Journal of Computer Science

Research Article

Annotation of Facial Expressions (Navarasa) in Videos Using Deep Learning

[1]Reshma M R, [1]Kannan B, [2]Jagathy Raj V P and [3]Shailesh S

[1]Department of Computer Applications, Cochin University of Science and Technology, Kerala, India
[2]School of Management Studies, Cochin University of Science and Technology, Kerala, India
[3]Department of Computer Science, Cochin University of Science and Technology, Kerala, India

Article history
Received: 24-09-2024
Revised: 24-12-2024
Accepted: 06-01-2025

Corresponding Author:
Reshma M R
Department of Computer Applications, Cochin University of Science and Technology, Kerala, India
Email: reshemmar@gmail.com

Abstract: Facial expression plays a significant role in understanding human behavior, alongside other relevant elements such as body posture, gait, and hand gestures. In the context of Indian Classical Dance (ICD), these elements work together to convey the stories depicted by characters from Indian mythology. Accompanying songs and shlokas evoke emotions, reflected in performers' facial expressions known as the performance. In recent years, computer vision techniques have garnered Navarasa, or "nine emotions". Recognizing these expressions is vital for appreciating significant attention due to their applications in the active research field of emotion recognition in humans. This paper aims to develop a deep-learning algorithm to detect facial expressions, thereby enhancing the comprehension of ICD performances. Our approach decodes the meanings of Navarasas represented in images and videos by recognizing and annotating these emotions as performed. By utilizing image processing and transfer learning techniques, we achieved an accuracy of 92% in classifying and annotating the Navarasas.

Keywords: Transfer Learning, Classification, Annotation, Facial Expression Recognition, Navarasa, Indian Classical Dance

Introduction

Facial Expression Recognition (FER) is essential to analyze human behavior (Newmark, 2022; Tian *et al.*, 2001). The emotional nature of humans helps to activate or motivate their behaviors. Human beings are guided by emotions that lead us to categorize things as significant or unimportant. Both verbal and nonverbal communication is employed to express feelings. Verbal communication uses spoken words or spoken words (language) to exchange information, emotions, and thoughts (Nikolaus *et al.*, 2016). Languages are a social and behavioral phenomenon and are a clue to the organization of cognitive processes that deal with observing, understanding, envisioning, and communicating. In nonverbal communication, facial expressions, body postures, or gestures convey emotions. A human can communicate very efficiently using nonverbal means. Among those, facial expressions play a critical role in understanding emotion. This is not only in the daily social interactions of a human but also in cases of creative self-expression, such as performing arts, which shows an insight into how the auditory and motor modalities are connected. The world's cultures allude to a profound relationship to human creativity and self-expression through the power of music and dance. The popularity of such music and rhythm-oriented performing art forms throughout the world further demonstrates the prevalence of dance as a tool for self-expression. Art forms help connect different people and help them gain insight into each other's culture, helping to evaluate each other in their diversity. Thus, learns to understand the behavior and lifestyle of various groups and ethnicities. Some essential ingredients of cultural tourism are living culture and performing arts (Cros and McKercher, 2020).

Indian society is an example of one among the rich cultural heritage, where their social values are knit with music and dance. Its extensive and legendary past has persisted in the shape of living arts practiced in practically every region of India, both in rural and urban places. The performing arts depict stories, emotions, Love, and devotion from the myths and religious scriptures found in the country and become part of their life: Many Indian dances, theatre, and music forms enlisted in intangible cultural heritage compiled by UNESCO. India's rich 'Intangible Cultural Heritage' (ICH) and varied tangible heritage make it valuable, and we appreciate the significance of using these to promote domestic and foreign tourist visits. Most art forms are still actively performed as part of religious ceremonies or tourist activities, which accumulate a lot of tourist attractions. The live presentation to an audience or spectators will be delightful when they can decipher the play or story performed. These artistic mediums are rich in human movement, which presents new challenges for computer science. Technological breakthroughs have opened up a world of opportunity for acquiring, categorizing, and archiving ICH content. These open, active research projects cover various computer vision tasks, including pose estimation, image categorization, face recognition, object recognition, and many others (Joshi and Chakrabarty, 2021; Aristidou *et al.*, 2022; Reshma and Kannan, 2019). Also, it functions as a vital key for numerous real-world applications, such as animation, human-computer interface, human behavioral detection, etc.

The Indian classical dance forms follow the classical manual on dance called Natyastra, which includes a clear and detailed account of dance organized and written by Bharata Muni (1996). Abhinaya Darpan. , a Similar significant work in dance, is also prominently considered (Nandikesvara *et al.*, 1970). Both these manuals present the principles of dance. Indian aesthetics pursue them to constitute the characteristics of the dance. The main three aspects of dance are Natya, Nritta, and Nritya. The imitation of characters or the dramatic element in the dance 'Natya.' The primary form of dance movements is Nritta, and the mudras or gestures that are the expressional component is Nritya. According to Abhinaya Darpanam, the dance related to Sentiment (Rasa) and Psychological States (bhava) is called nritya. India's classical literature, dance, and theatre are rooted in the Rasa chapter of Bharata Muni's Natyastra. One could easily argue that the Rasa prakarana, or the chapter on Rasa, has been an imperative component of all Indian performing arts. Rasa's exact English equivalent does not exist. Human emotions or feelings are the closest in meaning in this context. There are nine types of Rasa, collectively called "Navarasa".

Understanding the feelings (emotions) or facial expressions is essential to enjoy the art performance thoroughly. The technique of employing a computer to analyze different emotion data, extract feature values that describe emotions, create a mapping link between feature values and emotions, and then categorize emotions to determine the emotional state is known as emotion recognition. At the same time, emotional computing has garnered a lot of interest and study since it centers on theories and technology that advance knowledge of human emotions (Guo *et al.*, 2020). Humans primarily use physical movements to communicate and express their feelings in daily routines. This type of movement is used prominently in a dancer's emotional expression, and facial emotion recognition is an area of research not frequently evaluated in the domain of performing art forms like Indian classical dance, as dance emotions are expressed externally through dance moves (Shikanai and Hachimura, 2015). The dancer's emotional movement style (Fan *et al.*, 2012) and the association between actions and feelings were confirmed by Kipp and Martin (2009), which also examined the relationship between gestures and emotions in drama.

In the years, most studies in the terrain of computational methodologies in dance or automated dance motions examined a specific dance form's pose or gesture like in Bhavanam and Iyer (2020), Jain *et al.* (2021), and Shailesh and Judy (2020). Moreover, there are no publicly accessible datasets to evaluate facial expressions or Navarasam in Indian classical dance. For a better understanding of facial expressions in Indian classical dance, this research suggests a framework that classifies and annotates facial expressions. The framework uses deep learning techniques and image preprocessing to identify facial emotions. The nine forms of facial emotions or navarasa considered in this research on the theoretical aspects of Indian classical dance, Shringar(Erotic), Hasya (Humorous), Karuna (Compassion), Raudra (Anger), Veera (Heroic), Bhayanaka (Fear), Bibhatsa (odious), Adbhuta (Wonderous), Shanta (Peaceful). This research examines a revolutionary technique that categorizes and annotates facial expressions by combining technology and facial emotion identification in dance. Other diverse fields where facial expression analysis is beneficial include e-learning in education (Wu, 2016), the creation of friendly robots (Cid *et al.*, 2013), the monitoring of driver fatigue (Whitehill *et al.*, 2013), interactive video games (Khalifa *et al.*, 2017), psychiatry and medical treatment (Bazrafkan *et al.*, 2017), and numerous other systems involving human-computer interaction. These kinds of studies also make meaningful contributions to the other two broad fields of artistic communication (Dan, 2012) and cultural heritage preservation (Reshma *et al.*, 2023).

Table 1 briefly describes Navarasam (nine facial emotions) in Indian Classical Dance forms. The paintings illustrating the nine navarasa, or facial expressions, in Indian classical dance are shown in Figure (1). (Sources for Figure (1) - the book "Mohiniyattam: Ariyendath Ellam" (Vijayan, 2012)).

The paper is comprised of four main sections. The first section presents a Literature Review. The second section, titled Methodology, offers an overview of the proposed framework for recognizing and annotating facial expressions in Indian classical dance Forms. This section includes detailed subsections on the dataset, data preprocessing, and data Augmentation. The third section, Implementation, describes how the proposed approach is applied to classify and annotate facial expressions. The section on Experimental Results and Analysis follows it. Finally, the paper concludes with a summary in the Conclusion section.

DOI: 10.3844/jcssp.2025.1364.1378

Table 1: Description of Navarasa in ICD

Rasa	Descriptionn
Shringar (Erotic)	A supreme feeling that can cure anything. An intrinsic, universal, egoless condition that permeates everything.This Rasa brings us together and enables us to encounter heavenly Love.
Hasya (Humorous)	Hasyam ties us to happiness, laughing, and humor. We become stress-free and forget about everything when we are joyful. It is a mentality, a thrill that comes from inside, that keeps us content.
Karuna (Compassion)	Karuna is a feeling triggered by suffering, grief, or sorrow connected to melancholy. It connects us all, enabling us to understand and relate to one another's life circumstances. Karuna encompasses all tragic and depressing emotions: Heartache, separation, grief, loss of a lover, and pain brought on by a loved one's passing.
Raudra (Anger)	This emotion, which means a roar caused by fury, can cause annoyance, hatred, and violence. A single instant of rage can ruin years of goodwill; it is a property of fire. Raudra Rasa can take many forms, such as royal wrath, anger brought on by wrongdoing, rage at injustice, and contempt.
Veera (Heroic)	This Rasa is linked to a sensation of courage since it embodies a heroic attitude. It relates to boldness, bravery, self-assurance, confidence, and resolve. The Veera rasa can rouse the courageous from their slumber and instill them with vibrant energy. Aspects of this Rasa include bravery in warfare, the mindset of warriors during the fight, and the mindset of warriors during their final moments.
Bhayanaka (Fear)	This Rasa represents a critical mind indulging in self-pity, hate, and self-hatred. When one has this Rasa, they feel contempt or discontent with themselves and others. A terrible sensation called Beebhatsam evokes impolite and vulgar behavior through poor language and etiquette.
Bibhatsa (odious)	It is an emotion linked to anxiety, concern, unease, insecurity, and self-doubt, among other things. When confronted by something more powerful than themselves, bhayanakam is an emotion that arises.It's also a sensation of hopelessness and helplessness that we could experience in an uncontrollable circumstance.
Adbhuta (Wonderous)	This Rasa represents our innocence and the kid still living inside, which means wonder, amazement, and surprise. It is a feeling that inspires awe for the seemingly little objects around us. It feeds our curiosity and encourages us to live an adventurous life. Adbhutam is the sensation experienced when encountering something celestial and extraterrestrial that has never been witnessed or imagined.
Shanta (Peaceful)	This Rasa represents our complete tranquility and relaxation. It denotes a state of equilibrium, contentment, or calm. Shanta symbolizes the total unification of the mind, body, and universe.

Related Work

Research on human emotion recognition through facial expression analysis is ongoing. Analyzing automatic facial expressions comprises three key components: Face capture, facial data extraction, and expression recognition. In recent years, there have been significant advancements in face detection and tracking, feature extraction techniques, and expression categorization methods. Academicians have always been interested in the interactions between technology and art. Computer scientists and dancers collaborated to arrange, capture, analyze, and create a dance, provide templates for choreographic composition, and enhance comprehension of dance steps due to computer science and information technology developments in the 1960s; notable such work was "Choreography and computers" (Noll and Hutchinson, 1967). Currently, emotion recognition systems exist to detect emotions in speech, gestures, and text (Koolagudi and Rao, 2012; Batbaatar *et al.*, 2019).

Fig. 1: Facial expressions (Navarasa) in ICD

In the performing arts like dance or drama, the actors/dancers typically portray the characters with the help of the expression of human emotions. One of the foundational works related to the current field of automatic facial expression identification is Charles Darwin's research on emotions in humans (Newmark, 2022). He categorized different types of expressions into related groups and established broad principles regarding expressions and ways of expressing emotions in both humans and animals. The relationship between facial expressions and emotions and the information conveyed by these emotions is examined in (Ekman, 1993). Pantic and Rothkrantz (2000) investigate the challenges of designing and implementing an automated facial expression analysis system. Samal and Iyengar (1992) propose a comprehensive five-step approach for representation, detection, identification, expression analysis, and classification based on physical features. Donato *et al.* (1999) present an exhaustive survey and comparison of recent techniques for facial expression

DOI: 10.3844/jcssp.2025.1364.1378

recognition within automated Facial Action Coding Systems (FACS). A signal-processing method for analyzing an individual's emotional state is suggested by Wang *et al.* (2014). This approach involves analyzing EEG signals after smoothing and removing unrelated noise, then tracking changes in emotional states using manifold learning. In Black and Yacoob (2000), a local parameterized model for picture motion was developed using an optical flow-based cue. Black and Yacoob's study examined both non-rigid (affine-plus-curvature model) and rigid (planar model) face motion. Cohn *et al.* (1998) used a hierarchical optical flow technique to track facial fiducial points automatically. Valstar and Pantic (2006) examined the minute variations in face muscular action units (AUs) to investigate temporal behavior. Shen and Ji (2008) used geometric facial features to classify individuals by age.

Srimani *et al.* present a method for detecting navarasa facial expressions in Bharatanatyam using variations in facial expressions concerning entropy, skewness, and kurtosis (Srimani and Hegde, 2012; Luck *et al.*, 2014) define a dance emotion expression approach utilizing the Laban Movement Analysis (LMA) method (Luck *et al.*, 2014). They convey the emotions in the dance movements entirely, transforming the original dance movement data into three characteristic expression parameters to produce dance emotion data, which comprises the typical parameters of the three elements: body structure, spatial orientation, and force effect. Barros *et al.* obtained better accuracy using the sequence-dependent hierarchical features and multimodal information for recognizing emotions (Barros *et al.*, 2015). Zhao *et al.* (2016) presented a model in the survey that illustrates a novel technology that can determine an individual's sentiments through RF signals radiated off his body. (Gupta, 2018) suggested a model that automates the recognition of live facial emotion using Image processing and AI techniques. Mehta *et al.* used AUs to measure spontaneous and posed facial emotion intensities from facial expressions in publicly available databases (Mehta *et al.*, 2019).

Sravya *et al.* (2019) proposed face detection algorithms that range from simple edge-based algorithms to hybrid high-level approaches utilizing advanced pattern recognition methods. Tian *et al.* (2021) examine three distinct techniques to systematically evaluate bias and fairness in facial emotion detection: a baseline, an attribute-aware method, and a disentangled approach. They accomplish this by employing both well-known datasets, RAF-DB and CelebA. (Xu *et al.*, 2020). They concluded that Data augmentation increases the accuracy of the underlying model but cannot eliminate the bias effect. The convolutional neural network used in the deep learning framework to identify emotions that match the Navarasas related to ICD by Mohanty and Sahay, (2018) they proposed two datasets, CVLND-RGB and CVLND-D, comprising RGB images and the corresponding depth maps taken in controlled laboratory settings. They also used professional dancers to capture a sizable real-world dataset while they performed the Navarasas in situations with no restrictions. Kale and Rege (2019) utilized the LBP texture descriptor for feature extraction and classified the expressions into seven classes, namely happy, sad, angry, fearful, surprised, disgusted, and neutral, using linear SVM and achieved a recognition rate of 81.38%. Also, they tested to identify the difference between neutral faces and other expressions and claimed to get a 90% recognition rate. They also used a dataset with images captured in a controlled environment. Selvi *et al.* presented an approach to detect Kathakali facial expressions using deep learning and attained a Ninety percent accuracy using the CNN algorithm (Selvi *et al.*, 2021).

Li *et al.* (2021) use artificial intelligence techniques to examine dance therapy's psychological and perceptual elements. Dance therapy uses dance and movement to encourage emotional, social, cognitive, and physical integration. Li and colleagues use artificial intelligence tools to assess and comprehend the psychological impacts of dance therapy interventions. It could result in more individualized and successful therapeutic approaches. (Jiang and Yan, 2024) created a sensor-based dance coherent action-generating model using a deep learning framework. They aim to develop a model that uses information from dancers' wearable sensors to produce logical dance motion sequences. Jiang and Yan hope to generate realistic and expressive dance sequences by employing a deep learning framework to identify intricate patterns and correlations in the sensor data. High-level dancing movements are analyzed by (Wang and Tong, 2022) using deep learning and the Internet of Things (IoT). They investigate combining deep learning methods with Internet of Things sensors to analyze and understand complex dancing movements. Wang and Tong hope to learn more about the dynamics and patterns of dance moves by utilizing data from the Internet of Things devices, like motion trackers or cameras. It could help with choreography, performance evaluation, and interactive installations. Pandeya *et al.* (2021) use deep learning to concentrate on multimodal emotion classification for music videos. Although this study focused on music videos, dance performances could benefit from the methods and strategies created. To categorize the emotions expressed by dance and music movements, Pandeya and colleagues probably look at ways to collaboratively analyze textual, visual, and aural clues in music videos.

Using deep learning approaches and feature extraction methods, respectively, Wang *et al.* (2020) emphasize the identification and categorization of emotional states expressed through dance movements. Others, like Mallick *et al.* (2022), enhance our comprehension of aesthetic emotion and choreographic structures by delving into comparative analysis and posture recognition in certain dance styles. Aristidou *et al.* (2017) and Jiang and Yan (2024) are two works that investigate the creation of emotionally expressive motions using sensor-based models and algorithmic control. Furthermore, studies by Maret *et al.* (2018) and Jili *et al.* (2019) explore the extraction of emotional states from movement data using genetic algorithms and statistical techniques, while (Zacharatos *et al.*, 2014; Pandeya *et al.*, 2021) explore automatic emotion recognition in dance using body movement analysis and multimodal approaches. Sun *et al.* (2020) also present novel music-synchronized choreography generation methods, illustrating how adversarial learning can fuse dance and music. Together, these studies demonstrate the multidisciplinary character of this field of study, bridging the gap between artistic expression and technology to enhance our comprehension of the subtle emotional undertones present in dance moves.

Preliminaries

Transfer Learning for Feature Extraction

Humans can analyze and apply the knowledge we gathered in one task to another, which is similar, by using relevant knowledge gained from encountering the previous task perfectly. The concept of Transfer learning techniques in Deep Learning employs this principle. A peculiar phenomenon shared among many deep neural networks trained on images: Early layers of a deep learning model attempt to learn low-level information, such as identifying edges, colors, fluctuations in intensities, etc. No matter what image we analyze, these traits are not distinctive to any dataset or task. The models that use previously trained models' information (features, weights, etc.) already comprehend the features. It expedites the process compared to building neural networks from scratch. These layers are finally connected to a last layer (usually a fully connected layer, in the case of classification) to get the final output. We may use a pre-trained network (like Inception V3, VGG, and ResNet) without its final layer as a feature extractor for different tasks, as they have a tiered architecture.

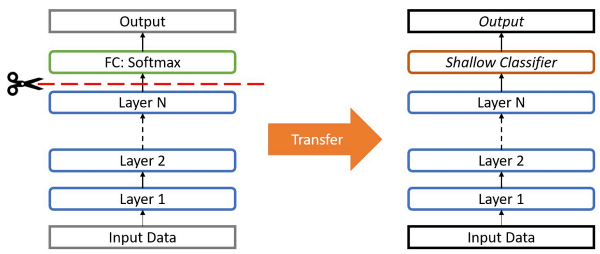

Fig. 2: Transfer learning with pre-trained models as feature extractors

Deep transfer learning involves using pre-trained deep learning models as feature extractors, as illustrated in Figure (2) (adapted from the book (Silaparasetty, 2020)). These pre-trained models serve as generic models because they are trained on large, diverse datasets. To leverage their capabilities, we utilize the outputs from one or more network layers, which have been trained on different tasks, as general feature detectors. These extracted features can then train a new, simpler model. The aim is to use the weighted layers of the pre-trained model to extract features while keeping the model's weights unchanged. It allows us to train effectively with new data for different tasks.

Transfer learning can be mathematically defined when using a pre-trained neural network as a feature extractor by breaking down the mathematical formulation step by step. Neural Network Representation: Consider a pre-trained neural network with parameters θ, which consists of multiple layers, including an input layer, hidden layers, and an output layer. Consider the input to the network as x and the output as y. Feature Extraction: Transfer learning involves using the pre-trained neural network to extract useful features from the input data. Let's denote the feature extraction function as $f(x; \theta)$, where f() represents input x's mapping to extracted features. Feature Representation: The extracted features, denoted as $h = f(x;\theta)$, capture important information about the input data. Task-Specific Model: A task-specific model generated to solve a particular problem using the extracted features after feature extraction. This model consists of additional layers specific to the target task, and $g(h; \varphi)$, where g() represents the task-specific mapping, and φ denotes the parameters of the task-specific layers.

Transfer Learning Model: The transfer learning model comprises feature extraction and task-specific functions. Let's denote this model as $F(x; \theta, \varphi)$, where F() represents the overall transfer learning model:

$$F(x;\theta,\varphi) = g(f(x;\theta);\varphi)$$

where $f(x; \theta)$ represents the feature extraction function that maps the input x to extracted features using the pre-trained neural network with parameters θ. $g(h;\varphi)$ represents the task-specific layers that take the extracted features h as input and produce the desired output for the target task, with parameters φ. In summary, transfer learning as a feature extractor involves utilizing the pre-trained network to extract informative features from the input data, followed by constructing a task-specific model that operates on these features to solve a specific problem.

Extra Tree Classifier for Feature Selection

In developing a machine learning model, selecting the right features becomes crucial, especially for data sets with many variables and properties. Selecting the most essential features for the model is known as feature selection. A feature is a characteristic that affects or aids in resolving the problem. Feature selection, a feature engineering technique, is necessary for creating a

DOI: 10.3844/jcssp.2025.1364.1378

machine learning model. Only a few of the dataset's variables are employed to develop the model; the remaining features are either redundant or unimportant. Imagine adding all these pointless and unnecessary attributes to the dataset. In that instance, it can negatively affect the model's overall performance and accuracy. The feature selection process in feature engineering resolves this issue. Due to its robustness in effectively managing continuous and categorical data, tree-based models have become prominent over the past decade. With this feature selection method, judgments regarding the significance of the characteristics are made using tree-based supervised models. The well-known model's Decision Trees and Random Forest are often the go-to tree-based models, but here we are choosing a lesser-known one, ExtraTree. The Extra Trees Classifier is a machine-learning technique typically employed for classification jobs. Although it is not a feature selection strategy directly, it can give hints about the significance of features. The Extra Tree Classifier, also known as the Extremely Random Tree Classifier, is an ensemble technique that uses the training dataset to seed many tree models built randomly and sorts the most popular features. Instead of using a bootstrap replica, it fits each decision tree on the entire dataset and divides the nodes at random split points. Depending on the level of randomness or entropy in the sub-nodes, nodes are split at every level of the constituent decision trees. The nodes were divided based on the dataset's variables, which led to selecting the constituent tree models' most homogeneous sub-child. It reduces variance and lessens the likelihood of overfitting the model. An Illustration of the Extra Trees Classifier is given in Figure (3).

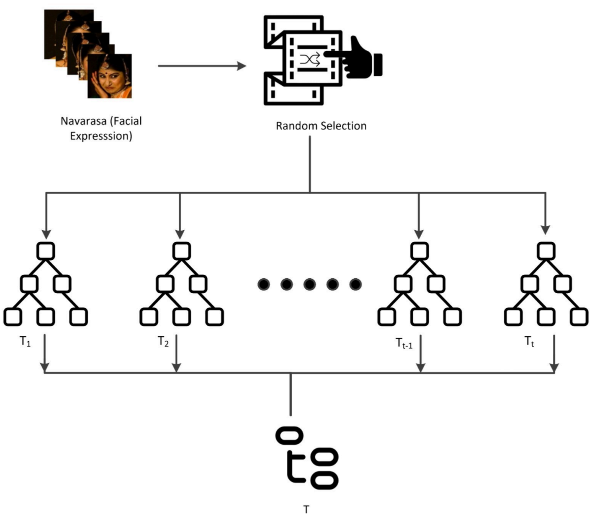

Fig. 3: Illustration of extra trees classifier

A mathematical definition of Extra Trees Classifier's step-by-step feature selection process follows. The task for Feature Selection: Consider a dataset with n samples, m features, represented by $X = \{x_1, x_2, ..., x_n\}$ and associated class labels as $y = \{y_1, y_2, ..., y_n\}$. Each x_i is a feature vector of length m. Extra Trees Classifier: The Extra Trees Classifier is a decision tree-based ensemble machine-learning technique. By choosing feature subsets at random, it builds several decision trees and utilizes majority voting to create predictions. The Extra Trees Classifier's feature importance scores can be computed based on how frequently a feature is used for splitting in the ensemble's decision trees during training. The relative relevance of each feature in the classification task is reflected in these essential ratings. We may choose a subset of the most pertinent features to the job using the Extra Trees Classifier's feature relevance ratings. The top-k characteristics with the highest scores determined, or the importance scores, are thresholded to define this subset. Mathematically, using the Extra Trees Classifier for feature selection can be defined as follows: To create an ensemble of decision trees and determine the feature importance scores, fit the Extra Trees Classifier on the training data X, y. The feature significance scores are given by the notation $I = \{i_1, i_2, ..., i_m\}$, where i_j denotes the importance score of the j^{th} feature. Applying a threshold or selecting the top-k features will allow you to choose a subset of features depending on the relevance rankings. Obtain the subset of the specified features with the formula $X^{'}X' = x'_1, x'_2, ..., x'_k$, where $x_i{}'$ is a feature vector of the selected features.

Materials and Methods

In this work, we are developing a framework to detect and label facial expressions, referred to as "navarasas" in Indian Classical Dance (ICD). Navarasa is a fundamental and essential dance element, representing the sentiment or emotional flavor that captivates the audience's interest. Identifying and understanding facial expressions plays a crucial role in interpreting the performer's roles and enhancing the overall enjoyment of the performance. Our framework will incorporate traditional object detection methods like YOLO and Detectron. However, as a real-time application, these conventional methods can introduce latency during inference, which is a significant concern. To address this, we propose a lightweight model that improves the preprocessing steps to optimize performance. The proposed design consists of two main components: A Navarasa classification model and an annotation model. These models work together to predict and identify the facial expressions displayed by the performer. The Navarasa classification model identifies all nine facial emotions, or navarasa, using transfer learning techniques in deep learning, allowing the framework to categorize facial expressions accurately. The detection model predicts the output along with the corresponding annotations—the architecture of the proposed framework illustrated in Figure (4).

In Figure (4), we illustrate the process in two phases. The first phase, Phase 1, involves Navarasa classification. This phase utilizes a facial expression (Navarasa) dataset and applies a series of data preprocessing techniques, including RGB conversion, resizing, and reshaping, followed by image augmentation. To enhance our model, we resourcefully employ off-the-shelf pre-trained models as feature extractors, forming the foundation for deep learning transfer learning. Next, we apply the Extra Trees classifier, a gain-based feature selection technique, to select the extracted features obtained through transfer learning. The chosen features serve as the input for the deep neural network classifier. The second phase focuses on the annotation of Navarasa, where the predictions from the deep neural network classifier are utilized for this purpose. We will provide a detailed and comprehensive discussion of both phases in the following sections, ensuring that you are fully informed and prepared for the upcoming content.

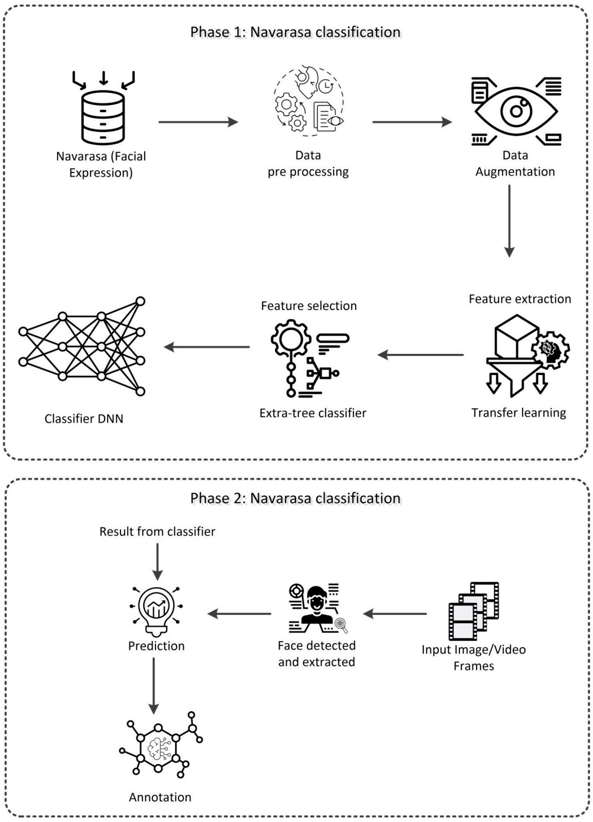

Fig. 4: The proposed architecture

Data Collection

As part of our research, we have focused on four South Indian classical dance forms. Due to the lack of publicly available datasets for facial expressions or Navarasa in Indian classical dance, we compiled a collection of handcrafted images from the Internet. We gathered 689 photos for our study by extracting frames from YouTube videos and other publicly accessible sources. The dataset features nine facial expressions in Indian classical dance: Shringar, Hasya, Karuna, Raudra, Veera, Bhayanaka, Bibhatsa, Adbhuta, and Shanta. We have evenly distributed the collected images across these nine classes. Although the dataset is limited, the classification tasks may be complex due to the varying number of dancers in each photo and the diverse backgrounds. Please refer to Figure (5) for the data distribution among the different classes.

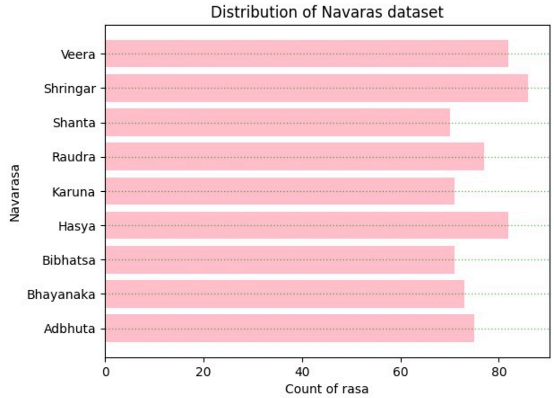

Fig. 5: Distribution of images in different classes

Data Preprocessing

The quality of the data used to train a neural network significantly affects the performance of an image categorization model. Image preprocessing allows us to focus on the features we want the neural network to learn by eliminating noise and distracting elements from the input images. The effectiveness of image recognition techniques relies heavily on the dataset's quality. In this study, we implement fundamental preprocessing steps, including RGB conversion, resizing, and reshaping. Resizing is particularly important in computer vision preprocessing, as deep learning models tend to train more efficiently on smaller images. In this research, we resize all input images to 128x128 pixels.

Data Augmentation

We are using OpenCV for image data augmentation to artificially expand our training set, providing our model with significantly more data for training. This technique generates new images from existing ones by applying minor modifications, such as adjusting brightness, rotating images, or shifting subjects horizontally or vertically. This process helps improve our model's accuracy by enhancing its ability to recognize new variations of the training data and also assists in preventing overfitting. Our study's standard image augmentation methods include vertical shifts, horizontal shifts, vertical flips, rotations, and zooming.

Implementation

Navarasa Classification Model

The initial component of our proposed architecture is the Navarasa classification model. The implementation

DOI: 10.3844/jcssp.2025.1364.1378

stages in the suggested pipeline for this model are in Figure (6). The process includes several steps: First, images from the dataset are preprocessed and augmented before being fed into the transfer learning model for feature extraction. Pre-trained models address significant challenges in domains such as computer vision, with applications in tasks like image recognition, object detection, and natural language processing. In this case, we utilize pre-trained transfer learning models specifically for feature extraction. Once the features are extracted, we apply an Extra Trees classifier for feature selection, which enhances predictive accuracy and helps control overfitting. These selected features are then used in a Deep Neural Network to produce the final classification results in our proposed pipeline.

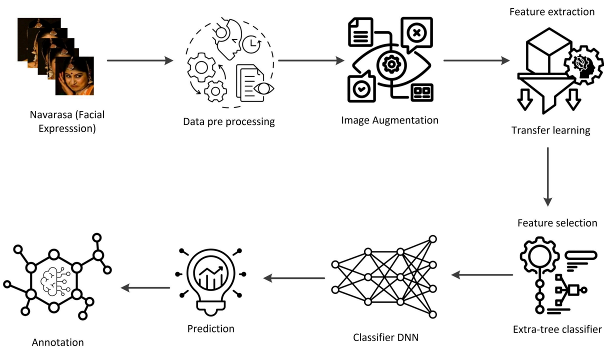

Fig. 6: Navarasa classification model

Feature Extraction

Feature extraction is one of the crucial tasks in the suggested framework. The secret to making classification challenges more successful is feature extraction. The most distinctive and instructive set of attributes is weighted to identify each image separately. In the case of manual feature extraction, creating and evaluating those takes a lot of work. We employ the Deep Learning method known as Transfer learning (Hussain *et al.*, 2019) for feature extraction in the suggested framework. A model developed through training on an extensive dataset, like ImageNet (Deng *et al.*, 2009), can be applied to an application with a smaller dataset. It is beneficial to eliminate the requirement for a more extensive dataset. In addition, it requires less training time than one from scratch. Here, pre-trained models VGG16 (Simonyan and Zisserman, 2014), ResNet50 (He *et al.*, 2015), DenseNet121 (Huang *et al.*, 2017), and EfficientNetB7 (Tan and Le, 2019) are employed. These models are used to learn features from the Navarasa (Facial expression) dataset after being trained on the ImageNet dataset.

Feature Selection

Feature selection methods help reduce the number of input variables while creating a predictive model. In some scenarios, reducing the number of input variables might enhance the model's efficiency while lowering the computing cost. Here, for feature selection, we employ an extra-tree classifiergeurts2006extremely. An extra-tree classifier applies to both classification and regression tasks. It is a meta-classifier that uses averaging to increase predictive accuracy and reduce over-fitting. It fits several randomized decision trees (also known as extra trees) on different sub-samples of the dataset. In scenarios where computational cost is a concern and features have to be carefully selected and analyzed, an ExtraTrees Classifier can be used.

Classification

The features selected using the ExtraTrees Classifier are fed into the classifier for this study, which addresses a multiclass classification problem focused on identifying nine emotions known as Navarasa in Indian classical dance. The nine classes are Shringar, Hasya, Karuna, Raudra, Veera, Bhayanaka, Bibhatsa, Adbhuta, and Shanta. We utilize a deep learning classifier called the Deep Neural Network (DNN). The proposed DNN model consists of a stack of layers, where the output of the first layer serves as the input to the next layer, continuing in this manner through the subsequent layers. The dense layers use the ReLU (Rectified Linear Unit) activation function. Since this is a multiclass classification problem, the last layer employs the softmax activation function, with categorical cross-entropy as the loss function and the Adam optimizer.

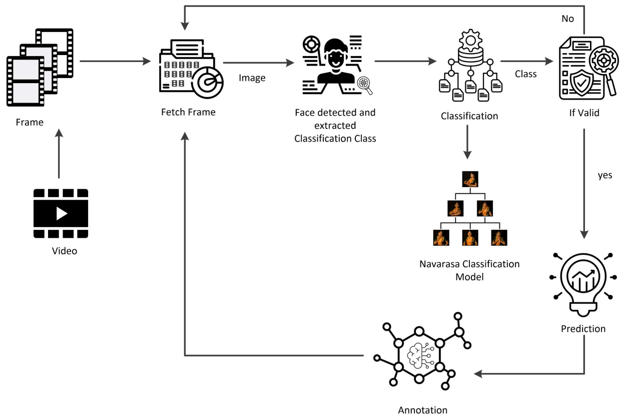

Fig. 7: Annotation model

Navarasa Recognition and Annotation

This phase intends to recognize and annotate the facial expression (Navarasa) from a video and image input. The pipeline following this stage is in Figure (7). A video is an image sequence made ready for processing after slicing the input videos into image frames. The relation between frames in a video is essential. The frame-fetch module draws a frame for processing from the image sequence, dissimilar to the previously processed image. It is rendered by finding the relative difference through frame-to-frame subtraction (He *et al.*,

DOI: 10.3844/jcssp.2025.1364.1378

2019). Those image frames will undergo additional processing if the discrepancy is relatively large. The frame selected up the frame-fetch module passed to the next step to check for the face of the performer or detect the face and extract it. The chosen frames are then fed into our Navarasa classification model to obtain the prediction. The classifier, in turn, predicts the input image into one of the nine classes. We then annotate the performer's expressions using the prediction findings. If the frame does not contain any identifiable navarasa class, then the frame is neglected, and the next frame is fetched from the video.

Results and Discussion

Tools and Implementation

In this study, we employed a variety of sophisticated tools and techniques to enhance the processes of data preprocessing, feature extraction, feature selection, and model evaluation. For data manipulation and visualization, we utilized well-known packages such as Matplotlib, NumPy, OpenCV, Pandas, and os, which allowed us to handle and analyze our datasets efficiently. To extract and select relevant features from the data, we leveraged the power of DenseNet121, VGG16, ResNet50, andEfficientNetB7, various deep learning architecture renowned for its efficacy in image processing tasks, alongside the Extra Trees Classifier, known for its robust performance in feature selection. This combination enabled us to identify and isolate the most informative characteristics of our data. For the transfer learning aspect of our study, we implemented Keras and TensorFlow. These deep learning libraries facilitated the use of pre-trained models, allowing us to build upon existing knowledge and optimize the learning process significantly. This approach improved the speed of our training and enhanced the model's overall accuracy. To ensure a comprehensive assessment of our model's performance, we generated a variety of evaluation metrics, including a confusion matrix. It enabled us to gain deeper insights into the model's predictive capabilities and identify areas for potential improvement. Overall, this meticulous approach to utilizing advanced tools and techniques contributed to the robust results of our study.

Experimental Results

This work introduces a novel deep-learning framework for identifying and annotating Navarasas in Indian Classical Dance. The sequence followed for the implementation is data collection, preprocessing input images with image processing methods, and applying pre-trained or transfer learning models to extract the features. The features are pegged using an extra tree classifier, and facial expressions are classified using a deep neural network. These predicted outcomes of the neural network are used to annotate the dancer's facial expression or Navarasa in that frame of the video or image. The images of Navarasa in various sizes, collected from the Internet and performed solo in Indian classical dance styles, comprise the dataset. Our framework's primary goal is to appropriately annotate and classify incoming photos and videos into the appropriate classes. Using the suggested Navarasa classification and annotation model, multiclass classification on the dataset begins. The outcomes that the recommended framework produced for the dataset are organized in this section.

Table 2: Comparison of the accuracy of pre-trained models + DNN (data augmentation)

Pre-trained Models	Accuracy (%) without data augmentation	Accuracy (%) with data augmentation
DenseNet121	73	82
VGG16	85	91
ResNet 50	87	91
EfficientNetB7	86	92

Table 3: Comparison of the accuracy of pre-trained models + DNN (feature selction)

Pre-trained Models	Accuracy (%) without feature selection	Accuracy (%) with feature selection
DenseNet121	69	79
VGG16	80	90
ResNet 50	82	90
EfficientNetB7	83	91

Table 4: Comparison of the accuracy of various image sizes

Image Sizes	Accuracy (%)
80X80	65
128X128	92
220X220	82
300X300	80

In the suggested framework, features are extracted from the dataset using pre-trained models; we engaged the task with various available models but got commendable results only with a few models. They are VGG16, ResNet50, DenseNet121, and EfficientNetB7; the feature selection task is meticulously performed using an extra-tree classifier. The Deep Neural Network (DNN) classifier was employed to classify the images into nine categories. The transfer learning model EfficientNetB7 for feature extraction and an extra-tree classifier for feature selection and classification using Deep Neural Network (DNN) achieved the best result of 92% accuracy among the various pre-trained models. A similar combination of transfer learning models with DNN VGG16 and ResNet50 achieved 91% % accuracy, whereas DenseNet121 achieved 82% % accuracy. Table 2 provides an overview of the accuracy of all pre-trained models plus DNN with and without image augmentation, which can be used to determine whether adding it to the dataset is meaningful. We can observe that the augmented outcome is superior to the natural one. Figure (8) also shows a summary of the outcome of Navarasa

DOI: 10.3844/jcssp.2025.1364.1378

Classification using the various pre-trained models in terms of accuracy, precision, recall, and f1-Score. Table (3) provides a similar overview of the accuracy of all pre-trained models plus DNN with and without feature selection, which can be employed to determine whether feature selection brought meaningful outcomes. Using the feature selection outcome is superior to the one without feature selection. Preprocessing techniques are used to remove noise from the data. Image resizing was done with Keras's help. The output obtained when performing our experiments with different image sizes is shown in Table (4).

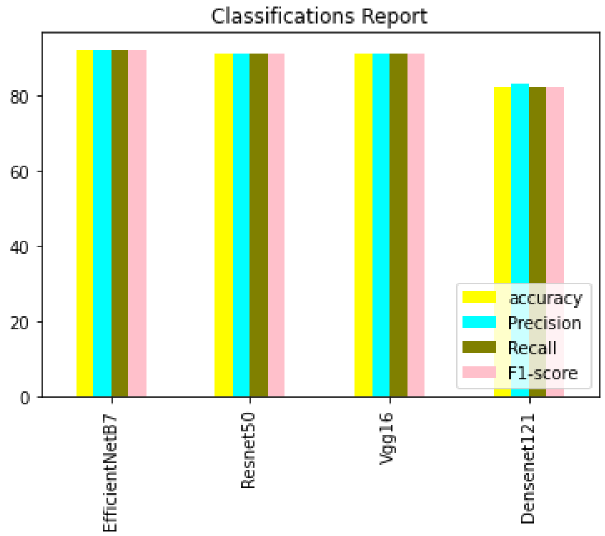

Fig. 8: Classification Results with pre-trained model plus DNN with data augmentation

Table 5: Comparison of proposed framework with related works

Year	Related Works	Method	Classes	Dataset	Accuracy
2018	Mohanty and Sahay	CNN model	8	Real-world Navarasa	83 . 9%
2018	Mohanty and Sahay	CNN on the augmented real-world database.	8	Real-world Navarasa	88%
2019	Kale *et al.*	LBP + linear SVM	7	Handcrafted images captured in a controlled environment	81.38%
2021	Selvi *et al.*	CNN	9	Kathakali Face expression	90%
	Proposed Framework	EfficientNetB7 + extra-tree classifier + DNN	9	Handcrafted Navarasa dataset	92%

Table (5) compares the associated studies in Navarasa classification or recognizing Navarasa facial expressions. While comparing with other similar works found in related works in identifying facial expressions or Navarasa in Indian Classical dance, our proposed framework attained a slighter, better accuracy of 92%. Mohanty and Sahay (2018) did a detailed study with eight facial expressions but obtained better results with the dataset they collected in the controlled laboratory environment rather than the real-world navarasa dataset. As ours is a handcrafted dataset complied with images from real-time performances in publically available sources, we choose only those related works with such datasets. In the comparison table, two methods by Mohanty and Sahay (2018). tested on the real-world navarasa dataset and augmented real-world navarasa dataset, which obtained 83.9 and 88%, respectively. Kale and Rege (2019) got a result of 81.38% with handcrafted images captured in a controlled environment. They classified the expressions into seven classes using LBP for feature extraction and linear SVM. Selvi *et al.* (2021) classified Kathakali facial expressions into nine categories using CNN and achieved an accuracy of 90%.

In Indian classical dance, the facial expressions on the performer's face are termed Navarasa, the nine rasas, namely Shringar, Hasya, Karuna, Raudra, Veera, Bhayanaka, Bibhatsa, Adbhuta, and Shanta. In the Navarasa classification model, these Rasas are the nine classes. A summary of the per-class classification of these rasas in terms of precision, recall, and f1-Score these metrics for evaluating the performance of our classification model shown in the graphs in Figure (9) the classification result with EfficientNetB7, ResNet50, VGG16, and DenseNet121.

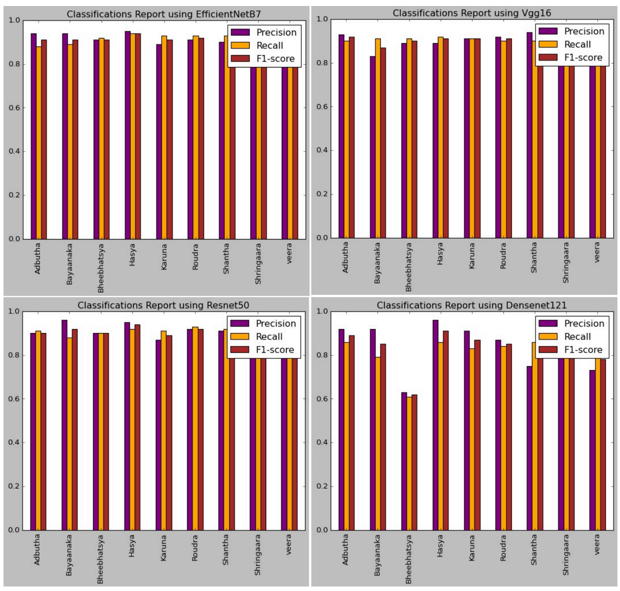

Fig. 9: Per-class Classification Results with different pre-trained models plus DNN with data augmentation

Figure (10) shows the output with annotation from the video, which gets sliced into images as in a frame and then identifies the Rasa depicted.

Figure (11) shows the (a) accuracy and (b) loss curves for the best accuracy achieved combo in our framework, EfficientNetB7, with extra tree and DNN for classification. A loss curve during training is one of the most often used graphs to troubleshoot a neural network. It gives us an overview of the training procedure and how the network learns. It is superior to iteration for plotting loss across epochs. An additional graph often used curve to understand the evolution of neural networks is the accuracy curve. The loss function is computed over all data items throughout an epoch and guaranteed to deliver the quantitative loss measure at the designated epoch.

Fig. 10: Navarasa detection in image

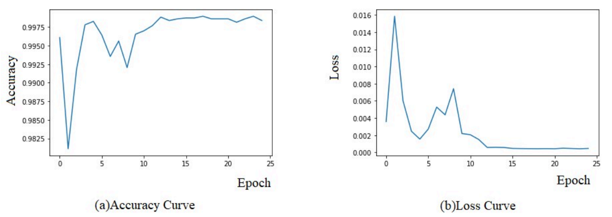

Fig. 11: Accuracy and loss curve

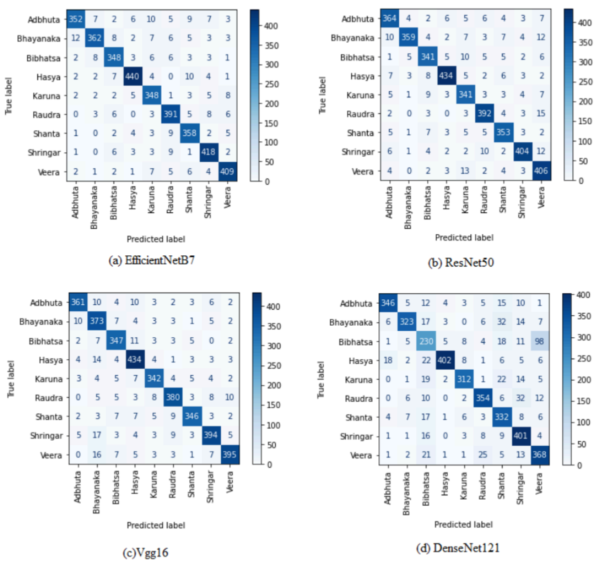

Fig. 12: Confusion matrix of pre-trained models

The confusion matrix can offer a holistic understanding of the neural network model predictions. We have the predicted labels in the confusion matrix on the x-axis; on the y-axis, we have the correct results. The diagonal represents the actual categorization values of this nine-class problem, with values in the boxes representing each class's count—the outcome of integrating various pre-trained models with DNN in Figure (12). The confusion matrices for four deep learning models—EfficientNetB7, ResNet50, VGG16, and DenseNet121—evaluated on a multiclass classification task involving nine emotion categories: Adbhuta, Bhayanaka, Bibhatsa, Hasya, Karuna, Raudra, Shanta, Shringar, and Veera. Each confusion matrix visualizes the distribution of accurate and predicted labels, with diagonal values representing correct predictions. EfficientNetB7 and ResNet50 exhibit higher accuracy across most categories, with a notably strong performance in the "Hasya" and "Karuna" classes, as seen by high diagonal values. VGG16 demonstrates good classification accuracy but shows slightly more misclassifications in certain classes, such as "Raudra" and "Shanta." DenseNet121 has a relatively higher confusion rate among the "Bibhatsa" and "Karuna" categories, as indicated by off-diagonal values. These matrices highlight the comparative effectiveness of the models, with EfficientNetB7 and ResNet50 demonstrating the most consistent performance across the categories. Table 6 shows the result created by the video annotation model of three sample videos.

Table 6: Result of video annotation model

File name	Time	Navarasa
Video1.mp4	00:02:44	Shringar
	00:02:21	Hasya
	00:02:04	Karuna
	00:01:40	Raudra
	00:01:25	Veera
	00:01:07	Bhayanaka
	00:00:43	Bibhatsa
	00:00:28	Adbhuta
	00:00:10	Shanta
Video2.mp4	00:02:00	Shringar
	00:01:48	Veera
	00:01:37	Karuna
	00:01:32	Adbhuta
	00:01:06	Hasya
	00:01:55	Bhayanaka
	00:00:33	Bibhatsa
	00:00:22	Raudra
	00:00:05	Shanta
Video.mp3	00:01:26	Bibhatsa
	00:01:17	Bhayanaka
	00:01:09	Raudra
	00:01:02	Veera
	00:00:48	Adbhuta
	00:00:35	Karuna
	00:00:31	Hasya
	00:00:23	Shringar
	00:00:18	Shanta

DOI: 10.3844/jcssp.2025.1364.1378

Conclusion

Emotion recognition is a domain that spans diverse fields like psychology, cognitive science, and computer science. Most researchers focused on the conventional approach of identifying emotions from Text-based data. But now, with the advances in artificial intelligence and computer vision, other types of data input like images, videos, and audio are highly analyzed for emotion recognition. In this study, we designed a framework capable of recognizing and annotating navarasa (nine facial expressions) or emotions from Indian Classical Dance using deep learning techniques. This framework will aid in digitally understanding and preserving our cultural heritage, which helps safeguard this knowledge for future generations. One major challenge for the study was the dataset; we created a navarasa dataset with handcrafted images. However, we have another problem: Less data for training. To overcome this, we used pre-trained models and data augmentation methods, which helped improve the classifier's accuracy in classifying images in this multiclass classification task. This framework gained commendable accuracy in organizing and annotating facial expressions in Indian Classical Dance.

Acknowledgment

We express our heartfelt gratitude to Cochin University of Science and Technology for providing us with this academic opportunity, and to the Department of Computer Applications for their continuous support and invaluable services throughout our work. We are particularly thankful to the editorial team for their diligent efforts in reviewing, refining, and facilitating the publication of our study. We also wish to acknowledge the resources and platforms made available to us, which have allowed our research to reach a wider audience. Through this publication, we hope to contribute meaningfully to the digital preservation of cultural heritage. Our proposed framework, aimed at improving the accuracy of organizing and annotating facial expressions in Indian Classical Dance, aspires to enhance digital understanding and inspire future research in this significant and nuanced domain, safeguarding this knowledge for future generations.

Funding Information

This study received no external financial support, grants, institutional funding, or contributions from third parties for its conception, execution, or publication.

Author's Contributions

Reshma M R: Developed the research idea and study framework, collected and organized the dataset, and contributed to the methodology. Performed data analysis and validation, interpreted the results, and visualized the findings. Drafted the manuscript, conducted the literature review, and coordinated revisions based on reviewer feedback.

Kannan B: Devised the research idea, designed the study framework, coordinated the data analysis, and supervised the overall project execution.

Jagathy Raj V P: Designed the study, coordinated the data analysis, and supervised the overall project execution.

Shailesh S: Envisioned the research idea, performed data analysis and experimental validation, and supervised the overall project execution.

All authors reviewed, contributed to, and approved the final version of the manuscript.

Ethics

The authors affirm that this manuscript is original, authored by them, contains unpublished material, has not been submitted elsewhere, accurately represents their research, is effectively contextualized within existing literature, and does not involve any ethical issues.

References

Ajili, I., Ramezanpanah, Z., Mallem, M., & Didier, J.-Y. (2019). Expressive motions recognition and analysis with learning and statistical methods. *Multimedia Tools and Applications*, *78*(12), 16575-16600. https://doi.org/10.1007/s11042-018-6893-5

Aristidou, A., Chalmers, A., Chrysanthou, Y., Loscos, C., Multon, F., Parkins, J., Sarupuri, B., & Stavrakis, E. (2022). Safeguarding our Dance Cultural Heritage. *In Eurographics 2022-43nd Annual Conference of the European Association for Computer Graphics*, 1-6. https://doi.org/10.2312/egt.20221055

Aristidou, A., Zeng, Q., Stavrakis, E., Yin, K., Cohen-Or, D., Chrysanthou, Y., & Chen, B. (2017). Emotion control of unstructured dance movements. *Proceedings of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 1-10. https://doi.org/10.1145/3099564.3099566

Barros, P., Jirak, D., Weber, C., & Wermter, S. (2015). Multimodal emotional state recognition using sequence-dependent deep hierarchical features. *Neural Networks*, *72*, 140-151. https://doi.org/10.1016/j.neunet.2015.09.009

Batbaatar, E., Li, M., & Ryu, K. H. (2019). Semantic-Emotion Neural Network for Emotion Recognition From Text. *IEEE Access*, *7*, 111866-111878. https://doi.org/10.1109/access.2019.2934529

Bazrafkan, S., Nedelcu, T., Filipczuk, P., & Corcoran, P. (2017). Deep learning for facial expression recognition: A step closer to a smartphone that knows your moods. *2017 IEEE International Conference on Consumer Electronics (ICCE)*, 217-220. https://doi.org/10.1109/icce.2017.7889290

Bharata, & Rangacharya, A. (1996). *English translation with critical*.

Bhavanam, L. T., & Neelakanta Iyer, G. (2020). On the Classification of Kathakali Hand Gestures Using Support Vector Machines and Convolutional Neural Networks. *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, 1-6. https://doi.org/10.1109/aisp48273.2020.9073398

Black, M. J., & Yacoob, Y. (2002). Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. *Proceedings of IEEE International Conference on Computer Vision*, 374-381. https://doi.org/10.1109/iccv.1995.466915

Cid, F., Prado, J. A., Bustos, P., & Nunez, P. (2013). A real time and robust facial expression recognition and imitation approach for affective human-robot interaction using Gabor filtering. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2188-2193. https://doi.org/10.1109/iros.2013.6696662

Cohn, J. F., Zlochower, A. J., Lien, J. J., & Kanade, T. (2002). Feature-point tracking by optical flow discriminates subtle differences in facial expression. *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 396-401. https://doi.org/10.1109/afgr.1998.670981

Cros, H., & McKercher, B. (2020). *Cultural Tourism*. https://doi.org/10.4324/9780429277498

Dan, N. V. (2012). Artistic Communication and Dance Sport Particularities. *Procedia - Social and Behavioral Sciences*, *46*, 4869-4873. https://doi.org/10.1016/j.sbspro.2012.06.351

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255. https://doi.org/10.1109/cvprw.2009.5206848

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(10), 974-989. https://doi.org/10.1109/34.799905

Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist*, *48*(4), 384-392. https://doi.org/10.1037//0003-066x.48.4.384

Fan, R., Xu, S., & Geng, W. (2012). Example-Based Automatic Music-Driven Conventional Dance Motion Synthesis. *IEEE Transactions on Visualization and Computer Graphics*, *18*(3), 501-515. https://doi.org/10.1109/tvcg.2011.73

Guo, F., Li, F., Lv, W., Liu, L., & Duffy, V. G. (2020). Bibliometric Analysis of Affective Computing Researches during 1999~2018. *International Journal of Human-Computer Interaction*, *36*(9), 801-814. https://doi.org/10.1080/10447318.2019.1688985

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *ArXiv Preprint Arxiv:1512.03385*.

He, Z., Jin, T., Basu, A., Soraghan, J., Di Caterina, G., & Petropoulakis, L. (2019). Human Emotion Recognition in Video Using Subtraction Pre-Processing. *Proceedings of the 2019 11th International Conference on Machine Learning and Computing*, 374-379. https://doi.org/10.1145/3318299.3318321

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700-4708. https://doi.org/10.1109/cvpr.2017.243

Hussain, M., Bird, J. J., & Faria, D. R. (2019). A Study on CNN Transfer Learning for Image Classification. *Advances in Computational Intelligence Systems*, *840*, 191-202. https://doi.org/10.1007/978-3-319-97982-3_16

Jain, N., Bansal, V., Virmani, D., Gupta, V., Salas-Morera, L., & Garcia-Hernandez, L. (2021). An Enhanced Deep Convolutional Neural Network for Classifying Indian Classical Dance Forms. *Applied Sciences*, *11*(14), 6253. https://doi.org/10.3390/app11146253

Jiang, H., & Yan, Y. (2024). Sensor based dance coherent action generation model using deep learning framework. *Scalable Computing: Practice and Experience*, *25*(2), 1073-1090. https://doi.org/10.12694/scpe.v25i2.2648

Joshi, M., & Chakrabarty, S. (2021). An Extensive Review of Computational Dance Automation Techniques and Applications. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *477*(2251), 20210071. https://doi.org/10.1098/rspa.2021.0071

Kale, M. R., & Rege, P. P. (2019). Classification of expressions in indian classical dance using lbp. *In 2019 Ieee 16th India Council International Conference (Indicon). IEEE*, 1-4. https://doi.org/10.1109/INDICON47234.2019.9029006

Khalifa, H., Babiker, B., Goebel, R., & Cheng, I. (2017). Facial expression recognition using SVM classification on mic-macro patterns. *2017 IEEE International Conference on Image Processing (ICIP)*, 1757-1761. https://doi.org/10.1109/icip.2017.8296583

Kipp, M., & Martin, J.-C. (2009). Gesture and emotion: Can basic gestural form features discriminate emotions? *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 1-8. https://doi.org/10.1109/acii.2009.5349544

Koolagudi, S. G., & Rao, K. S. (2012). Emotion Recognition from Speech: A Review. *International Journal of Speech Technology*, *15*(2), 99-117. https://doi.org/10.1007/s10772-011-9125-1

Li, X., Karuppiah, M., & Shanmugam, B. (2021). Psychological Perceptual Analysis Based on Dance Therapy Using Artificial Intelligence Techniques. *International Journal on Artificial Intelligence Tools*, *30*(06n08), 2140012. https://doi.org/10.1142/s0218213021400121

Lin-Lin Shen, & Zhen Ji. (2008). Modelling geiometric features for face based age classification. *2008 International Conference on Machine Learning and Cybernetics*. 2008 International Conference on Machine Learning and Cybernetics (ICMLC), Kunming, China. https://doi.org/10.1109/icmlc.2008.4620909

Luck, G., Saarikallio, S., Burger, B., Thompson, M., & Toiviainen, P. (2014). Emotion-driven encoding of music preference and personality in dance. *Musicae Scientiae*, *18*(3), 307-323. https://doi.org/10.1177/1029864914537290

Mallick, T., Das, P. P., & Majumdar, A. K. (2022). Posture and sequence recognition for Bharatanatyam dance performances using machine learning approaches. *Journal of Visual Communication and Image Representation*, *87*, 103548. https://doi.org/10.1016/j.jvcir.2022.103548

Maret, Y., Oberson, D., & Gavrilova, M. (2018). Identifying an Emotional State from Body Movements Using Genetic-Based Algorithms. *Artificial Intelligence and Soft Computing*, *10841*, 474-485. https://doi.org/10.1007/978-3-319-91253-0_44

Mehta, D., Siddiqui, M. F. H., & Javaid, A. Y. (2019a). Recognition of emotion intensities using machine learning algorithms: A comparative study. *Sensors*, *19*(8), 1897. https://doi.org/10.1109/ICISC.2018.8398861

Mehta, D., Siddiqui, M. F. H., & Javaid, A. Y. (2019b). Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study. *Sensors*, *19*(8), 1897. https://doi.org/10.3390/s19081897

Mohanty, A., & Sahay, R. R. (2018). Rasabodha: Understanding Indian classical dance by recognizing emotions using deep learning. *Pattern Recognition*, *79*, 97-113. https://doi.org/10.1016/j.patcog.2018.01.035

Nandikesvara, Coomaraswamy, A. K., & Duggirala, G. K. (1970). *The mirror of gesture: being the Abhinaya darpaṇa of Nandikeśvara*.

Newmark, C. (2022). Charles Darwin: The Expression of the Emotions in Man and Animals. *Schlüsselwerke Der Emotionssoziologie*, 111-115. https://doi.org/10.1007/978-3-658-37869-1_12

Nikolaus, J., Thomas, R., & Thomas, P. (2016). 3. Effects of Verbal and Non-Verbal Elements in Communication. *Verbal Communication*, 39-54. https://doi.org/10.1515/9783110255478-004

Noll, A. M., & Hutchinson, A. (1967). Choreography and Computers. *Dance Magazine*, *41*(1), 43-45.

Pandeya, Y. R., Bhattarai, B., & Lee, J. (2021). Deep-Learning-Based Multimodal Emotion Classification for Music Videos. *Sensors*, *21*(14), 4927. https://doi.org/10.3390/s21144927

Pantie, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1424-1445. https://doi.org/10.1109/34.895976

Reshma, M. R., & Kannan, B. (2019). Approaches on Partial Face Recognition: A Literature Review. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 538-544. https://doi.org/10.1109/icoei.2019.8862783

Reshma, M. R., Kannan, B., Jagathy Raj, V. P., & Shailesh, S. (2023). Cultural heritage preservation through dance digitization: A review. *Digital Applications in Archaeology and Cultural Heritage*, *28*, e00257. https://doi.org/10.1016/j.daach.2023.e00257

Samal, A., & Iyengar, P. A. (1992). Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. *Pattern Recognition*, *25*(1), 65-77. https://doi.org/10.1016/0031-3203(92)90007-6

Selvi, C., Anvitha, Y., Asritha, C. H., & Sayannah, P. B. (2021). Kathakali face expression detection using deep learning techniques. *Journal of Physics: Conference Series*, *2062*(1), 012018. https://doi.org/10.1088/1742-6596/2062/1/012018

Shailesh, S., & Judy, M. V. (2020). Automatic Annotation of Dance Videos Based on Foot Postures. *Indian Journal of Computer Science and Engineering*, *11*(1), 89-98. https://doi.org/10.21817/indjcse/2020/v11i1/20110 1047

Shikanai, N., & Hachimura, K. (2015). Evaluation of impressions and movements related to negative emotional expressions in dance. *2015 15th International Conference on Control, Automation and Systems (ICCAS)*. 2015 15th International Conference on Control, Automation and Systems (ICCAS), Busan, Korea (South). https://doi.org/10.1109/iccas.2015.7365000

Silaparasetty, V. (2020). *Deep Learning Projects Using TensorFlow 2*. https://doi.org/10.1007/978-1-4842-5802-6

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv*, 1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Sravya, A. S. L., Chinmayi, K. S., Yoshita, L. N., V, S., & Manitha, P. V. (2019). A Smart Aid for Visually Challenged in Face Detection and Recognition. *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 704-708. https://doi.org/10.1109/icssit46314.2019.8987853

Srimani, P., & Hegde, R. (2012). Analysis of facial expressions with respect to navarasas in bharathanatym styles using image processing. *International Journal of Knowledge Engineering*, *3*(2), 193.

Sun, G., Wong, Y., Cheng, Z., Kankanhalli, M. S., Geng, W., & Li, X. (2021). DeepDance: Music-to-Dance Motion Choreography With Adversarial Learning. *IEEE Transactions on Multimedia*, *23*, 497-509. https://doi.org/10.1109/tmm.2020.2981989

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105-6114. https://doi.org/10.48550/arXiv.1905.11946

Tian, Y.-I., Kanade, T., & Cohn, J. F. (2001). Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(2), 97-115. https://doi.org/10.1109/34.908962

Valstar, M., & Pantic, M. (2006). Fully Automatic Facial Action Unit Detection and Temporal Analysis. *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA. https://doi.org/10.1109/cvprw.2006.85

Vijayan, K. (2012). Mohiniyattam: Ariyendath ellam. *DC Publications*.

Wang, S., Li, J., Cao, T., Wang, H., Tu, P., & Li, Y. (2020). Dance Emotion Recognition Based on Laban Motion Analysis Using Convolutional Neural Network and Long Short-Term Memory. *IEEE Access*, *8*, 124928-124938. https://doi.org/10.1109/access.2020.3007956

Wang, S., & Tong, S. (2022). Analysis of high-level dance movements under deep learning and internet of things. *The Journal of Supercomputing*, *78*(12), 14294-14316. https://doi.org/10.1007/s11227-022-04454-y

Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional State Classification from EEG Data Using Machine Learning Approach. *Neurocomputing*, *129*, 94-106. https://doi.org/10.1016/j.neucom.2013.06.046

Whitehill, J., Bartlett, M. S., & Movellan, J. R. (2013). Automatic Facial Expression Recognition. *Social Emotions in Nature and Artifact*, 88-109. https://doi.org/10.1093/acprof:oso/9780195387643.003.0007

Wu, C.-H. (2016). New technology for developing facial expression recognition in e-learning. *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*. 2016 Portland International Conference on Management of Engineering and Technology (PICMET), Honolulu, HI, USA. https://doi.org/10.1109/picmet.2016.7806648

Xu, T., White, J., Kalkan, S., & Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. *Computer Vision - ECCV 2020 Workshops*, 506-523. https://doi.org/10.1007/978-3-030-65414-6_35

Zacharatos, H., Gatzoulis, C., & Chrysanthou, Y. L. (2014). Automatic Emotion Recognition Based on Body Movement Analysis: A Survey. *IEEE Computer Graphics and Applications*, *34*(6), 35-45. https://doi.org/10.1109/mcg.2014.106

Zhao, M., Adib, F., & Katabi, D. (2016). Emotion recognition using wireless signals. *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. MobiCom'16: The 22nd Annual International Conference on Mobile Computing and Networking, New York City New York. https://doi.org/10.1145/2973750.2973762