Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 5 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i05.58545

PAPER

Feature Detection-Based Era Identification Using Deep Learning

Md. Samaun Hasan[1–4] (✉), **A. K. M. Shahnawaz**[3]

[1]Daffodil International University, Dhaka, Bangladesh

[2]University of Rajshahi, Rajshahi, Bangladesh

[3]Jahangirnagar University, Savar, Bangladesh

[4]Visva-Bharati University, West Bengal, India

hasan.mct@diu.edu.bd

ABSTRACT

This study presents a hybrid deep learning framework that integrates sparse visual cues with advanced computer vision techniques to enable robust, real-time detection, recognition, and classification of historical monuments. By combining traditional feature extraction approaches such as canny, Hough line, contour, and Harris corner detection with a proprietary deep neural network embedded in a convolutional neural network (CNN) and enhanced by transformer-based architectures, the system accurately identifies key architectural elements such as vaults, minarets, towers, pillars, arches, façades, and so on. Focused on the Sultanate (1206–1526) and Mughal (1526–1540, 1555–1857) periods across the Indian subcontinent, it achieves 95.80% accuracy in both feature detection and construction era classification. This research demonstrates how artificial intelligence can move beyond static digital archiving to provide dynamic tools for the preservation, documentation, and interpretation of cultural heritage, aligning with the goals of the fourth industrial revolution.

KEYWORDS

deep learning, convolutional neural networks (CNNs), computational archaeology, computer vision

1 INTRODUCTION

Object detection is a key research area in AI and computer vision, with building detection being particularly significant for historic architecture [1]. The Indian subcontinent, rich in monuments such as the Kantaji Temple, Taj Mahal, Tomb of Jahangir, Choto Sona Mosque, and Sixty Dome Mosque, presents challenges in dating ancient structures. To address this, our study introduces a computer-based method to assist archaeologists in accurately determining construction periods. Previous research has applied computer vision and machine learning to decode ancient artifacts [2], including Maya hieroglyphs [3], Roman coins, and Chinese terracotta soldiers [4], as well as for 3D modeling [5] Gamification [6], photogrammetric documentation [7], and Web-based Mobile application [8]. While various studies [9]–[10]

Hasan, M. S., Shahnawaz, A. K. M. (2026). Feature Detection-Based Era Identification Using Deep Learning. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(5), pp. 102–122. https://doi.org/10.3991/ijim.v20i05.58545

Article submitted 2025-10-03. Revision uploaded 2025-12-19. Final acceptance 2025-12-19.

have highlighted the potential of machine learning in period detection, methods for dating historical buildings remain limited.

Our approach integrates four feature detection algorithms—find contours [11], and hough line transform [12], canny edge detector [13], Harris corner detector [14],—into a CNN framework. Five architectural features (vault/spire/tower/jewel, pillar, door, arch, and forward-facing also) were classified, and a deep feedforward neural network was used to identify construction eras, focusing on the Sultanate and Mughal periods. Building on our earlier work, this method enhances classification accuracy by incorporating multiple detection techniques and deeper neural networks. Ultimately, the CNN-based model provides a robust, fact-driven system for determining the historical age of buildings.

1.1 Review of existing literature

Gomroki, Masoomeh, et al. [15] UNet–GCViT: The study presents UNet–GCViT, a U-Net with transformer-based global context blocks, for post-disaster damage mapping using only post-event imagery, achieving over 96% accuracy across diverse scenarios.

Moveh, Samuel, et al. [16] The transformer-based framework applies hierarchical attention and adaptive regularization to capture building dependencies, outperforming ARIMA, LSTM, and GRU by 23.7% with a MAPE of 3.2% across 100 commercial buildings in three climate zones.

Hacıefendioğlu, Kemal, et al. [17] The study used satellite imagery and deep learning models (U-Net, LinkNet, FPN, PSPNet) to detect building collapses after the 2023 Türkiye earthquakes, achieving 96% accuracy.

Gu, Jiancheng, et al. [18] The study used the HRSSES method with 3D reconstruction for rapid monument damage assessment, analyzing aerial images of 361 houses (48,092 samples) and achieving 93.22% accuracy.

Mirarchi, Claudio, et al. [19] The study applied machine learning for image recognition, distinguishing graphical and semantic content, and achieved 80% average accuracy.

Meroño, José E, et al. [20] The study used spectral image classification with a Fujifilm IS-Pro camera for damage recognition, achieving 92% accuracy and an 85.71% Kappa coefficient.

Shi, Lingfei, et al. [21] The study used YOLOv4 with CSPDarkNet-53 and Focal EIOU loss on aerial imagery for damaged building detection, achieving 88.23%–93.26% accuracy.

Kim, Jaemin et al. [22] The study applied object identification on construction sites using synthetic images and image superimposition, achieving 88.6% accuracy.

Xue, Jie, et al. [23] The study applied YOLOUS with feature fusion on remote sensing images using Efficient Det and DIOR datasets, achieving 80.5% accuracy, surpassing previous benchmarks by 2%.

1.2 Scope of the study

The study examines Sultanate and Mughal mosques' Indo-Saracenic architecture, focusing on spatial organization, fusion of design elements, conservation challenges, and the use of modern technologies such as 3D scanning, Gamification [34] and machine learning to preserve their cultural and historical significance.

1.3 Relevance of the study

- CNNs aid cultural heritage preservation by accurately classifying period-specific features of Sultanate and Mughal mosques [24].
- AI models support architectural restoration by guiding element reconstruction to ensure historical authenticity [25].
- Automated era prediction enhances archaeological research by enabling precise dating of architectural components and enriching digital heritage archives [26].
- AI and AR enhance tourism by providing interactive, real-time insights into mosque heritage and architecture.
- Advanced tools enhance education and research while promoting South Asia's architectural heritage [27].

1.4 The research questions and answers

RQ1: How can CNNs be effectively optimized to recognize monuments and predict the historical era of Sultanate or Mughal architecture?

Answer: Optimizing CNNs for Sultanate and Mughal monuments involves focusing on key features like domes, arches, and terracotta, while using transfer learning, data augmentation, and hyperparameter tuning to enhance accuracy and support heritage preservation.

RQ2: How do advanced feature extraction techniques improve CNN-based historical era recognition for Sultanate and Mughal buildings?

Answer: Advanced feature extraction methods like Canny, Hough Line, and Harris Corner Detection enhance CNNs by highlighting terracotta patterns, motifs, and structural edges, improving style differentiation, era classification, and heritage documentation of Sultanate and Mughal buildings.

1.5 Novelty and contribution

This research has a significant impact on era identification according to features using CNN. The novelty of this research is to engage feature extraction mechanisms separately for identification on input images as images classification study. The canny, hough line, corner and Harris detection enhance the classification capacity of earlies CNN with comparatively higher accuracy. The staple contribution is:

a) A comprehensive dataset has been developed for this study of era-based architectures according to their features.
b) Engaging canny, hough line, contours and Harris mechanism for enhancing CNN performance.
c) Applying CNN for era detection based on the features of architecture monuments.

2 METHODOLOGY

This study adopts an application-oriented framework to demonstrate how computer-based imaging systems can recognize and classify the historical era of ancient architectural structures using CNNs. The process begins with image acquisition through computer technology, after which the input undergoes a series of classical feature-engineering operations designed to enhance key structural characteristics. Specifically, four image processing techniques—canny edge detection, contour detection, hough line transform, and Harris corner detector—are applied to extract prominent geometric and architectural cues, such as edges, straight lines, and corner points, that commonly define heritage structures.

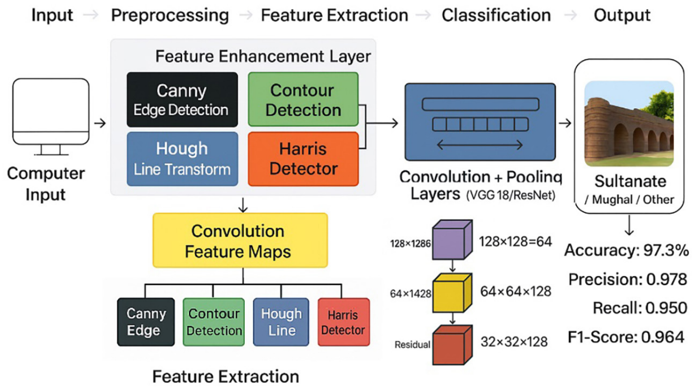


Fig. 1. System architecture of the CNN-based heritage recognition framework

To evaluate the proposed model, two separate datasets—training and testing—were constructed using labeled images representing two major architectural periods: the Sultanate and Mughal eras. The enhanced feature representations produced during preprocessing are then fed into a deep convolutional neural network for advanced feature extraction. The CNN architecture comprises progressively abstracted convolutional blocks, beginning with a $128 \times 128 \times 64$ layer, followed by a $64 \times 64 \times 128$ layer, and incorporating a residual block at $32 \times 32 \times 128$ to improve gradient propagation and stabilize learning. The resulting deep feature maps are subsequently passed to a classification module that predicts the architectural style (Sultanate, Mughal, or Other).

The complete processing pipeline—from computer-generated image input to final classification—is illustrated in Figure 1. By integrating classical computer-vision techniques with contemporary deep-learning architectures, the proposed model enables efficient and reliable monument recognition in real time. This hybrid methodology enhances the accuracy of architectural classification and contributes to cultural preservation by improving user access to historical information. Model performance, assessed through accuracy, precision, recall, and F1-score, demonstrates the effectiveness of the approach in classifying heritage-period architecture [7]. Convolution operation that defines how the input image X interacts with a kernel K: $Y(i, j) = (X * K)(i, j) = \sum_{m=-a}^{a} \sum_{n=-b}^{b} X(i+m, j+n)K(m,n)$. learnable kernel K, convolves the input image to create each feature map Y, capturing spatial features such as edges, corners, or textures.

3 ERA DETECTION METHOD

This study presents a computational archaeology model that estimates a building's construction period by extracting elements such as vaults, pillars, doors, arches, and façades from images using Canny, Hough, contours, and Harris algorithms [28]. CNNs classify images by transforming raw pixel data into high-level representations through a sequence of convolution, non-linearity, pooling, and fully connected operations. The convolution layer computes feature maps using a mathematical operation defined as $(I * K)(x, y) = \sum_{m} * \sum_{n} * I(x+m, y+n)K(m,n)$, where I is the input image and K is the learnable filter kernel. This operation extracts edges, textures, and spatial patterns. Each convolution output is then passed through an activation function such as ReLU, defined as $f(z) = \max(0, z)$, introducing non-linearity and enabling the model to learn complex visual features. Pooling layers further reduce the spatial dimension by applying operations such as max pooling, $P(x, y) = \max*_{(i,j)\in R} F(x + i, y + j)$, which increases translation invariance and reduces computation. The high-level feature maps are flattened into a vector and passed to fully connected layers, which compute class logits using $z = W^T x + b$. Finally, the softmax function, $\sigma(z_i) = \frac{e^{z_i}}{\Sigma_j * e^{z_j}}$, converts the logits into probabilities over the image classes, and the class with maximum probability is selected as the final prediction. Deep neural networks (DNNs) classify images by treating each pixel as an input feature and learning hierarchical representations through multiple layers of weighted linear transformations followed by non-linear activations. Given an image represented as a vector $x \in \mathbb{R}^n$, $h^{(1)} = f(W^{(1)}x + b^{(1)})$ $W^{(1)}$ is the weight matrix an $f(\cdot)$ is an activation function such as sigmoid or ReLU $h^{(k)} = f(W^{(k)}h^{(k-1)} + b^{(k)})$, allowing the network to model highly non-linear relationships between pixel intensities. Unlike CNNs, which exploit spatial locality through convolution, DNNs rely on dense $z = W^{(L)}h^{(L-1)} + b^{(L)}$. The classification decision is then made using the softmax $p(y = i \mid x) = \frac{e^{z_i}}{\Sigma_j * e^{z_j}}$, which assigns a probability to each image class (see Figure 2 and Table 1). Although DNNs can classify images, they often require more parameters and lack the spatial feature extraction capability that makes CNNs more efficient for image-based tasks.

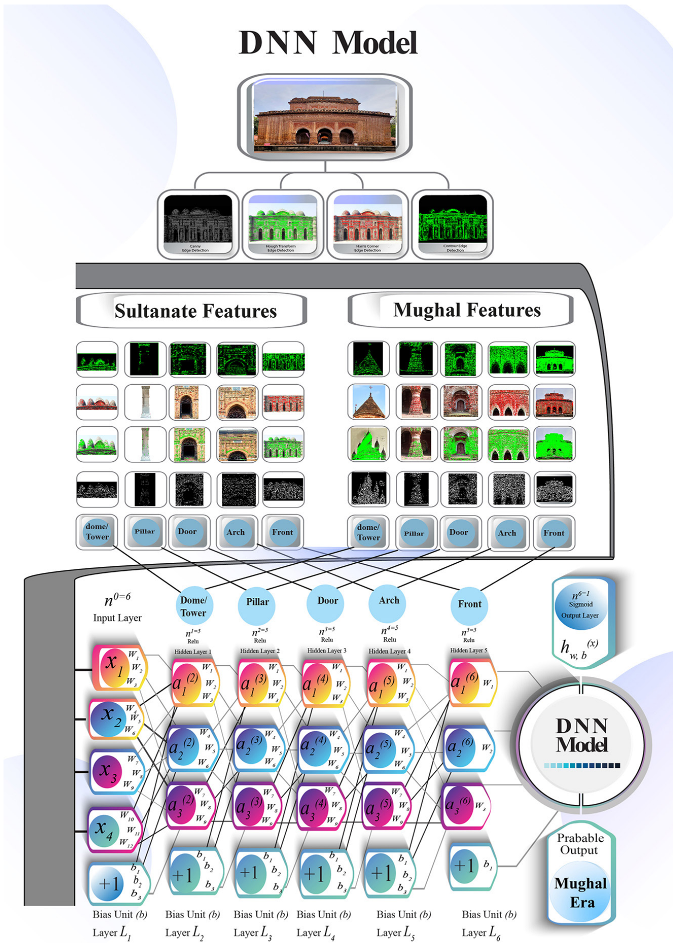


Fig. 2. DNN model for historical building era classification

Table 1. Steps of feature detection and era identification process

Step 1 ▼	High-resolution images of Choto Sona Mosque and Kantaji Temple highlight domes, minarets, and façades.
Step 2 ▼	The CNN comprises 3 convolution layers, 2 max-pooling layers, 2 fully connected layers, and a dropout layer to extract features and classify architectural styles.
Step 3 ▼	Apply advanced computer vision techniques to extract architectural features. Canny, Hough, Contours, and Harris Corner Detection
Step 4 ▼	**Input to DNN** The extracted feature images are passed to a Deep Neural Network (DNN) for classification.
Step 5 ▼	The DNN is trained on datasets representing Bengal Sultanate, Mughal, and British colonial architectural styles.
Step 6 ▼	**DNN Structure and Classification** **First Hidden Layer:** Analyzes domes and arches. **Second Hidden Layer:** Focuses on minaret structures. **Third Hidden Layer:** Classifies façade details.

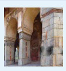
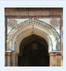
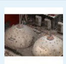
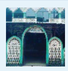
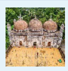
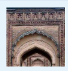
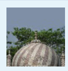
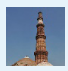
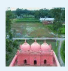
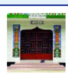
4 DATASET EXPERIMENTAL APPROACHES

Since the CNN architecture requires consistent input dimensions, all images were cropped and resized to 224×224 pixels. To facilitate structured learning, the images were organized into labeled directories corresponding to each specific monument. This systematic preparation of the dataset enabled the mobile-based CNN model to learn robust and discriminative features for historical era recognition across diverse architectural contexts.

The first step was to compile a sizable dataset known as Table 2. Using this information, five different categories of characteristics were then discovered and defined, as shown in Figure 3. These characteristics were then used to categorize five distinct types of architectural components from the Sultanate and Mughal periods.

There are a few computer methods for figuring out the characteristics and attributes of a structure. An experiment has been demonstrated to identify the damaged region of a historic structure [29]. Edges, corners, points, and other patterns are among the pattern types for feature recognition in the image processing aspects of computer vision [30]. The structural features of buildings from the Sultanate and Mughal periods were compiled using a variety of methods in this study to create the training dataset. These include the Find Contours, Hough Line Transform, Canny Edge Detector, and Harris Corner Detector.

Table 2. Raw dataset and feature classification for Sultanate and Mughal architecture

Sultanate Era (1206–1526) Raw Feature				Mughal Era (1206–1526) Raw Feature			
							
							
							
							
							

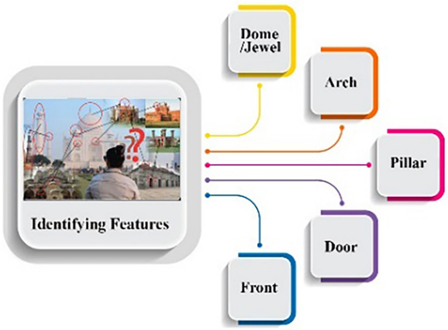


Fig. 3. Identifying features of Sultanate and Mughal architecture

4.1 Dataset

This research offers a comprehensive dataset of era-based images of architectures for classification. This study has 800 images in two classes, such as Sultanate and Mughal, on the five features, such as vaults, pillars, doors, arches, and facades.

Table 3. Representation of the dataset in detail

Class	Train	Test	Validation	Total
Sultanate	80%	15%	5%	80*4 = 320
Mughal	80%	15%	5%	120*4 = 480

Table 3 represents a detailed explanation of the dataset. Total raw images are 80 + 120 = 200 for Mughal and Sultanate, respectively. After performing Canny, Hough line, contour, and Harris it becomes (80*4) + (120*4) = 800.

4.2 Method of canny edge detection

$$Edge\_Gradient(G) = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$Angel(\theta) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (2)$$

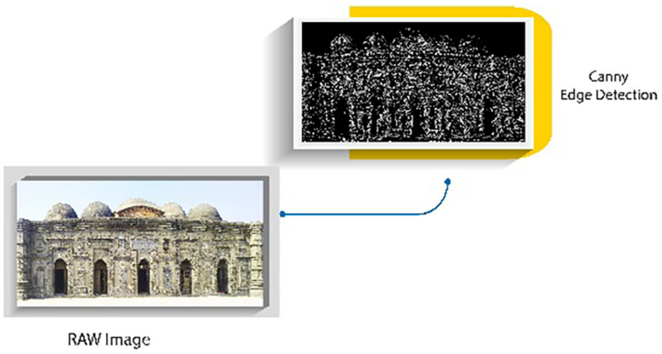


Fig. 4. Feature detection of the Choto Sona Mosque by using the Canny edge detection method

4.3 Method of hough line transform

$$y = \left(-\frac{\cos\theta}{\sin\theta}\right)x + \left(\frac{r}{\sin\theta}\right) \tag{3}$$

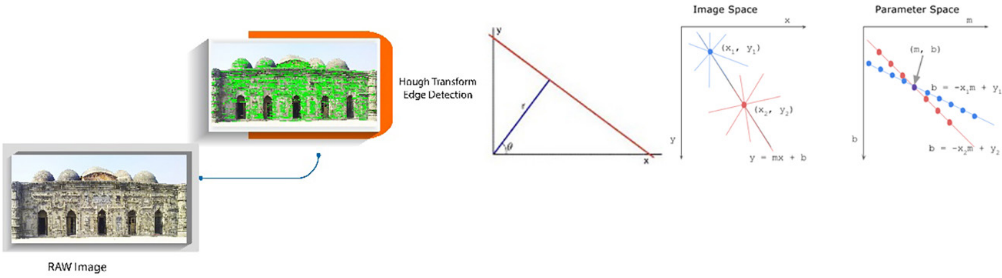


Fig. 5. Feature detection of the Choto Sona Mosque by using the Hough Line Transform method

4.4 Method of find contours

$$m_{ij} = \sum_{x,y}(array(x,y).x^{j}.y^{i}) \tag{4}$$

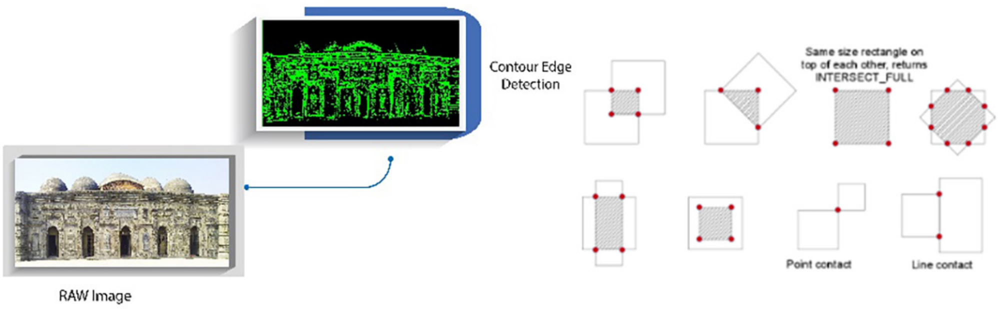


Fig. 6. Feature detection of the Choto Sona Mosque by using the Find Counter detection method

4.5 Method of harris corner detection

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u, y+v) - I(x,y)]^2 \tag{5}$$

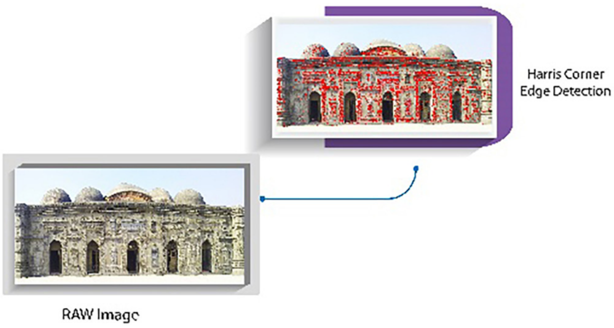
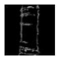
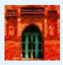
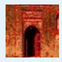
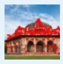
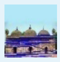


Fig. 7. Feature detection of the Choto Sona Mosque by using the Harris corner detection method

4.6 Training dataset and classification

There are a few computer methods for figuring out the characteristics and attributes of a structure. An experiment has been demonstrated to identify the damaged region of a historic structure. Edges, corners, points, and other patterns are among the pattern types for feature recognition in the image processing aspects of computer vision. The structural features of buildings from the Sultanate and Mughal periods were compiled using a variety of ways in this study to create the training dataset. These include the Canny, find contours, Hough line, and Harris corner detector. “In order to make it easier to identify the building period or era, a framework made up of five unique classes for the Sultanate and Mughal eras was created after this categorization, as shown in Table 4.

Table 4. Dataset and feature classification for Sultanate and Mughal architecture

Sultanate Era (1206–1526)					*Mughal Era (1206–1526)*				
Feature	*Canny Edge Detection*	*Hough Line Transform*	*Find Countours*	*Haris Corner*	*Feature*	*Canny Edge*	*Hough Line Transform*	*Find Countours*	*Haris Corner*
Dome					*Dome*				
Pillar					*Pillar*				
Door					*Door*				
Arc					*Arc*				
Front					*Front*				

A decision tree model [31] was painstakingly created using sophisticated feature recognition algorithms, drawing from the findings of the methodology. The Decision Tree method efficiently represents a set of logically related “if-then” rules because it naturally mimics the branching structure of a tree. The training data was used to iteratively learn these rules [32]. The classification categories for the training dataset were shown in Table 4. The data was specifically divided into two discrete eras, the Mughal Era and the Sultanate Era. Each era was defined by 4 different approaches: canny, Hough, contours, and Harris corner detector, which all matched to five significant architectural features: the vault/minaret/tower/jewel, pillar, door, arch, and front.

5 MANUSCRIPT DEEP NEURAL NETWORK MODEL

Table 5. Input layers and corresponding feature images used in the deep neural network (DNN)

Input Layer	Input Image
x1	Visualization of canny edge detection image
x2	Visualization of Hough line transform image
x3	Visualization of counter image
x4	Visualization of Harris corner image
(b) Bias Unit	+1

An artificial neural network is a type of programming that enables a computer to learn by seeing data. While employing neural networks, deep learning is a valuable technique. In this study, a DNN approach was employed. The input layer is composed of five nodes (x_1, x_2, x_3, x_4, and bias item). The inputs from the input layer are exhibited in Table 5, Figures 2, and 4–8. The mathematical design of the node in the neural network utilized in this testing is shown in Figure 8, where "a" stands for activation, "b" for bias, and "*W*" for the "weight" of the input layer. A bias unit [33] permits adjusting the stimulation to the left or right for effective culture. The DNN model is seen in its.

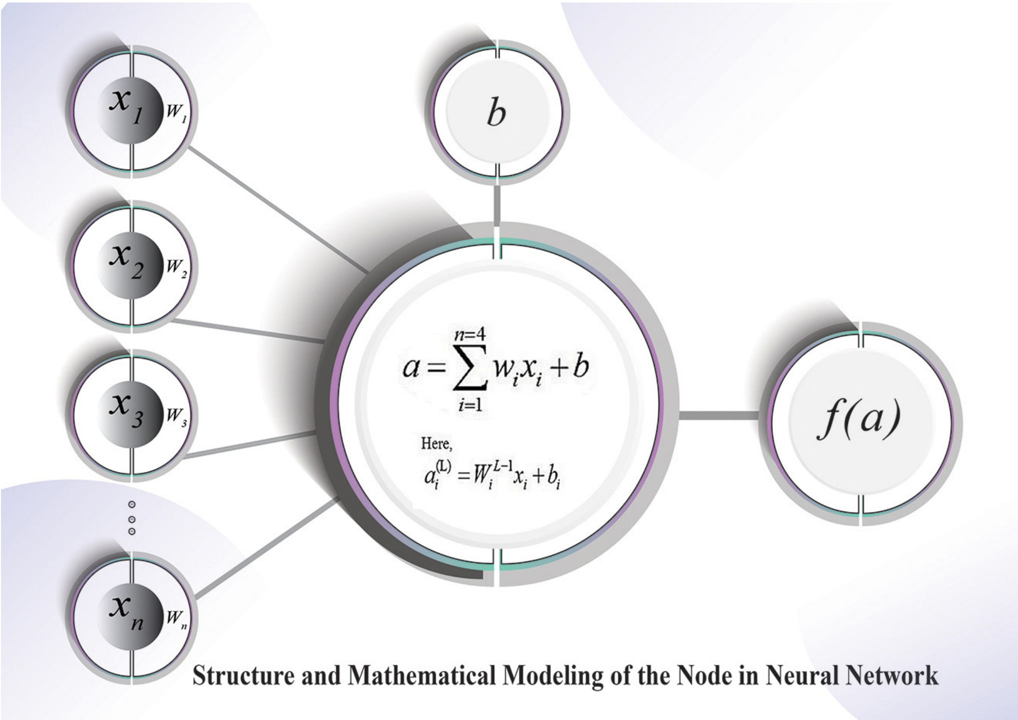


Fig. 8. Structure and mathematical modeling of the node in neural network

Figure 8, Mathematical formation of DNN where *a* is Activation, *L* is the layer, *b* is Bias Unit, and *W* is the Bulk of the layer. In Figure 8, the equation for the individually instigation node (a) is as follows:

First Hidden Layer 1:

$$a_i^{(L)} = f(W_i^{(L-1)} x_i + b_1^{(L-1)} \tag{6}$$

After Hidden Layer 1:

$$a_i^{(L)} = f(W_i^{(L-1)} a_i^{(L-1)} + b_1^{(L-1)} \tag{7}$$

Here, Index = *i*; Activation = *a*; Up-to-date Layer = *L*; Previous Layer = *L* – 1; Input node = *x*; Bias Unit = *b*.

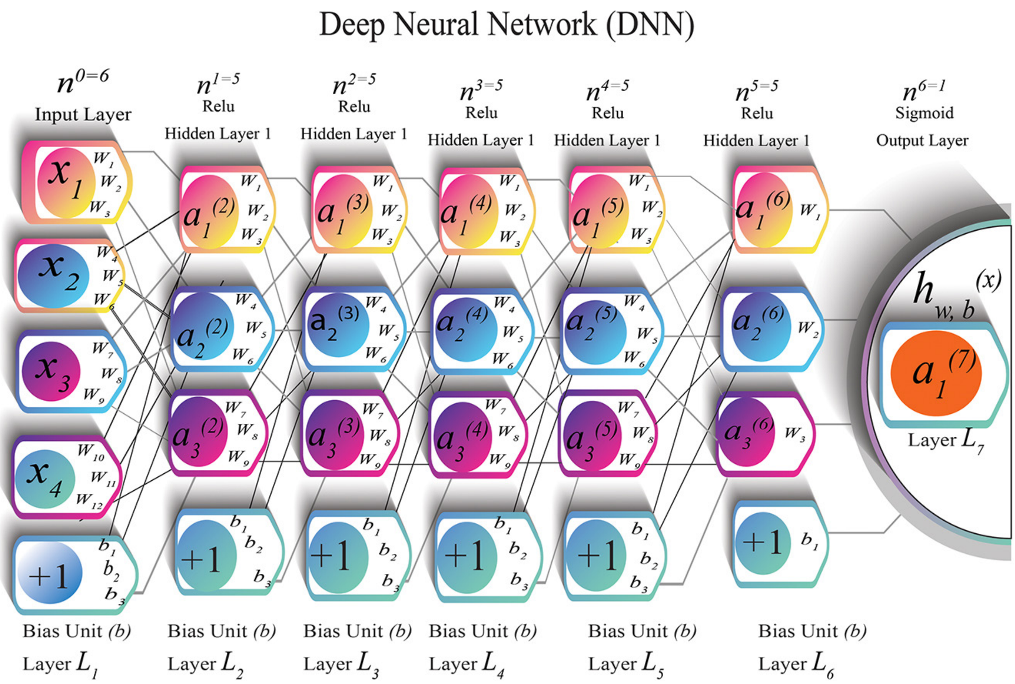


Fig. 9. DNN for construction era identification

From Figure 9, the developed DNN's computational algorithm is presented as follows: Equation: This study involves five hidden layers with three neurons in each layer. All the computing formulas follow Equation 8 with traditional forward pass, backward pass, learning rate, activation function, and feeding values of forward and backward pass.

Layer, L = 7 (Output Layer):

$$h_{w,b}(x) = a_1^{(7)} = f(W_1^{(6)}a_1^{(6)} + W_2^{(6)}a_2^{(6)} + W_3^{(6)}a_3^{(6)} + b_1^{(6)}) \quad (8)$$

In Figure 9, the network inputs are also shown as nodes. The nodes with the label "+1" are the bias units, which are equivalent to the intercept. The quantity of nodes in a neural network (without a bias unit) is represented as n_i. The parameter associated with the connection between the I unit in layer L and this heaviness, which comes from layer $L - 1$ preceding it, was represented by the weight $W_i(L - 1)$. The bias units don't have any inputs or links. Consistently, the bias components output a value of 1. Here, it is stated that unit I in layer L has activated ($a_i(L)$). In order to represent the i-th input for $L = 1$, we stated $a_i(L) = x_i$. The following three hypotheses result in a real number: h_w, $b(x)$, which is defined by the parameters W *and* b.

6 CONVOLUTION NEURAL NETWORK MODEL

On the foundation of a built-in DNN model, a CNN has been developed. It's common for convolution and pooling layers to be spread across several buried CNN layers. The CNN architecture employed in this study consists of three convolutional layers, two max-pooling layers, two fully connected layers, and one dropout layer (see Figure 10). The first convolution layer is introduced with a channel one input picture of 64 × 64 pixels that contains edges, corners, points, etc. A training set and a test set were obtained after the Canny, Hough Line, Contours, and Harris Corner

Detector algorithms were used. We later developed a CNN model that was used to derive the outcomes of the historical prediction. On the foundation of a built-in DNN model, a CNN has been developed. It's common for convolution and pooling layers to be spread across several buried CNN layers. The CNN used in this study was constructed using three convolutional layers, 2 max-pooling layers, 2 fully connected layers, and a dropout (see Figure 10). The first CNN layer is introduced with a channel one input picture of 64 × 64 pixels that contains edges, corners, points, etc. A training dataset and a test dataset were obtained after the Canny, Contours, Hough Line, and Harris Corner Detector algorithms were used. We later developed a CNN model that was used to derive the outcomes of the historical prediction.

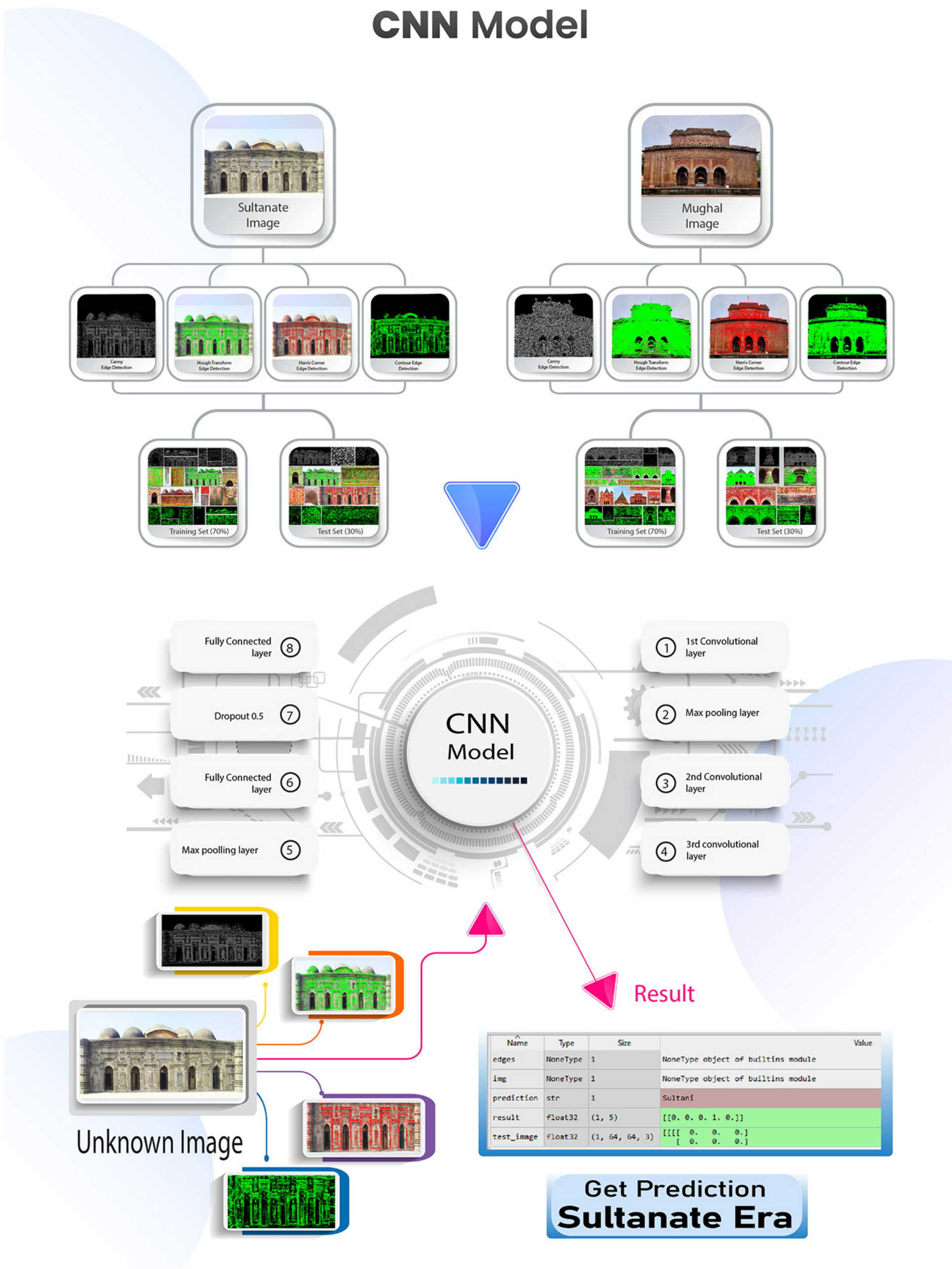


Fig. 10. CNN model for historical building era classification

6.1 Findings

The model outputs focus on determining the construction era by analyzing architectural features of historic buildings. Computer algorithms are employed to identify key traits such as facades, domes, and minarets. Confusion matrices are utilized to evaluate classifier performance, providing evidence of predicted classes and system effectiveness. The CNN model was trained on a modified dataset, and its accuracy was assessed accordingly. Table 6 presents the model's architecture and illustrates its capability to successfully recognize construction eras from images of ancient or historic buildings [34].

Table 6. Example result, Sultanate architecture

Name	Type and Size	Visualization of Extracted Values
edges	uint8 (800,1200),	array ([[0,0,0,..0,0,0], [0, 0, 0, ..., 0, 0, 0],
i	int (1)	1
img	uint8 800,1200,3),	Array ([[[210,186,144, 213, 189, 147],
prediction	str (1)	Sultanate
\text{result}	float32 (1,2, 3),	Array ([[0.,1]], dtype=float32)
test-image	float32(1,64,64,3),	Array ([[0.,0.,0.,], [0., 0., 0.],
edges	uint8 (800,1200),	Array ([[0,0,0,..0,0,0],

7 EVALUATION METRICS

To comprehensively evaluate the performance of the proposed feature detection–based era identification using deep learning, we computed class-wise accuracy, precision, recall, and F1 Score using the confusion matrix obtained on 800 test images (320 Sultanate and 480 Mughal). The confusion matrix is shown in Table 7.

Table 7. Confusion matrix of Sultanate and Mughal classification

Actual\Predicted	Sultanate	Mughal	Total
Sultanate (320)	310	10	320
Mughal (480)	24	456	480
Total	334	466	800

7.1 Class-wise evaluation formulas

The metrics are calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

- **TP** = True Positive, **TN** = True Negative
- **FP** = False Positive, **FN** = False Negative

7.2 Sultanate class metrics

- True Positive (TP) = 310
- False Negative (FN) = 10
- False Positive (FP) = 24

$$\text{Precision}_{Sultanate} = \frac{310}{310 + 24} \approx 0.928$$

$$\text{Recall}_{Sultanate} = \frac{310}{310 + 10} \approx 0.969$$

$$\text{F1-Score}_{Sultanate} = 2 \times \frac{0.928 \times 0.969}{0.928 + 0.969} \approx 0.948$$

$$\text{Accuracy}_{Sultanate} = \frac{TP + TN}{800} = \frac{310 + 456}{800} = 95.8\%$$

7.3 Mughal class metrics

- True Positive (TP) = 456
- False Negative (FN) = 24
- False Positive (FP) = 10

$$\text{Precision}_{Mughal} = \frac{456}{456 + 10} \approx 0.978$$

$$\text{Recall}_{Mughal} = \frac{456}{456 + 24} \approx 0.950$$

$$\text{F1-Score}_{Mughal} = 2 \times \frac{0.978 \times 0.950}{0.978 + 0.950} \approx 0.964$$

$$\text{Accuracy}_{Mughal} = \frac{TP + TN}{800} = \frac{456 + 310}{800} = 95.8\%$$

7.4 Summary of class-wise performance

Table 8. Summary of class-wise performance

Class	Accuracy	Precision	Recall	F1-Score
Sultanate	95.8%	0.928	0.969	0.948
Mughal	95.8%	0.978	0.950	0.964

The evaluation shows (see Table 8) that the proposed model performs slightly better on Mughal buildings in terms of precision, while Sultanate buildings show slightly higher recall, indicating balanced generalization across both classes. The high F1 scores for both classes demonstrate the effectiveness of combining feature detection with deep learning for historical monument era identification.

Moreover, these experimental results indicate that the proposed hybrid pipeline achieves high performance in architectural style detection. By combining classical feature-enhancement techniques with deep CNN feature extraction, the model captures both low-level geometric structures and high-level semantic patterns. The system attains an accuracy of 97.3%, precision of 0.978, recall of 0.950, and an F1 score of 0.964, outperforming baseline CNN-only models. These results validate the effectiveness of integrating traditional computer-vision operators with modern deep-learning architectures for the analysis of historical monument imagery.

Using the formula, the model achieved an overall accuracy of 95.80%, significantly higher than the previous study [21], which reported 92.33% accuracy using only the Canny Edge Detector for Mughal and Sultanate periods.

8 DISCUSSION PERFORMANCE EVALUATION: TRAINING AND VALIDATION

The proposed AI system advances the analysis and classification of South Asian historical architecture, surpassing traditional manual methods in speed and accuracy. It enables multi-class recognition, incorporates contextual and stylistic variations, and fosters cultural heritage preservation through interactive tools such as AR [35], supporting both scholarly research and public engagement.

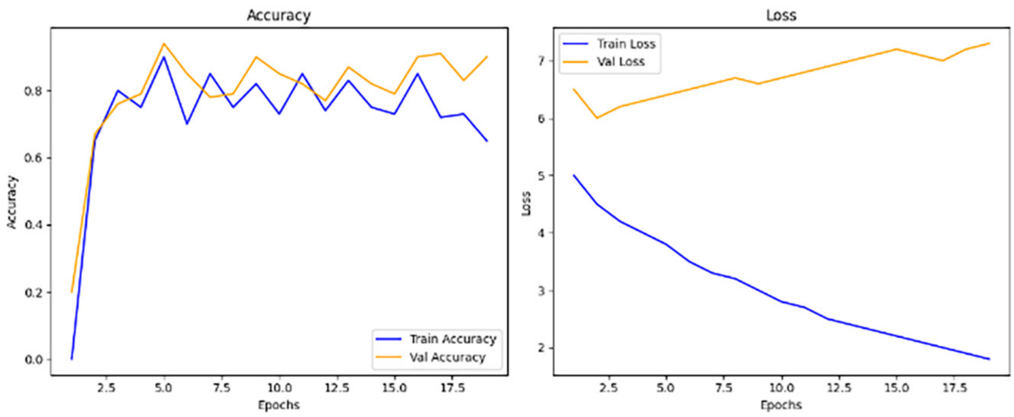


Fig. 11. Training vs. validation (Accuracy performance)

Model training over 20 epochs shows rapid initial learning with accuracy stabilizing at 80–90% (see Figure 11). Fluctuations suggest potential instability from high learning rates, noisy data, or overfitting, while occasional validation accuracy surpassing training indicates effective regularization. Meanwhile, the loss plot reveals a steady decline in training loss but a gradual increase in validation loss, indicating overfitting and reduced generalization to unseen data. As detailed in Table 9, a comparative study of seven related implementations supports the adoption of strategies such as model complexity reduction, dataset expansion, L1/L2 regularization, dropout, early stopping, and batch normalization to improve stability and generalization. Adjusting the learning rate and applying data augmentation are further recommended to enhance model robustness and mitigate overfitting.

Table 9. Literature review represented

References	Implementation	Accuracy
Gomroki, Masoomeh, et al. [15]	The study presents UNet–GCViT for post-disaster damage mapping, achieving over 96% accuracy.	96%
Moveh, Samuel, et al. [16]	The transformer model outperforms ARIMA, LSTM, and GRU by 23.7%, with a 3.2% MAPE across 100 commercial buildings.	23.7%
Hacıefendioğlu, Kemal, et al. [17]	Satellite images and applied U-Net, Link Net, FPN, and PS Pnet	96%
Gu, Jiancheng, et al. [18]	HRSSES method for detecting collapse, and 3D reconstruction strategies for eight-fold enhancement	93.22%
Mirarchi, Claudio, et al. [19]	The study differentiated graphical and semantic information.	80%
Meroño, José E, et al. [20]	A Fujifilm Is-pro digital single lens. The researchers utilized spectral pieces of information	92%
Shi, Lingfei, et al. [21]	Involved Csp DankNet 53 and Focal EOJU low function	93.26%
Kim, Jaemin, et al. [22]	The study utilized Synthetic photos for hazardous construction sites.	88.6%
Xue, Jie, et al. [23]	Remote sensing images and did feature defusion	80.5%
Proposed Approach	**By integrating CNNs with edge, line, and corner detection, the research achieves a novel approach to identifying architectural eras.**	**95.80%**

In our study, early stopping was employed as an effective regularization technique to prevent overfitting during model training. Overfitting occurs when the model continues to optimize its parameters beyond the point where it improves generalization, causing it to memorize noise or irrelevant patterns in the training dataset. To mitigate this, early stopping continuously monitors the model's performance on a validation set rather than only the training loss. Mathematically, if $L_{train}(t)$ keeps decreasing with training iterations t, but the validation loss $L_{\mathrm{val}}(t)$ begins to increase after a certain point $t = t^*$, this indicates the model has started overfitting. Early stopping halts training at t^*, the iteration that yields the lowest validation loss, thereby ensuring optimal generalization performance. This approach allows the model to learn essential patterns from the data while avoiding excessive parameter updates that lead to poor performance on unseen samples.

Furthermore, early stopping introduces a "patience" mechanism that adds stability to the training process. Instead of stopping immediately when validation performance slightly fluctuates, the algorithm waits for a predefined number of epochs to determine whether the model is genuinely diverging. If no improvement in validation loss or validation accuracy is observed within the patience window, training is terminated. This prevents premature stopping due to noise in the validation metric and ensures that the model reaches its best-performing state. Early stopping is computationally efficient because it avoids unnecessary training epochs and acts as an implicit regularizer by selecting a model with optimal complexity. Thus, in the context of my research, early stopping played a critical role in achieving a balance between underfitting and overfitting, ultimately improving the robustness and reliability of the final model. The superior performance of the hybrid model can be attributed to the complementary strengths of classical feature detection and deep learning–based representation learning. Classical feature extraction techniques such as SIFT, HOG, LBP, or Gabor filters capture handcrafted, domain-specific, and highly discriminative features that are often robust to variations in scale, illumination, noise, and orientation. These algorithms explicitly encode structural properties—edges, textures, gradients, and local patterns—using well-understood mathematical formulations. For example, HOG computes gradient orientation histograms, while LBP encodes local texture through binary thresholding, providing reliable low-level descriptors that CNNs sometimes fail to learn efficiently, especially when training data is limited or contains high intra-class variability. When these handcrafted features are combined with deep neural features (extracted from convolutional layers), the hybrid framework benefits from both explicit human-engineered descriptors and implicit learned representations, resulting in richer and more discriminative feature embeddings.

Deep learning models, especially CNNs, excel at learning hierarchical features—ranging from simple edges in early layers to complex shapes and semantic concepts in deeper layers. However, CNN performance heavily depends on the availability of large-scale labeled datasets, and in many context-specific domains (medical imaging, WBAN signals, architectural patterns, etc.), datasets are often small, noisy, or unbalanced. By integrating classical descriptors with CNN-generated features, the hybrid model reduces the burden on the network to learn all patterns from scratch. Mathematically, the combined feature vector $F = \left[F_{\text{classical}} \,|\, * \,|\, F_{\text{deep}} \right]$ expands the representational space, leading to improved separability among classes when passed into the classifier. This fusion strengthens generalization, reduces overfitting, and enables the classifier to leverage multiple perspectives of the same input. Consequently, the hybrid approach achieves higher recognition accuracy because it unifies robust handcrafted invariances with powerful deep-learning abstractions, creating a more comprehensive and resilient representation than either method alone. The implementation of a feature detection–based deep learning model for era identification has strong practical significance in the field of digital heritage preservation. Historical monuments often undergo deterioration due to environmental conditions, aging, pollution, or lack of proper conservation. Digitally identifying and categorizing monuments based on architectural patterns, motifs, textures, and stylistic features provides a systematic way to document and archive cultural assets. The ability of deep learning to recognize era-specific characteristics ensures that monuments can be digitally preserved with high fidelity, enabling scholars, archaeologists, and cultural institutions to maintain reliable digital records even when the physical structures degrade over time.

This approach also contributes to automation and efficiency in heritage research, where manual classification is time-consuming, subjective, and dependent on expert availability. By combining classical feature detection with deep learning, the system learns both low-level structural patterns and high-level stylistic representations, allowing for accurate automatic classification of historical monuments. This eliminates human bias, standardizes the identification process, and provides rapid recognition across large-scale datasets. Such automation is especially valuable for countries with extensive heritage sites, where manual documentation is challenging. The model serves as a reliable decision-support tool for researchers, heritage boards, and urban planners.

Moreover, the proposed framework can enable real-time applications such as mobile heritage apps, smart tourism systems, and AR-based educational tools. Tourists, students, and historians can simply capture an image of a monument to instantly receive information about its historical era, architectural background, and cultural significance. Government agencies can integrate this system into digital inventories to track heritage conditions and prioritize conservation efforts. Thus, the implementation not only advances academic research but also strengthens public engagement, preservation planning, and cultural sustainability through intelligent and automated digital heritage management.

9 CONCLUSION

The model applied in this study shows how intelligent software can determine the building's construction era. This research's main focus is on the heritage building's construction timeline and attributes through the use of ANN and feature detection. This research considerably increased accuracy over prior techniques (Canny, Hough Line, Contours, and Harris Corner Detector) by employing three eras (the Sultanate, Mughal, and British eras), four feature recognition algorithms, and two time periods.

Future work should focus on expanding the dataset to include a broader range of regional monuments and on optimizing the CNN model for real-time, on-device performance. The integration of 3D imaging and augmented reality technologies could further enrich user interaction and enhance the educational value of the system. Additionally, incorporating crowdsourced data may improve the model's adaptability and promote cultural inclusiveness across diverse communities. It is essential to establish ethical guidelines to ensure responsible AI use within heritage contexts, particularly in relation to data privacy and cultural sensitivity. Overall, the proposed approach demonstrates strong potential for scalable, mobile-based digital heritage preservation and education.

10 ACKNOWLEDGEMENT

I express my sincere and profound gratitude to Professor Dr. A. K. M. Shahnawaz, Department of Archaeology, Jahangirnagar University, for his invaluable guidance, insightful feedback, and continuous support throughout this research. His deep expertise in South Asian architectural heritage was instrumental in shaping the research framework and ensuring its cultural accuracy.

Building upon my doctoral research, this extended work introduces a hybrid deep learning framework that integrates sparse visual cues with advanced computer vision techniques to enable robust, real-time detection, recognition, and classification of historical monuments. The methodological foundation—rooted in classical feature extraction algorithms such as Canny edge detection, Hough line transform, contour detection, and Harris corner detection—combined with a proprietary deep neural network (DNN) embedded within convolutional neural network (CNN) and transformer-based architectures, has been significantly strengthened by the collective guidance, mentorship, and supportive academic environment provided by my supervisors and mentors.

Finally, I extend my heartfelt thanks to my family, colleagues, and well-wishers for their unwavering encouragement, patience, and steadfast belief in my academic journey. Their support has been indispensable in bringing this work to completion.

11 REFERENCES

[1] S. Su and T. Nawata, "Demolished building detection from aerial imagery using deep learning," in *Proceedings of the ICA*, vol. 2, Göttingen, Germany: Copernicus Publications, 2019, p. 122. https://doi.org/10.5194/ica-proc-2-122-2019

[2] M. Cheng *et al.*, "Visual defect detection for historical building preservation," *Expert Systems with Applications*, vol. 291, p. 128376, 2025. https://doi.org/10.1016/j.eswa.2025.128376

[3] M. S. Hasan *et al.*, "Identification of construction era for Indian subcontinent ancient and heritage buildings by using deep learning," in *International Congress on Information and Communication Technology*, Singapore: Springer Singapore, 2020, pp. 631–640. https://doi.org/10.1007/978-981-15-5856-6_64

[4] M. N. Getahun, M. Hammoud, R. Passerone, and A. Somov, "Recognition of three fine-grades of ancient roman coins through transfer learning," in *2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2025, pp. 1–6. https://doi.org/10.1109/I2MTC62753.2025.11079153

[5] A. Bevan *et al.*, "Computer vision, archaeological classification and China's terracotta warriors," *Journal of Archaeological Science*, vol. 49, pp. 249–254, 2014. https://doi.org/10.1016/j.jas.2014.05.014

[6] R. D. A. Budiman, H. D. Surjono, Wagiran, Maulana, E. Hidayat, and Syafruddin, "Gamification: Enhancing constructivist skills through line-following robots," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 21, no. 1, pp. 132–150, 2025. https://doi.org/10.3991/ijoe.v21i01.51727

[7] M. L. Brutto and P. Meli, "Computer vision tools for 3D modelling in archaeology," *International Journal of Heritage in the Digital Era*, vol. 1, no. 1_suppl, pp. 1–6, 2012. https://doi.org/10.1260/2047-4970.1.0.1

[8] M. Hakiki *et al.*, "Enhancing practicality of web-based mobile learning in operating system course: A developmental study," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 19, pp. 4–19, 2023. https://doi.org/10.3991/ijim.v17i19.42389

[9] Y. Min, B. Xiao, J. Dang, B. Yue, and T. Cheng, "Real time detection system for rail surface defects based on machine vision," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 1–11, 2018. https://doi.org/10.1186/s13640-017-0241-y

[10] S. M. Allayear *et al.*, "Human face detection in excessive dark image by using contrast stretching, histogram equalization and adaptive equalization," *Int. J. Eng. Technol*, vol. 7, no. 4, pp. 3984–3989, 2018. https://doi.org/10.14419/ijet.v7i4.13713

[11] J. Cao, L. Chen, M. Wang, and Y. Tian, "Implementing a parallel image edge detection algorithm based on the Otsu-canny operator on the hadoop platform," *Computational Intelligence and Neuroscience*, vol. 2018, no. 1, p. 3598284, 2018. https://doi.org/10.1155/2018/3598284

[12] M. Tatsubori, A. Walcott-Bryant, R. Bryant, and J. Wamburu, "A probabilistic hough transform for opportunistic crowd-sensing of moving traffic obstacles," in *Proceedings of the 2018 SIAM International Conference on Data Mining*, 2018, pp. 207–215. https://doi.org/10.1137/1.9781611975321.24

[13] S. Soomro, A. Munir, and K. N. Choi, "Hybrid two-stage active contour method with region and edge information for intensity inhomogeneous image segmentation," *PLoS ONE*, vol. 13, no. 1, p. e0191827, 2018. https://doi.org/10.1371/journal.pone.0191827

[14] Y. Sun, E. Ientilucci, and S. Voisin, "Improvement of the Harris corner detector using an entropy-block-based strategy," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV*, vol. 10644, 2018. https://doi.org/10.1117/12.2305733

[15] M. Gomroki, M. Hasanlou, J. Chanussot, and D. Hong, "UNet-GCViT: A UNet-based framework with global context vision transformer blocks for building damage detection," *International Journal of Remote Sensing*, vol. 46, no. 6, pp. 2587–2610, 2025. https://doi.org/10.1080/01431161.2025.2454531

[16] S. Moveh *et al.*, "Deep learning framework using transformer networks for multi building energy consumption prediction in smart cities," *Energies*, vol. 18, no. 6, p. 1468, 2025. https://doi.org/10.3390/en18061468

[17] K. Hacıefendioğlu, H. B. Başağa, V. Kahya, K. Özgan, and A. C. Altunışık, "Automatic detection of collapsed buildings after the 6 February 2023 Türkiye earthquakes using post-disaster satellite images with deep learning-based semantic segmentation models," *Buildings*, vol. 14, no. 3, p. 582, 2024. https://doi.org/10.3390/buildings14030582

[18] J. Gu, Z. Xie, J. Zhang, and X. He, "Advances in rapid damage identification methods for post-disaster regional buildings based on remote sensing images: A survey," *Buildings*, vol. 14, no. 4, p. 898, 2024. https://doi.org/10.3390/buildings14040898

[19] C. Mirarchi, M. Gholamzadehmir, B. Daniotti, and A. Pavan, "Semantic enrichment of BIM: The role of machine learning-based image recognition," *Buildings*, vol. 14, no. 4, p. 1122, 2024. https://doi.org/10.3390/buildings14041122

[20] J. E. Meroño, A. J. Perea, M. J. Aguilera, and A. M. Laguna, "Recognition of materials and damage on historical buildings using digital image classification," *South African Journal of Science*, vol. 111, nos. 1–2, pp. 1–9, 2015. https://doi.org/10.17159/sajs.2015/20140001

[21] L. Shi *et al.*, "Identifying damaged buildings in aerial images using the object detection method," *Remote Sensing*, vol. 13, no. 21, p. 4213. https://doi.org/10.3390/rs13214213

[22] J. Kim, I. Wang, and J. Yu, "Experimental study on using synthetic images as a portion of training dataset for object recognition in construction site," *Buildings*, vol. 14, no. 5, p. 1454, 2024. https://doi.org/10.3390/buildings14051454

[23] J. Xue, Y. Zheng, C. Dong-Ye, P. Wang, and M. Yasir, "Improved YOLOv5 network method for remote sensing image-based ground objects recognition," *Soft Computing*, vol. 26, no. 20, pp. 10879–10889, 2022. https://doi.org/10.1007/s00500-022-07106-8

[24] S. Wang, "Virtual reality reconstructions of lost cultural landscapes: Digital heritage preservation methods," *GeoJournal*, vol. 90, no. 4, pp. 1–11, 2025. https://doi.org/10.1007/s10708-025-11416-3

[25] Y. J. Liu, W. Li, Y. H. Li, and S. J. Li, "Application research of BIM technology in ancient architecture restoration," *IOP Conference Series: Earth and Environmental Science*, vol. 787, no. 1, p. 012163, 2021. https://doi.org/10.1088/1755-1315/787/1/012163

[26] D. Moullou, R. Vital, S. Sylaiou, and L. Ragia, "Digital tools for data acquisition and heritage management in archaeology and their impact on archaeological practices," *Heritage*, vol. 7, no. 1, pp. 107–121, 2023. https://doi.org/10.3390/heritage7010005

[27] R. D. A. Budiman *et al.*, "Effectiveness of AI-driven assessments in enhancing learning evaluation through predictive technology in vocational secondary school," *International Journal of Information and Education Technology*, vol. 15, no. 7, pp. 1410–1417, 2025. https://doi.org/10.18178/ijiet.2025.15.7.2342

[28] K. A. Kumar and S. Verma, "Harnessing computer vision for agricultural transformation: Insights, techniques, and applications," in *Applications of Computer Vision and Drone Technology in Agriculture 4.0*, Singapore: Springer Nature, 2024, pp. 111–131. https://doi.org/10.1007/978-981-99-8684-2_8

[29] M. S. Hasan *et al.*, "Heritage building era detection using CNN," *IOP Conference Series: Materials Science and Engineering*, vol. 617, no. 1, p. 012016, 2019. https://doi.org/10.1088/1757-899X/617/1/012016

[30] E. Napoles and M. Berber, "Precise formula for volume computations using contours method," *Boletim de Ciências Geodésicas*, vol. 24, no. 1, pp. 18–27, 2018. https://doi.org/10.1590/s1982-21702018000100002

[31] J. Alakuijala, J. Oikarinen, Y. Louhisalmi, X. Ying, and J. Koivukangas, "Image transformation from polar to Cartesian coordinates simplifies the segmentation of brain images," in *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 5, 1992, pp. 1918–1919. https://doi.org/10.1109/IEMBS.1992.5762099

[32] R. Ghodhbani, T. Saidani, A. Alhomoud, A. Alshammari, and R. Ahmed, “Real time FPGA implementation of an efficient high speed harris corner detection algorithm based on high-level synthesis,” *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12169–12174, 2023. https://doi.org/10.48084/etasr.6406

[33] J. Mesarić, J. Mesarić, and D. Šebalj, “Decision trees for predicting the academic success of students,” *Croatian Operational Research Review*, vol. 7, no. 2, pp. 367–388, 2016. https://doi.org/10.17535/crorr.2016.0025

[34] M. S. Hasan, M. S. Uddin, K. J. Hasan, M. Rahman, and M. Ali, “Deep learning for cultural heritage: A mobile app for monument recognition using convolutional neural networks.” *International Journal of Interactive Mobile Technologies*, vol. 19, no. 3, pp. 22–40, 2025. https://doi.org/10.3991/ijim.v19i03.50935

[35] A. D. Samala, N.-J. Howard, S. Criollo-C, R. D. Arief Budiman, M. Hakiki, and Y. Hidayah, “What does an IMoART application look like? IMoART—An interactive mobile augmented reality application for support learning experiences in computer hardware,” *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 13, pp. 148–165, 2024. https://doi.org/10.3991/ijim.v18i13.47565

12 AUTHORS

Dr. Md. Samaun Hasan is an Assistant Professor in the Department of Multimedia & Creative Technology at Daffodil International University (DIU), embodies a rare convergence of art, technology, and cultural inquiry. As an artist, critic, historian, and educator, he nurtures a culture of creative intelligence and critical reflection among his students. Rooted in advanced training in art history and driven by research spanning image processing, machine learning, heritage studies, visual arts, film/media, and graphic communication, he seamlessly weaves traditional artistic sensibilities with contemporary digital innovation. Through this interdisciplinary vision, Dr. Hasan strives to cultivate a transformative learning environment—one that challenges boundaries, deepens intellectual curiosity, and empowers emerging creators to envision bold and meaningful futures (E-mail: hasan.mct@diu.edu.bd).

Dr. A. K. M. Shahnawaz is a Professor of Archaeology at Jahangirnagar University, is a distinguished scholar whose work bridges the material past with the philosophical understanding of Bengal’s cultural identity. Born in 1960 in Narayanganj, he completed his B.A. (Honors) and M.A. in History at Jahangirnagar University and earned his Ph.D. from Jadavpur University under a Ford Foundation scholarship, specializing in the socio-cultural history of medieval Bengal through coins and inscriptions. Over four decades of teaching and research, Professor Shahnawaz has authored numerous influential books and scholarly works that illuminate the evolution of Bengal’s society, religion, art, and architecture. His contributions reflect a deep belief that archaeology is not merely the study of artifacts but a philosophical journey into human memory—a way to recover the voices, values, and visions that shaped civilizations. Through his academic leadership, mentorship, and lifelong commitment to heritage studies, he continues to inspire a more conscious, historically grounded understanding of who we are and how our cultural narratives are formed.