The Eight Great Stupa Image Classification Using Machine Learning

Pasang Tenzin[1], Sonam Wangmo[2*], and Tashi Wangchuk[3]

[1-3] *Department of Information Technology, Jigme Namgyel Engineering College, Dewathang*

[*] *Corresponding author: sonamewangmo.jnec@rub.edu.bt*

Published: December 2025

Abstract

In this project, a machine learning based system was developed for classifying the stupa images into eight distinct categories namely, Descent from God Realm, Enlightenment Stupa, Heart Lotus, Miraculous Display, Nirvana Stupa, Reconciliation Stupa, Turning of the Wheel and Victory Stupa. For feature extraction combination of Local Binary Pattern (LBP), and Color Histogram (HSV) was applied. Principal Component Analysis (PCA) was employed for dimensionality reduction to improve computational efficiency and model performance. Three machine learning classifiersSupport Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF)were trained and evaluated. The SVM classifier, using the RBF kernel, achieved the highest performance with a cross-validation score of 90.38% and a test accuracy of 94%, both KNN and Random Forest models also demonstrated strong performance by achieving 92% accuracy. Per-class F1-score analysis showed consistently high classification performance across most classes, with SVM slightly outperforming the others. The experiment results demonstrate that traditional machine learning methods combined with effective feature extraction are highly suitable for stupa image classification.

Keywords— Stupa Image Classification, Machine Learning, Local Binary Pattern (LBP), Support Vector Machine (SVM), Bhutanese Stupas

1 Introduction

This project introduces a web-based application for classifying Bhutanese stupas, specifically focusing on the eight types representing key moments in Buddha's life (Desheg Chorten Gye). Currently, Bhutan lacks a systematic stupa classification system, leading to frequent misidentification due to reliance on often insufficient human knowledge and a general lack of public awareness regarding their distinct categories [1]. This absence of accurate information makes it challenging for individuals to find reliable data about these culturally and religiously significant structures.

2 Literature Review

Stupas (or chortens) are globally significant religious monuments, especially in Bhutan's Vajrayna tradition, where their unique designs symbolize specific spiritual concepts [2]. The tradition of building these monuments, which originally housed relics of enlightened beings, began during Buddha's time and evolved to include eight distinct types commemorating major milestones in his life, collectively known as Desheg Chorten Gye [1]. Image classification, a core computer vision task, labels images based on their content and underpins advanced applications like object detection. Historically, this relied on manually extracted features and traditional algorithms (SVM, k-NN, RF), which struggled with complex data. The field was revolutionized by deep learning, especially Convolutional Neural Networks (CNNs) like VGGNet, which automatically extract features and improved performance significantly. More recently, transformer models, adapted from natural language processing (e.g., Vision Transformer or ViT), have also proven highly effective in image classification by using self-attention mechanisms [2].

Image classification uses various machine learning methods, with traditional techniques like SVM and KNN remaining valuable despite deep learning's rise. Traditional methods offer advantages like lower computational needs, effectiveness with limited data, and easier accessibility. Deep learning, an evolution of traditional ML, uses neural networks (which excel in classification) to mimic brain-like learning; other ML methods include SVM, Naïve Bayes, Decision Trees, and KNN. For the current document, Random Forest, SVM, and k-NN were specifically employed as classification algorithms [3].

3 Methodology

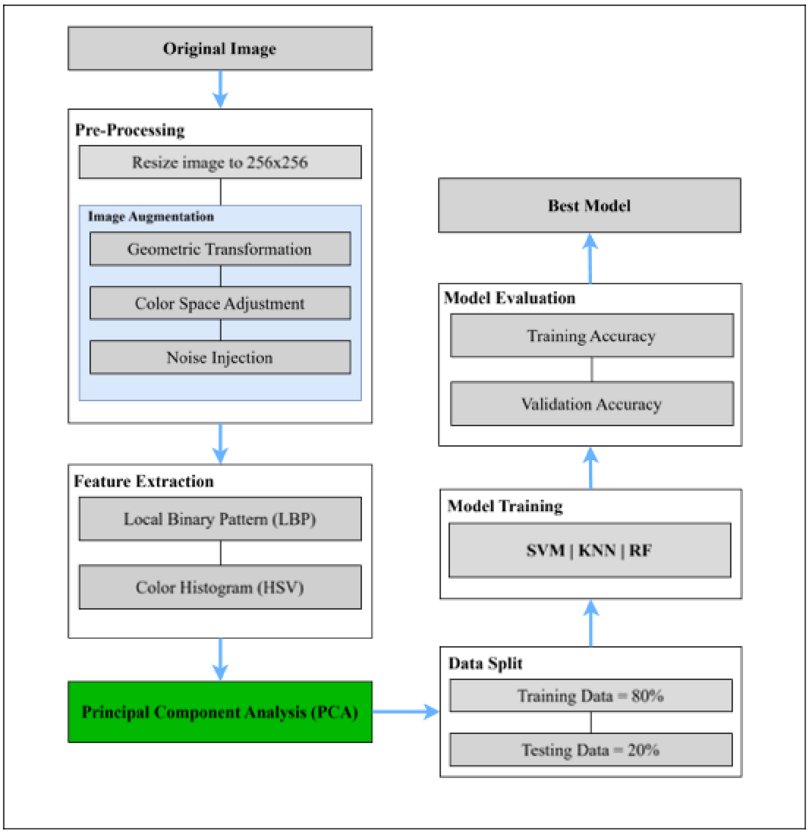

Figure 1: Methodology Flow diagram

3.1 Support Vector Machine (SVM)

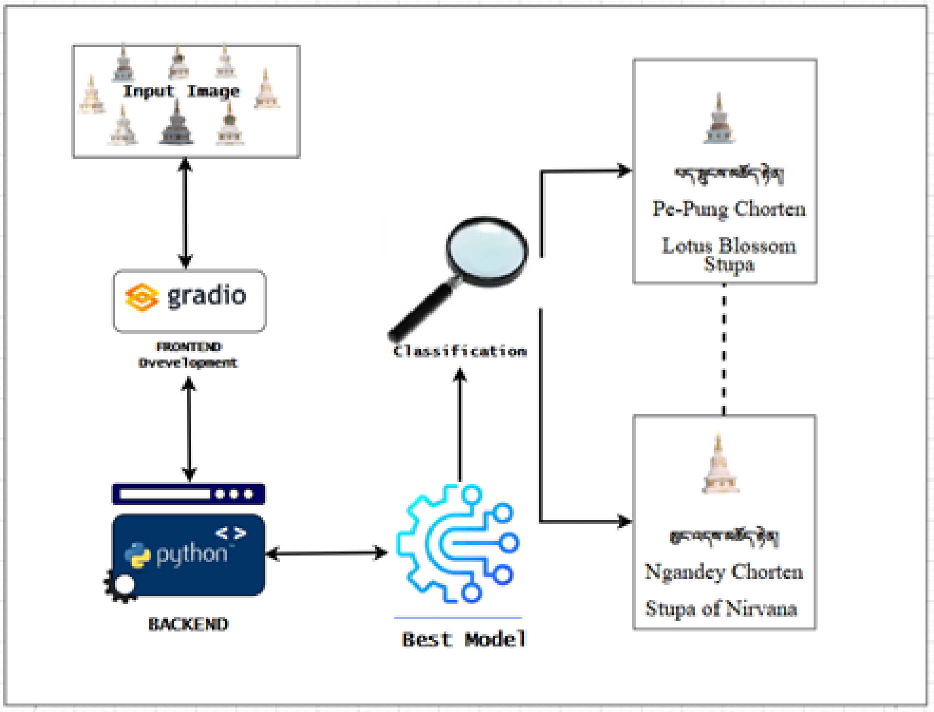

Figure 2: Support Vector Machine margin maximization for binary classification.

Support Vector Machines (SVMs) are a type of supervised machine learning algorithm that is used for both classification and regression problems. This is especially useful for binary classification tasks, such as distinguishing spam from non-spam emails or identifying whether an image shows a cat or a dog. The primary objective of an SVM is to maximize the margin between the two classes. A wider margin generally leads to better performance when the model is applied to new, unseen data (Support Vector Machine (SVM) [4].

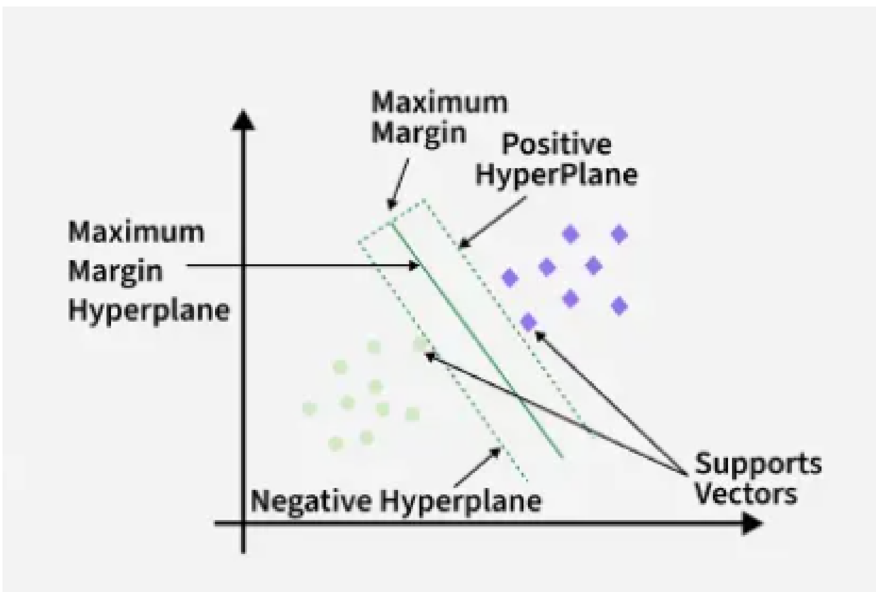

Figure 3: SVM Diagram

Key Concepts:

1. *Hyperplane*: A line or a plane that separates the classes. A decision boundary separating different classes in feature space.
2. *Support Vectors*: The closest data points to the hyperplane, crucial for determining the hyperplane and margin in SVM.
3. *Margin*: The distance between the hyperplane and the support vectors. The distance between the hyperplane and the support vectors.

3.1.1 Mathematical Computation of SVM

Lets consider a binary classification problem with two classes, labeled +1 and -1. We have a training dataset consisting of input feature vectors X and their corresponding class labels Y. The equation of a linear hyperplane can be expressed as:

$$\mathbf{w}^T\mathbf{x} + b = 0 \tag{1}$$

where $\mathbf{w}^T$ is the normal vector to the hyperplane, indicating the direction perpendicular to it.

b is the bias or offset term, representing how far the hyperplane is from the origin along the direction of **w**.

3.2 Random Forest (RF)

Random Forest is a machine learning algorithm that builds multiple decision trees to improve prediction accuracy. Each tree is trained on a random subset of the data, and their individual results are combinedby majority vote for classification tasks or by averaging for regression tasks. This approach helps increase accuracy and reduce the likelihood of errors [5].

3.2.1 How does the Random Forest Algorithm works?

- Build Multiple Decision Trees.
- Select Random Features.
- Individual Prediction.
- Combine the Results.

3.3 K-Nearest Neighbor (KNN)

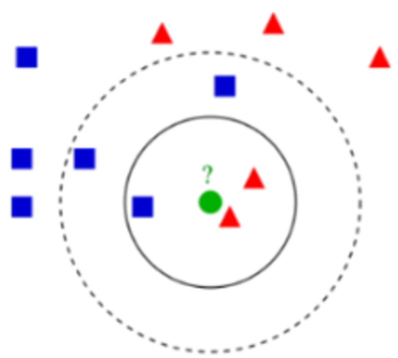

Figure 4: KNN Algorithm

The K-nearest neighbors (KNN) algorithm is a straightforward supervised machine learning technique that makes predictions by measuring how close a data point is to others. It is commonly used for both classification and regression tasks due to its simplicity and ease of use. The algorithm works by finding the K closest data points (neighbors) to the input based on distance [6].

For classification, it assigns the class that appears most frequently among the K neighbors to the input point. For regression, it predicts the value by taking the average or a weighted average of the target values of the K neighbors [6].

3.4 Feature Extraction

In LBP images, sharp stupa edges show high texture from varied local patterns, while smooth areas like ground and sky have fewer variations; boundaries with trees or sky also show significant LBP changes. HOG captures object appearance via gradient orientation distribution, while HSV color histograms (hue, saturation, value) describe pixel color distribution. For example, a red line's peak at bins 10-20 suggests dominant gold/red hues on the stupa, while saturation indicates color intensity and the blue line's value shows brightness [7].

3.4.1 Local Binary Pattern (LBP)

Local Binary Patterns (LBP) is a visual descriptor commonly used in computer vision for classification tasks. It represents a specific example [8].

3.4.2 Mathematical Definition of Basic LBP

The LBP value of the pixel is computed as:

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)\, 2^p \tag{2}$$

Where,

- P: number of neighbors (usually 8)
- R: radius from the center pixel (usually 1)
- g_p: gray-level value of neighbor p
- g_c: gray-level value of the center pixel
- $s(g_p - g_c)$: thresholded binary result (0 or 1)

3.4.3 Augmentation

Augmentation refers to artificially increasing the diversity of a training dataset by applying transformations to existing images. This helps improve model robustness, generalization, and prevents overfitting, especially when the original dataset is limited [9].

In this image given in figure 5, it includes custom parameters that we have used during augmentation.

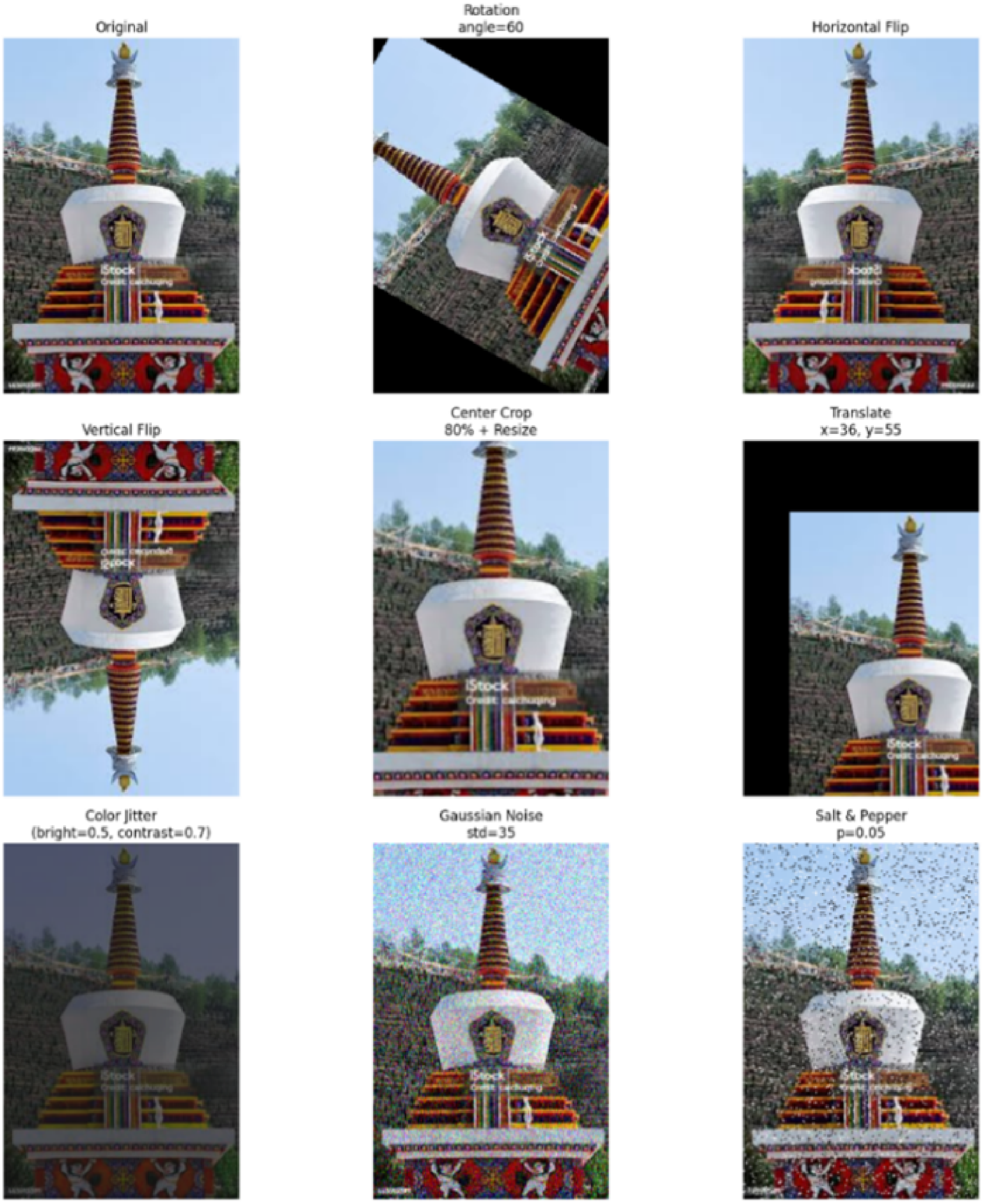

Figure 5: Showing Augmented Images

3.5 Model Evaluation

Table 1: Model Evaluation showing the best accuracy

Model	Best CV Score	Test Accuracy	Precision	Recall	F1-score
SVM	0.9038	0.94	0.94	0.94	0.94
KNN	0.8755	0.92	0.92	0.92	0.92
Random Forest	0.8788	0.92	0.92	0.92	0.92

Table 2: Showing FI score

Model	Best Accuracy (Validation)	Best Parameters Included
SVM	94%	rbf kernel, $C = 100$, PCA $= 0.9$
KNN	92%	$k = 3$, distance weighting, PCA $= 0.9$
Random Forest	92%	300 trees, max depth $= 20$, PCA $= 0.95$

4 Results

4.1 Image Classification Process

After training the model and rating them we have found SVM performs the best from KNN and Random Forest.

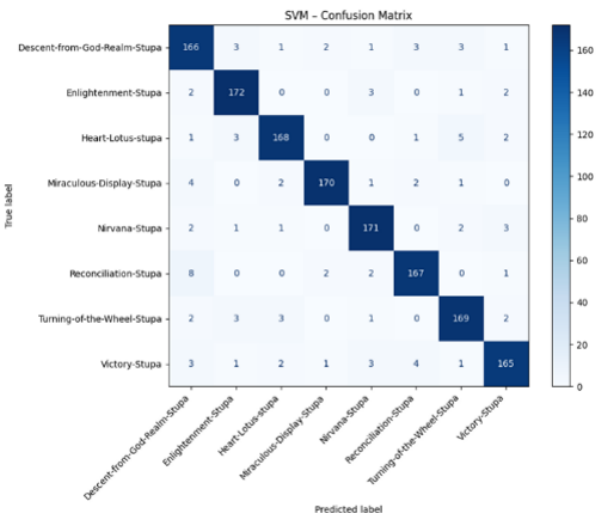

Figure 6: SVM confusion matrix

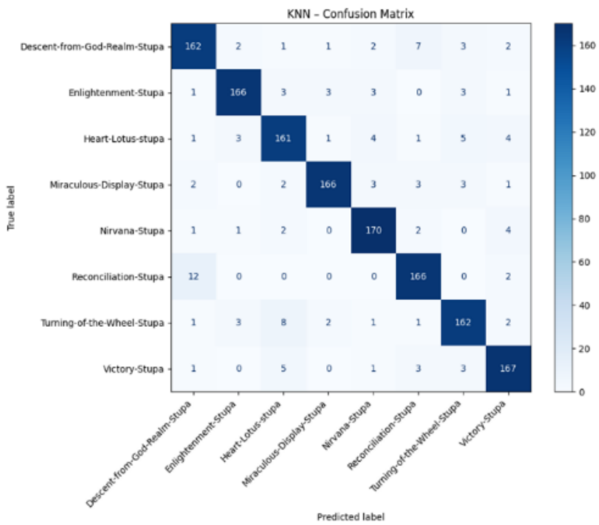

Figure 7: KNN confusion matrix

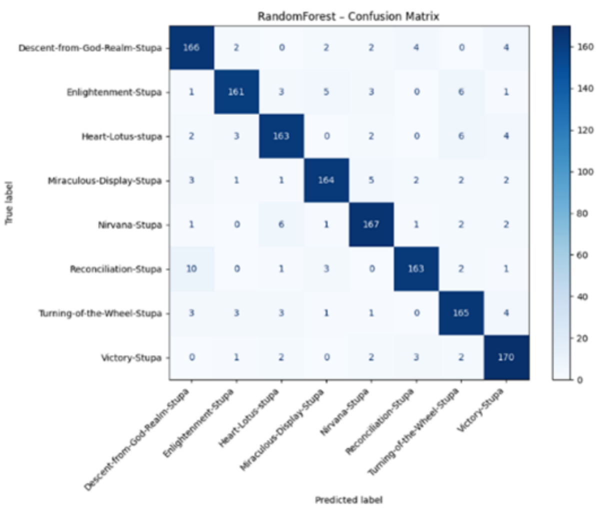

Figure 8: Random Forest confusion matrix

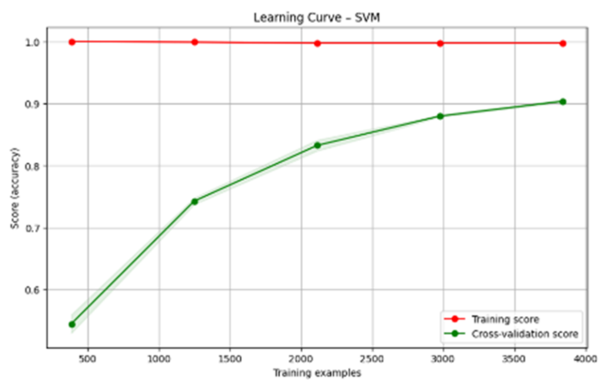

Figure 9: SVM confusion matrix

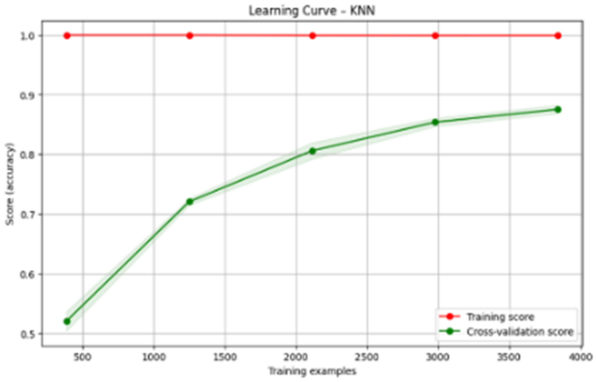

Figure 10: KNN learning curve

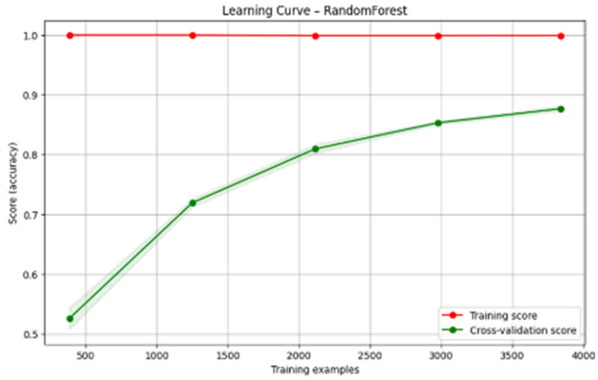

Figure 11: Random Forest learning curve

5 Conclusion

Initially, our image classification project used HOG, LBP, and HSV features with SVM, RF, and KNN models, but yielded low accuracy and long computation times due to limited resources [10]. Data collection and ambiguous image scenes (e.g., white stupa on white background) were significant challenges. We found that for traditional monuments like stupas, a combination of LBP and HSV features worked better than including HOG. Ultimately, traditional machine learning proved effective for our smaller dataset [11].

6 Acknowledgment

The attainment of any accomplishment, whether academic or otherwise, is not solely reliant on one's personal endeavors but also on the guidance, support, and collaboration of intellectuals, mentors, and friends.

First and foremost, we would like to extend our sincere appreciation to Jigme Namgyel Engineering College, the Royal University of Bhutan, for including the Project Work (CPR201) as a module because it has helped me learn a variety of life skills, such as public speaking, presentation skills, the importance of teamwork, and the necessity for researching various research papers. The module has helped me develop professionally and more competently.

My heartfelt gratitude goes to my friend Ms. Rinchen Zangmo, Ms. Chimi Lhamo Lama, Mr. Sonam Wangyel Tshewang, Mr. Karma Yonten, Mr. Tshering Dendup and Ms. Denchen Choden for helping me in data collection.

we owe a great deal to our two mentors, Ms. Sonam Wangmo and Mr. Tashi Wangchuk for always guiding me and inspiring me to carry out the work more effectively. we are deeply grateful for their continuous guidance and steadfast encouragement throughout the entire duration of my project.

References

[1] D. Phuntsho, *Buddhist stupas and their symbols in Vajrayana Buddhism in Bhutan*, International Journal of Science and Innovative Research (IJESIR), 2021.

[2] A. Dosovitskiy et al., *An image is worth 16Œ16 words: Transformers for image recognition at scale*, arXiv preprint arXiv:2010.11929, Oct. 2020.

[3] B. Pearl, *Image classification without neural networks*, ResearchGate, Jul. 6, 2024.

[4] Y. R., *Support vector machine (SVM) algorithm*, GeeksforGeeks, May 28, 2025.

[5] J. S. M. Sillanpaa, *Random forest algorithm in machine learning*, GeeksforGeeks, May 30, 2025.

[6] T. Srivastava, *Guide to K-nearest neighbors algorithm in machine learning*, Analytics Vidhya, May 1, 2025.

[7] D. Huang, *Local binary patterns and its application to texture classification*, French National Center for Scientific Research (CNRS), 2024.

[8] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, World Journal of Engineering and Technology, 2015.

[9] M. Akbiyak, *Data augmentation in training CNNs: Injecting noise to images*, ResearchGate, Oct. 1, 2019.

[10] M. Sabha, *Performance analysis of deep learning models*, IAES International Journal of Artificial Intelligence (IJ-AI), vol. 13, no. 4, pp. 4722–4735, Dec. 2024.

[11] E. Halsaunet, *Histograms of oriented gradients for human detection*, Norwegian University of Science and Technology (NTNU), 2015.