www.ijumes.com



CASE STUDY RESEARCH PAPER

An Integrated Approach for Analyzing and Evaluating Mix Design Proportions and Compressive Strength of Geopolymer Concrete Sections Lacking Technical Data by Integrating Digital Image Processing and Machine Learning

Mansour Keshavarzi[1], Towhid Pourrostam[2*], Mohammad Reza Javaheri Tafti[1]

1 Department of Civil Engineering, Taft. C., Islamic Azad University, Taft, Iran

2 Department of Civil Engineering, CT.C., Islamic Azad University, Tehran, Iran*

ARTICLE INFO

Article History:
Received 2026-05-04
Revised 2026-07-08
Accepted 2026-08-25

Keywords:
Geopolymer concrete; Digital image processing; Image segmentation; Gene expression programming; Compressive strength; Ground granulated blast furnace slag

ABSTRACT

Geopolymer concrete is a promising low-carbon alternative to Portland cement concrete; however, its compressive strength is governed by complex, nonlinear interactions among mix design parameters, activator characteristics. This study proposes a framework that combines digital image processing of concrete cross-sections with gene expression programming to support the assessment of slag-based geopolymer concrete when technical data are limited. Cross-sections of a slag-based geopolymer concrete specimen were imaged and processed through grayscale conversion, histogram equalization, noise removal, Canny edge detection, thresholding, and conversion from the RGB to the HSV color space. HSV color ranges were then used to generate binary masks for coarse aggregate, fine aggregate, and the slag-rich binder phase, and the coverage of each phase was quantified by pixel counting. Within the classified solid phase, the coverage percentages were 35.67% for coarse aggregate, 34.25% for fine aggregate, and 30.67% for the slag-rich binder. In parallel, a GEP model was developed from 117 compressive strength records of slag-based geopolymer concrete, using the coarse and fine aggregate contents, specimen age, NaOH molarity, Na_2SiO_3/NaOH ratio, and GGBS content as inputs. The model achieved R = 0.97, RMSE = 5.62 MPa, and MAE = 1.87 MPa for the training set and R = 0.96, RMSE = 2.66 MPa, and MAE = 1.26 MPa for the testing set. The results show that image-derived phase descriptors and an explicit GEP expression can be combined into a low-cost, partially non-destructive assessment framework, whereas full inference of mix proportions from images requires calibration against specimens of known composition.

DOI: 10.22034/ijumes.2026.743389

Running Title: Mix Design Proportions and Compressive Strength of Geopolymer Concrete Sections

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
26	11	03

*Corresponding Author:
Email: *t.pourrostam@iau.ac.ir*
Phone: 00983532628000
ORCID: https://orcid.org/0000-0002-3981-4111

INTRODUCTION

The development of sustainable construction materials and the reduction of the construction industry's dependence on resource-intensive raw materials and energy-intensive processes have become central research themes in civil engineering and materials science (Wong, 2022). Concrete is the most widely used construction material in the world owing to the availability of its raw materials, its formability, and its favorable mechanical performance; however, the production of Portland cement is associated with considerable energy consumption and substantial carbon dioxide emissions. In rapidly growing cities, where the construction and renewal of infrastructure account for a large share of material and energy demand, low-carbon binders are therefore an important component of urban sustainability strategies. Alkali-activated systems and geopolymer concretes have attracted attention as a promising pathway towards such materials (Ma et al., 2018; Zhang, 2024). Geopolymer concretes based on fly ash, ground granulated blast furnace slag (GGBS), and other aluminosilicate precursors can make use of industrial by-products while providing adequate mechanical properties and durability (Wasim et al., 2021). Nevertheless, their mechanical behavior is governed by complex interactions among the chemical composition and proportions of the precursors, the type and concentration of the alkaline activator, the liquid-to-solid ratio, the content and grading of the aggregates, the curing conditions, and the specimen age (Parathi et al., 2021; Tan & Gjorv, 1996). Consequently, the relationship between mix design parameters and compressive strength is neither linear nor simple: a change in one constituent can alter the geopolymerization process, the formation of reaction products, the porosity, the interfacial transition zone, and the continuity between phases. Compressive strength should therefore be regarded as the outcome of multiscale interactions among composition, structure, and processing conditions rather than as a function of a single constituent. A fundamental challenge in assessing hardened geopolymer concrete is obtaining information on the composition and internal structure of a specimen after curing. Under normal circumstances, the mix design can only be determined from construction records, production documents, or specialized chemical and physical tests. When a hardened specimen without baseline information is available, determining the relative proportions of its constituents and relating them to mechanical performance becomes a complex inverse problem (Chowdhury et al., 2021). Advanced laboratory techniques such as scanning electron microscopy, X-ray diffraction, spectroscopy, and X-ray computed tomography provide valuable information on the structure and composition of the material (Zelelew et al., 2013), but their cost, equipment requirements, analysis time, and need for specialized personnel limit their routine use. In this context, digital image processing offers a low-cost and scalable tool for extracting quantitative information from concrete sections (Khan et al., 2023). An image of a properly prepared section of hardened concrete contains information on the distribution of the phases, the size and shape of the aggregate particles, the proportion of voids, and the dispersion pattern of the constituents. Through appropriate preprocessing, contrast enhancement, noise removal, segmentation, and feature extraction, this information can be converted into numerical variables (Han et al., 2016; Waris et al., 2022).

The significance of this approach increases when the objective is not only to identify the phases geometrically but also to convert image-derived features into engineering variables that can be used in computational models. An image of a concrete section can thus be regarded as a primary data source for reconstructing part of the information on the structure and composition of the specimen. Descriptors such as the areal fraction of aggregate, the particle size and shape distribution, the perimeter-to-area ratio, porosity, texture, and the gray level of different regions can be used for this purpose (Ozen &

Guler, 2014; Han et al., 2016). By calibrating these descriptors against specimens with known mix designs, a computational relationship can be established between image-based characteristics and equivalent composition parameters (Guo et al., 2015). Recent advances in machine learning have created new opportunities for modeling the nonlinear relationships between the constituents of concrete and its mechanical performance. The main advantage of these methods over simple empirical relationships is their ability to capture complex interactions among many variables (Ahmed et al., 2021; Ahmed et al., 2022). Image-based approaches combined with computational models have also been used to estimate the compressive strength and other mechanical properties of concrete (Başyiğit et al., 2012; Dogan et al., 2017; Waris et al., 2022). However, most machine learning models for predicting compressive strength rely on information available at the mixing stage, such as the contents of binder, aggregate, water, and admixtures, the activator characteristics, the age, and the curing conditions. When the mix design is unknown, such models cannot be applied directly. This limitation is a significant gap in the practical use of data-driven models, because in the assessment of existing materials the original composition is frequently unavailable and only the hardened specimen, or an image of its section, can be examined. Accordingly, the problem can be reformulated from a pure strength-prediction problem into a multistage inverse problem (Damadipour et al., 2019). In the first stage, the information contained in the image of a concrete section is converted into quantitative variables through image processing and feature extraction. In the second stage, these variables are related, through calibration data, to equivalent composition parameters. In the third stage, the estimated parameters are used as inputs to a computational model that predicts compressive strength. The present study implements and demonstrates the first and third components of this framework for slag-based geopolymer concrete: (i) an HSV-based image analysis procedure that quantifies the coverage of coarse aggregate, fine aggregate, and the slag-rich binder phase in a concrete section, and (ii) an explicit gene expression programming (GEP) model that predicts compressive strength from mix design parameters and specimen age. The calibration link between image-derived descriptors and mix proportions (the second stage) is identified as the next step and is discussed as a limitation. In this framework, image processing serves to extract information from the specimen, whereas machine learning serves to infer hidden relationships and predict performance, providing a basis for faster, lower-cost, and partially non-destructive assessment of geopolymer concrete when the original mix design is unavailable.

MATERIALS AND METHIODS

This section reviews studies that have applied image processing to the characterization of aggregates and concrete and to the estimation of concrete properties. Barbosa et al. (2011) applied image processing to analyze segregation in lightweight aggregate concrete. Because of the density difference between the lightweight aggregates and the mortar, the aggregates tend to float during vibration, which may affect the mechanical properties of the concrete. The authors proposed an image-based classification tool and an index for classifying concretes according to the uniformity of the particle distribution, and validation tests showed that the approach can be used to evaluate the overall particle distribution in concrete specimens. Bruno et al. (2012) used image analysis to detect the aggregate gradation of asphalt mixtures from planar images. By combining different digital image classification methods applied to images of asphalt cores, they developed a reliable classification procedure based on 24 images from six asphalt surfaces. The gradations estimated for 22 images correlated strongly with the actual gradation, and the remaining images reproduced the gradation with minor errors. Ozen and Guler (2014) assessed the optimum threshold and particle shape pa-

rameter for image analysis of the aggregate size distribution of concrete sections. Two-dimensional images of the lateral surfaces of specimens prepared with six mix designs of different gradations, aggregate sources, and maximum aggregate sizes (five replicates each) were analyzed. The optimum threshold and pixel distribution functions differed between cross-sections, and the maximum Feret diameter, with an average error of less than 10%, was identified as the most suitable shape descriptor for estimating the percentage of aggregate passing each sieve. Guo et al. (2015) developed a stereological method for estimating the three-dimensional aggregate gradation of asphalt mixtures from digital images. The aggregates were idealized as ellipsoids, and the corresponding stereological coefficients were determined computationally. The aspect ratio of the aggregates had a significant influence on the stereological coefficients, and refining this assumption considerably reduced the estimation error compared with existing methods.

Han et al. (2016) proposed a two-dimensional image analysis method for evaluating the characteristics and distribution of coarse aggregate in concrete. The aggregate characteristics included particle size, angularity, texture, and orientation, whereas the distribution was described by the homogeneity of the particle distribution, the inter-particle spacing, and the mortar-to-aggregate area ratio (MAAR). The method was applied to one conventional and four reinforced concretes, and comparisons of the gradation curves and of the MAAR with the mortar-to-aggregate volume ratio (MAVR) confirmed the consistency between two- and three-dimensional information. The method also revealed the effect of random factors, such as vibration and non-uniform sampling, on concrete homogeneity. Başyiğit et al. (2012) and Dogan et al. (2017) investigated the estimation of concrete compressive strength from images of concrete specimens and showed that features extracted by image processing can be related to the measured strength, offering an alternative to conventional destructive testing. More recently, Waris et al. (2022) combined image processing with machine learning to estimate the mechanical properties of hybrid concrete and reported that image-derived features can serve as useful predictors of mechanical performance. Damadipour et al. (2019) evaluated the size distribution of coarse aggregates using image processing and introduced a self-adjusting size parameter, defined as the minimum dimension of the equivalent shape of a particle. Two coarse aggregate samples from the Aji-Chay River, with D50 values of 12.5 and 19 mm, were analyzed, and the gradations predicted by the proposed algorithm agreed well with the measured gradations. The reviewed studies show that image analysis can reliably quantify aggregate gradation, shape, and distribution, and that image-derived features can be related to concrete strength. However, most of these studies addressed asphalt mixtures or Portland cement concrete, and few have examined geopolymer concrete. Moreover, image-based phase quantification and explicit, interpretable strength models have generally been developed separately. The present study addresses this gap by combining HSV-based phase quantification of geopolymer concrete sections with an explicit GEP model for compressive strength within a single assessment framework.

The geopolymer concrete was prepared from an aluminosilicate precursor, alkaline activator solutions of sodium hydroxide and sodium silicate, natural sand, and natural gravel. Ground granulated blast furnace slag (GGBS), supplied by Sepahan Cement Company (Isfahan, Iran), was used as the sole aluminosilicate precursor; its chemical composition is given in Table 1. The sodium silicate solution was a gel containing 54.1% water, 32.5% SiO_2, and 1.12% Na_2O. The sodium hydroxide solution, with a concentration of 12 M, was prepared by dissolving NaOH pellets of 95% purity in distilled water 24 h before casting, so that the solution could return to ambient temperature after the exothermic dissolution. A sodium silicate-to-sodium hydroxide mass ratio of 1.5 was used. (Tab. 1)

Table 1: Chemical composition of GGBS (wt.%)

Oxide	Content (%)
SiO2	32.57
Al2O3	1.26
Fe2O3	3.3
CaO	34.07
MgO	0.89
SO3	0.5
Na2O	1.12
K2O	0.56
Loss of ignition	2.28

Natural dolomitic sand and gravel from Amol, Iran, were used as the fine and coarse aggregates, respectively. Their grading curves are shown in Figure 1, and both were graded in accordance with ASTM C33 (ASTM International, 2003). The water absorption of the fine and coarse aggregates was 2.8% and 2.1%, and their specific gravities were 2.565 and 2.580, respectively. The maximum aggregate size (Dmax) was 12.5 mm. (Fig. 1)

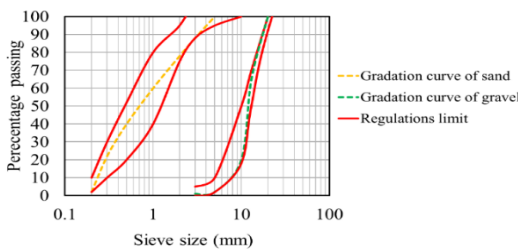


Figure 1: Grading curves of the aggregates.

mixed until a uniform dry blend was obtained. The alkaline activator, consisting of the sodium hydroxide and sodium silicate solutions, was then added, and mixing continued for 2 min. The fresh concrete was cast in pre-oiled cylindrical molds, after which curing of the specimens commenced (Tab. 2 and Fig. 2).

Methodology

Mix Design

The mix proportions are given in Table 2. The mixture contained 400 kg/m³ GGBS, 675 kg/m³ fine aggregate, 850 kg/m³ coarse aggregate, 150 kg/m³ sodium hydroxide solution, and 225 kg/m³ sodium silicate solution (Na_2SiO_3/NaOH = 1.5), giving a fresh unit weight of 2300 kg/m³. The mixture was prepared in a 50-L laboratory mixer. The coarse and fine aggregates were first dry-mixed, after which the GGBS was added and

ID	Material consumption (kg/m3)				
	FA	CA	NaOH solution	GGBS	Na2SiO3/NaOH
GPC	675	850	150	400	1.5



Figure 2: Preparation of the geopolymer concrete specimens.

Digital Image Processing

Digital image processing (DIP) enables classification tasks, including the classification of soil and rock materials, to be performed with reasonable speed and accuracy. Many scientific and industrial fields require the analysis of particle shape (Ozen & Guler, 2014), because particle characteristics govern the structure, gradation, and mechanical properties of granular materi-

als. Numerous descriptors have been proposed to characterize particle shape, some describing the overall form and others describing finer surface details (Guo et al., 2015). Barrett proposed describing particle shape at three scales form, roundness (angularity), and surface texture as illustrated in Figure 3 (Han et al., 2016). (Fig. 3)

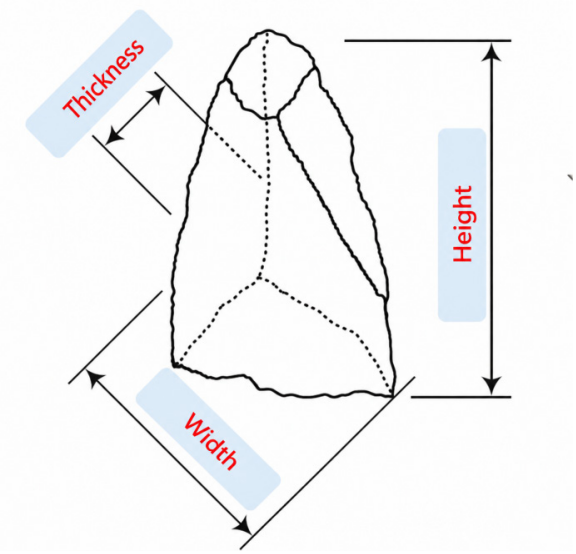


Figure 3: Particle shape characteristics at different scales (Guo et al., 2015).

In particle morphology, Wadell defined sphericity as a measure of how closely a particle approaches a sphere, and roundness as the ratio of the average radius of curvature of the corners to the radius of the maximum inscribed circle; Krumbein and Rittenhouse later proposed visual charts for estimating roundness and sphericity from two-dimensional projections (Zelelew et al., 2013). Sphericity and roundness are independent descriptors: a particle may be well rounded yet elongated, and therefore of low sphericity (Waris et al., 2022). Particle size is routinely measured with sieves, but descriptors such as elongation, angularity, roundness, and sphericity cannot readily be obtained with conventional laboratory equipment, and their experimental determination is costly and time-consuming. Image-based approaches have therefore been proposed for extracting particle shape characteristics. Because the coarse aggregate forms the principal load-bearing skeleton of concrete and its size and grading must fall within the limits of ASTM C33 (ASTM International, 2003), the present study uses Python-based digital image processing to quantify the coarse aggregate, fine aggregate, and slag-rich binder phases in geopolymer concrete sections.

Fundamentals of Digital Images

In artificial intelligence applications, image processing refers to the set of operations applied to digital images to enhance their quality and to extract information from them; it is an essential preprocessing step for tasks such as object detection and classification. A digital image is stored as a two- or three-dimensional numerical matrix in which each element represents a pixel; in 8-bit images, pixel values range from 0 to 255 (Dogan et al., 2017). Pixel intensity is described by a function I(x, y) for single-channel images and I(x, y, c) for multichannel color images, where c denotes the channel. Based on this representation, images can be divided into four general categories:

- Binary Images
- Grayscale Images
- RGB Color Images
- RGBA Images

In binary images, each pixel takes one of two values, 0 (black) or 1 (white); such images are commonly used to isolate specific regions of an image (Bovik, 2009). Grayscale images have a single channel with 256 intensity levels, ranging from 0 (black) to 255 (white), with intermediate values representing shades of gray (Kamel & Zhao, 1993; Kumar & Verma, 2010). RGB color images consist of three channels red, green, and blue each stored as an 8-bit matrix of the same size; a 24-bit RGB image can therefore represent about 16.7 million colors. A pixel with values (0, 0, 0) is black, a pixel with values (255, 255, 255) is white, and a pixel with values (255, 0, 0) is pure red (Danahy et al., 2007). RGBA images contain a fourth (alpha) channel that stores transparency information.

FINDINGS AND DISCUSSION

Machine Learning Modeling

Database

To develop the compressive strength model, a database of 117 experimentally measured compressive strength records of slag-based geopolymer concrete with NaOH concentrations of 4–8 M was used (Waris et al., 2022). The data were preprocessed in MATLAB, and 75% of the records (88 data points) were assigned to training and 25% (29 data points) to testing. The output variable was the compressive strength. The workflow of the research is shown in Figure 4. (Fig. 4)

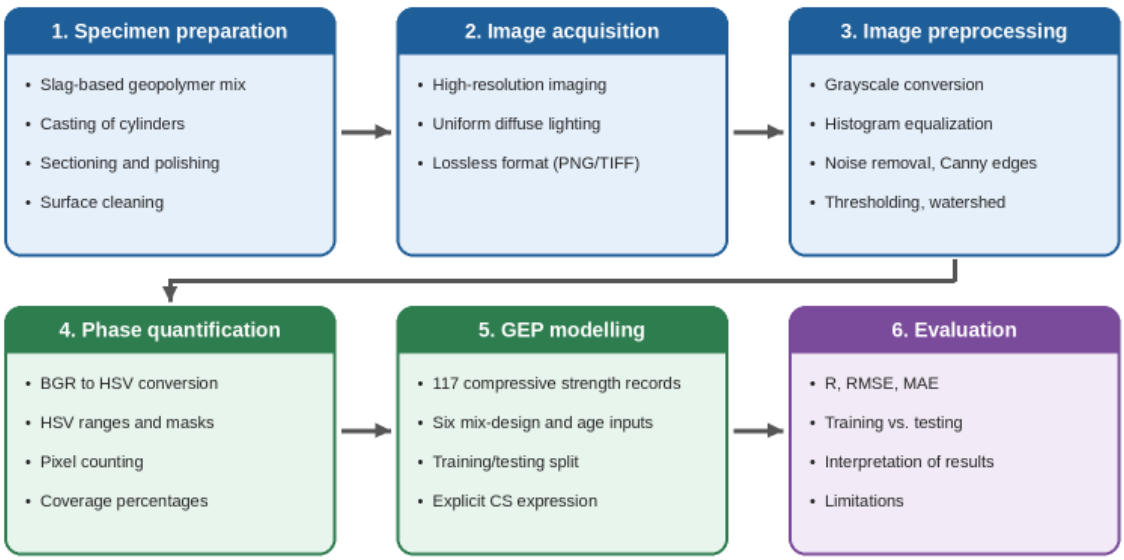


Figure 4: Workflow of the research

As expressed in Equation (1), the input variables of the compressive strength (CS) model were the coarse aggregate content (CA), fine aggregate content (FA), specimen age (AS), NaOH molarity, sodium silicate-to-sodium hydroxide ratio (Na_2SiO_3/NaOH), and GGBS content.

$$CS = f(CA, FA, AS, Molarity, Na_2SiO_3/NaOH, GGBS) \quad (1)$$

The descriptive statistics of the variables (minimum, maximum, mean, and standard deviation) are reported in Table 3. The standard deviation indicates the spread of the data around the mean; the relatively large standard deviations show that the database covers a wide range of mix designs and ages, which is favorable for model development. Figure 5 presents the Pearson correlation coefficients between the input and output variables. The coefficient ranges from −1 to 1, where 1 indicates a perfect positive correlation, 0 no linear correlation, and −1 a perfect negative correlation. Specimen age showed the strongest positive correlation with compressive strength, whereas the remaining variables showed weak to moderate correlations, confirming that compressive strength cannot be described by any single variable. (Tab. 3 and Fig. 5).

Gene Expression Programming Method

Gene expression programming (GEP) was introduced by Ferreira (2001). Like the genetic algorithm (GA) and genetic programming (GP), GEP evolves a population of candidate solutions by selecting individuals according to a fitness function and applying genetic operators such as mutation, transposition, and recombination. GA

Table 3: Descriptive statistics of the model variables

Variable	Mean	Standard Deviation	Minimum	Maximum
CA	827.50	88.02	518.96	960.30
FA	854.2	114.8	468.4	1035.6
AS	41.67	35.39	7.00	90.00
Molarity	6.000	1.640	4.000	8.000
Na_2SiO_3/NaOH	2.0171	0.3880	1.5000	2.5000
GGBS	318.04	35.95	265.52	379.31
CS (MPa)	62.93	10.95	37.70	90.11

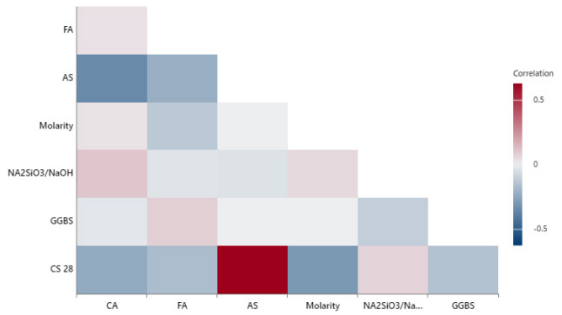


Figure 5: Pearson correlation matrix of the input and output variables.

encodes solutions as fixed-length linear strings and GP as parse trees of variable size and shape; GEP combines both features by encoding individuals as fixed-length linear chromosomes that are expressed as expression trees of different sizes and shapes (Ferreira, 2001). The algorithm starts by generating an initial population of chromosomes, either randomly or using prior knowledge of the problem. Each chromosome is expressed as an expression tree and evaluated using the fitness function; the evolution continues until a satisfactory solution is found or a predefined number of generations is reached, and the best individual is returned as an explicit mathematical expression. This explicit output is the main advantage of GEP over black-box models such as artificial neural networks.

Model performance was evaluated using the correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE), defined in Equations (2)–(4), where y_i and $\hat{y}_i$ are the measured and predicted values of the i-th data point, $\bar{y}$ and the overbar on $\hat{y}$ denote the corresponding mean values, and n is the number of data points.

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}} \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (4)$$

Image Analysis Procedure

The image analysis procedure used to quantify the constituents of the concrete specimens consisted of the following stages.

Preparation and Preprocessing

Step 1: Preparation of the specimens

Before imaging, the specimens were prepared as follows (Fig. 6):

•Sectioning: the cylindrical specimens were cut to expose a cross-section suitable for imaging.

•Surface polishing: the cut surface was ground and polished to obtain a smooth, uniform surface and a clear image of the aggregates.

•Surface cleaning: dust and loose particles were removed from the surface to avoid image artifacts.



Figure 6: Preparation of the specimens prior to image analysis.

Step 2: Imaging of the specimens

Images of the prepared cross-sections were captured as follows:

•Camera: a high-resolution digital camera was used.

•Lighting: uniform, shadow-free, diffuse lighting was used to enhance contrast.

•Image format: the images were saved in lossless formats (TIFF or PNG) to avoid compression artifacts.

Step 3: Image preprocessing

The preprocessing stage, which aimed to improve image quality and reduce noise, is illustrated in Figure 7 and comprised the following operations (Ozen & Guler, 2014; Han et al., 2016):

•Grayscale conversion: the color image was converted into a grayscale image.

•Contrast enhancement: histogram equalization was applied to improve the distinction between the aggregates and the binder matrix.

•Noise removal: smoothing filters were applied to reduce image noise.

•Edge enhancement: the Canny edge detector was applied to detect the aggregate boundaries.

The aggregates were then separated from the binder matrix as follows:

•Thresholding: adaptive thresholding was used to separate regions of different intensity.

•Segmentation: the watershed algorithm was used to separate touching aggregate particles.

•Boundary refinement: the separated boundaries were examined and refined.

•Connected component labeling: the number and position of the aggregate particles were determined.

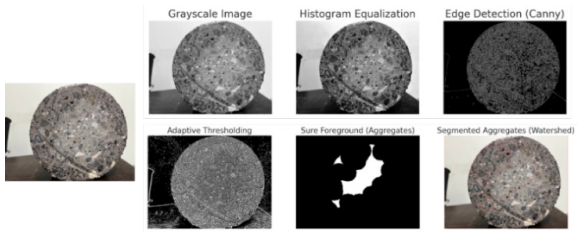


Figure 7: Image preprocessing stages.

Extraction of Aggregate Features

The quantification of the aggregates and of the slag-rich binder phase involved identifying and separating the material phases (aggregate, binder matrix, and voids) in the cross-sectional images, followed by:

•Connected component analysis to count the aggregate particles;

•Determination of the area of each particle from its connected pixels;

•Calculation of the average particle dimensions from the particle size distribution.

Phase Segmentation Stages

Conversion of the Image to the HSV Color Space

The color image was converted from the BGR (blue, green, red) color space used by OpenCV to the HSV (hue, saturation, value) color space. HSV separates chromatic information (hue and saturation) from brightness (value), which makes color-based classification less sensitive to variations in illumination.

Determination of Color Ranges

Color ranges were defined in the HSV space for each constituent (gravel, sand, and slag-rich binder) and adjusted according to the colors observed in the images (Fig. 8). On the OpenCV scale (H: 0–179; S and V: 0–255), the following

ranges were used:

•Gravel: H = 20–30, S = 100–255, V = 100–255.

•Sand: H = 10–20, S = 50–255, V = 50–255.

•Slag-rich binder: S = 0–50 and V = 100–255 (bright, low-saturation regions) for all hues.

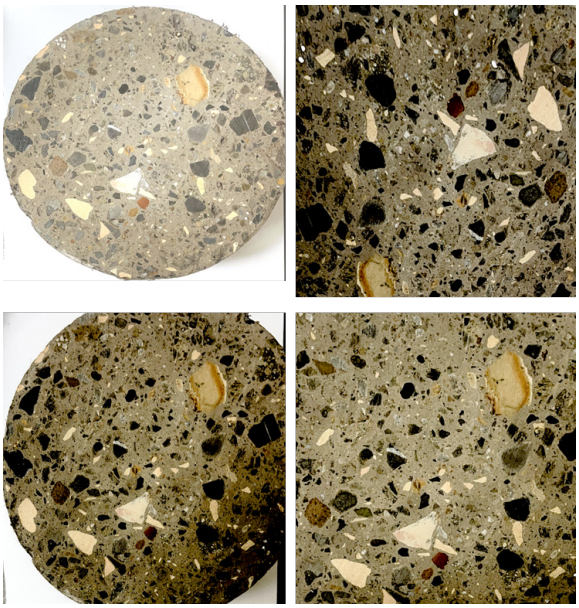


Figure 8: Cross-sections of the geopolymer concrete specimens used to define the HSV color ranges.

Generation of Masks

A binary mask was generated for each constituent using the cv2.inRange function in Python. In each mask, the pixels belonging to the constituent are white (255) and all other pixels are black (0).

Pixel Counting

The number of white pixels in each mask, which represents the area of the corresponding constituent in the image, was counted using cv2.countNonZero. Figure 9 illustrates the intermediate outputs:

•Original image (a): a grayscale rendering of the cross-section in which the aggregates (gravel and sand), the slag-rich phase, and the matrix can be distinguished.

•Processed image (b): the image after contrast enhancement and intensity transformation, in which the constituents are separated according to their visual features.

•Binary image (c): the final output, in which the pixels are assigned to bright regions (aggregates and slag-rich phase) or dark regions (matrix); this image is used to quantify the area fraction and distribution of the bright phases.

Calculation of Coverage Percentage

The coverage percentage of each constituent was calculated using Equation (5), where N_{total} is the number of pixels assigned to the constituent and N_{total} is the number of pixels in the reference area:

$$\text{Coverage } (\%) = \frac{N_{\text{phase}}}{N_{\text{total}}} \times 100 \qquad (5)$$

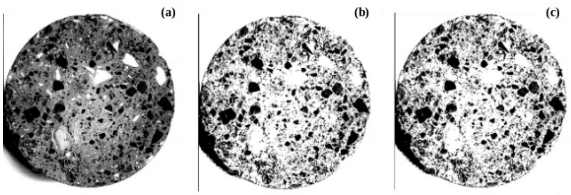


Figure 9: Pixel classification of the geopolymer concrete section: (a) original image, (b) processed image, and (c) binary image.

RESULTS AND CONCLUSION

Results of Image Analysis

The coverage percentages of gravel, sand, and the slag-rich binder phase were calculated by color segmentation and pixel counting using the Python code shown in (Fig. 10).

The binary classification of the cross-section (Fig. 9c) yielded the following coverage percentages:

•Bright regions (gravel, sand, and slag-rich phase): 15.71% of the image area.

•Dark regions (matrix): 84.29% of the image area. These regions correspond to the reacted geopolymer binder formed from the alkaline activator and the precursor.

These values describe the relative distribution of the bright and dark phases in the section.

The bright regions were then classified using the HSV masks. Expressed as percentages of the

```python
import cv2
import numpy as np

def analyze_concrete_image(image_path):
    # Load the colour image and convert it to HSV
    image = cv2.imread(image_path)
    hsv = cv2.cvtColor(image, cv2.COLOR_BGR2HSV)

    # HSV ranges of the three solid phases
    lower_gravel = np.array([20, 100, 100])   # gravel: Hue 20-30
    upper_gravel = np.array([30, 255, 255])
    lower_sand = np.array([10, 50, 50])       # sand: Hue 10-20
    upper_sand = np.array([20, 255, 255])
    lower_slag = np.array([0, 0, 100])        # slag-rich phase:
    upper_slag = np.array([180, 50, 255])     # low S, Value 100-255

    # Binary masks (255 = phase, 0 = other)
    mask_gravel = cv2.inRange(hsv, lower_gravel, upper_gravel)
    mask_sand = cv2.inRange(hsv, lower_sand, upper_sand)
    mask_slag = cv2.inRange(hsv, lower_slag, upper_slag)

    # Pixel counting
    gravel_px = cv2.countNonZero(mask_gravel)
    sand_px = cv2.countNonZero(mask_sand)
    slag_px = cv2.countNonZero(mask_slag)
    solid = cv2.bitwise_or(cv2.bitwise_or(mask_gravel, mask_sand), mask_slag)
    solid_px = cv2.countNonZero(solid)

    # Coverage of each phase relative to the classified solid pixels
    return (100 * gravel_px / solid_px,
            100 * sand_px / solid_px,
            100 * slag_px / solid_px)

g, s, b = analyze_concrete_image('concrete_section.png')
print(f"Gravel = {g:.2f}%, Sand = {s:.2f}%, Slag = {b:.2f}%")
```

Figure 10: Python code used for HSV-based phase segmentation and pixel counting.

classified solid-phase pixels, the coverage of each constituent was as follows:

•Gravel: 35.67%

Gravel covered 35.67% of the classified solid phase, reflecting the substantial contribution of the coarse aggregate to the load-bearing skeleton of the concrete.

•Sand: 34.25%

Sand covered 34.25% of the classified solid phase; the fine aggregate fills the spaces between the coarse particles and contributes to the compactness of the matrix.

•Slag-rich binder: 30.67%

The slag-rich phase covered 30.67% of the classified solid phase. Because GGBS is the precursor of the geopolymer binder, this phase represents the binder-rich regions that provide cohesion between the aggregates. The sum of the three values (100.59%) slightly exceeds 100% because the HSV ranges are inclusive at their limits (H = 20 for gravel and sand, and S = 50 for sand and the slag-rich phase), so that a small number of boundary pixels are assigned to two classes.

Development of the Compressive Strength Model

The 117 records were introduced into GeneXpro-Tools 5.0 to derive an explicit expression relating the compressive strength to the input variables (CA, FA, AS, Molarity, Na_2SiO_3/NaOH, and GGBS). The resulting GEP expression is given in Equation (6), where f'c is the compressive strength (MPa); CA, FA, and GGBS are in kg/m³; AS is the specimen age (days); and M is the NaOH molarity.

$$f'_c = \left(\frac{\mathrm{AS} + \mathrm{GGBS} - 4{,}23}{(\mathrm{Na_2SiO_3/NaOH}) \times (\mathrm{GGBS} - \mathrm{M})} + 2{,}42\right) \times \left(\sqrt{4{,}23 - \frac{\sqrt{\mathrm{GGBS}}}{\mathrm{FA}}} - \sqrt{\mathrm{Na_2SiO_3/NaOH} + 10{,}87} + 6{,}65\right) \times \sqrt[4]{10{,}87(\mathrm{AS} + \mathrm{GGBS} + 4{,}23) + \mathrm{GGBS} - \mathrm{CA}} \quad (6)$$

Figure 11 compares the predicted and measured compressive strengths for the training and testing sets. For the training set, the model achieved R = 0.97, RMSE = 5.62 MPa, and MAE = 1.87 MPa; for the testing set, the corresponding values were R = 0.96, RMSE = 2.66 MPa, and MAE = 1.26 MPa. The high correlation coefficients and the similar performance on the training and testing sets indicate that the model captured the nonlinear relationships between the input variables and compressive strength without evident overfitting. The data points are distributed closely around the line of equality over a range of approximately 35–90 MPa.(Fig. 11).

Interpretation of Results

The primary objective of the image analysis was to transform the visual information contained in a cross-section of geopolymer concrete into quantitative data that can be used to identify its constituents and, subsequently, to estimate the parameters affecting compressive strength. Unlike conventional assessments based mainly on visual inspection or direct measurement, the present method extracts information on the internal structure of the specimen through digital processing and converts it into quantitative variables. Such an approach is particularly valuable when baseline information on the mix

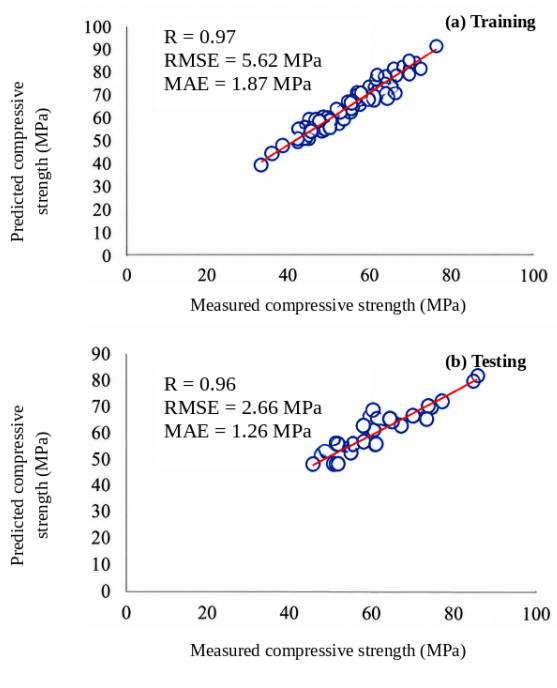


Figure 11: Predicted versus measured compressive strength of the GEP model: (a) training set and (b) testing set.

design of a hardened specimen is unavailable. The preprocessing results showed that color space conversion, contrast enhancement, noise removal, boundary detection, and segmentation enabled the relative discrimination of the constituents in the cross-section. The HSV space is advantageous because the color differences between the constituents can be examined largely independently of the overall brightness; consequently, the color ranges defined for gravel, sand, and the slag-rich phase enabled a dedicated mask to be generated for each phase.

From an engineering perspective, the value of this stage lies not in the binary image itself but in its ability to provide information on the relative proportion and distribution of the constituents. Studies based on two-dimensional analysis of concrete sections have shown that the areal fraction of aggregate, particle size, distribution, shape, and uniformity of dispersion provide useful information about concrete structure (Ozen & Guler, 2014; Han et al., 2016). However, two-dimensional results must be interpreted with due regard to the distinction between areal and volumetric quantities: the areal fraction obtained from an image should not be equated directly with the volumetric or mass fraction of the mixture constituents without stereological correction and validation (Guo et al., 2015). The performance of the proposed algorithm can therefore be assessed at two levels: its ability to separate the phases visually and its capacity to convert this separation into quantitative indicators. The pixel-counting results show that the method converts information on the principal constituents of the cross-section into numerical values, which forms the basis for linking image-derived features to computational models of compressive strength.

Quantitative Analysis of the Distribution of Constituent Phases

After segmentation, the coverage of each constituent was calculated as the ratio of the number of pixels assigned to that constituent to the total number of reference pixels, thereby transforming the qualitative structure of the section into quantitative data. In the binary classification, the bright regions (aggregates and slag-rich phase) accounted for 15.71% of the image area, and the dark matrix regions for the remaining 84.29%. Within the classified solid phase, the coverage percentages were 35.67% for gravel, 34.25% for sand, and 30.67% for the slag-rich binder phase, indicating that all three phases contribute substantially to the observed structure. The coarse aggregate forms the load-bearing skeleton and reduces the volume of binder required, whereas the fine aggregate and the binder fill the spaces between the coarse particles and govern the continuity and compactness of the matrix. These percentages should nevertheless be interpreted with caution, because the areal fraction in a two-dimensional section is not necessarily equivalent to the mass fraction in the mix design: the size, shape, and orientation of the particles and the position of the cutting plane all

affect the apparent contribution of each phase (Guo et al., 2015). The main contribution of this stage is therefore not the individual percentage values but the demonstration that a hardened specimen without compositional records can be transformed, through its image, into a set of quantitative descriptors that can support the assessment of existing concrete, quality control, and the examination of specimens with incomplete records.

Engineering Interpretation of Aggregate and Binder Distribution

The distribution of aggregates affects stress transfer, compactness, and the mechanical behavior of concrete. A uniform distribution of aggregates within the matrix reduces local stress concentrations and provides a more continuous load path, whereas clustering of aggregates or paste-rich zones leads to heterogeneity and increases the likelihood of stress concentration. Image-based studies have used the uniformity of distribution, inter-particle spacing, and mortar-to-aggregate area ratio to characterize concrete structure (Han et al., 2016). In the examined specimen, the segmentation results allowed the aggregate phases and the binder phase to be identified separately. The proposed method can therefore also serve as a basis for extracting more advanced indicators such as particle count, particle area, equivalent diameter, aspect ratio, sphericity, perimeter, inter-particle spacing, and distribution uniformity.

Information obtained from a single image can thus be extended from a simple coverage percentage to a set of structural features, in line with image-based studies that characterize concrete through particle size, shape, roundness, orientation, and spacing (Ozen & Guler, 2014; Waris et al., 2022). At the same time, the mechanical performance of geopolymer concrete depends not only on the aggregate content but also on the precursor composition, activator concentration, ratio of the alkaline solutions, specimen age, and curing conditions. An image of the cross-section can reveal part of the structural information, but a complete interpretation of compressive strength requires that image-derived features be used together with mix design variables and processing conditions.

Relationship Between Image-Derived Features and Compressive Strength

A key component of the proposed framework is the computational model for estimating compressive strength. The 117 records were divided into training (88 records, 75%) and testing (29 records, 25%) sets, and the inputs comprised the coarse aggregate, fine aggregate, specimen age, molarity, Na_2SiO_3/NaOH ratio, and GGBS content. These inputs are justified by the behavior of geopolymer concrete, whose compressive strength results from the simultaneous interaction of the aggregate skeleton, the quantity and nature of the binder, and the conditions of the geopolymerization reaction; a simple linear relationship between any single constituent and compressive strength is therefore not expected, as also reflected in the correlation matrix (Figure 5). GEP is well suited to such problems because it searches for nonlinear relationships among multiple variables and returns them as an interpretable mathematical expression (Equation 6). In the proposed framework, the aggregate contents estimated from image analysis are intended to replace the corresponding mix design inputs once the relationship between areal coverage and mix proportions has been calibrated.

Generalizability of the Proposed Approach

A potential advantage of the proposed method is that it reduces dependence on baseline mix design information. Conventional data-driven prediction of the compressive strength of geopolymer concrete requires the raw-material and production variables to be specified in advance, whereas the ultimate objective of the present framework is to extract part of this information from the hardened specimen itself. This distinction is important in practice. In the assessment

of existing structures, information on the original mix design, material proportions, binder type, or production conditions may be unavailable, and imaging of the cross-section with extraction of geometric and phase-related features can provide information that is not available from project documentation. The generalizability of the method, however, depends strongly on image quality, lighting conditions, surface preparation, differences in the color and texture of the materials, and the procedure used to define the HSV ranges. Segmentation studies have shown that threshold parameters may vary with the type of cross-section and the imaging conditions (Ozen & Guler, 2014); the use of a fixed set of thresholds for all specimens must therefore be accompanied by adequate validation. Accordingly, the present results should be interpreted as a demonstration of the applicability of the framework to the studied specimen rather than as conclusive proof of its performance for all geopolymer concretes.

This study presented a framework that combines digital image processing and gene expression programming for the assessment of slag-based geopolymer concrete sections with limited technical data. The main conclusions are as follows:

•HSV-based segmentation and pixel counting separated the coarse aggregate, fine aggregate, and slag-rich binder phases of the geopolymer concrete section, with coverage percentages of 35.67%, 34.25%, and 30.67% of the classified solid phase, respectively; in the binary classification, the bright regions accounted for 15.71% and the matrix for 84.29% of the image area.

•The GEP model developed from 117 compressive strength records provided an explicit expression for compressive strength in terms of the aggregate contents, specimen age, NaOH molarity, Na_2SiO_3/NaOH ratio, and GGBS content, with R = 0.97, RMSE = 5.62 MPa, and MAE = 1.87 MPa for the training set and R = 0.96, RMSE = 2.66 MPa, and MAE = 1.26 MPa for the testing set.

•Image-derived phase descriptors and the explicit GEP expression together form a low-cost, partially non-destructive basis for assessing existing geopolymer concrete whose construction records are incomplete. Full inference of mix proportions from images requires calibration against specimens of known composition, stereological correction, and validation over a wider range of materials and imaging conditions, which are recommended for future work. By supporting the assessment and reuse of low-carbon geopolymer concrete in urban infrastructure, the framework contributes to more sustainable and energy-efficient construction practice.

REFERENCES

Ahmed, H. U., Mahmood, L. J., Muhammad, M. A., Faraj, R. H., Qaidi, S. M., Sor, N. H., ... & Mohammed, A. A. (2022). Geopolymer concrete as a cleaner construction material: An overview on materials and structural performances. Cleaner Materials, 5, 100111.

Ahmed, H. U., Mohammed, A. A., Rafiq, S., Mohammed, A. S., Mosavi, A., Sor, N. H., & Qaidi, S. M. (2021). Compressive strength of sustainable geopolymer concrete composites: a state-of-the-art review. Sustainability, 13(24), 13502.

ASTM International. (2003). Standard specification for concrete aggregates (ASTM C33-03). ASTM International.

Barbosa, F. S., Beaucour, A. L., Farage, M. C. R., & Ortola, S. (2011). Image processing applied to the analysis of segregation in lightweight aggregate concretes. Construction and Building Materials, 25(8), 3375-3381.

Başyiğit, C., Comak, B., Kılınçarslan, Ş., & Üncü, İ. S. (2012). Assessment of concrete compressive strength by image processing technique. Construction and Building Materials, 37, 526-532.

Bovik, A. C. (2009). Basic binary image processing. In The Essential Guide to Image Processing (pp. 69-96). Academic Press.

Bruno, L., Parla, G., & Celauro, C. (2012). Image analysis for detecting aggregate gradation in asphalt mixture from planar images. Construction and Building Materials, 28(1), 21-30.

Chowdhury, S., Mohapatra, S., Gaur, A., Dwivedi, G., & Soni, A. (2021). Study of various properties of geopolymer concrete–A review. Materials Today: Proceedings, 46, 5687-5695.

Damadipour, M., Nazarpour, M., & Alami, M. T. (2019). Evaluation of particle size distribution using an efficient approach based on image processing techniques. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 43, 429-441.

Danahy, E. E., Agaian, S. S., & Panetta, K. A. (2007, February). Algorithms for the resizing of binary and grayscale images using a logical transform. In Image Processing: Algorithms and Systems V (Vol. 6497, pp. 305-314). SPIE.

Dogan, G., Arslan, M. H., & Ceylan, M. (2017). Concrete compressive strength detection using image processing based new test method. Measurement, 109, 137-148.

Ferreira, C. (2001). Gene expression programming: A new adaptive algorithm for solving problems. Complex Systems, 13(2), 87–129.

Guo, Q., Bian, Y., Li, L., Jiao, Y., Tao, J., & Xiang, C. (2015). Stereological estimation of aggregate gradation using digital image of asphalt mixture. Construction and Building Materials, 94, 458-466.

Han, J., Wang, K., Wang, X., & Monteiro, P. J. (2016). 2D image analysis method for evaluating coarse aggregate characteristic and distribution in concrete. Construction and Building Materials, 127, 30-42.

Kamel, M., & Zhao, A. (1993). Extraction of binary character/graphics images from grayscale document images. CVGIP: Graphical Models and Image Processing, 55(3), 203-217.

Khan, M. A. M., Kee, S. H., Pathan, A. S. K., & Nahid, A. A. (2023). Image processing techniques for concrete crack detection: a scientometrics literature review. Remote sensing, 15(9), 2400.

Kumar, T., & Verma, K. (2010). A theory based on conversion of RGB image to gray image. International Journal of Computer Applications, 7(2), 7–10.

Ma, C. K., Awang, A. Z., & Omar, W. (2018). Structural and material performance of geopolymer concrete: A review. Construction and Building Materials, 186, 90-102.

Ozen, M., & Guler, M. (2014). Assessment of optimum threshold and particle shape parameter for the image analysis of aggregate size distribution of concrete sections. Optics and Lasers in Engineering, 53, 122-132.

Parathi, S., Nagarajan, P., & Pallikkara, S. A. (2021). Ecofriendly geopolymer concrete: A comprehensive review. Clean Technologies and Environmental Policy, 23, 1701-1713.

Tan, K., & Gjorv, O. E. (1996). Performance of concrete under different curing conditions. Cement and Concrete Research, 26(3), 355-361.

Waris, M. I., Plevris, V., Mir, J., Chairman, N., & Ahmad, A. (2022). An alternative approach for measuring the mechanical properties of hybrid concrete through image processing and machine learning. Construction and Building Materials, 328, 126899.

Wasim, M., Ngo, T. D., & Law, D. (2021). A state-of-the-art review on the durability of geopolymer concrete for sustainable structures and infrastructure. Construction and Building Materials, 291, 123381.

Wong, L. S. (2022). Durability performance of geopolymer concrete: A review. Polymers, 14(5), 868.

Zelelew, H. M., Almuntashri, A., Agaian, S., & Papagiannakis, A. T. (2013). An improved image processing technique for asphalt concrete X-ray CT images. Road Materials and Pavement Design, 14(2), 341-359.

Zhang, B. (2024). Durability of sustainable geopolymer concrete: A critical review. Sustainable Materials and Technologies, e00882.

COPYRIGHTS
©2023 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.

HOW TO CITE THIS ARTICLE

Keshavarzi, M., Pourrostam, T. & Javaheri Tafti, M. R. (2026). An Integrated Approach for Analyzing and Evaluating Mix Design Proportions and Compressive Strength of Geopolymer Concrete Sections Lacking Technical Data by Integrating Digital Image Processing and Machine Learning. (e743389). International Journal of Urban Management and Energy Sustainability, 7(3), e743389

DOI: 10.22034/ijumes.2026.743389