CNN-Based Novel Deep Learning Ensemble Model for Kathakali Hand Mudra Recognition and Classification

SREENA. S[1], SHIBILI.T[2], AKHILA. M[3], LALJI CYRIAC[4]
Assistant Professor Dept. of CSE
College of Engineering, Vadakara, Kerala, India

***Abstract*—Kathakali, a classical dance form originating from Kerala, India, is characterized by intricate hand gestures (mu- dras), body movements, facial expressions, and background mu- sic. These mudras play a pivotal role in conveying emotions and storytelling, but their complexity often makes them challenging for non-experts to interpret. This paper explores the use of deep learning techniques, specifically Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) to automatically classify Kathakali hand gestures across five distinct classes. We investigate various methods for data preprocessing and classification, leveraging pre-trained models such as ResNet50, VGG16 and InceptionV3 for feature extraction. Our aim is to enhance the understanding and recognition of Kathakali mudras, making them more accessible to the public and aiding in their preservation.**

***Index Terms*—Feature extraction, Image classification, CNN, Deep learning**

I. INTRODUCTION

Kathakali, a classical dance form originating from the southern Indian state of Kerala is known for its intricate hand gestures, body movements, facial expressions, and background music. The complexity of its hand-gesture language makes it challenging for audiences to interpret the Kathakali mu- dras. This paper explores various approaches for recognizing Kathakali dance mudras performed by artists using deep learning techniques. The goal of this research is to investigate different methods for classifying Kathakali hand gestures. Due to the intricate nature of these mudras, they are often difficult for the general public to understand.

In this study, we propose the use of machine learning techniques for data preprocessing, followed by the application of Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) for the classification of Kathakali hand gestures across five distinct categories: Pathaaka, Mu- draakhyam, Katakam, Mushti, and Kartharee Mukham. Specifically, deep learning models, including pre-trained networks such as ResNet50, VGG16, and InceptionV3, are employed for their ability to automatically extract relevant features from images, offering significant advantages in terms of accuracy and efficiency. By using a deep learning ensemble model with ResNet50, VGG16 and InceptionV3, we achieved a 93% classification accuracy. Through the application of these deep learning techniques, this project aims to bridge the gap be- tween traditional art forms and modern technology, enriching both the practice and appreciation of Kathakali.

The related work is discussed in Section II, the problem statement is outlined in Section III, the proposed methodology is explained in Section IV, experimental results and discussions are provided in Section V, conclusions are drawn in Section VI, future work is discussed in Section VII, and references are listed in Section VIII.

II. RELATED WORK

The reviewed papers highlight various approaches to gesture recognition in Kathakali and other South Indian classical dance forms. Techniques like few-shot learning, data augmentation, and hybrid models combining traditional classifiers with CNNs are promising, especially when labeled data is limited. These methods advance the automation of Kathakali hand gesture recognition and help preserve traditional dance knowledge.

Kathakali and similar traditional art forms are not widely popular, resulting in limited technological research. Most existing works in this field build deep learning models from scratch, starting with the creation of large datasets—an inherent challenge. In contrast, our approach leverages minimal data and pre-trained models, offering a novel and efficient solution to the task of gesture classification.

In their 2020 work on Kathakali hand gesture recognition, Bhavanam et al. [1] (Bhavanam and Iyer, 2020) achieved 74% accuracy using Convolutional Neural Networks (CNNs) on a dataset of 654 images, each cropped to show a single hand. The dataset, published by Iyer and Bhavanam (2019), consists of 56x56 pixel color images. The authors also compared the performance of an SVM classifier on the same dataset, with CNNs outperforming SVM in accuracy.

In their 2023 study, Malavath et al. [2] (Malavath and De- varakonda, 2023) achieved 79% accuracy using CNN models on the same dataset. They also compared CNN performance with a Naive Bayes model.

In a recent study, Niveditha Parthasarathy et al. [3] (2023) published a benchmark dataset of 5 Bharatanatyam mudra classes and proposed a classification method using fine-tuned MobileNetV2 models. In offline tests with 6000 images, they achieved an accuracy of 86.45%. We have also included experiments on this dataset as part of our work.

Other works on Indian classical dance mudra recognition, such as Pradeep et al. [4] (2023) and Haridas et al. (2022), also use CNN models. Haridas et al. employed YOLO for Bharatanatyam mudras and reported 73% accuracy. Pradeep et al. evaluated multiple CNN models, with their approach—locating hands in larger images before classification—resulting in a lower accuracy of 50%. Similarly, Nandeppanavar et al. (2023) used VGG-19 and ResNet50V2, achieving 96.44% accuracy, following a similar deep learning- based approach, either fine-tuning or training from scratch.

Kavitha Raju et al. [5] addresses mudra recognition as a 24-class classification problem and introduces a novel vector- similarity-based approach that leverages pose estimation techniques. This method eliminates the need for extensive training or fine-tuning, effectively overcoming the challenge of limited data availability that is common in similar AI applications. Achieving an accuracy rate of 92%, our approach demonstrates performance that is either comparable to or surpasses existing model-training-based methods in this field.

Kavitha Raju et al. [6] treats mudra recognition as a 24-class classification task and proposes a vector-similarity- based approach using pose estimation, bypassing the need for extensive training or fine-tuning. Achieving 92% accuracy, it performs on par with or better than existing methods, while handling small datasets (1–5 samples) with slightly reduced performance. The system can process images, videos, and real- time streams, and works with both hand-cropped and full-body images.

Gayathri Vadakkot et al [7] uses eigen mudra projections, a novel technique, to capture the intrinsic patterns of hand gestures. They employ CNNs for feature extraction and classification, which has proven effective in recognizing the intricate and subtle variations of Bharatanatyam mudras.

Kalaimani M & Sigappi, A. N [8] explores the use of VGG (Visual Geometry Group) models for recognizing Bharatanatyam mudras. VGG models, which are well-known deep learning architectures for image classification, are lever- aged for mudra recognition in Bharatanatyam.

Saba Naaz1 & K.B. Shiva Kumar [9] propose an integrated deep learning approach for classifying Bharatanatyam mudras. The focus is on using convolutional neural networks (CNNs) and other deep learning techniques to automate the recognition of hand gestures.

K.V.V. Kumar & P.V.V. Kishore [10] focuses on an alternative approach for mudra recognition in Bharatanatyam by utilizing Histogram of Oriented Gradients (HOG) features, which capture the spatial structure of hand gestures, combined with a Support Vector Machine (SVM) classifier.

III. PROBLEM STATEMENT

Develop a deep learning model to classify Kathakali hand gestures or ”mudras” which are integral to this traditional Indian dance form. These gestures convey specific meanings and emotions, making accurate recognition essential for enhancing performance and training. The project aims to contribute to the preservation and promotion of Kathakali by providing tools for performers and educators using CNN and ANN-based deep learning techniques.

A. objective

By leveraging deep learning techniques, this project aims to bridge the gap between traditional art forms and modern technology, enriching both the practice and appreciation of Kathakali.

IV. PROPOSED METHODOLOGY

A. *CNN Architecture Models*

In this project, we use Convolutional Neural Networks (CNNs), a powerful architecture widely employed in computer vision tasks such as image classification and object detection. Specifically, we implement an ensemble model combining VGG16, ResNet50 and InceptionV3. This ensemble approach leverages the strengths of each model, improving accuracy and robustness in classifying Kathakali hand gestures.

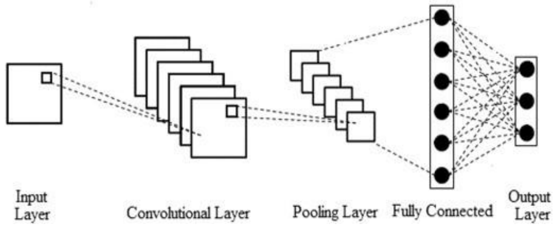

Fig. 1: Basic CNN Architecture

Preprocessing : An Image Data Generator is a crucial technique in deep learning, especially when working with image data. It is used to automatically augment and preprocess images during the training process.

- Data Augmentation: Image data generators allow for real-time data augmentation. This means the generator can modify the training images by applying random transformations such as:
  - Rotation : Rotating the image by random angles.
  - Flipping : Horizontally or vertically flipping the images.
  - Scaling : Zooming in or out on the image.
  - Translation : Shifting the image along the x or y axis.
  - Shearing : Applying a shear transformation to distort the image.
  - Brightness/Contrast adjustments : Randomly altering the brightness and contrast of images.

- Normalization : Image data generators normalize pixel values by scaling them to a consistent range (e.g., [0, 1]), improving model training. This accelerates convergence and helps the model learn more effectively by ensuring all input data has a uniform scale.
- Improved Robustness and Accuracy : Augmenting the data exposes the model to varied inputs, making it more robust and less prone to overfitting. This leads to better generalization and improved accuracy on unseen data.

Classification: In this work, we use various pre-trained Convolutional Neural Network (CNN) architectures to classify images of Kathakali Mudras (5 classes). An ensemble model combining ResNet50, VGG16 and InceptionV3 will be trained on the same dataset to evaluate performance in terms of accuracy and loss. The five Kathakali Mudras used for classification are Pathaaka, Mudraakhyam, Katakam, Mushti and Kartharee Mukham. This approach aims to identify the best-performing model for accurate gesture classification.

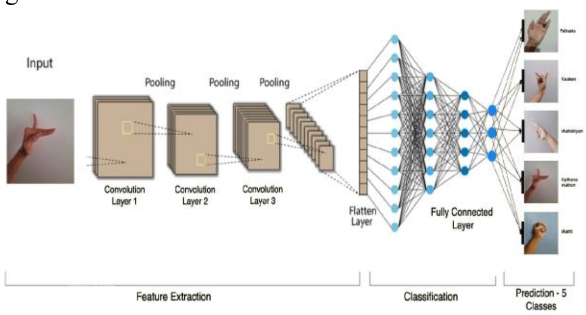

Fig. 2: Proposed Architecture

B. *Ensemble Model - Resnet50, VGG16 & InceptionV3*

The Ensemble Model combines the strengths of multiple deep learning architectures — ResNet50, VGG16 and Incep- tionV3 — to improve classification accuracy for Kathakali Mudra recognition. By leveraging these models together, the ensemble approach benefits from the individual strengths of each architecture:

- ResNet50 has 50 layers (49 convolution and 1 fully con- nected) excels in handling deep networks with residual connections, improving training efficiency.
- VGG16 consists of 16 layers of learnable parameters, including 13 convolutional layers and 3 fully connected layers offers a simple yet deep architecture, providing reliable feature extraction for image classification.
- InceptionV3 has 48 layers of learnable parameters (47 Convolutional Layers and 1 Fully Connected Layer) uses multi-sized convolutions, capturing features at different scales to enhance accuracy.

C. *Dataset*

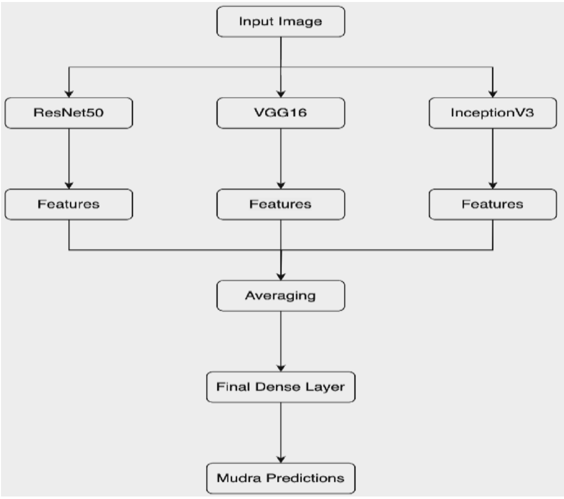

Fig. 3: Work Flow

A 5-class dataset is available on Kaggle for Kathakali mudras. It contains 5 mudras, namely 'mudrakhya', 'pataka', 'kataka', 'kartari mukha', and 'musti', which is a small subset of the complete set of 24 mudras. It contains 2 folders, Train and Test with 1147 images in Train folder and 100 images in Test folder. This dataset contains images with higher pixel clarity and more hand orientations. In certain image samples, due to the shape of the mudra and the hand orientation, not all fingers are visible and hence mudra is not clear, though. This dataset was also used and with an 80:20 split on the data for the training and the test sets, we applied our method for mudra classification.

Mudra Type	Train	Test
Kartharee Mukham	225	26
Katakam	223	21
Mudraakhyam	237	14
Mushti	217	22
Pathaaka	240	18

TABLE I: Dataset (Number of Samples)

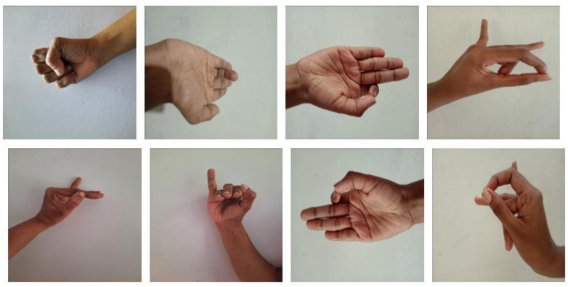
Fig. 4: Sample Images from Kathakali Mudra Dataset

V. EXPERIMENTAL RESULTS

A. *Evaluation Metrics for Classification*

The following equations describe the common evaluation metrics used to assess the performance of classification mod-els.

- True Positive (TP): Correctly predicted positive cases
- True Negative (TN): Correctly predicted negative cases
- False Positive (FP): Incorrectly predicted positive cases
- False Negative (FN): Incorrectly predicted negative cases

Accuracy measures the overall correctness of the model. It calculates the proportion of correctly predicted instances (True Positives, TP and True Negatives, TN) out of all predictions (including False Positives, FP and False Negatives, FN). Accuracy is a general indicator of the model's performance.

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

Precision evaluates the proportion of true positive predictions out of all the positive predictions made by the model. It answers the question: "Of all the instances predicted as positive, how many were actually positive?" Precision is useful when the cost of false positives is high.

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

Recall (or Sensitivity) calculates the proportion of true positives correctly identified by the model out of all actual positive instances. It measures the ability of the model to find all relevant positive cases. Recall is important when the cost of missing positive instances (false negatives) is significant.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1 Score is the harmonic mean of Precision and Recall. It combines both metrics into a single value, balancing the trade- off between them. F1 Score is particularly useful when dealing with imbalanced datasets, where one class is significantly more frequent than the other.

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

B. *Resnet50*

ResNet50, introduced by Microsoft Research in 2015, is a deep CNN designed to tackle the vanishing gradient problem using residual (skip) connections. With 50 layers, it enables efficient training of deeper networks. ResNet50 achieved 87% classification accuracy for Kathakali Mudra recognition.

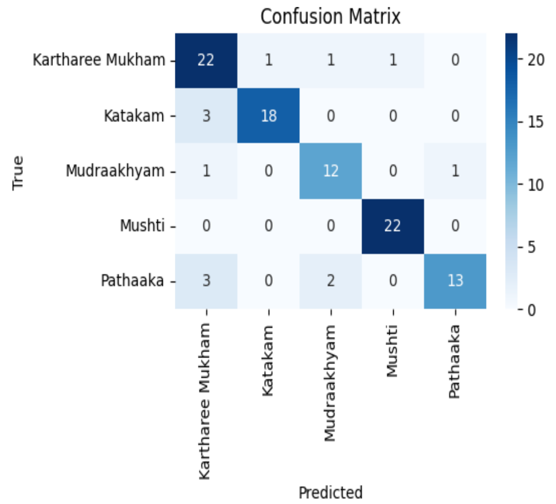

Fig. 5: Resnet50 - Confusion Matrix

Mudra Type	Precision	Recall	F1-Score	Support
Kartharee Mukham	0.76	0.88	0.81	25
Katakam	0.95	0.86	0.90	21
Mudraakhyam	0.80	0.86	0.83	14
Mushti	0.96	1.00	0.98	22
Pathaaka	0.93	0.72	0.81	18
Accuracy			0.87	100
Macro Avg	0.88	0.86	0.87	100
Weighted Avg	0.88	0.87	0.87	100

Overall Accuracy : 0.87

TABLE II: Classification Report - ResNet50

C. VGG16

VGG16 is a widely used Convolutional Neural Network (CNN) architecture, developed by the Visual Geometry Group (VGG) at Oxford University. Known for its simplicity and depth, VGG16 consists of 16 layers: 13 convolutional layers and 3 fully connected layers. This architecture has been effective in various image classification tasks and, in our case, achieved an 84% classification accuracy for recognizing Kathakali Mudras.

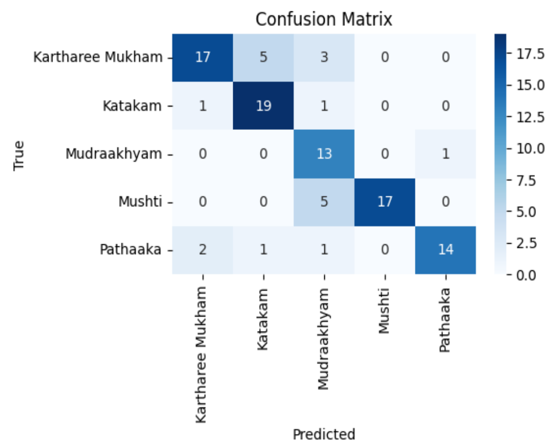

Fig. 6: VGG16 - Confusion Matrix

Mudra Type	Precision	Recall	F1-Score	Support
Kartharee Mukham	0.78	0.72	0.75	25
Katakam	0.78	0.86	0.82	21
Mudraakhyam	0.81	0.93	0.87	14
Mushti	0.95	0.95	0.95	22
Pathaaka	0.88	0.78	0.82	18
Accuracy			0.84	100
Macro Avg	0.84	0.85	0.84	100
Weighted Avg	0.84	0.84	0.84	100

Overall Accuracy: 0.84

TABLE III: Classification Report - VGG16

D. InceptionV3

Inception V3, part of the Inception (GoogLeNet) series introduced by Google, is a highly efficient deep learning model for image classification. It balances depth, computational cost and accuracy by employing techniques like factorization into smaller convolutions, asymmetric convolutions, and auxiliary classifiers. The architecture uses multi-sized convolutions (1x1, 3x3, 5x5) within the same layer to capture features at different scales. InceptionV3 achieved 80% classification accuracy for Kathakali Mudra recognition.

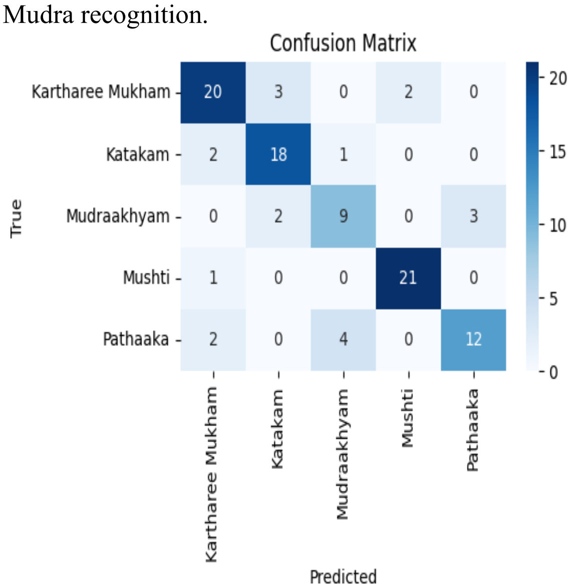

Fig. 7: InceptionV3 - Confusion Matrix

Mudra Type	Precision	Recall	F1-Score	Support
Kartharee Mukham	0.80	0.80	0.80	25
Katakam	0.78	0.86	0.82	21
Mudraakhyam	0.64	0.64	0.64	14
Mushti	0.91	0.95	0.93	22
Pathaaka	0.80	0.67	0.73	18

Accuracy			0.80	100
Macro Avg	0.79	0.78	0.78	100
Weighted Avg	0.80	0.80	0.80	100

Overall Accuracy: 0.80

TABLE IV: Classification Report - InceptionV3

E. Ensemble Model - Resnet50, VGG16 & InceptionV3

When combined in an ensemble model, the predictions of each architecture are aggregated, through averaging, leading to more accurate and robust results. This ensemble model achieved 93% classification accuracy, outperforming individual models and demonstrating superior performance in recognizing Kathakali Mudras.

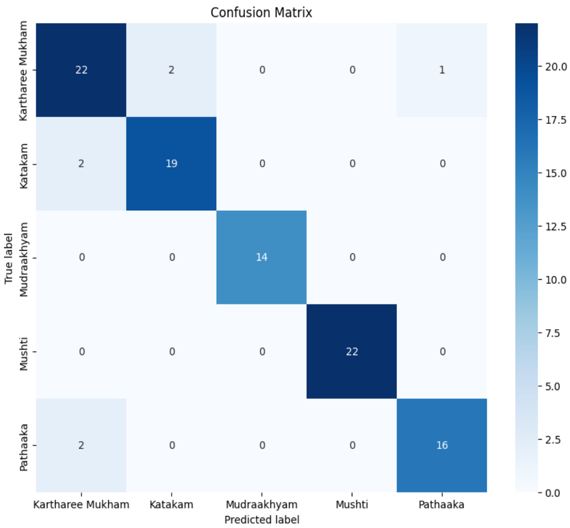

Fig. 8: Ensemble Model - Confusion Matrix

Mudra Type	Precision	Recall	F1-Score	Support
Kartharee Mukham	0.85	0.88	0.86	25
Katakam	0.90	0.90	0.90	21
Mudraakhyam	1.00	1.00	1.00	14
Mushti	1.00	1.00	1.00	22
Pathaaka	0.94	0.89	0.91	18
Accuracy			0.93	100
Macro Avg	0.94	0.93	0.94	100
Weighted Avg	0.93	0.93	0.93	100

Overall Accuracy: 0.93

TABLE V: Classification Report - Ensemble Model - Resnet50, VGG16 & InceptionV3

The following table compares the performance of ResNet50, VGG16, InceptionV3 and an Ensemble Model on a classification task. The Ensemble Model outperformed individual models with the highest accuracy (0.93) and balanced averages (Macro: 0.94, Weighted: 0.93). This demonstrates the effectiveness of combining multiple models for improved performance.

Model name	Accuracy	Macro Avg	Weighted Avg
ResNet50	0.87	0.87	0.87
VGG16	0.84	0.84	0.84
InceptionV3	0.80	0.78	0.80
Ensemble Model - Resnet50, VGG16 & InceptionV3	0.93	0.94	0.93

TABLE VI: Comparison of Models

VI. CONCLUSION

This study demonstrates the successful classification of Kathakali hand gestures using deep learning, specifically Convolutional Neural Networks (CNNs). By combining pre- trained models like ResNet50, VGG16 and InceptionV3 in an ensemble approach, we achieved 93% classification accuracy, showcasing deep learning's effectiveness in identifying Kathakali mudras. The findings underscore the potential of computational methods in preserving performing arts, particularly in making Kathakali gestures more accessible worldwide. This research paves the way for applications in dance education, interactive tools, and cultural preservation, effectively bridging traditional Indian art with modern technology and encouraging further innovation.

VII. FUTURE WORK

Using an ensemble of pre-trained models like ResNet50, VGG16 and InceptionV3, we achieved 93% accuracy in Kathakali hand gesture recognition, demonstrating the potential of deep learning for gesture recognition in classical Indian dance. Future work will focus on incorporating facial gesture recognition, improving efficiency with models like Efficient- Net and exploring multimodal recognition (body movements and voice). Expanding the dataset to include more dance forms will enhance generalizability and support the preservation of traditional art.

REFERENCES

[1] Lakshmi Tulasi Bhavanam and Ganesh Neelakanta Iyer. 2020.On the "Classification of kathakali hand gestures using support vector machines and convolutional neural networks".In 2020 International Conference on Artificial

Intelligence and Signal Processing (AISP), pages 1–6. IEEE.
[2] Pallavi Malavath and Nagaraju Devarakonda. 2023.”Classification of kathakali asamyuktha hasta mudras using naive bayes classifier and convolutional neural networks”.In Applied Computing for Software and Smart Systems: Proceedings of ACSS 2022, pages 131–144. Springer.
[3] Niveditha Parthasarathy and Yogesh Palanichamy. 2023.”Novel video benchmark dataset generation and real-time recognition of symbolic hand gestures in indian dance applying deep learning techniques”.ACM Journal on Computing and Cultural Heritage, 16(3):1–19.
[4] Radhika Pradeep, R Rajeshwari, VR Ruchita, Radhika Bubna, and HR Mamatha. 2023.”Recognition of indian classical dance hand gestures”.In 2023 International Conference on Inventive Computation Technologies (ICICT), pages 814–820. IEEE.
[5] Kavitha Raju, Nandini J. Warrier , ”Pose2Gest: A Few-Shot Model-Free Approach Applied In South Indian Classical Dance Gesture Recognition, May 2024” -, IIIT-Kottayam
[6] Kavitha Raju, Nandini J. Warrier, ”Kathakali Hand Gesture Recognition With Minimal Data, April 2024” - , IIIT-Kottayam
[7] Gayathri Vadakkot, Karthi Ramesh, and Govind Divakaran ”Auto- matic one-hand gesture (mudra) identification in bharatanatyam using eigenmudra projections and convolutional neural networks,” Journal of Electronic Imaging 32(2), 023046 (27 April 2023).
[8] Kalaimani M & Sigappi, A. N. (2023).” Implementation of VGG Mod- els for Recognizing Mudras in Bharathanatyam Dance”. International Journal of Intelligent Systems and Applications in Engineering, 12(3s), 306–319
[9] Saba Naaz1 & K.B. Shiva Kumar (2023) . ”Integrated deep learning classification of Mudras of Bharatanatyam: A case of hand gesture recognition”. The Scientific Temper (2023) Vol. 14 (4): 1374-1380
[10] K.V.V. Kumar, P.V.V. Kishore (2017) ”Indian Classical Dance Mudra Classification Using HOG Features and SVM Classifier”. (IJECE) Vol. 7, No. 5, October 2017, pp. 2537 2546 ISSN: 2088-8708, DOI:10.11591/ijece.v7i1.pp2537-2546