Power System Technology
ISSN:1000-3673

Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

Indian Classical Dance Recognition using Convolutional Neural Networks

Dr.Bhavana.R.Maale[1] , Mrs.K R.Soujanya[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Visvesvaraya Technological University-Belagavi, Centre for Post-Graduate Studies, Muddenahalli, Chikkaballapur, India, sg.bhavana@gmail.com,

ORCID ID: https://orcid.org/0000-0002-6515-6116

[2]Research Scholar, Department of Computer Science and Engineering, Visvesvaraya Technological University-Belagavi, Centre for Post-Graduate Studies, Muddenahalli, Chikkaballapur, India, soujanya33.sonu@gmail.com,

ORCID ID: https://orcid.org/0009-0002-9553-1375

Correspondense E-mail: sg.bhavana@gmail.com

Abstract

The extraction and recognition of human movements in performing arts, particularly dance, represent a complex and captivating area of research. In the modern era of globalization, the creative expression and production strategies of classical dance have evolved significantly, creating a need for advanced technological systems to preserve and analyze these art forms.

The primary objective of this study is to develop an automated machine learning framework capable of detecting and recognizing the movements of Indian classical dancers from video sequences. Understanding the semantics of dance movements not only aids in safeguarding India's rich cultural heritage but also supports the development of digital tools for dance education and performance analysis.

The proposed system begins with dancer detection in video frames using background subtraction and Histogram of Oriented Gradients (HOG) techniques. Subsequently, body part segmentation—including the hands, legs, and face—is carried out through HAAR-based and skin color detection methods. From each segmented frame, essential features such as Image Color Coherence Vector (ICCV), Gray Level Run Length Matrix (GLRLM), and Shift Invariant descriptors are extracted to interpret the dancer's movements effectively. Finally, the system employs a Convolutional Neural Network (CNN) for the classification and recognition of dance sequences.

Experimental validation demonstrates that the proposed approach achieves promising accuracy in recognizing complex dance patterns, offering a valuable contribution toward digital preservation and intelligent interpretation of Indian classical dance.

Keywords: HOG, HAAR,Image color coherence vector (ICCV), Level Run Length Matrix (GLRLM), and Shift Invariance and convolutional neural networks (CNN).

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

I. INTRODUCTION

In World, many centuries have been passed beyond recognition, but the man continues to dance. In the human body system,the organism is a more complex and self-healing system. Based on few physiological elements, rhythm plays an important role. The body's organs present inside and process mental function by adopting a rhythmic manner. The natural-movement of, natural and un-realistic, the small and big microcosm, also adopts a few rhythmic senses. Man belongs to the Universe;hence the rules for its existence are also applicable to him. Moreover, presences of internal rhythm synchronize with nature's rhythm;happier humans turn to be.

As the part of digitization locates its place into the world's art, the dancer's body is not available as a rigid resemble of identity. The collection of the living beings and the digital bodies becomes the platform for malleable factors, where these actors are visualized and re-visualized. The digitized body is usually utilized as a scheme over the platform for coheren textinguish feelings of aloofness, luminal identity, and longing. For increasing multimedia information accessing using the internet, multimedia information, especially video streaming indexing,plays very important. Apart from being used for removal can be utilized for the digitization process of cultural heritage;this rises to be an exciting issue. Another application is used for analyzing a specific dance language. Analysis of classical dance includes video, audio, text, and their relationship makes up a complete dance form, as shown in figure1.

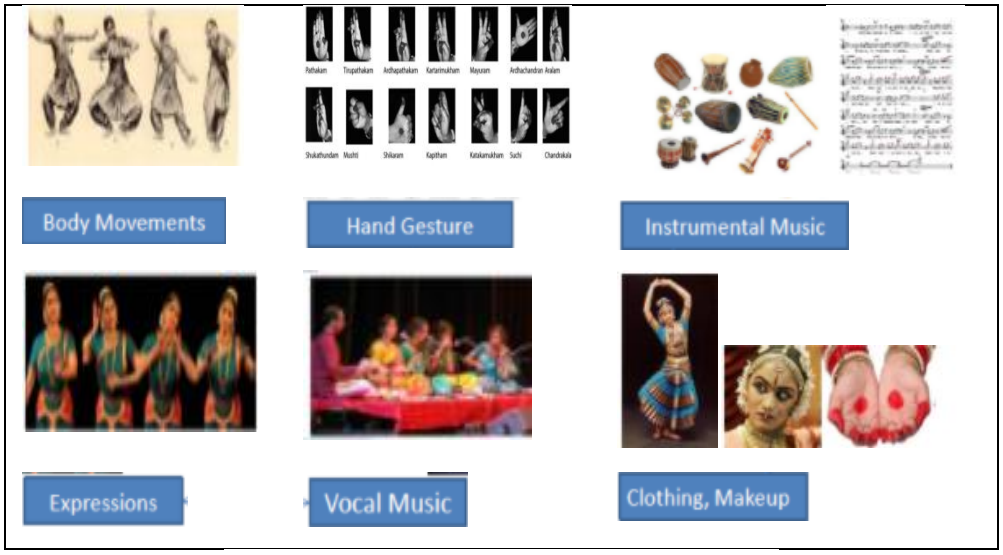

Figure 1: Analysis of Indian Classical Dance

Indian classical dance forms include many complicated body signs from rotating, twisting fingers, hands, and body-bending w.r.t their motion trajectory and spatial points. This paper has extracted these complex movements from video sequences and classified them as Indian classical dance forms Khattak and Kuchipudi. This paper proposes using the background subtraction method, human object detection, segmentation of different parts of the dancer, and texture features representing the dance

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

sequence. Three features are combined and computed for representing for exacting shape presence in the video frames. We have trained a convolutional neural network to categorize query videos of dance taken from the dataset for recognition.The remaining parts of the paper are grouped as follows: Literature Survey is written in the second section,the proposed method is explained in the third section. In the fourth section, we show the experimental outcomes and at last fifth section includes the conclusion of the paper.

II. LITERATURE SURVEY

Kahol [1] etal. Explains a step-by-step producer for user-centric gesture segmentation. This algorithm was tested for around 23 different sequences of motion, like moving and running. From these examples,segmentation is performed with the help of five various performers, and a classifier naïve Bayesian is utilized in encoding each performer segmentation format. The strategy was expandable to any motion series, as not depending on a group of unique poses, as encoded pose limits as a binary vector in the form of hierarchically coupled body part activities. The method reached89.2% for generic motion serial accuracy but unable to validate on semantically higher series like modern dance.

Ankita Bisht[2] etal. proposed a novel framework to identify seven various Indian classical dance types. The presented method depends on the concatenation of 3 level lower-frameworks which uses the formats from trained CNN and optical flows. The experiment created a database that includes videos of 620 for 7 types of classical dance formats and obtained an efficiency of 75.83%.

Anuja P. Parameshwaran[3] etal., Their main work is to build a 2D database of single-hand gestures from 27 different forms collected using YouTube videos, Google search engine, live performances by professional artists under staged environment. At the second level, they have tried to identify the effectiveness of Convolutional Neural Networks by optimizing the hyperparameters to recognize and classify single hand gestures. Finally, to achieve a higher level of accuracy, they have evaluated transfer learning results and double transformation learning.

Tanwi Mallick [4] etal. has generated a model for examining Adavus and focused on Bharatanatyam dance. In their work to extract critical frames, they have used audio events. The essential poses are detected within K-frames. At last, using a sequence of poses, they have recognized the Adavu. Three different categories of classifiers like GMM, SVM, and CNN are studied to analyze Posture recognition. The proposed framework has been achieved 90% accuracy. From key frames, they have achieved 83.5% accuracy. To recognize signature posture, they have built 23 classes of posture present in Natta Adavus and achieved 83% accuracy utilizing GMM and the angular type of features extracted from skeletons with 98% efficient utilizing SVM and HOG based features through RGB video frames. They also merged 23 different groups into 15 forms for structuring a CNN classifier and achieving a performance of 99% for estimating posture identification using CNN.

Lakshmi Tulasi et al. [5] conducted research on the recognition of Kathakali hand gestures (mudras) using machine learning and deep learning approaches. They developed a specialized dataset of Kathakali hand gestures and employed Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) for image recognition and classification. The comparative analysis between the two

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

models demonstrated an accuracy of up to 74%, highlighting the effectiveness of deep learning in gesture interpretation.

Surbhi Gautam et al. [6] proposed an automatic recognition system for classical Indian dance movements by utilizing a video database. Since each video comprises multiple frames representing various dance actions, they extracted Histogram of Oriented Gradients (HOG) features to interpret and classify dance poses across diverse backgrounds. Their method was evaluated on 50 poses derived from nine Bharatanatyam dance videos with varying environmental conditions, achieving reliable performance in pose recognition.

Gnana Priya et al. [7] examined multiple machine learning approaches and implemented a Deep Convolutional Neural Network (DCNN) to identify different dance poses without the need for explicit feature extraction. Initially, they trained a base model and subsequently refined it for specific recognition tasks. The network demonstrated high efficiency and adaptability by learning hierarchical features through convolutional, pooling, and fully connected layers. Convolutional layers used kernels to detect image features through element-wise operations, while pooling layers reduced spatial dimensions. The model achieved a 62% accuracy in pose classification, showing potential for further optimization.

Shailesh et al. [8] introduced a deep learning-based framework for automatic video recognition focusing on dance postures and foot positions. Their approach was structured into two stages. In the first stage, a classifier was trained to identify various stances based on postures extracted from video frames. The second stage processed videos as sequences of images and applied classification techniques to recognize specific stands. A meta-annotation file was generated to record the occurrence of particular postures with corresponding timestamps. The proposed method exhibited superior accuracy and efficiency compared to conventional machine learning models, demonstrating robustness in classifying dance stands.

Vinay Kaushik et al. [9] developed a classical dance classification model emphasizing body posture, hand gestures, facial expressions, and head orientations. Their study proposed a posture-oblivious shape signature model integrated into a sequence learning framework. They utilized symmetric Spatial Transformer Networks (STNs) to extract poses from the initial frames of dance videos and applied CNN-based posture estimation. Subsequently, pose sequences were converted into posture flow representations by calculating similarity measures between consecutive poses and employing non-maximal suppression. The resulting normalized vectors—comprising distance features, flow, and joint angles—effectively captured the spatio-temporal structure of body movements. This approach successfully modeled complex skeletal dependencies across video frames, enhancing dance style classification accuracy.

Shweta Mozarkar et al. [10] focused on the identification of dance mudras that convey specific meanings using pattern recognition and image processing techniques. Their study presented a Computer-Aided Detection (CAD) system for recognizing Bharatanatyam mudras by employing the saliency detection method based on hyper complex quaternion Fourier transform phase (FTTP) to isolate the hand gesture from its background. The salient features extracted from static double-handed

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

mudra images were classified using the K-Nearest Neighbour (KNN) algorithm. The system demonstrated effective recognition performance in differentiating between various mudras, contributing to the digitization and preservation of Indian classical dance forms.

III. METHODOLOGY

The primary objective of this study is to develop an automated machine learning framework capable of identifying a performer's movements within a video frame. Since a video comprises a continuous sequence of images representing specific dance actions, each frame captures unique shape and color features that characterize distinct dance movements.

This section outlines the proposed methodology and its key components in detail. The approach focuses on detecting and classifying dance movements from a series of video frames using feature extraction techniques and a Convolutional Neural Network (CNN) architecture.

As illustrated in Figure 2, the video sequence is first decomposed into individual frames to locate and isolate the presence of the dancer. The methodology begins with dancer detection, followed by segmentation of various body parts such as the head, hands, and legs. From these segmented regions, three types of features are extracted—ICCV, GLRLM (Gray Level Run Length Matrix), and Shift Invariant features—to represent the movement patterns accurately.

Prior to classifier training, dimensionality reduction is applied using Kernel Principal Component Analysis (KPCA) and Singular Value Decomposition (SVD) to enhance recognition accuracy and computational efficiency. The reduced feature set is then fed into a multiclass CNN classifier, which is trained to recognize and categorize the performer's dance movements within each frame of the video.

This systematic approach enables precise identification of dance actions while optimizing model performance through effective feature engineering and deep learning integration.

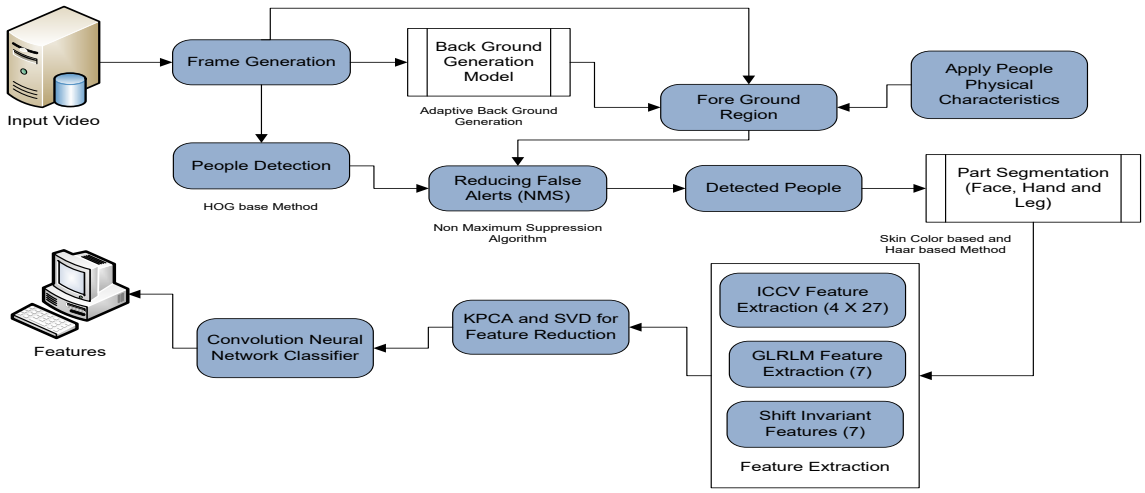

Figure 2: Proposed System Architecture.

3.1 Back Ground Generation Model

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

Identifying stationary and non-stationary objects separating moving objects named “foreground” and static information named “background” by video streaming is the primary step in various video-based instances. The stepsincludedin the background subtraction. As shown in the below block diagram figure 3 for the background generation model, video frames are used to create a background system and then extract foreground by adopting background modeling, with subsequent upgrading.

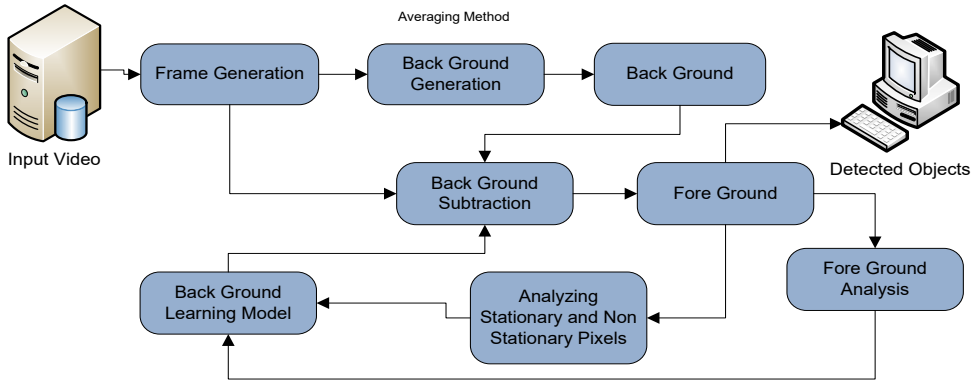

Figure 3: Background Ground Generation Model

The background subtraction methods initially include computing the first background image using which a background model is represented and a mechanism required to update the background model whenever some changes are noticed automatically. Finally, every pixel is divided into background and foreground objects. The algorithm is written below**:**

Algorithm of Background Ground Generation Model

Inputs: Input Frames

Output: Foreground Detection

Step.1 Initialization Background

The initial frame initiates the subtraction method through N-trained frames that include or do not include foreground objects.

Bt and It representing background and existing image at the duration of t.

Step.2 Maintaining Background

To upgrade the background image of B_t *and* I_t *Denotes background and the existing image at the duration of t. Including in repeating for new background image with various learning rates based on the previously classified pixel as foreground or as background:*

$$B_{t+1}(x,y)=(1-\alpha)B_t(x,y)+\alpha I_t(x,y)$$

If(x,y) *is background*

$$B_{t+1}(x,y)=(1-\beta)B_t(x,y)+\beta I_t(x,y)$$

If (x,y)*is foreground*

(2) As the next step, the detection of moving objects is done by identifying foreground, which includes separating pixels into foreground or background by differentiating the

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

background with the current frame.

Step.3 Foreground detection

An identifying object moving with the help of foreground detection includes separating pixels belonging to the foreground or background by observing the background image with the current frame. This includes labeling pixels, namely background pixels or foreground pixels. This process is termed a classification task.

Step 4The 2nd and 3rd steps are performed repeatedly as time increases.

3.2 Dancer Detection

After performing foreground localization of regions by adopting an adaptive background model of candidate objects for extracting and classifying utilizing HOG descriptors, at the stage of classification,every object is named,whether human or not. The HOG descriptors are adopted for this process due to classification outcomes through the requirement to reject every object which is moving and absent in the ROI of the model, and together for accelerating further computations. To get the HOG descriptors, initially, the gamma and input color is normalized. Secondly, oriented gradients are computed at different directional filters. In the third step, the images are partitioned as cells, and oriented gradients frequencies are counted w.r.t to each cell. The frequencies are shown on histograms. Sequentially, a group of cells is gathered more extensive and overlapping blocks that are square or rectangular (called R-HOG) or placed in the coordinate system of polar-logarithmic (C-HOG). In the end, representation is taken by combing oriented histograms of each specific cell. Figure 4 explains the flowchart of a HOG descriptor.

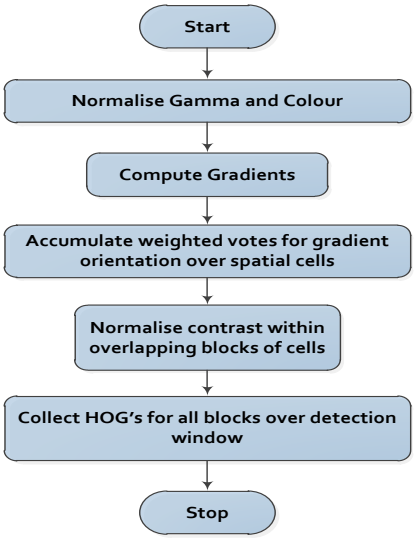

Figure 4: Flowchart of HOG Descriptor

3.3 Dancer Segmentation

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

As shown in the block diagram, the human image is segmented in two steps. Using the YCbCr algorithm, hands and legs are identified, and in the second part, using Haar Classifiers,the face is segmented.

3.3.1 Hand and Leg Segmentation

The algorithm accepts input in the RGB format and subjects to gray conversions for illumination. Transformations are carried out to compute the scale factor and convert the RGB image into YCbCr using a threshold value to detect skin color. After segmenting skin color, unwanted data, i.e., noise, is removed by filtering by applying the median filter. For more tunned images, morphological functions are operated on the image. The entire procedure of the YCbCr algorithm is explained in an algorithmic format.

Algorithm of color based Segmentation

Inputs: Human Segmentation Frames

Output: Segmented parts

Step.1 *Input Frame is initialized as an RGB image.*
Step.2 *I am initializing the binary format of the output image.*
Step.3 *I am applying the Gray world method for illumination compensation.*
Step.4 *Using RGB elements from the input image by inversing their average values and the smallest average value is selected.*
Step.5 *Computing the scaling parameters and scale the values.*
Step.6 *Convert RGB image into YCbCrutilizing the threshold values for identifying the color of skin.*
Step.7 *Label the skin-colored pixels.*
Step.8 *After performing segmentation of skin color, the noisy image filtered by applying the median filter.*
Step.9 *To obtain a tunned image, morphological operations are applied to the image.*

3.3.2. Haar based Face Segmentation

Viola and Jones derived an algorithm named Haar Classifiers for detecting any kind of object rapidly. Many include human faces by adopting AdaBoost classifier cascading depending on Haar properties and not based on pixels. The main idea of the Haar classifier for identifying an object is by the help of Haar-like features. These properties utilize the pixel's intensity and calculate contrast values in neighboring rectangular format sets of pixels. The variations in contrast in-between the pixel sets can be used to find relatively lighter and darker regions. More than one or three neighboring pixels groups having relative contrast variance form a Haar-like feature. Haar-like features are shown in Figure 5

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

are utilized for detecting an image. Scaling Haar features are done simply by lowering or decrementing the no.of pixel sets analyzed. By doing so, features can be applied for detecting objects of different sizes.

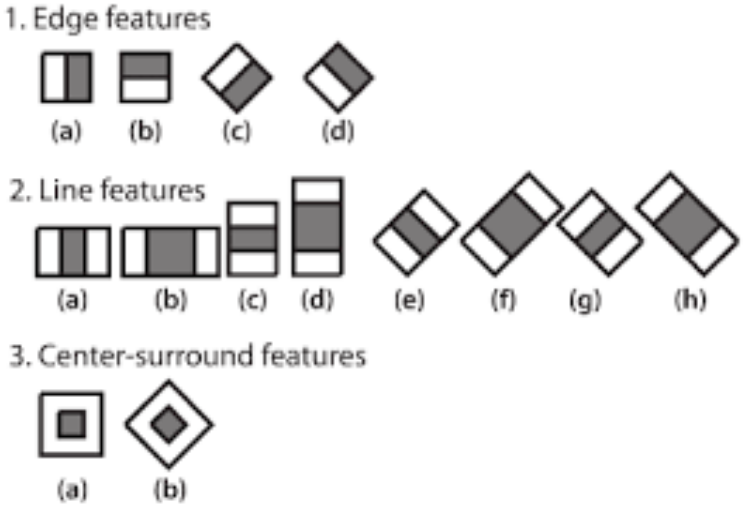

Figure 5 commonly usedHaar Features

The cascading scheme-based classifiers provide the sub-images with higher probability analyzed based on all Haar features that differentiate an object. It can also be used to change the efficiency of a classifier. We can raise both the false alarm rate and positive hit rate by reducing the number of levels. The inversion of the same is also valid. Viola and Jones reach a 95% accuracy rate for detecting the face of the human face with 200 efficient features.

3.4 Feature Extraction

3.4.1 Color Coherence Vector

The method of Color Coherence Vector, histogram of each bin is divided into 2 variants, coherent and incoherent. If pixel forms' value is present in the more extensive uniformly-colored area, then it is called to be coherent, or else it is termed incoherent. On the other hand, coherent pixels are a contiguous area of the image, where incoherent pixels do not. The vector of color coherence describes the categorization for all the colors present in the image. By comparing two pixels *a* and *b*,

C_i: no.of pixels which are coherently represented with color j.

N_i: no.of in coherent pixels in color j

$$D(a,b)=\sum_{i=0}^{n}(|a_{C_i}-b_{C_i}|+|a_{N_i}-b_{N_i}|) \quad (1)$$

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

3.4.2 Gray level run length method

GLRLM is a method for extracting more excellent order statistical features present in the mammogram images. GLRLM includes a group of pixels which are continuous with similar gray value. The run-length can be described as no.of neighboring gray values present in a specific direction. It can be computed by adding no.of times the correspondence run obtains in the image. For gray level run length matrix$P(i,j/\theta)$ and the $(i,j)th$ attributes explain the total runs along with gray level i and whose length j occur in the image along the angle. The gray level run length can be obtained as:

$$\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P(i,j/\theta) \text{ and } 1\leq N_Z(\theta)\leq N_P \quad (2)$$

Where, $N_g$represents no.of intensity value

N_rAs no.of run lengths in the image.

N_P as no.of pixels present in the image

$N_Z(\theta)$as total no.of runs in the image w.r.t angle

$P(i,j/\theta)$ be the run-length matrix at an arbitrary direction θ

The features are Short Run Emphasis, Long Run Emphasis, Gray Level Non-uniformity, Run-length non uniformity, Run percentage, Low Gray level Run Emphasis, and High Gray Level run Emphasis the accuracy of the classifier.

Table 1: GRLM features and formulas

FEATURES	Formula
SRE	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\mid\theta)}{j^2}}{N_Z(\theta)}$
LRE	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\mid\theta)j^2}{N_Z(\theta)}$
GLN	$\frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_r}P(i,j\mid\theta))^2}{N_Z(\theta)}$
RLN	$\frac{\sum_{j=1}^{N_r}(\sum_{i=1}^{N_g}P(i,j\mid\theta))^2}{N_Z(\theta)}$
RP	$\frac{N_Z(\theta)}{N_P}$
LGRE	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\mid\theta)}{i^2}}{N_Z(\theta)}$
HGRE	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\mid\theta)i^2}{N_Z(\theta)}$

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

3.4.3 SHIFT Invariant

The Algorithm of SIFT is utilized for detecting and explaining the local attributes of an image. This method is famous for linking two images with defining local descriptors, including information for match-making in-between them. The basic idea of the SIFT descriptor is done by converting the image into a format that includes points of interest. These key points include characteristic features of the face image. SIFT resists in variance for scaling and rotating. The algorithm is explained as below:

1. I am detecting max and minimum points from the space scale.
2. I am locating the characteristic vital points.
3. Assigning the orientation and
4. The descriptor fora characteristic key point.

3.5 Feature Reduction

3.5.1 Kernel Principal Component Analysis

The concept of KPCA is based on assumptions that various databases cannot separate linearly in their space. These are made separable linearly from projecting into a perfect dimensional space. The included dimensions are the applications of few arithmetic functions carried through the actual data dimensions.Therefore, projecting our database into a perfect dimensional feature space so that they are separable linearly, further PCA is applied to the new dataset. By undergoing these reductions of linear dimension in actual space matches anon-linear dimension reduction from the original space.Performing PCA in the new database requires more calculations. But adopting kernel approaches applying these calculations from the actual state space. This is carried out from the kernel function that raises a lane of calculating dot product among two vectors - in higher dimensional space- in the actual space.

Algorithm of KPCA

Inputs: Human Segmentation Frames

Output: Segmented parts

Step.1 *Initially select kernel functions* $k(x_i,x_j)$ *and T denotes the transformation to a greater dimension*

Step.2 *Similar to PCA, need to find covariance matrix from the input data. Then we need to utilize the kernel function for the calculation of the matrix. Therefore, computing the kernel matrix in matrix format resulted from applying kernel function to every data pair.*

$$k=T(x)T(x)^T$$

Step.3 *Central to kernel matrixes is the same as subtracting the mean value from the converted data and then divided through standard deviations.*

$$k_{new}=k-2(I)k+(I)k(I)$$

where matrix I and its elements are equal to i/d

Step.4 *As the next step, eigenvectors and eigenvalues are calculated from the*

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

matrix

Step.5 Arranging the eigenvectors depending on their corresponding eigenvalues in a descending format.

Step.6 By selecting the required dimensions of the dataset we need, denoted as m. next, Then first m eigenvectors are selected and added together to a single matrix.

Step.7 In the end, Compute the multiplication of that matrix with the input data,the outcome generates a reduced dataset.

3.5.2 Singular value decomposition (SVD)

With the higher-order, SVD shows the lesser no.of.dimensions needed for representing a matrix or linear transformation [18][27]. More regularly, multi-dimensional information is represented in lesser dimensions as the unwanted presence of data. If a group of n order dimension vectors relies on a k-dimensional subspace where k < n and each n-vector have only k degrees of freedom, which are uniquely explained by k numbers. The simple correlations if it gets natural to create SVD better candidates for reduction of the feature. However, if the reduction is not possible, it may show the magnitude for the singular values discovered by the SVD.SVD factorizes an input matrix into three sets of matrices such as $\sum$, U and V such that

$$A=\sum UV^T \tag{3}$$

Where,

- U as an ortho normal matrix, where columns are termed as left singular format, whose diagonals are called singular values,
- V is an orthonormal matrix in which columns are termed as suitable singular vectors.

A larger matrix generally does not require complete factorization but only higher singular values and associated singular vectors. Storage space is saved;noise can be removed and recovered in the lower rank format of the matrix.Having top kk singular values, with dimensions for results with lower rank matrix :

$$U=mxk \tag{4}$$

$$\sum=k\times k \tag{5}$$

$$V=n\times k \tag{6}$$

3.6 Convolutional Neural Networks

CNN is an instance of Deep Neural Networks. Such networks employ convolutions for extracting sensible data or critical points from the input features that are later utilized to build the sub-layers of neural network calculations.In our work, CNNis designed as a multistage approach for collecting information following [21, 25]. The approach is built having input layer, convolutional of four layers, rectified linear units (ReLu) of five, and stochastic pooling of two layers, single dense and SoftMax output layer.

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

Indian classical dance frames having sizes 640 × 480 are chosen to feed the system. Initially, the frames are pre-processed,reducing the size to 128 × 128 × 3. Reducing the size of input frames increment the calculation capacity for higher-performing computation (HPC) from where programmed coded. The HPC is utilizedfor training the CNN made up of 6-node added with CPU and GPU processing machine.

By assuming the frame of the input video is of size$I \in R^{w \chi h}$.Theconvolutionalkernelwithsize$K$isconsideredforconvolutionwithastrideof$S$and$P$padding for filling the input video frame boundary. The size of the output of the convolution layer is given by

$$S_{out}=\frac{(I-K+2P)}{S}+1 \tag{7}$$

The architecture of our CNN model consists off our convolutional layers. While the first two layers extract the low-level features (like lines, corners, and edges), the last two layers learn high-level features. The output of a convolutional layer is generally denoted with the following standard equationas

$$y^n{}_j=f(\sum_{i\in c_j} y_i^{n-1}*k^n{}_{ij}+\zeta_j^n \tag{8}$$

Where$n$ represents the layer,$k_{ij}$is the convolutional kernel,$\zeta_j$represents bias,and the input maps are represented by c_j. The CNN uses a tanh activation function with an additive bias formulated as

$$h_{ni}^{xy}=\tanh(\zeta_{mi}+\sum_{w=0}^{w_i-1}\sum_{h=0}^{h_j-1} w_{ij}^{wh}\, h^{(x+w)(y+h)}{}_{i-1}) \tag{9}$$

Z_{ij}represents feature map bias which is non-supervisory trained,andw_i,$h_j$are the kerne lwidth and height,respectively. $W^{wh}$is the weight of the kernel at position (w,h). Over a region,the max value of a feature is obtained using the pooling technique,which reduces the data variance.We implemented our architecture with a stochastic pooling technique by calculating the probability values for each region.For every feature map$c$,the probability is givenby:

$$\chi_{w,h}^{n,k}=Stochastic_{(w,h,i,j)\epsilon p}\left(\chi_{w,h}^{n-1,k}u(i,j)\right) \tag{10}$$

Where $\chi_{w,h}^{n,k}$is the neuron activation function at a point (w, h)in spatial coordinates,and(i,j)is the weighing function of the window.Compared toother pooling techniques, stochastic pooling makes CNN converge faster and improves generalization in processing in variant features.

This identification of Indian classical dance is a multi class classification problem. therefore, a SoftMax regression layer given by a hypothesis function$h_\phi(x)$is being used as

$$h_\phi(x)=\frac{1}{1-e^{(-\phi^T x)}} \tag{11}$$

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

ϕneed to be trained such as the cost function(ϕ)is to be minimized.

$$j(\phi)=-\frac{1}{m}[\sum_{i=1}^{m}\sum_{j=0}^{l}1\{y^i=j\}log_p(y^i=\frac{Z}{x^i};\phi)] \quad (12)$$

The probability of classification in the layer of SoftMax for categorizing an input$x$as a group$Z$ is obtained by

$$P\left(y^i=\frac{Z}{x^i};\phi\right)=\frac{e^{\phi_j^T x^i}}{\sum_{l=1}^{k}e^{\phi_j^T x^i}} \quad (13)$$

The CNN is trained for gathering the attributes of each type of dancing posture through supervised learning.The selected internal feature descriptors define similarity within training images.We have framed 200 postures from ICD(offline data)done by10 different classical dancers.The length of the entire database is around 2000 dancing postures where all the postures are captured from 2 secs or 60 frames per second, creating 120k frames. In the same way, web-based dancing data is recorded through YouTube, and all the postures are normalized to 60 fps. As a result of features describing which are trained by the CNN model, the max activation neuron is considered for accurately recognizing the dancing posture. Lastly, the feature-based maps are viewed by averaging the image patches with stochastic responses in higher layers.

IV. Experimental results

A selected set of videos was initially used to evaluate the effectiveness of the proposed method. For this study, a custom database of Indian classical dance performances was developed, comprising Kathak and Kuchipudi dance forms. The dataset includes videos collected from publicly available sources such as YouTube, along with recordings captured in a controlled offline environment to ensure consistency and clarity. Figure 17 illustrates sample images from the constructed database.

The same dancer's video sequences were utilized for both training and testing phases, where features were systematically extracted from the dance sequences. Subsequently, the performance of the classifier was validated based on its accuracy and efficiency in recognizing and categorizing dance gestures.

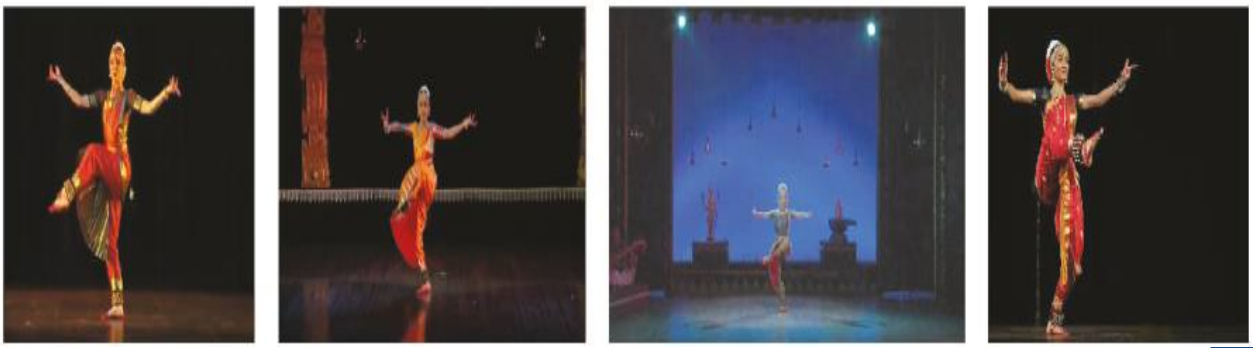

Figure: Dance dataset

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

From the first input video of the dance dataset, the process of dancer identification, feature extraction, and graphical representation is illustrated in Figure 18. The foreground regions of the video frames are detected using adaptive background modeling based on Gaussian Mixture Models (GMMs), which effectively isolate the dancer from the background. Once the dancer is identified, a bounding box is generated to define the dancer's spatial dimensions within the frame.

Subsequently, Shift Invariant and GLRLM (Gray Level Run Length Matrix) features are extracted from the segmented body regions of the dancer. To enhance the efficiency and accuracy of recognition, dimensionality reduction techniques are applied to retain only the most relevant features. The refined feature set is then used to train a Convolutional Neural Network (CNN) classifier, which performs the final stage of dance movement classification.

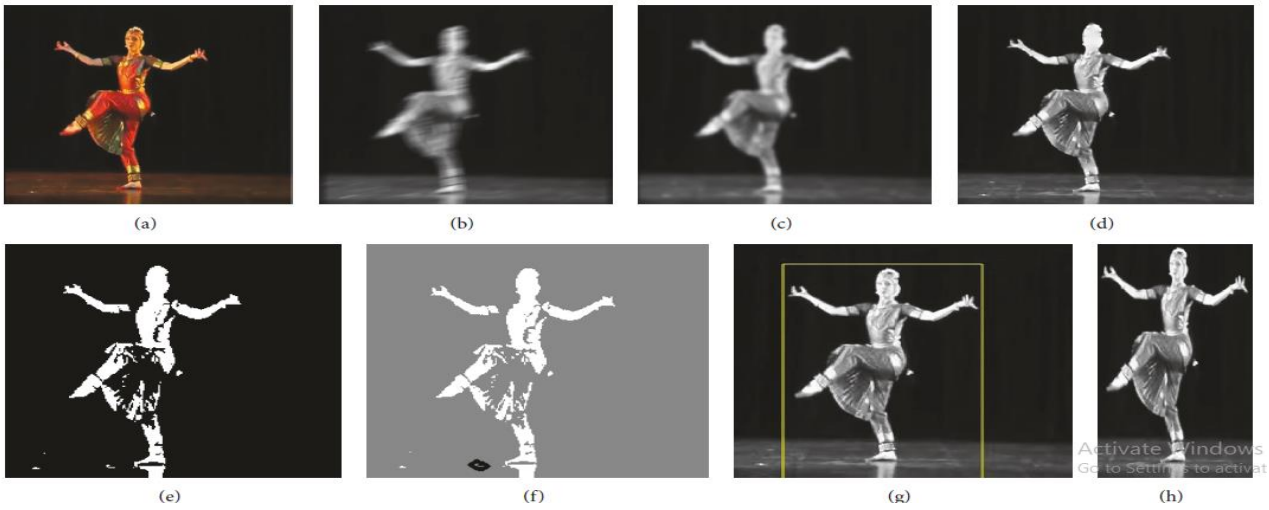

Figure : Dancer extraction. (a) Original frame, (b) mean filtered, (c) Gaussian filtered, (d) distance saliency map, (e) silhouette mask, (f) connected components labelling, (g) identified dancer, and (h) dancer extracted.

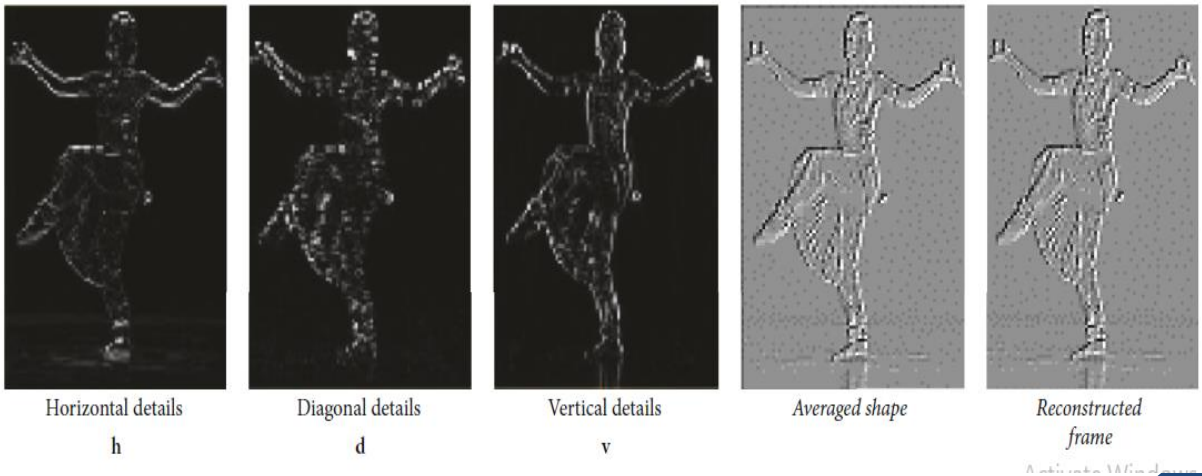

Figure: Harr feature, SHIFT, GLRLM representing shape in three different orientations.

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

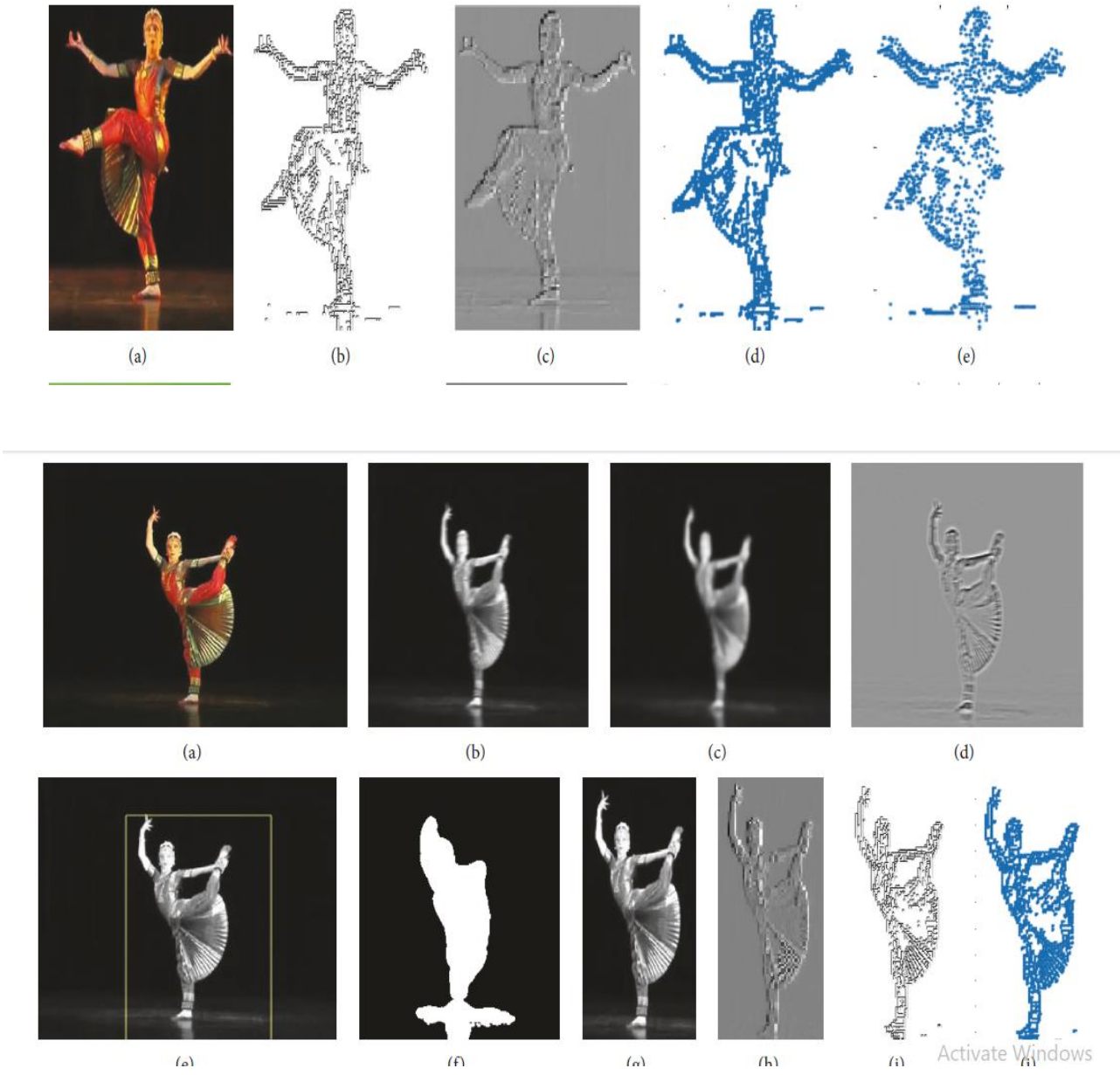

Figure : (a) Input frame (b) gray scale conversion (c) SHIFT, GLRLM, HOG, Background Subtraction , Dancer foreground extracted, dancer detection.

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

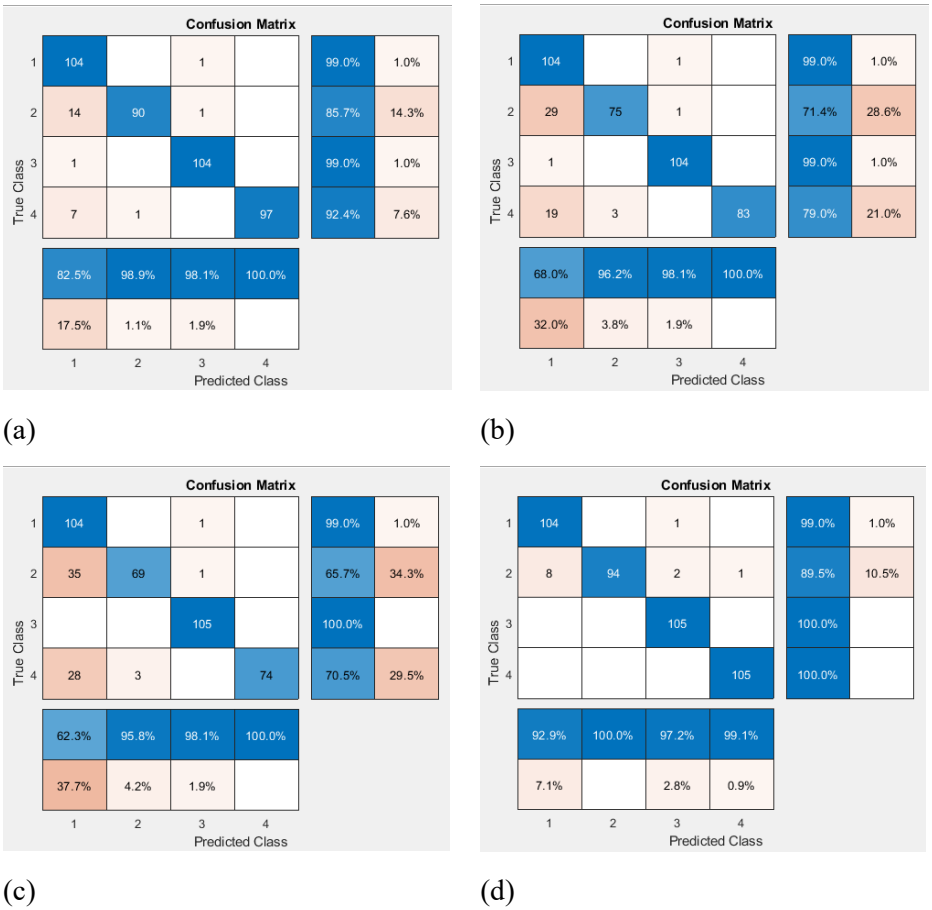

Figure: Confusion matrix with (a) 56feature of color features (b) 256 features of GDP(c) all 312(56+256) features, (d) 56 features of color, and only 128gradient features

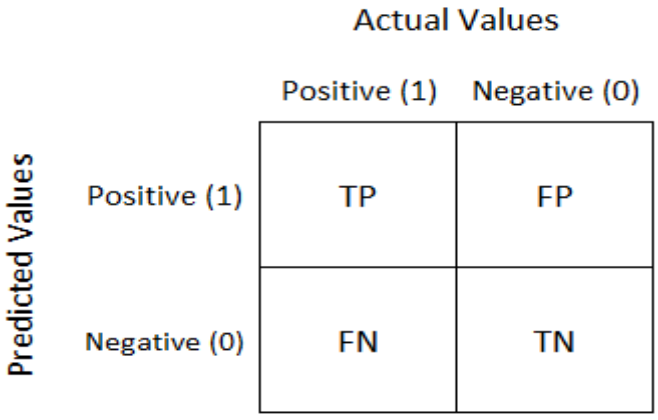

Figure 25:Confusion Matrix Table

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

$$Recall=\frac{TP}{TP+FN} \tag{8}$$

$$Precision=\frac{TP}{TP+FP} \tag{9}$$

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

The figure shows how to use the confusion matrix to calculate the overall accuracy of classifying the healthy leaves, bacterial affected leaves, mosaic affected leaves, and yellow spot affected leaves. Based on the overall accuracy calculation, it is found that the proposed system offers 97.1429% accuracy.

Recognizing and classifying Indian classical dance forms through machine learning presents a complex and multifaceted challenge due to the intricate nature of body movements, gestures, and expressions. Each performance involves coordinated actions of various body parts—such as the hands, legs, and face—which must be accurately detected, segmented, and analyzed for effective classification.

The present study successfully developed an automated framework that integrates advanced computer vision techniques with deep learning architecture, specifically Convolutional Neural Networks (CNNs). The proposed system demonstrated strong performance in identifying dancers and recognizing movement patterns through systematic feature extraction and dimensionality reduction methods, including ICCV, GLRLM, and Shift Invariant features optimized using KPCA and SVD.

Experimental results revealed that the CNN-based classification model achieved high accuracy in distinguishing dance postures and styles within video frames. This validates the efficiency of the proposed framework in handling the dynamic complexity of Indian classical dance sequences.

The findings highlight that deep learning, particularly CNN architectures, can effectively interpret complex visual and temporal information inherent in dance performances. Future research can extend this work by incorporating larger and more diverse datasets, exploring additional classical dance forms, enhancing real-time recognition capabilities, and improving robustness against background and lighting variations. Such advancements would contribute significantly to preserving India's cultural heritage through digital technologies and facilitate innovative tools for dance education, performance analysis, and archival applications.

References:

1. Vinay Kaushik, PreranaMukherjee_BrejeshLall," Nrityantar: Pose oblivious Indian classical dance
sequence classification system", arXiv, Dec 2020.
2. Tanwi Mallick, ParthaPratim Das, Arun Kumar Majumdar, "Posture And Sequence Recognition For Bharatanatyam Dance Performances Using Machine Learning APPROACH," Department of Computer Science and Engineering.

Power System Technology
ISSN:1000-3673
Received: 06-05-2025 Revised: 15-06-2025 Accepted: 25-07-2025

3. Surbhi Gautam, Garima Joshi, Nidhi Garg," Classification of Indian Classical Dance Steps using HOG Features," International Journal of Advance Research in Science and Engineering, Vol.6, Issue No.8, 2017.
4. B. Gnana Priya, M. Arulselvi, " Deep Learning for Human Pose Classification using Multi-View Dataset,"International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.
5. Ankita Bisht, Riya Bora, Goutam Kumar, Pushkar Shukla, and Balasubramanian Raman, "Indian dance form recognition from videos."
6. Shailesh S, Dr. Judy M V, "Automatic Annotation Of Dance Videos Based On Foot Postures", Indian Journal Of Computer Science And Engineering (Ijcse)
7. Anuja P. Parameshwaran1, Heta P. Desai1, RajshekharSunderraman, Michael Weeks," Transfer Learning for Classifying Single Hand Gestures on Comprehensive Bharatanatyam Mudra Dataset".
8. Rhonda D. Phillips, Layne T. Watson, Randolph H. Wynne, and Christine E. Blinn, "Feature Reduction using a Singular Value Decomposition for the Iterative Guided Spectral Class Rejection Hybrid Classifier" Departments of Computer Science and Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.
9. Nithya R &Santhi B, "Comparative Study on Feature Extraction Method for Breast Cancer Classification,"*Journal of Theoretical and Applied Information Technology,* Vol.33, No.2, (2011).
10. A.M. Abdalla, S. Dress & N. Zaki, "Detection of Masses in Digital Mammograms using Second-order Statistics and Artificial Neural Network,"*International Journal of Computer Science and Information Technology (IJCSIT),* Vol.3, No.3, (2011).
11. Suckling J, Parker J, Dance D, Astley S, Hutt I &Boggis, C, "The mammographic images analysis society digital mammogram database,"*Excerpta Medical International Congress Series,* Vol. 1069, (1994), pp.375–378.
12. Monika Sharma RB & Dubey Sujata SKG, "Feature Extraction of Mammograms,"*International Journal of Advanced Computer Research, (IJACSA),* Vol.2 No.3, (2012).
13. Preetha K & Jayanthi SK, "Breast Cancer Detection and Classification using Artificial Neural Network with Particle Swarm Optimization,"*International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST),* Vol.2, (2016).
14. Chidambaranathan S, "Breast Cancer Diagnosis Based on Feature Extraction by Hybrid of K-Means and Extreme Learning Machine Algorithms,"*ARPN Journal of Engineering and Applied Sciences*, Vol.11, No.7, (2016).