# Ensemble Methodology for Indian Subcontinental Festival Identification in Images

Minhajur Rahman Mahi
*dept. of Computer Science and Engineering*
*Ahsanullah University of Science and Technology*
Dhaka, Bangladesh
mrahman61142@gmail.com

DM Raffin
*dept. of Computer Science and Engineering*
*Ahsanullah University of Science and Technology*
Dhaka, Bangladesh
dmraffin10@gmail.com

Pritom Kumar Paul
*dept. of Computer Science and Engineering*
*Ahsanullah University of Science and Technology*
Dhaka, Bangladesh
pritompaul2000@gmail.com

Md. Abir Hossain Bony
*dept. of Computer Science and Engineering*
*Ahsanullah University of Science and Technology*
Dhaka, Bangladesh
abirbony@gmail,com

***Abstract*—Festivals showcase diverse customs and traditions, making them vital to our cultural heritage. The deep learning models, especially Convolutional Neural Networks (CNNs) employed in this study assists in automatically recognize and classify festival images. The goal is to preserve our rich cultural legacy, ensuring future generations can appreciate these celebrations. Through this research, we aim to enhance intercultural understanding by highlighting the diversity of festivals. The study evaluates the performance of an ensemble model which has achieved a high accuracy of 89.18%.**

## I. Introduction

Festivals are lively expressions of customs, traditions, and celebrations, creating a colorful picture of our cultural heritage. The need to explore deep learning models for image classification comes from a strong dedication to building a strong system that can automatically recognize and sort these culturally important moments. Preserving our cultural legacy targeted to assure that future generations can still identify and enjoy these celebrations.The current topic of utilizing deep learning models used for the identification and categorization of pictures representing Indian subcontinental festivities was explored in the paper. Festivals in this region include a variety of customs, traditions, and celebrations, creating a dynamic blend of cultural heritage. To create a strong framework that automatically recognizes, classifies, and categorizes these festival photos using deep learning techniques is at the heart of this research.

For two reasons, this study was carried out.Firstly this paper aims to safeguard and record the abundant cultural legacy that these celebrations showcase. Through the automation of the identification process,this research ensures that these traditions will continue to be identified and appreciated by future generations. Additionally, we want to advance intercultural understanding.Through showcasing the diversity of these festivals,it is anticipated that understanding of the unique celebrations enriching the culture of the Indian subcontinent will be promoted. The initial goal was to enhance mutual understanding and bridge cultural gaps among diverse groups of people. Festival images of different classes might have similar features or interconnection between features.The paper will also explore the performance of deep learning models on diverse features.To accomplish our objective, Convolutional Neural Network (CNN) models are employed. The focus is on determining the most suitable strategy for accurately recognizing and classifying the numerous festivals represented in our dataset through extensive study and testing. Transfer learning is applied in this research, where pretrained models serve as feature extractors. Artifical Neural Networks have shown a good advancement in performance regarding the area of computer vision ,specially Convolutional Neural Networks (reference).Hence, this study employs CNN due to its notable performance in these areas. The utilization of pretrained models is integral to saving resources and time, given the large dataset addressed in this research.

## II. Literature Review

Images from google was utilized to create a custom dataset by Yasmin et al.s [1], which included four public gathering classes in Bangladesh. It evaluated the Inception V3, VGG16, and basic sequential models in their unmodified and modified forms, both with and without data augmentation. With 87.5% test set accuracy, the modified VGG16 model without data augmentation was shown to be the top performance among all models for classifying these categories of public gatherings. Convolutional neural networks are the main tool Llamas et al.s [2] utilizes to categorize photos of architectural heritage. Ten classes were included in the dataset they created, and models such as Inception V3, ResNet, Inception-ResNet-v2, and AlexNet were tested. Among the fine-tuned models, Inception-ResNet-v2 produced the best accuracy (93.19%) using 128×128 image sizes. The dataset of Llamas et al.s [2] was utilized by Janković [3]. The author trained a basic CNN

model alongside four other machine learning algorithms: (i) MLP, (ii) Forest PA, (iii) AODE, and (iv) RSeslibKnn. They conducted feature extraction and attribute selection to enhance accuracy. In the final test without attribute selection, the CNN model performed best with an accuracy of 92.91%. However, after applying attribute selection, the MLP model achieved the highest accuracy of 98.9%. Ma et al.s [4] merges AI methods like Machine Learning and Natural Language Processing to preserve Vietnamese festivals in the Mekong Delta. It involves image processing, festival name classification using CNNs, and creating a Vietnamese festival ontology. With 1,588 Google-sourced festival images, future plans include nationwide expansion and a multi-language system.

Mayank Mishra [5] focuses on developing an image classification system for smartphones, categorizing images from social media into document-based, quote-based, and photograph categories. The objective is to implement a deep learning-based classification system capable of efficiently grouping images into these predefined categories.Using a self-created dataset,they trained various convolutional neural network CNN based models on the dataset to find the task's optimum model , including a baseline CNN, and fine-tuned models. The best-performing model involves transfer learning via feature extraction using VGG16.The approach holds potential for efficient image analysis and management on social media platforms. The paper recommends feature extraction using VGG16 as the most effective model for classifying images from social media.

Mohdsanadzakirizvi [6] provides an overview of image classification using Convolutional Neural Networks (CNNs). It covers the basics of CNNs, introduces popular datasets like MNIST, CIFAR-10, and ImageNet, and demonstrates building CNN models for each dataset using the Keras library. The article outlines the steps involved in constructing CNN models, including data loading, preprocessing, and model architecture. It discusses the importance of CNNs in computer vision tasks, particularly image classification, and showcases code examples for each dataset. Additionally, the article explores transfer learning using the VGG16 model on the Imagenette dataset, highlighting its effectiveness in improving accuracy.

Munlika Rattaphun [7]'s work is done with transfer learning which is a machine learning technique that allows a model trained on a large dataset to be fine-tuned for a specific task. It has been widely used in a variety of applications,collecting a Thai culture image dataset which consists of 1,000 high-quality images from 10 well-known Thai cultures and traditions; ii) Investigating the performance of three famous CNNs, namely MobileNet, EfficientNet, and residual network (ResNet) as pre-trained models for Thai culture image classification; iii) Exploring how pre-trained models can be utilized for Thai culture image classification by comparing training the models.Total of 1,000 images, split evenly into 10 classes.Comparing three different networks, EfficientNet performed best at 95.87% followed by ResNet at 95.04%, and MobileNet at 92.56%.

## III. Methodology

### A. Dataset

A custom dataset consisting images of 24 festivals from the Indian subcontinent is utilized.These images were collected from the internet using Google, utilizing an extension that downloaded images based on the festivals' names as query information. Some of the collected images were removed to avoid negative influences which might occur due to various factors, including image quality, background, and brightness.

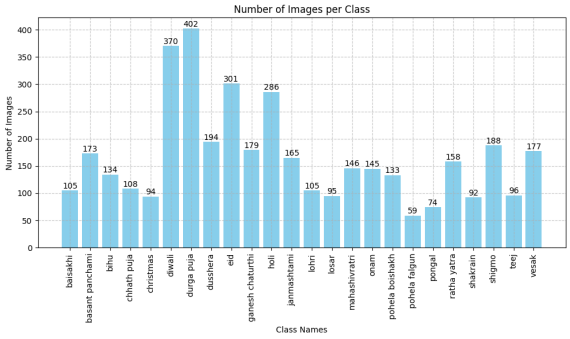


Fig. 1. Data Distribution

It is visible that there are some imbalances in out dataset. By assigning class weights for each class this issue this issue was addressed in the model training section.

Some sample images from the dataset are given below.


Fig. 2. Sample Images

The dataset has been partitioned into training and testing folders using an 80:20 split ratio.

### B. Preprocessing:

Every model which were employed in this paper necessitates its specific preprocessing steps. Original images typically have

pixel intensity ranging from 0 to 255. These values have been rescaled to the required range for each model.

### C. Transfer Learning :

The idea of ability to reuse forms the basis of the deep learning approach known as transfer learning. In this technique, a neural network previously trained and created for one task can serve as the foundation for another. All four models employed in this study consist of two components. Firstly, preprocessing layers were added, then pre-trained models are utilized for feature extraction, and in the classification part, fully-connected layers are employed. Secondly, a convolutional neural network was applied in this study to construct a model where feature engineering is conducted without human supervision. The feature extraction phase is executed by the pre-trained models. Four models have been employed in this thesis for the classification of festival images: 1) DenseNet-201 2) MobileNetV1 3) ConvNext-Base 4) EfficientNet-B7.

Dense Convolutional Network (DenseNet) is recognized for its unique architecture featuring dense connections between layers. Unlike traditional networks, each layer in DenseNet not only receives inputs from all preceding layers but also shares its own feature-maps as inputs for subsequent layers. This approach promotes effective information flow, enabling the network to leverage knowledge from both earlier and later layers. [8] EfficientNetB7 is part of the EfficientNet family, known for achieving state-of-the-art performance with relatively fewer parameters compared to traditional models. It employs a compound scaling method to balance model depth, width, and resolution, optimizing for both accuracy and computational efficiency. [9] ConvNeXt-Base is a convolutional neural network architecture that utilizes a combination of depthwise separable convolutions and spatially grouped convolutions to enhance feature representation. This design aims to capture both local and global dependencies in the input data efficiently. [10] MobileNets are based on a streamlined architecture that uses depthwise separable convolutions to build light weight deep neural networks. It employs depthwise separable convolutions to reduce computational complexity while maintaining effective feature learning. This makes MobileNetv1 particularly suitable for resource-constrained environments. [11]

### D. Hyper-parameters:

TABLE I
HYPER PARAMETERS

| Name | Values |
|---|---|
| Batch size | 8, 16, 32 |
| Activation Function | ReLU, Softmax |
| Regularizer | L2 |
| Loss Function | Categorical Cross-entropy |
| Optimizer | Adam, SGD |
| Learning Rate | 0.001, 0.00001 |

The hyper-parameters of TABLE I were tuned to get optimized results.

### E. Ensemble:

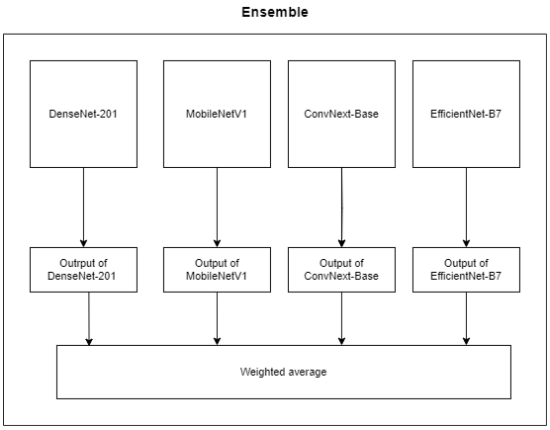


Fig. 3. Ensemble

The above mentioned 4 models were individually trained and tested initially. Then best three were ensembled by taking a weighted average of the outputs of each separate models. The suitable weights for each model was tuned to get the best result from the ensembled model.

### F. Implementation :

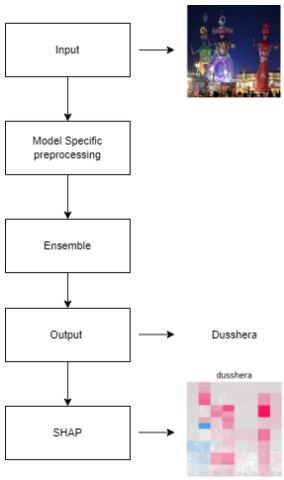


Fig. 4. Flowchart

Our dataset was turned into batched dataset where the batch size were varied and optimized. After that the preprocessing part was applied and fed to the ensemble model. The output generated by the model was then explained by shap.

## IV. EXPERIMENTAL RESULT

The deep learning models used in this study were evaluated on the test set of data consisting of 24 classes.The performance for each model can be observed in Table II.

TABLE II
THE PERFORMANCE FOR EACH OF THE APPLIED DEEP LEARNING MODELS

| Model Name | Accuracy |
|---|---|
| DenseNet-201 | 83.58 % |
| MobileNetV1 | 74.88 % |
| ConvNeXt-Base | **87.69 %** |
| EfficientNet-B7 | 83.46 % |

Based on the results, it can be observed that the ConvNeXt-Base model performed the best with 87.69 % of correctly classified instances.

Precision measures how accurate model is when it claims something belongs to a certain class. It is about being right when the model makes a positive prediction. On the other hand, recall looks at whether our model can identify and capture all instances of a specific class. Recall is about not missing any relevant objects when the model predicts positive.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (1)$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (2)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

TABLE III
PERFORMANCE MATRICES FOR ENSEMBLE MODEL

| Class | Precision | Recall | F1 Score |
|---|---|---|---|
| 1 | 0.95 | 0.86 | 0.90 |
| 2 | 0.77 | 0.57 | 0.66 |
| 3 | 0.88 | 0.85 | 0.87 |
| 4 | 1.00 | 1.00 | 1.00 |
| 5 | 0.94 | 0.89 | 0.92 |
| 6 | 0.84 | 0.96 | 0.89 |
| 7 | 0.96 | 0.96 | 0.96 |
| 8 | 0.97 | 0.92 | 0.95 |
| 9 | 0.86 | 0.98 | 0.92 |
| 10 | 0.97 | 0.83 | 0.90 |
| 11 | 0.95 | 1.00 | 0.97 |
| 12 | 0.75 | 0.82 | 0.78 |
| 13 | 1.00 | 0.90 | 0.95 |
| 14 | 0.86 | 1.00 | 0.93 |
| 15 | 0.76 | 0.87 | 0.81 |
| 16 | 0.81 | 0.86 | 0.83 |
| 17 | 0.96 | 0.89 | 0.92 |
| 18 | 0.86 | 0.50 | 0.63 |
| 19 | 1.00 | 0.93 | 0.97 |
| 20 | 0.94 | 0.91 | 0.92 |
| 21 | 0.88 | 0.79 | 0.83 |
| 22 | 0.92 | 0.92 | 0.92 |
| 23 | 0.76 | 0.80 | 0.78 |
| 24 | 0.88 | 0.81 | 0.84 |

For the ensemble model precision ,recall and f1-score have been stated respectively in Table III.

## V. RESULT ANALYSIS

The performance metrics table reveals notable discrepancies in the model's accuracy across different classes. While some classes exhibit satisfactory performance, others demonstrate significant inaccuracies. This variance in performance can be attributed to several factors, including class imbalance and color bias.

Classes with limited data availability like "Pohela Falgun" suffer from lower accuracy rates compared to classes with abundant samples. This imbalance in the dataset leads to disparities in the model's ability to accurately classify instances from underrepresented classes. As a result, the model may struggle to generalize effectively across all classes, impacting overall performance.

An intriguing observation is the influence of color bias on the model's predictions, particularly evident in classes with similar color distributions. For instance, the classes associated with "Basant Panchami" exhibit predominantly yellow hues.

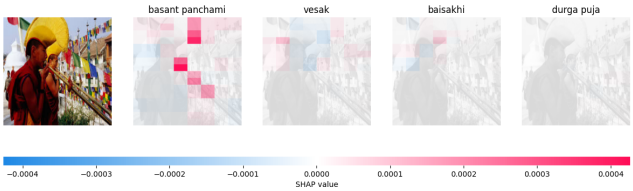


Fig. 5. Miss-classification due to presence of yellow color

Consequently, images containing significant yellow tones are more likely to be misclassified, with a tendency to be labeled as "Basant Panchami" due to its larger training data size. This bias introduces inaccuracies in the model's predictions, affecting the fairness and reliability of its output.

## VI. CONCLUSION

The misclassification of some festivals is attributed to both data imbalance among classes and the resemblance in features across different classes. Moving forward, the dataset can be explanded to encompass a broader array of festivals,number of images in each classes can be increased for a more comprehensive representation. Additionally, refining the deep learning models and introducing real-time image classification capabilities could further enhance the practical application of this technology in preserving and promoting cultural heritage. Convolutional Neural Networks (CNNs) were utilized to automate the identification and classification of festival images in the Indian Subcontinent. Through showcasing the diversity of these festivals, using automation in identification and classification of festival images,understanding of the unique celebrations enriching the culture of the Indian subcontinent was promoted. Notably, ensemble model exhibited better performance than the individual models.Foundation is established for future endeavors focused on leveraging deep learning models for the documentation and comprehension of diverse festivals.

## REFERENCES

[1] S. Yeasmin, N. Afrin, K. Saif, O.T. Imam, A. W. Reza,& Md. S. Arefin, "Image Classification for Identifying Social Gathering Types," Int. Conf. on Intell. Comput. & Optim., pp. 98–110, Springer, 2022.

[2] J. Llamas, P. M. Lerones, R. Medina, E. Zalama and J. Gómez-García-Bermejo, “Classification of architectural heritage images using deep learning techniques,” Appl. Sci., vol. 7, no. 10, pp. 992, Sep. 2017.
[3] R. Janković, “Machine learning models for cultural heritage image classification: Comparison based on attribute selection,” Inf., vol. 11, no. 1, pp. 12, Dec. 2019.
[4] NK. Chau, TT. Ma, Z. Bouraoui, TN. Do, “A Vietnamese Festival Preservation Application,” Proc. of Int. Conf. on Inf. Technol. and Appl.: ICITA 2021, pp. 449–460, Springer, 2022.
[5] M. Mishra, T. Choudhury, T. Sarkar, “CNN based efficient image classification system for smartphone device,” 2021.
[6] Mohd. S. Z. Rizvi, “Image Classification Using CNN (Convolutional Neural Networks),” Analyticsvidhya.com. https://www.analyticsvidhya.com/blog/2020/02/learn-image-classification-cnn-convolutional-neural-networks-3-datasets (accessed November 28th, 2023).
[7] M. Rattaphun, K. Songsri-in, “Thai Culture Image Classification With Transfer Learning,” Int. J. of Elect. & Comput. Eng. (2088-8708), vol. 13, no. 6, 2023.
[8] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, ”Densely connected convolutional networks,” in 2017 IEEE Conf. on Comput. Vision and Pattern Recognit. (CVPR), pp. 4700-4708, July 2017, doi: 10.1109/CVPR.2017.243.
[9] Mingxing Tan and Quoc V. Le, ”EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” IEEE Trans. on Pattern Anal. and Mach. Intell., vol. 43, no. 5, pp. 1781-1792, May 2021. doi: 10.1109/TPAMI.2019.2928085
[10] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, ”A ConvNet for the 2020s,” in 2022 IEEE/CVF Conf. on Comput. Vision and Pattern Recognit. (CVPR), pp. 11966-11976, doi: 10.1109/CVPR52688.2022.01167.
[11] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ”MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv:1704.04861 [cs.CV], 2017.