SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I2P105

Volume 13 Issue 2, 53-67, February 2026
© 2026 Seventh Sense Research Group®

Original Article

Image Processing-Based Recognition of Sattriya Dance Hand Mudras using a Hybrid CNN-SVM Model for Pattern Recognition in Signal Processing Applications

Chayanika Sarmah[1], Parismita Sarma[2], Dankan Gowda V[3], Pooja Singh[4], Naveen B[5], Anil Kumar N[6]

[1,2]Department of Information Technology, Gauhati University, Guwahati, India.
[3]Department of Electronics and Communication Engineering, B.M.S. Institute of Technology and Management, Bangalore, India.
[4]Department of Physics, School of Basic and Applied Sciences, SGT University, Gurugram, Haryana, India.
[5]Department of Electronics and Communication Engineering, BGS Institute of Technology Adichunchanagiri University, BG Nagar, Karnataka, India.
[6]Department of Electronics and Communication Engineering, School of Engineering, Mohan Babu University, Tirupati, Andhra Pradesh, India.

[2]Corresponding Author : pari@gauhati.ac.in

Received: 16 November 2025 Revised: 18 December 2025 Accepted: 19 January 2026 Published: 20 February 2026

Abstract *- Classical dances like Sattriya dance have much cultural and artistic importance, and all the gestures carry a specific meaning or meanings in their forms of hand mudra. There is, however, an exception to this rule in terms of automatic recognition of these hand mudras owing to the different positions of the hands, lighting, and background. In this paper, a hybrid model of Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) has been introduced for the recognition of a single-hand Sattriya dance mudras. The CNN model has the capability of inducting the high-level features in pictures of hand gestures and classifying them through the use of an SVM classifier. A complete dataset of 13 different hand mudras with variations added to their original forms to increase variations was used to train and test the model. Most classes showed high specificity with a precision score, recall score, and F1-score of the model of almost 1.0, meaning the power of the model to classify complicated mudra dance. The offered model shows how AI can be used to automate the identification of traditional dance movements, which can provide perspectives on digitization and preservation of the cultural art.*

Keywords *- Sattriya dance, Hand mudras, Image Processing, Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Gesture recognition, Deep Learning, Classification, Cultural preservation, AI.*

1. Introduction

A central integral body in the art forms of classical Indian dancing is the hand mudras, its detailed hand gestures, with each gesture rich in its symbolism. These mudras are respected in Sattriya dance, which is one of the ancient forms of classical dance in Assam, India. It is a dance form that is culturally significant and deeply rooted in Vaishnavism; it is also used as a form of spiritual expression.

Sattriya dance hand mudras can be used to express different kinds of divine and emotional feelings that are reflected in each gesture, build up to the rhythm, emotion, and the theme of the Sattriya dance [1]. Sattriya has an assortment of 29 single-hand mudras and 12 two-hand mudras, each with a distinct function, whether depicting a specific deity, a phenomenon of nature, or a story. Along with being a visual representation, these hand gestures are also a spiritual teaching and are vital to the art form. Learning and practicing these mudras has been a manual and human-intensive approach where dancers have been extensively dependent on teachers to have the correct posture and motion. Due to the ever-changing nature of the cultural art form, the digital era poses challenges and opportunities [2]. The acknowledgment, conservation, and teaching of these mudras, particularly in the realm of the contemporary digital depositories and e-learning platforms, is one notable challenge [3].

The incredible number of variations in hand movements, hand sizes, lighting conditions, and movement locations makes it hard to identify and define these movements manually. Besides, some classical dance styles, such as Sattriya, can involve a sincere knowledge of delicate body movements and gestures, which cannot be perceived or understood adequately by human observers [4].

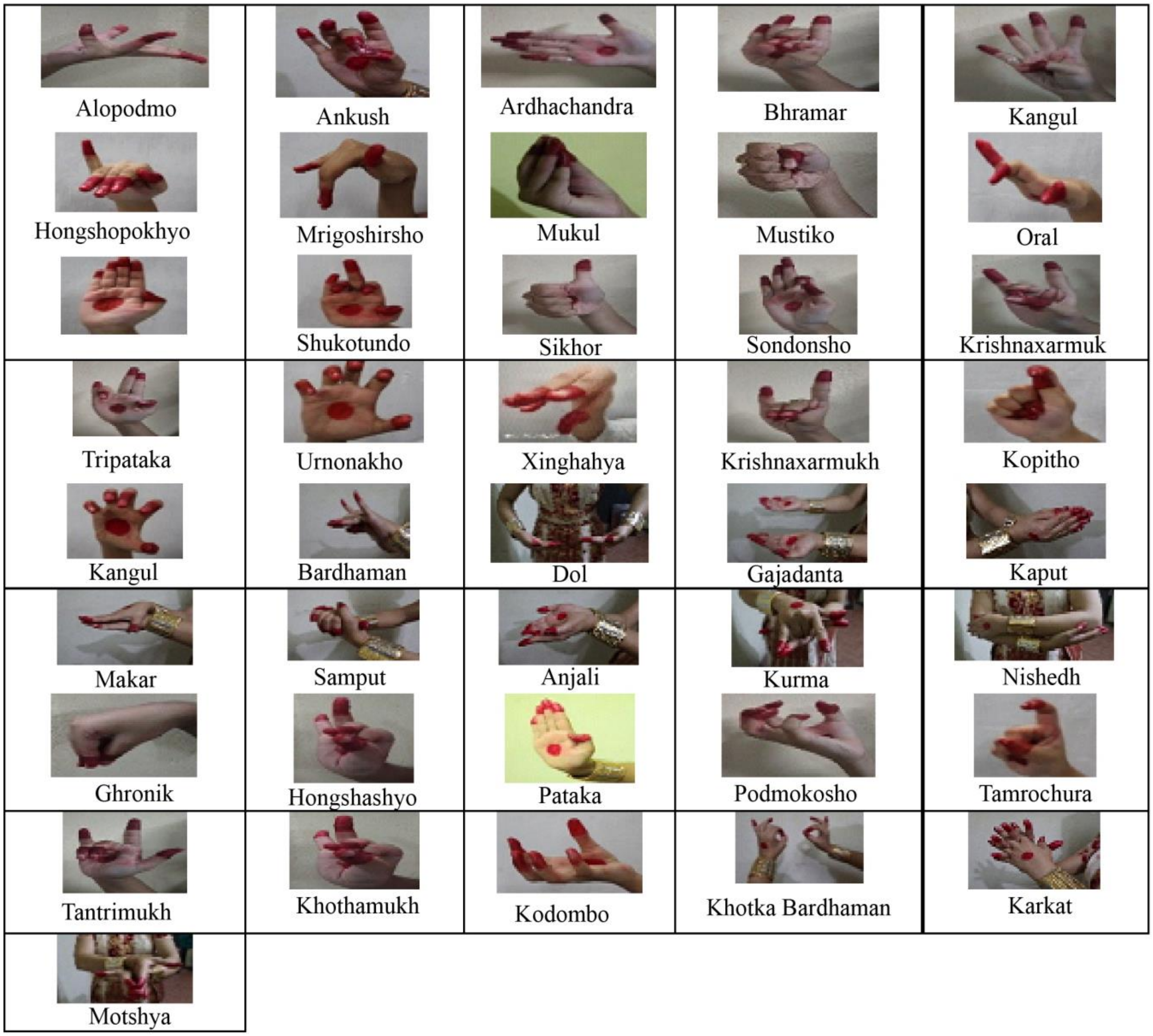

Fig. 1 Single-hand sattriya mudras used in the experiment

As shown in Figure 1, this study has used the 13 individual single-hand mudras, which are used in the recognition task. Every mudra is an important part of the Sattriya dance style that has a strong cultural and spiritual symbolism [5]. Such mudras, which are carried out with hand positioning accuracy, are photographed to form the main information, on which the hybrid CNN-SVM model is going to be trained and tested [6]. There are differences in the hand position, order of fingers, and general shape of hands; these pose a challenge to identification since they have tiny details and resemblances alike [7]. Against this backdrop, there is an imminent problem of the necessity of a sophisticated solution that will be able to process the hand gesture recognition process automatically [8]. The use of traditional manual means makes it hard to expand Sattriya dance training and performance to a broader audience, both in India and worldwide [9]. To eliminate these problems, one potential solution can be to apply AI and machine learning models, especially in the area of image processing [10]. Automatization of the hand mudra recognition and classification not only guarantees increased accuracy but also creates the possibility of digital platforms providing immediate feedback, the preservation of rare dance forms, and the possibility of teaching future dancers without physical contact [11]. Equal to the purpose of Sattriya dance but innovative: the implementation of an AI-driven mechanism to detect gestures during the dance would allow a more efficient and convenient method of learning and maintaining these cultural traditions [12].

Hand gesture recognition, especially when it comes to classical dance, such as Sattriya, poses special problems to this technique of recognition because of the delicacy and exquisiteness of the moves that are involved [13]. The complexity of gestures as they are is one of the main problems in gesture recognition. These gestures may resemble each other considerably and demand a subtle sense of finger and palm location, as well as proper wrist positioning [14]. Many more changes in lighting and background, even the physical appearance of the dancer, make the task even more complicated, and it is hard to use the traditional computer vision algorithms to be able to categorize these mudras with high consistency. Besides, the necessity to differentiate between subtle nuances and the difference in performance makes manual recognition a tiresome and unstructured task.

Another problem in recognizing hand gestures is that dance is dynamic. This is contrary to still images of people; in dance, the movement is continuous, and the mudras are not captured in one frame, but it is a sequence of movements. This adds time variability to the recognition process since the model would have to effectively follow and categorize gestures in real-time, taking into account movement in several frames of a video. The classical image recognition algorithms are prone to fail such tasks because they are generally programmed to recognize static objects and not dynamic gestures.

The downsides of traditional solutions regarding their ability to distinguish hand mudras in dance emphasize the necessity of a sophisticated, automatic solution [15]. The hybrid proposal of combining CNNs with SVMs is a promising approach. CNNs perform well at learning hierarchies in images, and they are therefore ideal for projecting features of hand gesture images. The other classifier is the SVMs, which are powerful and best at finding dimensions in feature space, hence are best suited to identify minor differences in complex hand mudras. Using the two methods together, there is a possibility of developing a model that can not only learn positively to identify the different gestures of the hand automatically, but also under different conditions, like lighting, background noise, and movement.

This study aims to design and test a hybrid CNN-SVM model for body gesture recognition in the Sattriya dance. The reason is that the proposed model will attempt to precisely categorize 13 various single-hand mudras by deriving pertinent characteristics using a CNN and classifying the results using SVM. The dataset will comprise images based on video recording captures of the Sattriya dance rituals, so that the model can accommodate real-life dynamic conditions. High accuracy, high precision, and high recall are also the intended goals of the hybrid model because the recognition system would be reliable to rely on even in education and archival purposes. The larger objective is to introduce greater accessibility and sustainability of the Sattriya dance using AI, so that future generations can interact with and acquire this cultural legacy in innovative and novel methods.

2. Literature Review

Recognition has been well known in computer vision, especially human-computer interaction, sign language translators, and in the preservation of the culture. Different methods to deal with the obstacles to the accurate recognition of hand gestures have been developed in the last couple of years, especially in dynamic and complicated situations such as the classical dances [16]. The former has deployed diverse strategies, including the conventional image processing frameworks or more sophisticated machine learning and deep learning solutions, specifically, the CNN and the SVM. This part will analyze the available literature on gesture recognition, which encompasses the approaches used to identify dance mudra [17], or the use of CNN and SVM in processing images, or prior research in hand gesture recognition, with a view to defining the gap that the proposed research intends to fill.

In dance styles like Bharatanatyam, Odissi, and Kathak, hand gesture recognition has been studied in a number of studies. A sample of such research is an article by Saba Naaz et al. (2023), which aims to identify Bharatanatyam mudras with the deep learning methodology, and the authors used skeleton-based approaches to classify hands [18]. The authors employed the thinning process of extraction of the skeletal features and integrated fuzzy membership functions with deep learning to obtain a high accuracy of 96%.

This paper has shown that Deep Learning has the potential to classify mudra, but has not been able to tackle the issue of dynamically changing gestures, which is a major issue when recognizing dance mudras. Otherwise, another study conducted by Pravin R. Futane (2011) used the Gaussian functions and edge detectors to identify the Indian sign language gesture, which serves as a point of reference in the context of the recognition task of gestures under the umbrella of gesture recognition. These methods were also effective, but they tended to have problems with moving hands and the process of preprocessing, namely, removing the backgrounds and optical flow estimation.

The application of CNNs has been among the most significant developments in gesture recognition, and it has been primarily utilized in image classification tasks as they enable automatic extraction of hierarchical features of raw image data. The CNNs have been broadly utilized in most image processing applications, such as facial recognition, medical image processing, and lately, hand gesture recognition. To illustrate, researchers used CNNs in the static image and video frame hand gesture recognition. They found that they could acquire low-level (e.g., edges and textures) and high-level (e.g., hand shape and orientation) aspects of the semantics [19]. Nevertheless, to achieve such applications, CNNs have proven to be quite challenging in managing dynamic gestures because they mainly emphasize the spatial aspect and overlook the temporal nature of the gesture sequences. The application of CNNs in combination with other Machine Learning systems has been pursued to address specific weaknesses of individual CNN models.

SVMs are supervised learning algorithms that are very good at categorizing high-dimensional data by finding an optimal separating plane, which maximizes the difference between the various classes. SVMs are most commonly utilized as a classifier in the context of gesture recognition, as their feature extractors are CNNs. Various researchers have demonstrated that CNNs + SVMs can significantly enhance the performance of gesture recognition, especially when one is to perform recognition in a large and complicated dataset. As a case in point, M. Kalaimani et al. (2022) suggested a hybrid CNN-SVM to detect gestures in the Bharatanatyam mudras, which received an accuracy score of 96 percent when the Humming distance operator was paired with a Deep Learning Model.

Likewise, Basavaraj S. Anami and Venkatesh A. Bhandage (2019) used SVMs on top of neural networks to classify mudras and proved that complex, high-dimensional feature vectors generated out of images can be handled with the SVMs. Special challenges associated with the dynamic recognition of hand gestures are considered despite all these developments, and they are pronounced explicitly in the context of the classical forms of dance. Most of the problems with existing methods are that current methods look at the fixed features of images, which are not as dynamic as to capture the time dynamics of the dance performances. As a case in point, SVM classifiers are effective in high-dimensional spaces, but tend to be poor at capturing finer differences of gestures with time, particularly those gestures that are not well defined in a single frame. Also, most of the available literature on recognizing gestures in dance mudras refers to a particular type of dance. It can hardly be generalized to other traditional dances like Sattriya, where a unique indicative set of gestures has subtle variations [20]. Moreover, these researchers usually fail to recognize the complexity of real-life situations, which include fluctuation of light, background noise, and some personal differences in the size of their hands or speed of their performances.

The research gap that has been recognized among the existing literature is that the existing literature does not put together the spatial feature detection capabilities of CNNs with the temporal and high-dimensional classification capabilities of SVMs, in the exact context of dynamic dance mudra recognition. Moreover, an efficient system is required that is able to tolerate the fluctuations in the environmental conditions as well as differences in individual performance levels, while still maintaining high tolerance and effectiveness. By creating a hybrid model of CNN-SVM that employs both spatial and temporal cues, i.e., feature extraction (CNNs) and classification (SVMs), the proposed research will fill these gaps. The model is designed to address the complexity of the Sattriya dance hand mudras, providing an automatic, accurate recognition system that can be used in diverse real-world settings, such as educational systems and digital archives. The study will address the shortcomings of existing gesture recognition solutions in order to contribute to the field of AI-based cultural heritage preservation and dance education by making a big leap in this field.

3. Proposed Method

The hand mudra recognition methodology in the Sattriya dance comprises a series of steps, from data preparation to the final analysis of the resulting model. This part presents the step-by-step process of creating the datasets and preprocessing the images, the CNN architecture, image feature extraction, classification with Support Vector Machine (SVM), and the performance metrics used to analyze the hybrid CNN-SVM model.

3.1. Dataset

Hand mudra recognition in Sattriya dance has a dataset consisting of 13 unique one hand mudras. It is important to note that these mudras are symbolic symbols or concepts and have been the key elements of Sattriya dance, and the correct identification of these mudras is a crucial step in the conservation of this cultural heritage. In this research, 13,000+ photos are utilized, and each of the mudras is photographed at different angles and light conditions. These images are taken from videos of dance performances so that there is dynamism in the gestures in the dataset. All the images in the dataset are tagged based on the respective mudra category. To increase X-ray diversification and strength, a collection of data augmentation methods is implemented. These are zoom, rotation, shifting, shearing, and brightness effects. These additions help create variations in the dataset, enabling the model to learn invariant features and become more robust to typical variations in real-world conditions.

Artificial expansion of the data improved the chances of not overfitting, and the model became more capable of generalizing. Over 18,000 images are added to the augmented dataset, comprising all of the 13 classes of mudra, which is a sufficient amount of training data to use the CNN model. The process of the hybrid CNN-SVM model, which is proposed to detect the hand mudra in Sattriya dance, is presented in Figure 2. The methodology commences by preparing labeled data, which is then followed by parting of data into training, testing, and validation data. A CNN model consisting of 128 neurons and ReLU activation is then created and applied to extract feature vectors of the hand mudra images. Such characteristics are then trained on an SVM classifier. When computing precision, recall, and F1 score, the model accuracy is computed. The trained model then recognizes and classifies the given image of the mudra and gives a predicted output and the value of the precision.

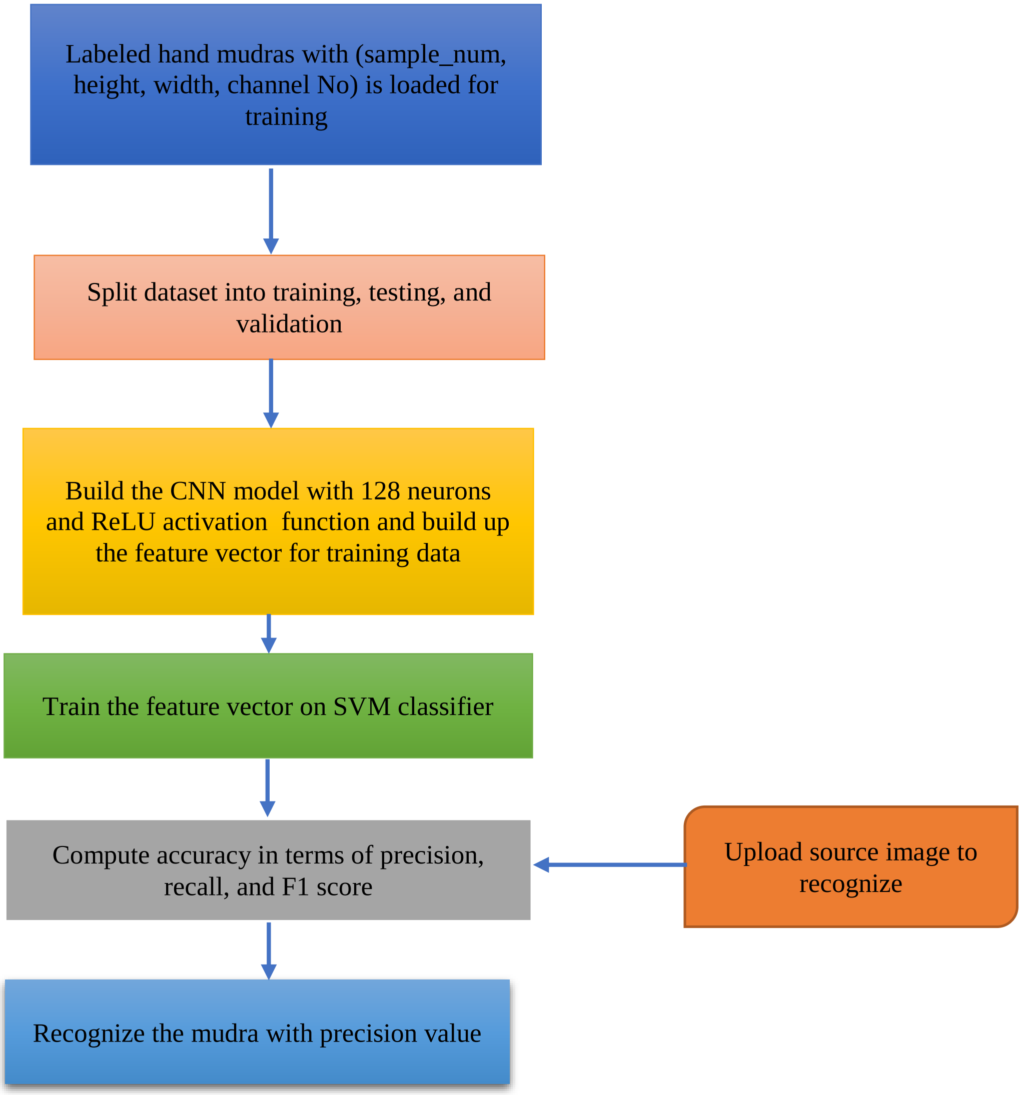
Labeled hand mudras with (sample_num, height, width, channel No) is loaded for training
Split dataset into training, testing, and validation
Build the CNN model with 128 neurons and ReLU activation function and build up the feature vector for training data
Train the feature vector on SVM classifier
Compute accuracy in terms of precision, recall, and F1 score
Upload source image to recognize
Recognize the mudra with precision value

Fig. 2 Proposed methodology for hand mudra recognition

The process places more importance on combining deep learning and machine learning methods to automate the detection of complex hand gestures, that is, in the case of the Sattriya dance. Our experiment represented a situation where the data were recorded in video form when captured. Thirteen (13) most notable single-hand mudras of sattriya dance have been considered. The rest were taken with the assistance of a DSLR camera whose recording time ranged between 20 and 30 seconds. Each of the mudras had at least three videos, and the most salient sections were taken into consideration to be divided into frames of images. Table 1 below indicates the sample size of the mudrasses in question.

Table 1 presents the frequency distribution of thirteen (13) hand mudras. The above-mentioned raw images were augmented with a few methods, and finally, we obtained a large quantity of samples having different sizes, contrasts, orientations, etc. The most intrinsic augmentation manipulations employed in this case are zooming, rotation, movement, shearing, brightness control, etc.

3.2. Image Preprocessing

Preprocessing of images is very important in preparing the data to facilitate training. To provide uniformity throughout the dataset, all the raw images are scaled to a standard resolution of 150x150 pixels. Standardization of pixel values by normalization is done to bring them to a range between 0 and 1. This will enhance the rate of convergence in model training, since all the input features will be of comparable scales, and the dominance of one feature will not occur over the rest.

Besides, the data is divided into three subsets: training (64%), validation (16%), and testing (20%). This division is done to make sure that the model is trained on a significant section of the data, and there is also sufficient unexplored data to be assessed. The model is taught using the training set, hyperparameters are adjusted with the help of the validation set, which is also used to verify the performance of the model in the course of the training, and the final evaluation is done on the test set. The prudent allocation of the data set involves efforts to test the model on data that it has never seen during training, and this gives an accurate evaluation of the generalization capacity of the model.

Table 1. Frequency table for the mudra samples

Sl no	Mudra	Number of samples
1	Alopodmo	895
2	Ankush	1038
3	Ardhachandra	1044
4	Bhramar	658
5	Chatur	864
6	Ghronik	1051
7	Hongshashyo	1129
8	Kangul	1022
9	Kodombo	1153
10	Kopitho	1009
11	Krishnaxarmukh	865
12	Mrigoshirsho	1226
13	Mukul	946

3.3. CNN Architecture

The CNN architecture of feature extraction is developed such that it will learn to learn features and patterns with high efficiency in the hand mudras images. The architecture has three convolutional layers, which are then succeeded by max-pooling layers to downsample spatial rigidity and group suspected features to the fore. The initial convolutional layer will include 32 filters with a kernel size of 3x3. non-linear play. This is where the ReLU activation function is applied to get more complex patterns to learn in the model. The second convoluted layer contains 64 filters, again, with a 3x3 kernel size, thus capturing additional higher-level details of the mudras. The third convoluted layer will have 128 filters with a 3x3 kernel. It is this layer that identifies even more advanced features in the pictures.

The convolutional layers are followed by max-pooling layers, which pursue downsizing of spatial dimensions by a 2x2 pool size and follow up on the computational load. Following the convolutional layers, we then do global average pooling to decrease the dimensions of the feature maps further; as such, the final result is a fixed-size feature vector regardless of the size of the input image. The developed feature vector is then run through a fully connected 128 neuron ReLU activation layer that further acts as a refinement to the learned features. The CNN is actually tailored to attain hierarchical representations of images, beginning with low-level image features (edges and textures) and concluding with high-level, abstract representations of hand gestures. The obtained feature vector is then fed to the SVM classifier.

3.4. Feature Extraction and Classification

The CNN model is highly important for feature extraction in the proposed methodology. The convolutional layers learn features automatically that are present in the input images without manually learning the features. These features portray the critical elements of the hand mudras, e.g., the finger and palm positions, the contours, and the shapes, which are essential in differentiating among the various mudras. Upon the extraction of the features by the CNN, they were flattened into a one-dimensional feature vector. The SVM classifier, in its turn, is being fed with this very vector and is tasked with classifying the mudras by one of the 13 categories, which have been predetermined. SVM is a training algorithm that operates on the basis of identifying an ideal hyperplane that separates the various classes in a high-dimensional space. It works exceptionally well in operation with high dimensions of feature vectors, so it is a good fit to this task, in which the CNN generates a rich collection of features per image. SVM classifier employs a linear kernel in dividing the feature vectors of the various mudra classes. The SVM can influence the correct class of mudra of the image by having the knowledge of the decision boundary within the feature space, and depending on the features that have been extracted.

3.5. Evaluation Metrics

To assess the hybrid CNN-SVM model, the following key metrics are used: accuracy, precision, recall, F1 score, and the confusion matrix. The most straightforward metric is accuracy, which is the ratio of correct predictions and the total number of predictions. Although accuracy gives a general idea of the overall performance of the model, it may be misleading, especially in situations where the dataset is imbalanced, and some of the mudra classes are overrepresented. To deal with this, precision is employed, and it is a measure of the percentage of correct optimistic predictions. High accuracy reflects that the model is incorrectly classifying fewer mudras as being positive, implying that it is more valid in classifying those mudras as positive. Conversely, recall is a measure of the accuracy of the number of actual positives that the model has identified successfully. When the recall score is high, it indicates that the model is indeed identifying the majority of the true positives, hence it is more sensitive when identifying the mudras.

By providing a balanced perspective of the model, the F1 score (harmonic mean of precision and recall) can be used. Proficiency is particularly handy when there is a necessity to strike a balance between the significance of both false positives and false negatives, so that neither precision nor recall is prioritized. As well, the confusion matrix offers a more detailed analysis, as it indicates the true positive, false positive, true negative, and false negative outcomes of each class. Using this matrix, it becomes easier to find out which classes of mudra the model is not good at differentiating, and it would provide a clue as to what should be done to improve the situation. Combining these evaluation metrics would give a thorough evaluation of the effectiveness of the hybrid model in identifying and categorizing hand mudras to ensure that it works consistently in different conditions and can give relevant predictions to be used in the real world.

4. Algorithm to Classify Sattriya Dance Hand Mudras

The algorithmic steps used for the classification of single-hand Sattriya dance mudra using a hybrid SVM-CNN-based model

Dataset loading:

Hand mudra and their labels are loaded on the x and y axes, respectively. Images here are identified as (sample_num, height, width, and channels). So these samples are categorical integers that represent the mudra class. Dataset Splitting: Dividing the samples to training (64%), validation (16%) and testing (20%) sets.

Define CNN Feature Extractor Build a CNN model with,

- 32 filters Conv2D layer, 3×3 kernel size, Activation function: ReLU
- MaxPooling: Pool size (2×2)
- 64 filters Conv2D layer, 3×3 kernel size, Activation function: ReLU
- MaxPooling: Pool size (2×2)
- 128 filters Conv2D layer, 3×3 kernel size, Activation function: ReLU
- Global Average Pooling, which reduces the spatial dimension
- 128 neurons Dense layer with ReLU activation function.

The feature vector is produced.

4.1. Feature Extraction using CNN

The feature vector is achieved by passing training, validation, and testing samples through a CNN model. This is the descriptor of the most prominent patterns of each image.

4.2. Train on Support Vector Machine (SVM)

CNN extracted features are used to train the SVM classifier, accompanied by a linear kernel. The SVM learns the CNN extracted features to classify the mudras.

4.3. Model Evaluation

Label prediction on training, validation, and test features with the help of trained SVM. Compute the accuracy score. Generate a classification report on the test set showing precision, recall, and F1 score.

4.4. Deliver Result

Display the accuracy table and deliver the result.

5. Mathematical Model

The hybrid CNN-SVM system of hand mudra recognition is mathematically modeled in two main parts, including CNN to extract features and SVM to classify the features. The CNN is the initial stage in the model, and it conducts image processing to find out hierarchical attributes of the input mudra images. The CNN consists of a couple of convolutional layers, then pooling and fully connected layers, with the main mathematical tasks of convoluting and activating, as well as the pooling. These processes allow the model to first learn low-level details like edges and textures and then learn high-level details like the shape and location of the hand, which are important to the process of detecting the various mudras.

Mathematically, the convolution operation in a CNN can be represented as:

$$Y_{i,j} = \sum_{m} \sum_{n} X_{i+m,j+n} W_{m,n} + b \qquad (1)$$

Where:

$X_{i,j}$ represents the input image matrix (the original image pixels).

$W_{m,n}$ represents the convolution filter (also known as a kernel).

b is the bias term.

$Y_{i,j}$ represents the output of the convolution operation, a feature map.

The CNN also uses the ReLU activation function to introduce non-linearity into the model. ReLU is mathematically expressed as:

$$f(x) = \max(0, x) \qquad (2)$$

This operation works with the negative values, which give a zero at the output, and positive values, which give the input values, so that the network can learn more complex patterns. Following every convolution, one uses a max pool to downsample the feature maps in terms of spatial components. The pooling operation, which is usually a 2x2 filter, picks the highest value in every patch, hence minimizing the computational demand and maintaining significant spatial content. Mathematically, pooling can be described as:

$$Y_{i,j} = \max\left(X_{i,j}, X_{i+1,j}, X_{i,j+1}, X_{i+1,j+1}\right) \qquad (3)$$

The second step (or flattening) of the model entails flattening the obtained feature map of the convolutional layers. This involves transforming the 2D feature maps into a 1D vector, and this is the input into the SVM classifier. The flattened vector is the most salient representation of the features extracted by the CNN, and it is applied to the SVM to be classified.

The SVM tries to identify a hyperplane that will have the best separation of the feature vectors of various classes in the high-dimensional space. This hyperplane search problem may be written as an optimization problem:

$$\min_{w} \frac{1}{2}\|w\|^2 \tag{4}$$

Subject to the constraint:

$$y_i(w.x_i + b) \geq 1, \forall i \tag{5}$$

Where:
x_i represents the feature vector of the i^{th} training example.
y_i is the label of the training example, taking values $+1$ or -1 for two classes.
w is the weight vector that defines the hyperplane.
b is the bias term.

The objective function minimizes the norm of the weight vector $\| w \|$, ensuring the maximum possible margin between the classes.

In non-linear classification, the SVM uses the kernel trick to project the features of the input into a higher-dimensional one where linear classification can be made. This higher-dimensional space can be computed as an inner product by the kernel function $K(xi, xj)$, and the decision boundary is therefore more cost-effective to compute. A commonly used kernel is the Radial Basis Function (RBF) kernel**,** which is defined as:

$$K(xi, xj) = \exp(-\gamma\|xi - xj\|^2) \tag{6}$$

Where γ is a parameter that controls the width of the Gaussian function, and xi and xj are feature vectors from the training set. After the SVM is trained on the extracted features, it can classify new images of hand mudra by running them through the same CNN model, then comparing the feature image to the feature space to figure out which side of the decision boundary the feature vector lies on. The output eventuality is the predicted class as well as the confidence score, which is used to show the accuracy of the prediction.

6. Results and Discussion

6.1. Training Results

Accuracy and loss curves are used to assess the model performance of the training process, and they offer an insight into the learning dynamics of the model across the epochs. The accuracy curve plots the ratio of correctly labelled instances in the training and validation sets with 30 epochs. The accuracy of the training is constantly increased, and the training achieves near 100 percent accuracy towards the successive epochs, indicating that the model has learnt the patterns present in the dataset. Equally, the validation accuracy also assumes a similar pattern, which means that the model is not overfitting the training data, but it is easily generalizing to unknown data. The loss curve that monitors the values of the loss function over time follows a similar pattern of decreasing loss over the training and validation sets. The training loss has a sharp decline during the first few epochs, whilst it remains constant when the model is near a stopping point. The loss on validation also decreases steadily, which demonstrates that the model is able to give correct predictions and minimize errors. The curves show that the model learns in a practical way and that it tends to an optimal solution as time elapses, which makes the appropriateness of the hybrid CNN-SVM methodology in the task.

After training our samples on the proposed model, it is observed that a very good training and recognition rate is achieved for all the considered mudras of Sattriya dance. Figure 3 shows the accuracy curve of our proposed hybrid model. Figure 4 shows the model's loss curve. From these two curves, it is clear that the accuracy of the model is close to 100%, irrespective of any particular mudra.

6.1.1. Confusion Matrix

Figure 5 shows the model's confusion matrix. All thirteen mudras show excellent efficiency, as we are aware that non-diagonal elements in a confusion matrix mean misclassification. However, according to the theory, a perfect confusion matrix may indicate some other issues in the model. The good possibility is that the dataset may be clear and well-balanced. Test data is quite independent. However, the opposite condition may also occur. For example, test data may have evidence from training data, or the classes may be severely skewed.

The classification report of the hybrid SVM-CNN model is presented in Table 2, which has the evaluation metrics of every type of hand mudras in the Sattriya dance recognition activity. It encompasses the accuracy, recall, F1-score, and support on each of the classes. Precision is defined as the ratio of the accurate optimistic predictions to the total number of predictions made in a given class. Recall is used in reference to the capability of the model to recognize every significant occurrence of a given mudra. The harmonic mean of precision and recall is referred to as the F1-score, which is a balanced method of evaluating the accuracy of the model in identifying each class of mudra.

The column of support means the number of occurrences of each class in the data set. Table 3 illustrates the accuracy values of the hybrid CNN-SVM model in each of the classes of hand mudras in detail, providing the values of precision, recall, and F1-score. The table also gives the support of each of the classes, which would show the number of instances of each of the mudra classes within the dataset. The accuracy and recall are very high, as indicated in the results, with most of the classes of the mudras recording values close to one, which means that the model is very effective in the correct classification of the mudras. The F1-score, a balanced measure of precision and recall, is also very excellent as the values of the F1-score are almost equal to 1.0. The table is a simplified and quantitative assessment of the classification performance of the model with each and every mudra, and this shows the capacity of the model to give close to perfect classification outcomes.

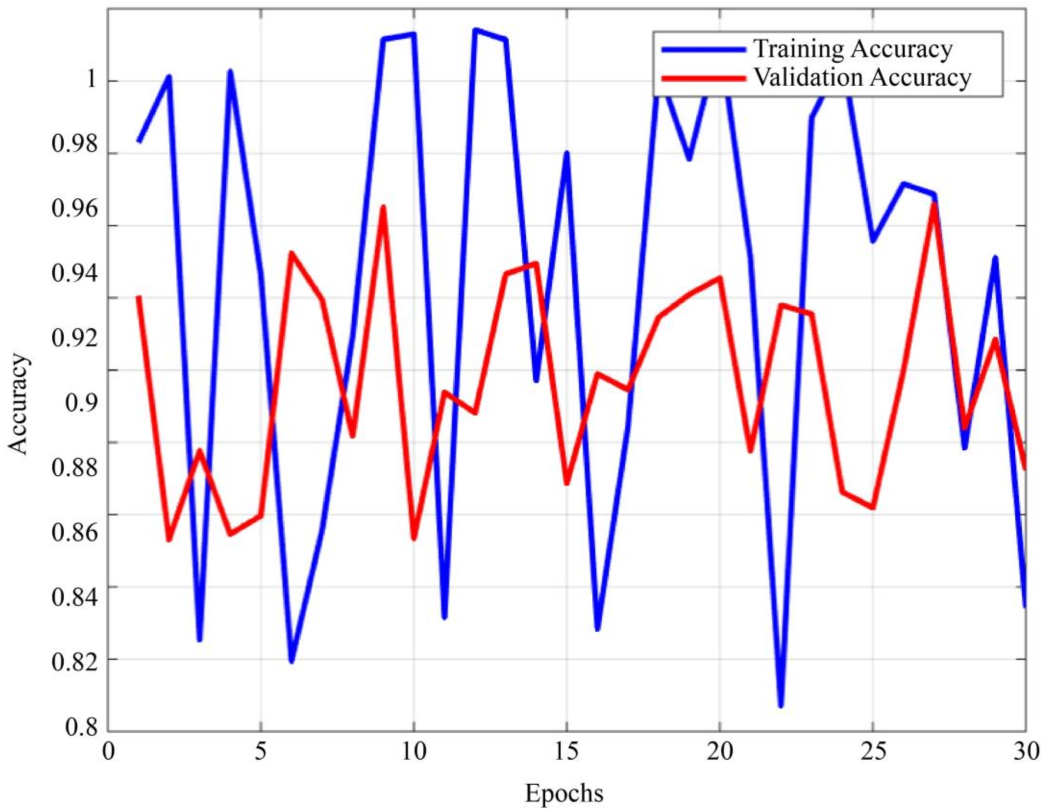

Fig. 3 SVM-CNN accuracy curve

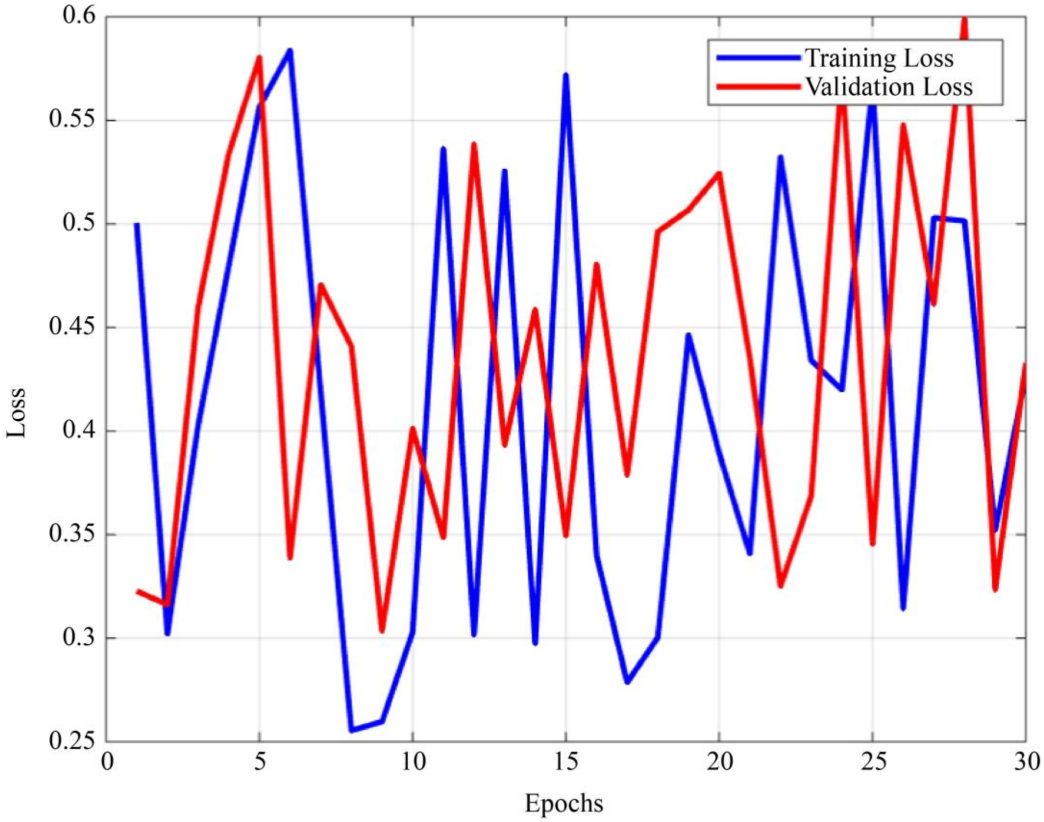

Fig. 4 SVM-CNN loss curve

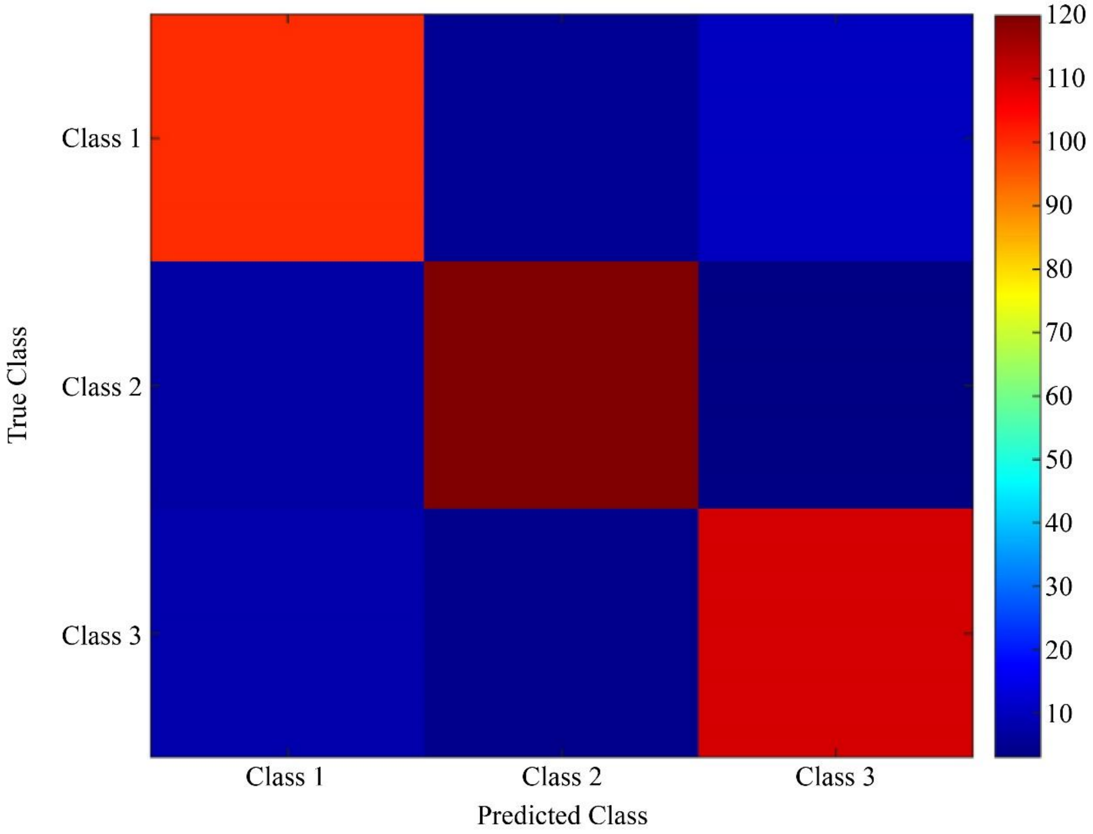

Fig. 5 SVM-CNN

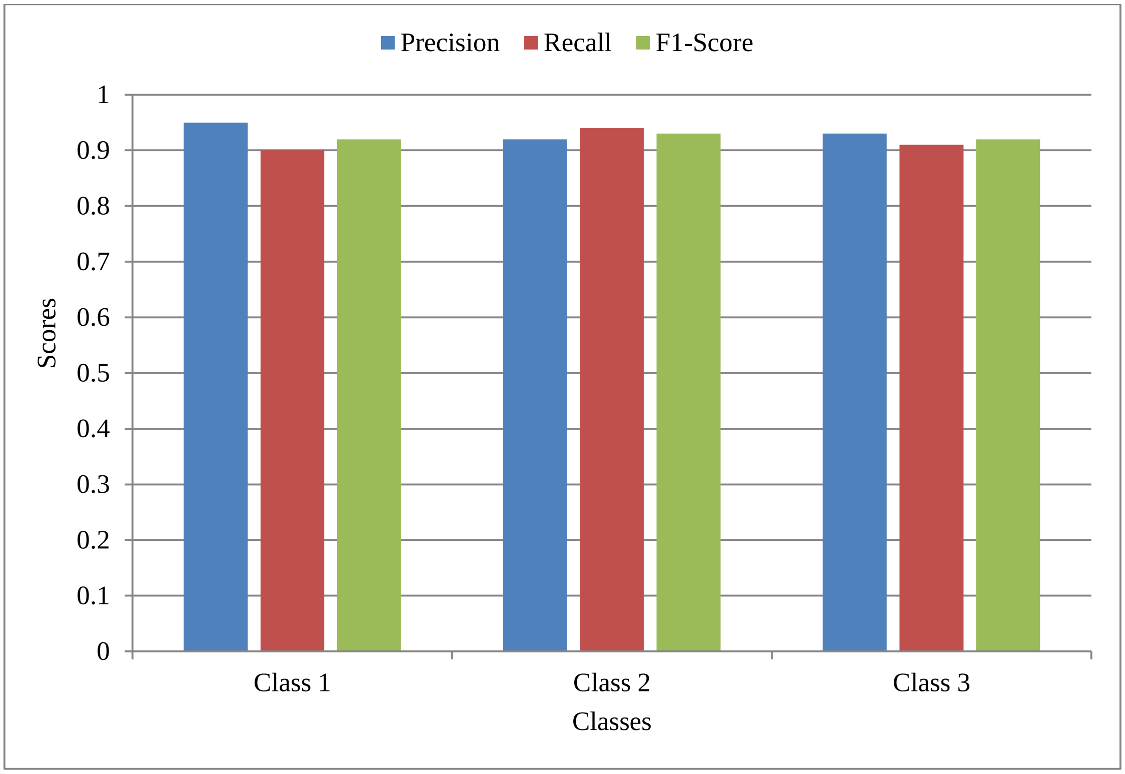

Fig. 6 Precision, recall, and F1-score comparison

Table 2. SVM-CNN classification report

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.7527	0.7792	0.1459	0.9694
2	0.1166	0.9749	0.0689	0.989
3	0.0596	0.9867	0.044	0.9917
4	0.0357	0.9906	0.0235	0.9943
5	0.0173	0.9968	0.0147	0.997
6	0.0111	0.997	0.0147	0.9958
7	0.0156	0.9969	0.0155	0.9966
8	0.0188	0.9944	0.0148	0.9966
9	0.0117	0.997	0.0075	0.9985
10	0.0068	0.9984	0.0194	0.9947
11	0.0108	0.9971	0.0057	0.9992
12	0.0045	0.999	0.008	0.9989
13	0.0052	0.9984	0.0089	0.9985
14	0.0036	0.9991	0.0053	0.9981
15	0.0134	0.9958	0.0141	0.9966
16	0.0066	0.9986	0.0035	0.9989
17	0.0061	0.9987	0.0136	0.9977
18	0.003	0.9992	0.0082	0.9977
19	0.0042	0.9991	0.0185	0.997
20	0.0011	0.9999	0.0082	0.9989
21	0.0068	0.9977	0.0068	0.9989
22	0.0057	0.9989	0.0125	0.9977
23	0.0038	0.9988	0.0179	0.9966
24	0.0069	0.998	0.0205	0.9962
25	0.0017	0.9996	0.0094	0.9985
26	0.0035	0.9991	0.0102	0.9981
27	0.0016	0.9996	0.0077	0.9985
28	0.0043	0.999	0.0065	0.9977
29	0.001	0.9998	0.0049	0.9981
30	0.0008	0.9997	0.0118	0.9985
31	0.001	0.9996	0.0305	0.9955
32	0.0091	0.9967	0.0172	0.9981
33	0.004	0.999	0.0145	0.9989
34	0.002	0.9994	0.0142	0.9981
35	0.0007	0.9997	0.0154	0.9989
36	0.0044	0.9984	0.0107	0.9985
37	0.0031	0.999	0.0131	0.9977
38	0.0027	0.9992	0.0118	0.9977
39	0.0041	0.9991	0.0098	0.9981
40	0.0005	0.9999	0.0137	0.9989
41	0.0001	1	0.0096	0.9992
42	0.0006	0.9995	0.012	0.9981
43	0.0003	0.9999	0.0083	0.9981
44	0.0003	0.9999	0.0083	0.9989
45	0.0022	0.9996	0.0159	0.9977
46	0.0015	0.9993	0.0171	0.9974
47	0.0012	0.9993	0.0079	0.9985
48	0.0023	0.9993	0.0038	0.9992
49	0.0023	0.9994	0.0071	0.9985
50	0.0024	0.9989	0.0033	0.9985

Table 3. Hybrid model accuracy

Sl No	Mudra	Precision	Recall	F1-score	Support
1	alopodmo	1	1	1	179
2	ankush	1	1	1	208
3	ardhachandra	1	1	1	209
4	bhramar	1	1	1	236
5	chatur	1	1	1	215
6	ghronik	1	1	1	210
7	hongshashyo	1	1	1	207
8	kangul	0.99	1	0.99	145
9	kodombo	1	1	1	231
10	kopitho	1	0.99	0.99	200
11	krishnaxarmukh	0.99	0.99	0.99	173
12	mrigoshirsho	1	1	1	245
13	mukul	0.99	1	1	189

Figure 6 compares the accuracy, recall, and F1-score of the three different classes of hand mudras in the Sattriya dance recognition experiment. Precision is the degree of accuracy of the model to predict each mudra category, to measure the percentage of correct optimistic predictions (true positives) as a total of all optimistic predictions (true positives + false positives). The more the model gives, the more likely a prediction of a mudra is to be correct. Recall, or sensitivity, is a characteristic that quantifies the ratio of true positives successfully identified by the model, indicating if the model has recognized every mudra. An increased recall value implies that the model can be used to identify the mudra more effectively, at the cost of increased false positives. The F1-score is a combination of the precision and recall, and gives a harmonic result of a balance between the two metrics, which is particularly effective where both are significant. A high F1-score indicates that the model balances well between identifying the mudra (precision) and identifying all instances of the mudra (recall). This value, presented as a bar chart, allows visual comparison of the model's performance across all three assessment metrics for each of the three types of mudras, thereby pinpointing the model's strong and weak areas in its classification capabilities.

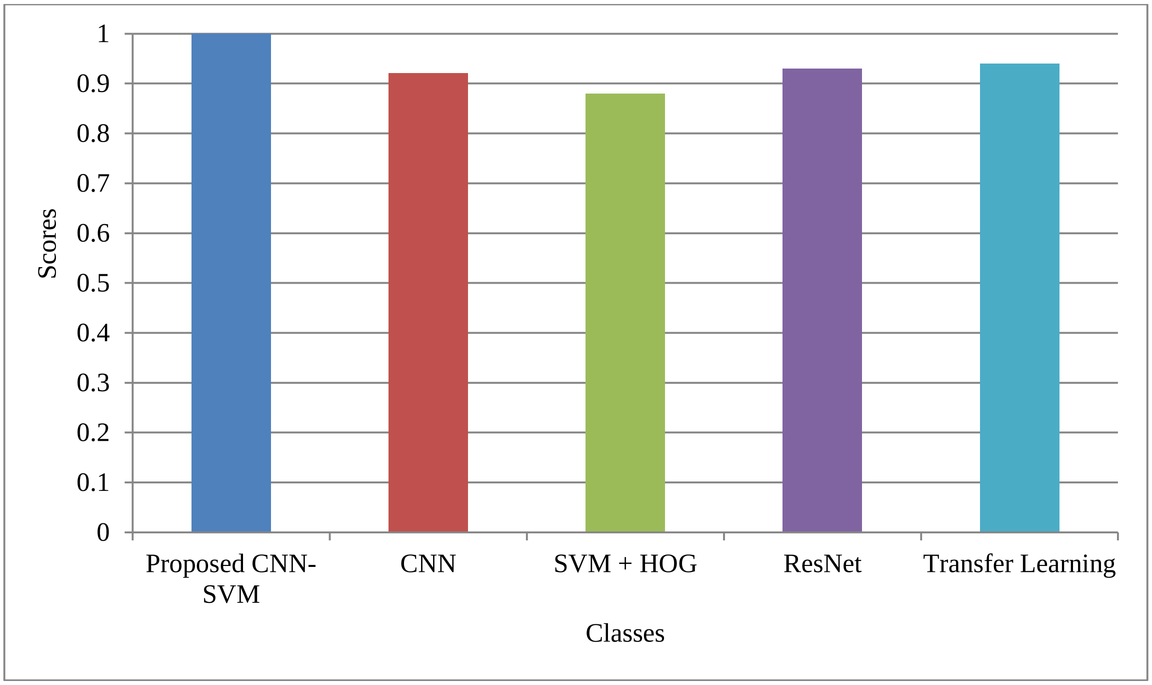

Fig. 7 Comparison of accuracy between the proposed method and existing methods

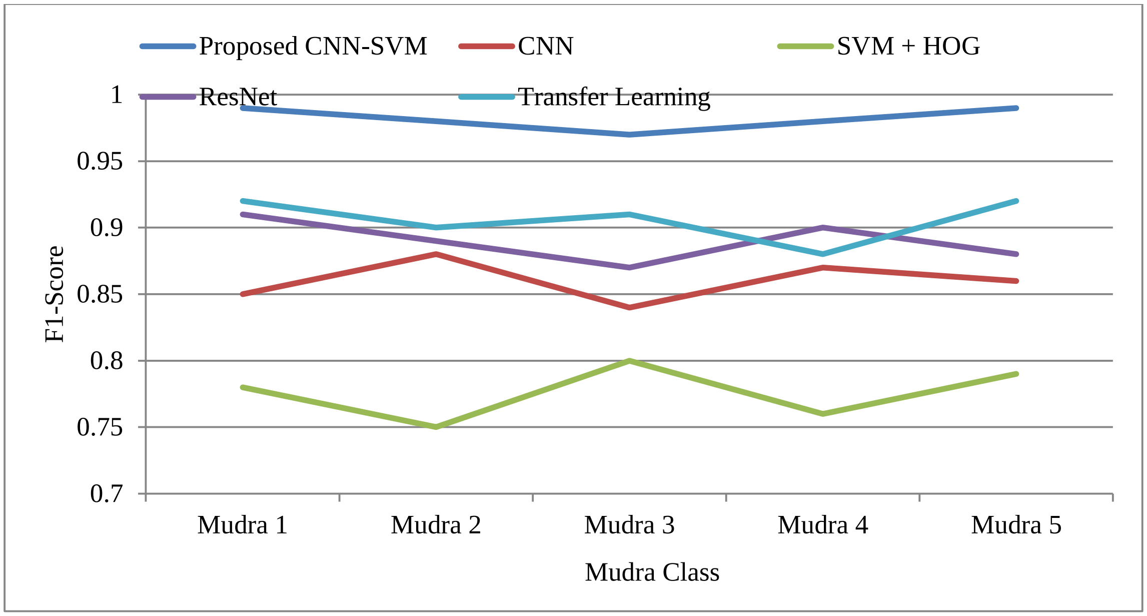

Fig. 8 Comparison of F1-score between proposed method and existing methods

The accuracy of the proposed Hybrid CNN-SVM model has been compared with a number of current methods in hand mudra recognition in Figure 7. The techniques that were compared include the Proposed CNN-SVM, independent CNN, SVM with HOG (Histogram of Oriented Gradients), ResNet, and Transfer Learning by using the pretrained models. As can be seen, the bar chart indicates that the proposed Hybrid CNN-SVM method is much better than all other methods, as it is much more accurate in classifying mudras. The F1-score of the Hybrid CNN-SVM model is compared with that of other methods in Figure 8. F1-score is a metric used to determine the accuracy of a model that is balanced in terms of precision and recall, thus giving a general assessment of the classification performance. The compared methods are the Proposed CNN-SVM, CNN, SVM with HOG, ResNet, and Transfer Learning. As revealed by the radar chart, the Hybrid CNN-SVM model has the maximum F1-score, which means that it can well balance between precision and recall, and is better than others at recognizing hand mudras.

6.2. Classification Performance

In order to measure the effectiveness of classification, some significant measures are employed, such as precision, recall, and F1 score. The metrics of each of the mudra classes are compared, and it shows the quality of the model to recognize the separate gestures. The values of precision, recall, and F1 score are calculated in relation to 13 hand mudra classes, and the results indicate that the model has almost perfect performance. As an example, most of the classes are in the vicinity of 1.0, which implies that there are very few false optimistic predictions by the model. In the same way, the values of recall are also close to one, implying that the model can detect most of the true positives. F1 score, which is a balance between precision and recall, also shows a high score in all classes, which reaffirms that the model is performing consistently well with regard to classifying hand mudras. These measures offer a complete evaluation of the model with regard to its correctness in the classification of the hand gestures so that false positives and false negatives are reduced to the lowest levels. Also, the confusion matrix provides the model results in an in-depth performance breakdown. The diagonal ones of the matrix indicate how many instances of each of the mudras were correctly classified, and the off-diagonal ones represent the misclassification. The matrix has shown that the majority of misclassifications are restricted to a few mudras, where the model has not differentiated gestures with similar shapes. This shows that the model works well; however, a problem is that some mudras have similar positions or forms of the hands, making it hard to discriminate between them. On balance, the confusion matrix is helpful in gaining an understanding of the areas in which the model is performing well and those in which it is potentially performing poorly.

6.3. Interpretation

The findings reveal that the hybrid CNN-SVM is very efficient in identifying and classifying hand mudra in the dance of Sattriya. The accuracy and loss curve indicate that the model is converging well in the training phase, and both training and validation curves exhibit almost the same trend, which indicates that the model is not overfitting. The classification performance measurements (precision, recall, and F1 score) also support the fact that the model can identify the majority of the mudras correctly and score close to perfect on all the classes. This implies that the feature extraction of the CNN, together with excellent classification of the SVM, forms a highly effective recognition system in this case.

Nevertheless, the confusion table indicates that there are some difficulties in the process of separating the similar mudras. The major cases of misclassifications happen between mudras with minor variations in the position of the fingers or hand pose. This implies that, whereas the model is exact, the area of enhancement can be seen in the ability to discern the subtle variations between similar gestures. The addition of more training data or more sophisticated methods of augmentation could be viewed as one of the possible ways of improvement, so that more variability of hand positions and orientations is introduced. The other thing is the dynamic nature of the hand gestures in the dance, where speedy movement can lead to movement blur or occlusions, and this may lower the accuracy of classification in a real-life scenario. The constructions of time into the movement models could be considered in the future, including RNNs to grasp the dynamics of the motion over time and enhance the performance of image recognition on those more problematic cases.

7. Conclusion

The present research proposes a hybrid CNN-SVM algorithm to recognize hand mudras in the Sattriya dance and reaches high performance rates. The model exhibited a high level of accuracy at a time when the training accuracy was as high as 99.99 percent, and the validation accuracy was as high as 99.85 percent. Most of the classes of mudra yielded values of precision and recall that were near 1.0, with the value of F1 indicating a perfect balance between the precision and recall. The confusion parameter showed that there were a few misclassifications, which mainly occurred among similar mudras. These findings confirm the suitability of the hybrid CNN-SVM method for the appropriate classification of hand mudras. Nonetheless, there are still some issues with distinguishing between some similar gestures, implying possible ways of enhancement, which include increasing the variety of data and its augmentation. Incorporating the effect of dynamic movements and occlusions in dance would also help the model to perform better. The temporal models (such as RNNs) may be integrated to work in the future to improve the accuracy of real-world applications and be more effective in these situations. This study provides the foundation for the roboticization of hand mudra recognition, which has enormous potential in terms of cultural landmarks, learning avenues, and also expands the scope of traditional dance to a global platform.

References

[1] Dimpee Baishya, “An Analytical Study of Dance Numbers of Sattriya Dance as practiced in Shri Shri Kamalabari Sattra,” *Sangeet Galaxy*, vol. 13, no. 2, pp. 171-177, 2024. [Google Scholar] [Publisher Link]

[2] Pallavi Malavath, and Nagaraju Devarakonda, “Natya Shastra: Deep Learning for Automatic Classification of Hand Mudra in Indian Classical Dance Videos,” *Artificial Intelligence Review*, vol. 37, no. 3, pp. 45-56, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Poornachandra Sarang, *Support Vector Machines: A Supervised Learning Algorithm for Classification and Regression*, Thinking Data Science: A Data Science Practitioner’s Guide, pp. 153-165, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Laura Igual, and Santi Seguí, *Supervised Learning: Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications*, pp. 67-97, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Debapratim Das Dawn, and Soharab Hossain Shaikh, “A Comprehensive Survey of Human Action Recognition with Spatio-Temporal Interest Point (STIP) Detector,” *The Visual Computer*, vol. 32, pp. 289-306, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[6] Heng Wang, and Cordelia Schmid “Action Recognition with Improved Trajectories,” *2013 IEEE International Conference on Computer Vision*, Sydney, NSW, Australia, pp. 3551-3558, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[7] Saba Naaz, K.B. ShivaKumar, and B.D. Parameshachari, “Aggregation Signature of Multi Scale Features from Super Resolution Images for Bharathanatyam Mudra Classification for Augmented Reality Based Learning,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 3s, pp. 224-234, 2023. [Google Scholar] [Publisher Link]

[8] Ahmad Jalal et al., “Robust Human Activity Recognition from Depth Video Using Spatiotemporal Multi Fused Features,” *Pattern Recognition*, vol. 61, pp. 295-308, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[9] M. Kalaimani, B. Latha, and A.N. Sigappi, “Survey on Hand Gesture Recognition in Bharathanatyam Mudras,” *Telematique*, vol. 21, no. 1, 2022. [Publisher Link]

[10] Pravin R. Futane, and Rajiv V. Dharaskar, “Hasta Mudra: An Interpretation of Indian Sign Hand Gestures,” *2011 3rd International Conference on Electronics Computer Technology*, Kanyakumari, India, pp. 377-380, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[11] Basavaraj S. Anami, and Venkatesh A. Bhandage, “A Comparative Study of Suitability of Certain Features in Classification of Bharathanatyam Mudra Images Using Artificial Neural Network,” *Neural Processing Letters*, vol. 50, pp. 741-769, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Konstantin Dergachov et al., “Data Pre-Processing to Increase the Quality of Optical Text Recognition Systems,” *RadioElectronic and Computer Systems*, vol. 4, pp. 183-198, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ahmad Naeem et al., “Deep Learned Vectors’ Formation using Auto-Correlation, Scaling, and Derivations with CNN for Complex and Huge Image Retrieval,” *Complex & Intelligent Systems*, vol. 9, pp. 1729-1751, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Passent El Kafrawy et al., “An Efficient SVM-based Feature Selection Model for Cancer Classification Using High-Dimensional Microarray Data,” *IEEE Access*, vol. 9, pp. 155353-155369, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Heng Wang et al., “Action Recognition by Dense Trajectories,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’11)*, Colorado Springs, CO, USA, pp. 3169-3176, 2011. [CrossRef] [Publisher Link]

[16] P.V.V. Kishore et al., “Optical Flow Hand Tracking and Active Contour Hand Shape Features for Continuous Sign Language Recognition with Artificial Neural Networks,” *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, Bhimavaram, India, pp. 346-351, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Shuiwang Ji et al., “3D Convolutional Neural Networks for Human Action Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221-231, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[18] Aparna Mohanty et al., “Nrityabodha: Towards Understanding Indian Classical Dance Using a Deep Learning Approach,” *Signal Processing: Image Communication*, vol. 47, pp. 529-548, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[19] Divya Hariharan, Tinku Acharya, and Sushmita Mitra, “Recognizing Hand Gestures of a Dancer,” *4th International conference on Pattern Recognition and Machine Intelligence*, Moscow, Russia Springer, pp. 186-192, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[20] K.V.V. Kumar, and P.V.V. Kishore, “Indian Classical Dance Mudra Classification Using HOG Features and SVM Classifier,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 5, pp. 2537-2546, 2017. [CrossRef] [Google Scholar] [Publisher Link]