Traditional Costume Image Classification for Indian States Using Deep Learning

Sahana R Koti
Dept. of Master of Computer Applications
K.L.E. Institute of Technology
Hubballi, India, 580 027
ksahanarj@gmail.com

Sahana Jatti
Dept. of Master of Computer Applications
K.L.E. Institute of Technology
Hubballi, India, 580 027
sahanajatti2017@gmail.com

Anupama S. Nandeppanavar
Dept. of Master of Computer Applications
KLE Institute of Technology
Hubballi, India
anupama.nandeppanavar@kleit.ac.in

Dr. Medha Kudari
Dept. of Master of Computer Applications
KLE Institute of Technology
Hubballi, India
medha.k@kleit.ac.in

Abstract*— Traditional Costume Image Classification for Indian States Using Deep Learning is an emerging field in computer vision that aims to identify and categorize traditional attires from various states based on their visual features. This research addresses the development of a robust image classification model to recognize and classify costumes from different Indian states. The project leverages a deep learning approach, utilizing convolutional neural networks (CNNs) to process and analyze costume images. The primary dataset comprises images of traditional costumes from Indian state, ensuring diversity in attire styles, colors, patterns, and cultural representations. Preprocessing steps include image resizing, normalization, and augmentation to enhance model generalization. To evaluate the model's performance, standard metrics such as accuracy, precision, recall, and F1-score are employed. The results demonstrate the model's capability to achieve high classification accuracy, with notable precision in distinguishing between similar attires from neighboring states. The model is trained using different DL algorithms, finding their accuracy and predicting the results. The analysis was made by considering the raw data. Models' performance was improved with normalization with the following accuracies. 95%, 98% and 100% accuracies are obtained for VGG16, MobileNetV2 and ResNet50V2 and DenseNet121 respectively.

***Keywords*— Deep Learning, VGG16, MobileNetV2, ResNet50V2, DenseNet121**

I. Introduction

The project on " Traditional Costume Image Classification for Indian States Using Deep Learning " aims to leverage advanced machine learning techniques to classify and recognize diverse cultural traditions based on images. This innovative endeavor seeks to capture the rich tapestry of traditions prevalent across different states, harnessing the power of state-of-the-art image classification models. The primary objective is to develop a robust system capable of accurately identifying and categorizing traditional practices, attire, rituals, and artifacts unique to each state. The states considered in the project work are Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharashtra, Manipur, Mizoram and West Bengal as shown in Fig. 1. The cultural diversity across states often manifests in distinct visual elements, making image classification an ideal approach for preserving and showcasing these traditions. The project begins by curating a comprehensive dataset that encapsulates a wide range of visual representations of various traditions from different states. Leveraging machine learning algorithms, particularly convolutional neural networks (CNNs) or similar architectures, the system learns to automatically extract features and patterns from the images, enabling it to discern the nuanced differences in traditional practices.

Fig. 1 List of States Considered

The significance of this project lies in its potential to contribute to cultural preservation, awareness, and education. By automating the process of tradition classification, the system can assist in identifying and cataloging cultural heritage, fostering a deeper understanding and appreciation of the diverse traditions that make each state unique. The project aligns with the broader goals of promoting cultural diversity, heritage conservation, and the application of cutting-edge technologies for meaningful societal impact.

The project is a groundbreaking initiative aimed at harnessing the capabilities of machine learning to classify and recognize diverse cultural traditions across different states. In a world characterized by rich cultural diversity, this project seeks to employ advanced image classification techniques to automatically identify and categorize traditional practices, artifacts, and ceremonies unique to each state. The primary focus is on creating a robust system that, through the power of state-of-the-art deep learning models, can analyze and classify images representing various traditions. By leveraging convolutional neural networks (CNNs) or similar architectures, the project aims to teach the system to discern intricate visual patterns and features specific to different cultural practices.

The ultimate goal is to contribute to the preservation and documentation of cultural heritage, fostering a deeper understanding and appreciation of the unique traditions that define each state. This project not only aligns with the technological advancements in machine learning but also carries significant implications for cultural preservation, education, and the celebration of diversity on a national scale.

The remaining part of the paper is organized as follows: section II contains the related work. Section III includes the methodology proposed along with the working model being presented. Section IV presents the results obtained using the proposed method. Section V describes the discussion based on results obtained and the inferences drawn. Section VI concludes the article.

II. RELATED WORKS

Conducting a literature survey to realize the state-of-the-art in Traditional Costume Image Classification for Indian States Using Deep Learning, the following is the gist of the articles related to the proposed work.

The study by Naikar, M. L., Nandeppanavar, A. S., & Kudari, M. (2023), explores the application of deep learning techniques for the detection of Lumpy Skin Disease (LSD) in cattle. This paper, spanning pages 182-187, highlights the development and implementation of a graphical user interface (GUI) integrated with a deep learning model to facilitate the early detection and diagnosis of LSD, a significant viral disease affecting cattle. The authors emphasize the importance of timely and accurate detection to mitigate the spread and impact of the disease on livestock health and the agricultural economy [2].

The paper by Thotad, P. N., Kallur, S., & Nandeppanavar, A. (2023), spans pages 1-6 and investigates the application of deep learning models for the early detection of plant diseases in agriculture. This study addresses a critical challenge in the agricultural sector, where timely and accurate identification of plant diseases is essential for crop health and yield optimization. The authors have developed an advanced deep learning model designed to analyze images of plants and accurately classify various diseases affecting them. By leveraging convolutional neural networks (CNNs), the model demonstrates high precision and reliability in detecting diseases from images, even in challenging scenarios with varying environmental conditions and plant types [3].

Khatchatourov, Artem, and Christoph Stamm explored the image-based classification of Swiss traditional costumes using contextual features. They discussed various image processing techniques, feature extraction methods, and machine learning algorithms applied to similar domains. The survey included a comprehensive review of existing work on Swiss traditional costumes, examining cultural aspects and challenges associated with their accurate digital representation. The authors also addressed the limitations and gaps in the existing literature that their research aims to address [5].

Sun, Yanan, et al. The study delves into the evolution of CNNs, emphasizing their role in image classification tasks. Researchers reviewed previous works on evolving neural network architectures, considering methods for adapting and optimizing CNN structures over time. The survey touched upon the historical development of deep learning techniques, including the rise of CNNs and their widespread success in computer vision applications. Furthermore, they examined literature on evolutionary algorithms applied to neural network design, investigating how genetic algorithms or other optimization techniques contribute to evolving effective CNN architectures for image classification [6].

Y. Sun, B. Xue, M. Zhang, and G. G. Yen likely focused on the application of evolutionary algorithms to the development and optimization of deep convolutional neural networks (CNNs) for image classification tasks. They explored how evolutionary methods can be employed to automatically adapt and evolve the architecture, parameters, or features of deep CNNs, aiming to enhance their performance in image classification. The authors investigated the evolutionary search space, assessing different configurations and structures of CNNs to identify those that are most effective for diverse image datasets [7].

Rawat, Waseem, and Zenghui Wang likely provided an extensive examination of the advancements and trends in the application of deep convolutional neural networks (CNNs) for image classification. Their survey spanned a range of publications in the field of computer vision and deep learning, exploring the evolution of CNN architectures, methodologies, and applications over time. They covered key breakthroughs and milestones in CNN development, such as the introduction of architectures like AlexNet, VGG, GoogLeNet, and ResNet, each contributing to improved image classification performance [8].

Liu, Z., Luo, P., Qiu, S., Wang, X., and Tang the study delved into the details of the Deep Fashion dataset, discussing its rich annotations and how it facilitates robust training and evaluation of algorithms for clothing-related tasks. The authors emphasized the challenges addressed by Deep Fashion, such as variations in clothing styles, poses, and occlusions commonly encountered in real-world scenarios. The paper also discussed the methodologies and techniques used to leverage Deep Fashion for developing or evaluating deep learning models, possibly focusing on convolutional neural networks (CNNs) or other advanced architectures [9].

Kalantidis, Y., Kennedy, L., and Li, L.-J. explored the existing research landscape concerning clothing recognition and segmentation for the purpose of generating automatic suggestions in everyday photos. The survey delved into studies related to computer vision, image analysis, and machine learning techniques applied to clothing recognition. Researchers reviewed methods and models employed for detecting and segmenting clothing items within images, considering the challenges posed by variations in pose, lighting, and background [10].

The paper authored by Akata, Z., Perronnin, F., Harchaoui, Z., and Schmid is likely to address best practices and methodologies in the domain of large-scale learning for image classification. The study may delve into the challenges associated with handling extensive datasets and propose effective strategies for training models at scale. The authors may discuss techniques for efficient feature extraction, model training, and optimization, considering the computational demands of working with large volumes of image data [11].

Chen, H., Gallagher, A., and Girod explored the domain of clothing description using semantic attributes. The literature in this area covered various techniques and methodologies employed in computer vision and image processing to effectively characterize clothing items. The authors investigated how semantic attributes, such as color, pattern, texture, and style, can be extracted and utilized for detailed clothing description. The survey delved into the challenges associated with recognizing and describing diverse clothing categories and variations in real-world images [12].

Shivakumara, Palaiahnakote, C. Pavan Kumar, Jagrut J. Nemade, Kshitiz Michael, Akash Kumar, Basavaraj S. Anami, and Umapada Pal, the paper introduces a novel system utilizing the U-Net architecture for the classification of wedding images from diverse cultural backgrounds. The authors focus on developing a robust method capable of accurately identifying and categorizing wedding images based on cultural characteristics, leveraging the U-Net's strengths in image segmentation and deep learning. This research contributes to the field of computer vision by addressing the unique challenges posed by the variability and richness of cultural wedding attire, decorations, and settings, aiming to improve automated image classification systems in multicultural contexts [13].

III. Methodology

The methodology for Traditional Costume Image Classification for Indian States involves collecting a diverse dataset of traditional costume images from various Indian states. Preprocessing steps like resizing and augmentation are applied to the images. A convolutional neural network (CNN) is then trained on this dataset to learn distinctive features of each state's traditional attire. The trained model is evaluated using metrics like accuracy, precision, and recall to ensure its effectiveness in correctly classifying the costumes into their respective states.

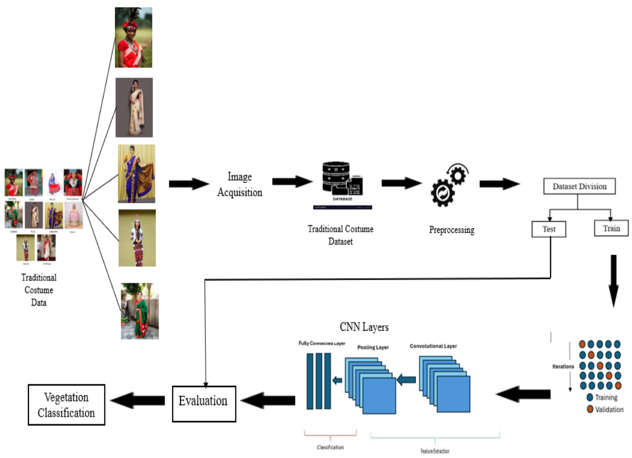

Fig. 2 Methodology of Traditional Costume Image Classification for Indian States

Fig. 2 provides a detailed workflow for Traditional Costume Image Classification for Indian States. It starts with the collection of traditional costume data, which includes images of traditional attire from different Indian states. These images are then acquired and compiled into a traditional costume dataset. The dataset undergoes preprocessing steps such as resizing, normalization, and augmentation to enhance the quality and variability of the images. Next, the dataset is divided into training and testing subsets to facilitate model training and evaluation. The core of the classification process involves a Convolutional Neural Network (CNN), which comprises convolutional layers, pooling layers, and fully connected layers. The CNN is trained on the training subset to learn intricate patterns and features specific to each state's traditional costumes. Once the model is trained, it is evaluated using the testing subset to assess its accuracy and effectiveness in classifying the traditional costumes. The evaluation phase includes calculating performance metrics such as accuracy, precision, and recall.

A. Dataset Preparation

Dataset Preparation involves several critical steps. Initially, a diverse collection of images representing traditional costumes from multiple sources, such as from google images, bing website, pinterest and cultural websites, and social media platforms. This ensures a wide range of styles and contexts.

Table 1 gives the list of ten states and number of images considered as the dataset. The images are augmented images where the images from each state range from 641 to 650 images. The total images from all the classes are 6817. The dataset is then divided where the 80% of images are considered for training the model and the remaining 20% images are part of validation. The number of images from training set ranges from 500 to 550 in each class and the total number of images in training set are 5450. The number of images from validation set ranges from 120 to 150 in each class and the total images in validation set are 1367.

Table 1. Image count of each state.

S.No.	States	Image Count
1	Chhattisgarh	641
2	Gujarat	894
3	Haryana	639
4	Himachal Pradesh	638
5	Karnataka	642
6	Kerala	690
7	Maharashtra	749
8	Manipur	642
9	Mizoram	644
10	West Bengal	638

B. Preprocessing

Preprocessing in the context of image classification involves preparing the raw image data to enhance the performance of machine learning models. This typically includes resizing images to a consistent dimension to ensure uniformity, which is crucial for feeding images into convolutional neural networks (CNNs). In addition to resizing, data augmentation techniques such as rotation, flipping, cropping, and color adjustments are applied to increase the diversity of the training dataset without collecting new images. This helps in reducing overfitting by exposing the model to various transformations of the input data. For the Traditional Costume Image Classification for Indian States using Deep Learning, preprocessing steps ensured that the dataset of 6,817 images was standardized and enriched, resulting in 5,450 training images and 1,367 validation images. These steps are critical for enhancing model generalization and robustness. The dataset is standardized and ready for subsequent testing machine learning deep learning tasks by performing the preprocessing techniques as listed below,

1) Image Resize: Image resizing is a crucial preprocessing step in image classification tasks, where images are adjusted to a uniform size to ensure consistency when fed into a convolutional neural network (CNN). All images were resized to 224x224 pixels. This specific dimension is widely used because it balances detail retention and computational efficiency, fitting well within the architectural requirements of popular CNN models like DenseNet, ResNet, VGG16, and MobileNet. By resizing all images to 224x224, the model can process each image uniformly, avoiding discrepancies that varying image sizes could introduce. This standardization is essential for maintaining the integrity of the feature extraction process, ensuring that the network learns and identifies patterns effectively across all input images.

2) Data Augmentation: Data augmentation is a vital preprocessing technique in image classification that enhances the diversity of the training dataset without the need for additional image collection. By applying various transformations to the existing images, data augmentation effectively increases the dataset size, helping to prevent overfitting and improve the model's generalization capabilities. Common augmentation techniques include rotation, flipping, cropping, scaling, translation, and color adjustments such as brightness, contrast, and saturation changes.

Data augmentation played a crucial role in enriching the dataset of 6,817 images. Each original image was transformed through these techniques, creating multiple variations that the model could learn from. For instance, rotating images helped the model recognize costumes from different angles. Cropping and scaling adjusted the images' framing, making the model adept at focusing on relevant features regardless of the image's position or scale. Color adjustments simulated various lighting conditions, ensuring the model could correctly classify images with different illumination and color profiles [4].

IV. Model Design

The process begins with image acquisition, where a dataset comprising images from ten classes Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharashtra, Manipur, Mizoram and West Bengal are collected. Each class contains between 600 to 750 images before dividing the dataset into 80% and 20%, resulting in a diverse and rich dataset. The acquired images undergo preprocessing, which includes resizing to a uniform dimension to maintain consistency and performing data augmentation techniques such as rotation, flipping, and scaling to enhance the dataset's variability and robustness. This preprocessing step ensures that the dataset is well-prepared for training the models. After preprocessing, the total number of images is 6817.

The dataset is then split into training and validation sets, with 80% (5450 images) allocated for training and 20% (1367 images) for validation. This split helps in evaluating the models' performance on unseen data. The training process employs several Convolutional Neural Network (CNN) architectures, including DenseNet121, ResNet50V2, VGG16, and MobileNetV2. These models are chosen for their proven effectiveness in image classification tasks. During training, DenseNet121 achieves the highest validation accuracy of 98.8%, followed closely by ResNet50V2 at 98.7%, MobileNetV2 at 98.2%, and VGG16 at 95.6%. For further evaluation, a separate test set comprising 50 images, with 10 images from each of the 10 classes, is used. On this test set, DenseNet121 and ResNet50V2 both achieve a perfect accuracy of 100%, demonstrating their robustness and reliability. MobileNetV2 and VGG16 achieve test accuracies of 90% and 92%, respectively.

A. Classifiers

In the state-wise costume image classification task, several advanced Convolutional Neural Network (CNN) architectures were employed as classifiers to distinguish between traditional costumes from ten different Indian states. The classifiers used include DenseNet121, ResNet50V2, VGG16, and MobileNetV2. Each of these models brings unique strengths to the table. DenseNet121 (Densely Connected Convolutional Networks) is known for its efficient layer connectivity, which helps in better gradient flow and feature reuse, leading to its superior performance in this task.

1) Visual Geometry Group-19 (VGG16): VGG16, developed by the Visual Geometry Group at the University of Oxford, is a deep convolutional neural network architecture that has significantly influenced the field of image classification. It consists of 16 layers, including 13 convolutional layers and 3 fully connected layers, with a distinctive pattern of using small 3x3 convolution filters throughout the network. This design choice allows for deep feature extraction while maintaining manageable computational complexity. VGG16 is known for its simplicity and effectiveness, achieving high accuracy in image classification tasks. Despite its relatively high computational requirements and memory consumption, VGG16 remains a popular choice due to its robust performance and ability to generalize well across different datasets. VGG16 is a widely recognized deep convolutional neural network (CNN) architecture developed by the Visual Geometry Group at the University of Oxford. It consists of 16 layers, including 13 convolutional layers and 3 fully connected layers.

VGG16 is praised for its straightforward and uniform architecture, which contributes to its robust performance in image classification tasks. The network utilizes convolutional layers to capture spatial features of input images, max-pooling layers for down-sampling, and fully connected layers for the final classification. ReLU activation functions introduce non-linearity, enhancing the model's capability to learn complex patterns. VGG16 has been pre-trained on large datasets like ImageNet, making it a popular choice for transfer learning in various computer vision applications. Despite its large number of parameters, which can be computationally intensive, VGG16 is favored for its deep network structure and ease of implementation.

2) MobileNetV2: MobileNetV2 is a streamlined convolutional neural network architecture designed for efficient performance on mobile and embedded vision applications. It builds on the success of the original MobileNet, introducing innovations like inverted residuals and linear bottlenecks, which enhance its ability to capture and represent image features while minimizing computational cost. The network employs depthwise separable convolutions, drastically reducing the number of parameters and multiplications compared to traditional convolutions. This efficiency allows MobileNetV2 to achieve high accuracy with significantly less processing power and memory usage, making it ideal for real-time image classification tasks on devices with limited resources.

Another crucial aspect of MobileNetV2 is the use of linear bottlenecks in conjunction with depth wise separable convolutions. Depth wise separable convolutions decompose the standard convolution operation into two distinct steps: the depth wise convolution, which applies a single convolutional filter per input channel, and the pointwise convolution, which uses a 1x1 convolution to combine these filtered outputs. This separation dramatically reduces the number of computations compared to traditional convolutions. The linear bottleneck layers further refine this process by maintaining a low-dimensional representation of the data, ensuring that the network does not lose important information while still achieving significant computational savings. Feature extraction in MobileNetV2 is accomplished through depth wise separable convolutions, which split the convolution operation into two distinct stages. The first stage, depth wise convolution, applies a single convolutional filter to each input channel independently, capturing spatial features while keeping the number of parameters low. The second stage, pointwise convolution (1x1 convolution), combines these outputs across channels to create new feature representations. This separation significantly reduces the computational burden compared to traditional convolutions, which operate on both spatial and depth dimensions simultaneously. MobileNetV2 incorporates batch normalization and ReLU6 activations after each convolutional layer, ensuring stable and efficient training by normalizing activations and providing non-linearity. The network also employs average pooling layers towards the end, which down-sample the spatial dimensions of the feature maps, aggregating the most critical information and reducing the feature map size before the final classification layer. This combination of depth wise separable convolutions, inverted residuals, and linear bottlenecks makes MobileNetV2 a powerful and efficient model for various real-time image classification and detection tasks, especially suited for mobile and embedded applications.

3) ResNet50V2: ResNet-50v2 is a convolutional neural network architecture that belongs to the ResNet (Residual Network) family. It improves upon the original ResNet-50 architecture by introducing various modifications aimed at enhancing performance and efficiency. These improvements include using a revised residual block design with bottleneck blocks, which helps in reducing the number of parameters while maintaining or improving accuracy. ResNet-50v2 has been widely used in computer vision tasks such as image classification and object detection, where its deep architecture and residual connections allow for effective learning of complex features from visual data [5].
The critical aspect of ResNet50V2 is the inclusion of residual blocks. Each residual block contains multiple convolutional layers and incorporates skip connections. These skip connections allow the network to learn residual functions by capturing the difference between the input and output of a block. By passing the input through the partnership and adding the residual connection, the network can capture fine-grained details while maintaining a high-level understanding of the image. During feature extraction, the input image passes through the convolutional layers and residual blocks in a forward propagation manner. As it progresses through the network, the image features are gradually refined and abstracted, capturing both low-level and high-level visual information. The deeper layers in the network capture more complex and abstract elements, while the shallower layers focus on local details. The outcome of the last convolutional layer or residual block represents the extracted features from the input image. These features can be further processed and used for various downstream tasks, such as classification, object detection, or segmentation. Including residual connections in ResNet50V2 helps address the vanishing gradient problem, enabling the network to extract and learn meaningful characteristics from the given images effectively [1].

4) DenseNet121: DenseNet-121 is a convolutional neural network architecture known for its densely connected layers. Unlike traditional architectures where each layer feeds into the next, DenseNet connects each layer to every subsequent layer in a feed-forward fashion. This dense connectivity pattern encourages feature reuse, facilitates gradient flow, and reduces the number of parameters. DenseNet-121 specifically refers to a variant with 121 layers, consisting of dense blocks where each layer receives input from all preceding layers within the same block. This design promotes feature propagation and enables efficient learning, making DenseNet121 effective for tasks like image classification and object detection in computer vision [2].

- Feature Extraction in DenseNet-121: In DenseNet-121, feature extraction occurs through a series of convolutional layers and dense blocks. Each dense block comprises multiple convolutional layers, where the output of each layer is concatenated with the inputs of all subsequent layers within the block. This dense connectivity allows the network to learn diverse and rich feature representations, capturing intricate patterns and details from the input images.

- Max Pooling: Max pooling layers in DenseNet-121 are used sparingly and typically appear in the transition layers between dense blocks. Max pooling layers perform down-sampling by selecting the maximum value from non-overlapping regions of the feature maps. This reduces the spatial dimensions of the feature maps, helping to control the computational complexity and memory usage while preserving important features.

- Average Pooling: Average pooling layers are often used towards the end of the network, particularly in the global average pooling layer before the fully connected layer for classification. Average pooling calculates the average value of the features within non-overlapping regions of the feature maps. This helps to reduce the spatial dimensions further while maintaining the overall distribution of the features.

V. RESULTS AND DISCUSSION

According to the collected data, the dataset consists of images representing 10 different classes of 10 different states costumes with 250 images per class. These images were used for training and validating different Convolutional Neural Network (CNN) models to classify the costume images to that particular state.

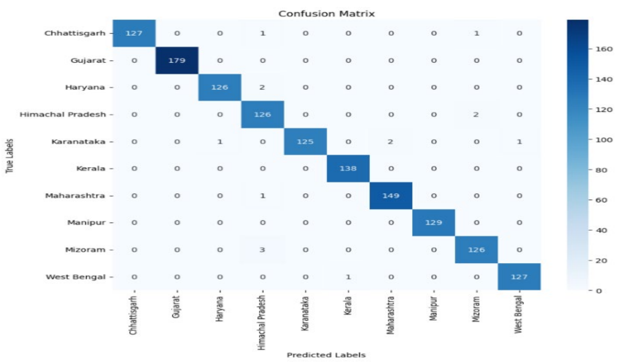

Fig. 3 Confusion matrix of Validation set of DenseNet121 Model

Fig.3 is the confusion matrix of the validation dataset of DenseNet121 Model, the confusion matrix provided illustrates the performance across ten Indian states: Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharashtra, Manipur, Mizoram, and West Bengal. Each cell in the matrix shows the number of correct and incorrect predictions made by the model. The diagonal elements represent correctly classified images for each state, with high values indicating strong classification accuracy. For example, Gujarat has 179 correct classifications out of the total images, while West Bengal has 127. Misclassifications, indicated by non-diagonal elements, are minimal, showing that only a few images from one state were incorrectly classified as another.

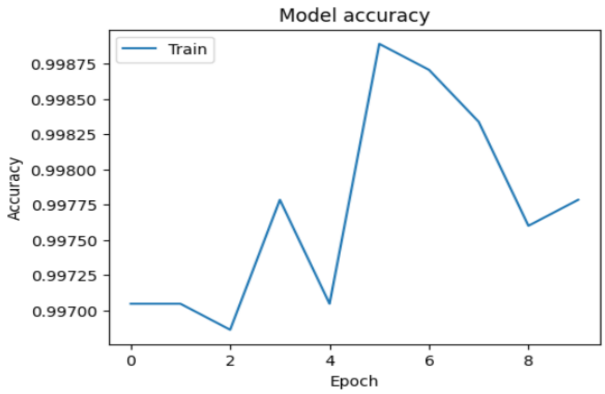

Fig. 4 Model accuracy of Validation set of DenseNet121 Model

Fig. 4 is the graph displays the model accuracy over ten training epochs, highlighting the training accuracy at each epoch. The y-axis represents accuracy, ranging from 0.9970 to nearly 0.9990, while the x-axis denotes the number of epochs, from 0 to 9. Initially, the model accuracy starts at approximately 0.9970 and remains steady for the first epoch. It then experiences fluctuations, with notable peaks and troughs indicating varying performance improvements and regressions during subsequent epochs. A significant peak occurs around epoch 5, where the accuracy nearly reaches 0.9990, followed by a decline and another fluctuation towards the end. These variations suggest that the model's training process involves fine-tuning and adjusting to the data, achieving high accuracy with some instability in intermediate epochs.

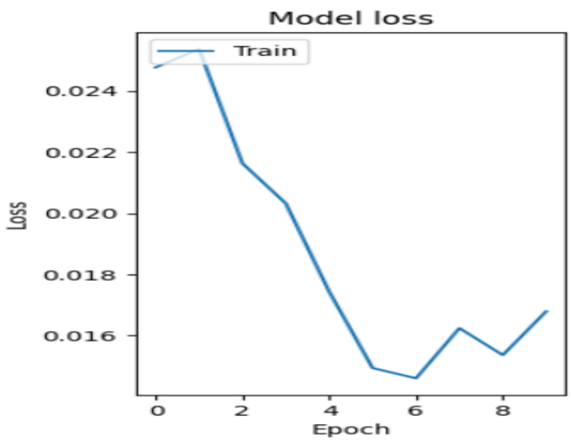

Fig. 5 Model loss of Validation set of DenseNet121 Model

Fig. 5 is the graph illustrates the model loss over ten training epochs, with the y-axis representing the loss value, which ranges from approximately 0.016 to 0.025, and the x-axis representing the number of epochs, from 0 to 9. Initially, the model loss starts at around 0.025 and shows a significant decline during the first few epochs. By epoch 3, the loss drops to about 0.018, indicating a considerable improvement in model performance. The loss continues to decrease, reaching its lowest point around epoch 5, at approximately 0.016. After this point, the loss begins to fluctuate slightly but remains relatively low, suggesting that the model has largely stabilized and is fine-tuning its performance. These variations in loss values indicate the model's ongoing adjustment and optimization during training.

To calculate accuracy, sum all true positives (diagonal elements), that is 46 images and divide by the total instances (50), resulting in an accuracy of 100% using the binary classification formula can be calculated as given in eq (1), precision can be calculated as given in eq (2).

Formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{eq (1)}$$

Formula:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{eq (2)}$$

Table 2 presents the performance metrics of four CNN models—DenseNet121, ResNet50V2, VGG16, and MobileNetV2—evaluated based on their accuracy and precision scores. DenseNet121 achieved the highest accuracy score of 98.8%, indicating its ability to correctly classify 98.8% of the costume images in the test set. DenseNet121 had a precision score of 0.99, meaning 99% of its predictions were accurate, highlighting its low false positive rate. ResNet50V2 followed closely with an accuracy score of 98.7% and a precision score of 0.99, demonstrating strong performance in correctly identifying costumes. MobileNetV2 V2 achieved an accuracy of 98.2% and a precision of 0.98, indicating that 98% of its predictions were accurate. VGG16, while still effective, had the lowest accuracy at 95.6% and a precision score of 0.96, suggesting that 96% of its predictions were correct.

Table 2: Performance of CNN models

Models	Accuracy Score	Precision Score
DenseNet121	**98.8%**	**0.99**
ResNet50V2	98.7%	0.99
MobileNetV2	98.2%	0.98
VGG16	95.6%	0.96

Fig. 6 is the confusion matrix displayed above provides a detailed evaluation of the performance of a Traditional Costume Image Classification classification model for the test dataset. In this matrix, the true labels are listed on the vertical axis, and the predicted labels are listed on the horizontal axis. Each cell in the matrix represents the number of instances where the true class corresponds to the predicted class. The diagonal cells, from the top left to the bottom right, show the number of correct predictions for each state. Off-diagonal cells indicate misclassifications, where the model's predicted label differs from the true label. The matrix reveals that the classifier correctly identified all 5 images for each of the ten classes, indicating perfect classification with no misclassifications.

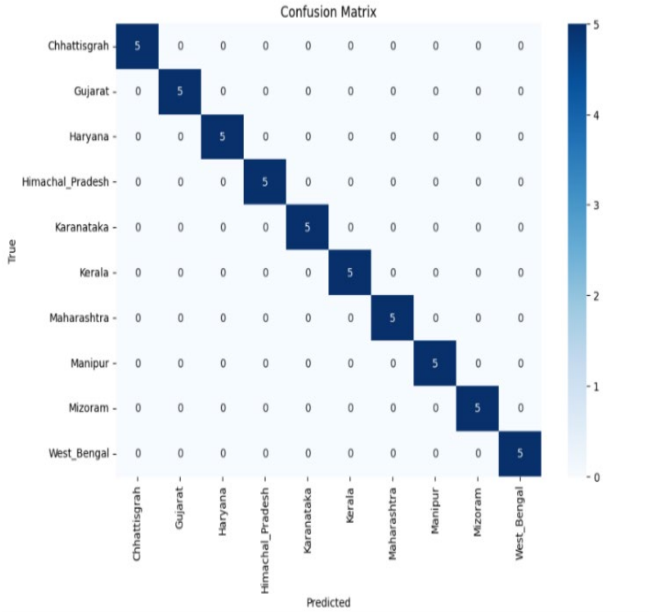

Fig. 6 Confusion matrix of tested data applying DenseNet121

CONCLUSION

The analysis of the confusion matrix demonstrates that the state-wise costume image classification model has achieved a perfect performance with a 100% accuracy rate. Each of the ten classes—Chhattisgarh, Gujarat, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharashtra, Manipur, Mizoram, and West Bengal—was correctly identified with no misclassifications, as evidenced by the zero values in all off-diagonal cells. This flawless classification indicates that the model effectively distinguishes between the unique features of traditional costumes from each state. The robust performance across all classes highlights the model's capability and reliability in accurately categorizing costume images. This success is particularly significant given the diversity and potential complexity in visual characteristics across different state costumes. The model's exceptional accuracy underscores its potential utility for applications in cultural heritage preservation, automated costume recognition in media, and educational tools for learning about traditional attire. The results of the state-wise costume image classification model, as depicted in the confusion matrix, are highly impressive, achieving perfect accuracy with no misclassifications. This indicates that the model is exceptionally proficient at distinguishing the unique characteristics and features of traditional costumes from the ten different Indian states included in the dataset.

REFERENCES

[1] Nandeppanavar, A. S., Kallur, S. S., Thotad, P., & Sankannavar, V. A. (2023, November). Bharatanatyam hasta mudra categorization using deep learning approaches. In *2023 IEEE North Karnataka Subsection Flagship International Conference (NKCon)* (pp. 1-6). IEEE.

[2] Naikar, M. L., Nandeppanavar, A. S., & Kudari, M. (2023, September). Lumpy Skin Disease Detection Using GUI Based Deep Learning Model in Cattle. In *2023 Third International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 182-187). IEEE.

[3] Thotad, P. N., Kallur, S., & Nandeppanavar, A. (2023, October). An Efficient Model for Plant Disease Detection in Agriculture Using Deep Learning Approaches. In *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). IEEE.

[4] Bajpai, C., Sahu, R., & Naik, K. J. (2023). Deep learning model for plant-leaf disease detection in precision agriculture. *International Journal of Intelligent Systems Technologies and Applications,* 21(1), 72-91.

[5] Khatchatourov, Artem, and Christoph Stamm. "Image-based Classification of Swiss Traditional Costumes using Contextual Features." *VISIGRAPP (4: VISAPP)*. 2020.

[6] Sun, Yanan, et al. "Evolving deep convolutional neural networks for image classification." *IEEE Transactions on Evolutionary Computation* 24.2 (2019): 394-407.

[7] Y. Sun, B. Xue, M. Zhang and G. G. Yen, "Evolving Deep Convolutional Neural Networks for Image Classification," in *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 2, pp. 394-407, April 2020, doi: 10.1109/TEVC.2019.2916183.

[8] Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation* 29.9 (2017): 2352-2449.

[9] Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1096-1104).

[10] Kalantidis, Y., Kennedy, L., and Li, L.-J. (2013). Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In Proceedings of the *3rd ACM conference on International conference on multimedia retrieval, pages 105– 112*. ACM.

[11] Akata, Z., Perronnin, F., Harchaoui, Z., & Schmid, C. (2013). Good practice in large-scale learning for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(3), 507-520.

[12] Chen, H., Gallagher, A., and Girod, B. (2012). Describing clothing by semantic attributes. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision – ECCV 2012, pages 609–623,* Berlin, Heidelberg. Springer Berlin Heidelberg.

[13] Shivakumara, Palaiahnakote, C. Pavan Kumar, Jagrut J. Nemade, Kshitiz Michael, Akash Kumar, Basavaraj S. Anami, and Umapada Pal. "A new U-Net based system for multi-cultural wedding image classification." Expert Systems with Applications 237 (2024): 121562.