Classification, Tagging, and Object Detection in Indian Folk Paintings

Thesis submitted in partial fulfillment
of the requirements for the degree of

*Master of Science in **Computer Science and Engineering** by Research*

by

Nancy Hada
2021701016
nancy.hada@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
July 2024

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Classification, Tagging, and Object Detection in Indian Folk Paintings " by Nancy Hada, has been carried out under my supervision and is not submitted elsewhere for a degree.

July 2024	Adviser: Prof. Kavita Vemuri

To my Friends and Family

Acknowledgments

I would like to express my heartfelt gratitude to all those that accompanied me during my research endeavor in the previous three years. This period has been an extraordinary experience filled with invaluable education, personal growth, and treasured memories.

First and foremost, I would want to convey my deep appreciation and gratitude to Prof. Kavita Vemuri. Her support was crucial in helping me achieve my dreams and reach unprecedented heights. I express my gratitude for the considerable amount of time and patience you have dedicated to addressing my inquiries, regardless of their trifling nature. Furthermore, I appreciate your attentive and patient listening to all the matters I wished to discuss.

I would like to express my appreciation to my exceptional squad, including Pragati, Ayushi, Supreeth, Tathagata, Aditya, Bhoomeendra, Amruth, Sravani, Dhruv, Dhaval, and Nayan, for transforming the challenging academic journey into an enjoyable experience. From the exclusive nocturnal talks to the tireless research submissions, we accomplished everything as a collective. They are not merely pals; they are my scholarly kin. We experienced amusement throughout both the positive and negative moments, commemorated successes, and transformed anxiety into a humorous performance. Thank you for being the source of positivity and inspiration during my challenging study process and for being the driving force behind my remarkable research adventure. You are the true protagonists of this academic narrative!

I express profound appreciation to my unwavering family - Raghav, Mom, and Dad; you served as my enthusiastic supporters, my emotional support system, and the driving force behind my perseverance. Your steadfast support and ceaseless encouragement served as the guiding compass that led me across unexplored realms of investigation. I am grateful for your unwavering support and guidance.

Abstract

Indian folk paintings are characterized by a rich mosaic of symbols, colors, textures, and stories. These serve as invaluable repositories of cultural legacy. This thesis presents a comprehensive approach to classifying these paintings into distinct art forms and tagging them with their unique salient features. Two datasets namely *FolkTalent* and *WarliScan*, are presented. 2279 digital photos of twelve distinct Indian folk art forms make up the *FolkTalent* dataset. Photos are taken from websites that are direct online sellers of these artworks. GPT-4 generated followed by an expert-reviewed tags, including a broad spectrum of characteristics like color, theme, artistic style, and patterns, are appended to each artwork. On fined-tuned Convolutional Neural Network (CNN) models, classification is carried out with a remarkable accuracy of 91.83% using the RandomForest ensemble method. Deeper insights into the paintings and improved search experiences depending on thematic and visual characteristics are made possible by tagging through well-adjusted CNN-based backbones with a proprietary classifier for multi-label picture classification.

Furthermore, *WarliScan* is presented, a unique dataset consisting of 250 digital scans of *Warli* paintings, each labeled and with exact coordinates for each unique object shown in the artwork. Originating in an Indian folk art style, *Warli* paintings are cultural guides as well as artistic expressions that narrate basic stories about the *Warli* community. *WarliScan* was created in order to help construct models in future for automatically verifying the authenticity of these artworks because there were no corpora with comprehensive annotations. A Mean Average Precision score of 0.585 was obtained by optimizing the *YoloV8n* model to create an object detection baseline that demonstrates the effectiveness of this dataset.

The suggested hybrid paradigm and the combined efforts in producing the *FolkTalent* and *WarliScan* datasets established new standards for the categorization and labeling of Indian folk paintings, therefore greatly advancing the cataloging and preservation of India's folk-art legacy.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Indian folk art is a rich source of cultural history, traditions, and social mores of many tribes nationwide. These art pieces are both creative manifestations and historical records of cultural tales. They have unique styles, symbols, colors, and textures. When examining computer vision, namely in the areas of object detection, labelling, and categorization, the intricate and varied characteristics of Indian folk art present both benefits and challenges. By creating thorough methods to accurately categorize, name, and identify objects in Indian traditional art, our work seeks to get beyond these challenges. This work achieved major progress by building two datasets, *FolkTalent* and *WarliScan*, and by devising novel techniques specially tailored to the unique characteristics of these artworks.

1.1 Classification of Indian Folk Paintings

1.1.1 Importance of Classification

Classification is a fundamental computer vision problem where a picture is assigned a label/name based on its contents. Precise categorization of Indian folk paintings is necessary because of the following reasons:

- **Preservation of Cultural Heritage:** Every folk art form reflects the cultural history of a certain area. This heritage is preserved in part because correct classification ensures that artworks are appropriately identified and cataloged.

- **Facilitation of Research:**Classification techniques that facilitate the access to and analysis of enormous painting collections by academics studying Indian folk art may be highly beneficial.

- **Improved User Experience:** Finding specific folk art forms becomes easier when accurate classification enhances the search and retrieval process for buyers, collectors, and art enthusiasts.

1.1.2 Challenges in Classification

The variety of forms and traits in Indian folk art makes classification fairly difficult. Folk art is so diverse in style, each having unique characteristics, that no one model can adequately classify it all. Classification gets harder since every style, like *Madhubani*, *Warli*, and *Pattachitra*, has different visual elements and background circumstances. This is because, although the main characteristics are different, many folk art forms have overlapping influences or aspects because of cultural interactions.

For example, the use of natural colors could be typical across several genres, while narrative themes from Hindu mythology could show up both in Madhubani and Pattachitra art. Particularly when artists combine aspects from several traditions, this overlap can make it difficult to categorize a work just depending on one or two visual traits. Moreover, many folk arts feature minute elements and subtle nuances with cultural relevance that could be absolutely important for comprehending the whole background and classification of the work. More general or simplified systems may overlook these subtleties.

1.2 Tagging of Indian Folk Paintings

1.2.1 Importance of Tagging

Tagging involves associating images with descriptive labels that capture key attributes and features. For Indian folk paintings, tagging is essential for several reasons:

- **Enhanced Search and Retrieval:** Detailed tags enable users to search for paintings using specific attributes, such as colors, themes, or patterns. This improves the search experience for buyers, collectors, and researchers.

- **Deeper Insight:** Tags provide a deeper understanding of the content and context of the paintings, highlighting the unique elements and narratives embedded in the artworks.

- **Facilitation of Authentication:** Tags can be used to verify the authenticity of paintings by checking for consistency with known attributes and features of specific art forms.

1.2.2 Challenges in Tagging

Indian folk paintings are complicated and intricate, which makes accurate and thorough tagging necessary to convey their spirit difficult. Many times, each picture has a plethora of complex patterns, symbols, and motifs that are difficult to tag all the way through but essential for accurate classification. Still, an additional degree of complication is added by semantic variance among various art forms. In *Madhubani*, *Warli*, and *Pattachitra*, same symbols or motifs may be utilized in several situations, hence sophisticated labeling is necessary to distinguish between them. To be sure these tags are accurate and applicable, professionals must manually verify them. This tedious manual verification preserves the accuracy and integrity of the tagging procedure and emphasizes the considerable effort needed to properly organize and categorize the vast and intricate universe of Indian folk paintings.

1.3 Significance of FolkTalent Dataset

The *FolkTalent* dataset is one important contribution to the solving of the categorization issues with Indian folk art. This dataset is a great source of training and evaluating classification algorithms with 2279 digitized images of twelve distinct Indian folk art genres. The dataset includes images collected from websites that directly sell these artworks. They are all categorized by color, theme, artistic style, and patterns. Upon generation from GPT-4 the tags are verified by expert. These tags are an essential instrument for the classification process and guarantee the dataset's correctness and relevancy.

This paper presents classification techniques that use the *FolkTalent* dataset to accurately classify Indian folk art. Combining a *RandomForest* ensemble technique with properly tuned Convolutional Neural Network (CNN) models, the models achieve an astounding classification accuracy of 91.83%. Given the perfectly balanced data distribution across classes shown in Figure 3.2), this high degree of accuracy validates the effectiveness of the categorization methods and the robustness of the *FolkTalent* dataset.

In addition to providing classification for Indian folk paintings, the *FolkTalent* dataset also provides a plethora of data for labeling these works of art. For each image in the dataset, attributes such as themes, colors, patterns, and salience components are annotated in the form of tags. Expert verification ensures that the generated tags are accurate and relevant.

The tagging methods used in this work have produced accurate tags for each painting in the *Folk-Talent* dataset. The models get a remarkable mean Average Precision (mAP) score of 84.15% on the validation dataset by using a multi-label classification strategy. This high level of tagging accuracy makes authentication simpler and offers deeper insights into the paintings, improving the overall search and retrieval user experience.

1.4 Object Detection in Warli Paintings

1.4.1 Importance of Object Detection

Finding and identifying objects in an image is known as object detection. Specifically for *Warli* paintings, which are well-known for their straightforward yet expressive style for a variety of reasons, object recognition is both vital and challenging. It is interesting to detect unique objects from *Warli* paintings which are created using basic geometric shapes. Several objects like a man and cow are make of triangles still having different meanings. The effort on identifying objects from *Warli* paintings was motivated by the following:

- **Authentication:** Confirming the authenticity of *Warli* paintings could require identifying and analyzing the unique elements and subjects they contain. With the evolution of AI, this can also help in detecting deep fakes.

- **Cultural Preservation:** The work precisely identifies and records the objects present in *Warli* paintings. This can help recording and preserving the culture of the *Warli* community by increasing focus on minute objects presented with such simplicity.

- **Enhanced Understanding:** Through the emphasis of its symbolic and narrative components, object detection offers a thorough picture of the composition and subject matter of *Warli* paintings.

1.4.2 Challenges in Object Detection

Various challenges arise with object recognition in *Warli* paintings due to their unique artistic aspects. *Warli* paintings pose a serious challenge for traditional object detection algorithms since they express landscapes and objects exclusively using basic geometric forms and lines. Typically trained on more complex, detailed images, and real world images, these models fail to identify and comprehend these fundamental patterns. Moreover, *Warli* paintings often show objects symbolically rather than literally, thus models have to understand abstract patterns and shapes, which is a problem that is outside the scope of standard detection techniques. A significant challenge to object detection model training and evaluation is the dearth of annotated materials, particularly for *Warli* paintings. This absence of specific data makes it impossible to build models that can accurately and reliably identify objects within *Warli* art, which emphasizes the need for tailored techniques and datasets to address these specific challenges.

1.4.3 Significance of WarliScan Dataset

The *WarliScan* dataset offers a complete item detection resource, therefore addressing the demand for annotated data in *Warli* paintings. With 250 digital scans of *Warli* paintings, each labeled and with precise coordinates for each unique object seen in the artworks, the *WarliScan* dataset is an invaluable resource for object identification model training and evaluation.

A high accuracy was achieved in object detection upon annotation in *Warli* paintings. It is achieved by the fine-tuning the object detection model created in using the *WarliScan* dataset. The model obtained a Mean Average Precision (mAP) score of 0.585 by fine tuning the *YoloV8n* model. This score proves the *WarliScan* dataset's robustness and the efficacy of the item detection techniques.

1.5 Key Contributions

The key contributions of this study can be summarized as follows:

- ***FolkTalent* Dataset Creation:** 2279 digital images of twelve distinct Indian folk art forms make up this massive resource for training and evaluating classification and tagging methods. The wealth of information tagged to every picture makes the dataset more valuable for research and practical applications.

- ***WarliScan* Dataset Creation:** The *WarliScan* dataset was created as a helpful resource for object detection model training and evaluation. It includes 250 digitally scanned *Warli* paintings together with detailed annotations. This dataset addresses the lack of annotated data for *Warli* paintings, therefore helping to build models appropriate for this unique art style.

- **Model development for classification:** This study uses a *RandomForest* ensemble strategy and well calibrated CNN models to reach a high classification accuracy of 91.83%. The classification methods developed in this work being modified to the unique characteristics of Indian folk paintings improve their effectiveness and durability.

- **Model development for Tagging:** A multi-label classification technique that produces detailed tags for each artwork results in an impressive *mAP* score of 84.15% on the validation dataset. Deeper understanding of the artworks is provided by the tagging methods developed in this work, which also enhance user search and retrieval experiences.

- **Model development for Object Detection:** Object detection technique was created by fine-tuning the *YoloV8n* model on the *WarliScan* dataset, which produced a *mAP* score of 0.585. With the object detection methods used in this work, the composition, content and the geometric and symbolic aspects of *Warli* paintings are completely covered.

1.6 Conclusion

In conclusion, this study represents a significant advancement in the classification, tagging, and object detection of Indian folk paintings. By creating comprehensive datasets and developing novel methods tailored to the unique characteristics of these artworks, this work contributes to the preservation and appreciation of India's rich folk-art heritage. The high accuracy achieved in classification, tagging, and object detection tasks demonstrates the effectiveness of the methods developed in this study, offering valuable tools for artists, buyers, and art enthusiasts.

Chapter 2

Exploration of Indian Folk Arts

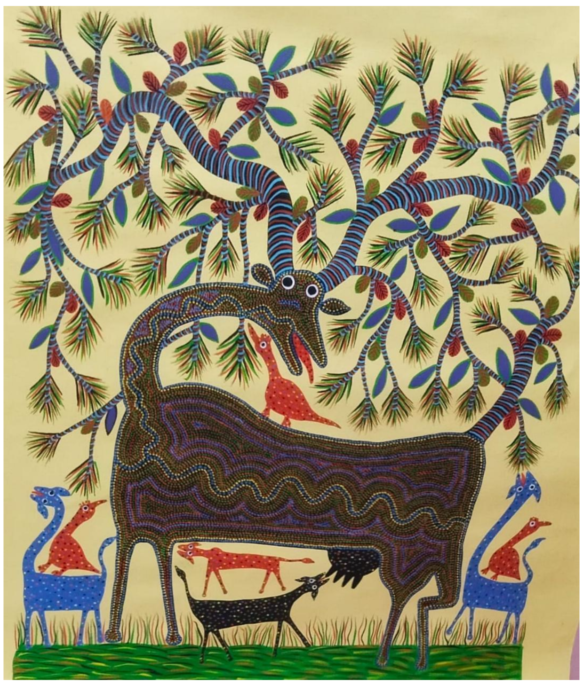
This chapter explores 12 different styles of Indian folk paintings, namely *Bhil*, *Gond*, *Kalamkari*, *Kalighat*, *Pichwai*, *Mata Ni Pachedi*, *Madhubani*, *Pattachitra*, *Phad*, *Tanjore*, *Rogan*, and *Warli*, to explain the richness of Indian culture. The sections below aim to provide an in-depth analysis of each art form. It highlights their unique characteristics, themes, techniques, and historical background. Through this exploration, the chapter reflects the intricate details in these art forms, thereby highlighting their importance in preserving India's rich cultural legacy.

2.1 Bhil

Bhil paintings have their origin in the *Bhil* tribe. It is one of India's major tribal populations and is mostly found in Madhya Pradesh, Gujarat, and Rajasthan. These artworks are strongly connected to the tribe's customs and metaphysical convictions. It is frequently crafted as a depiction of ceremonies or festivities. *Bhil* art is renowned for its vivid portrayals of mundane existence, natural surroundings, and traditional tales, all executed in a unique manner that prominently incorporates dots.

Bhil paintings typically centre on natural elements, encompassing fauna, flora, and depictions of the daily routines of tribal communities. These features possess not only visual appeal but also convey substantial cultural symbolism. Animals such as peacocks and elephants are frequently found, for example, in Figures 2.1a, 2.1b, representing reverence for the natural world and the intimate connection between tribal communities and their surroundings. The dots, delicately placed with bamboo sticks and natural dyes, symbolize seeds, denoting the concepts of development, fertility, and the perpetuity of life.

Bhil painters often paint during celebrations and weddings, imbuing their creations with blessings and wishes for prosperity and happiness. The vibrant color palettes used in the paintings are primarily made up of organic colors seen in fruits, flowers, and foliage. These hues give each painting's narrative life and vitality, creating dramatic canvases that successfully capture the spirit of the history of *Bhil*.



(a) The Immortal Tree of Life in Bhil Painting
Source: Bhil Painting by Bhuri Bai, Image link

(b) A Peacock in Bhil Art
Source: Bhil Art by Gita Bariya, Image link

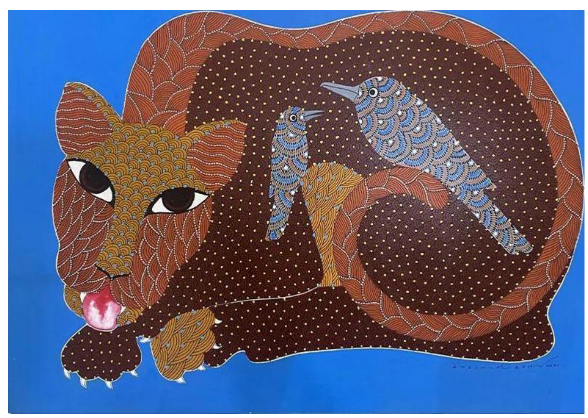
Figure 2.1: Various themes in Bhil paintings

2.2 Gond

Gond is one of the biggest tribes on India. They are mostly found in Madhya Pradesh. They create paintings categorized as Gond paintings, a form of tribal and folk art. The rich expressions and close ties to traditions and folklore set this art form apart. In *Gond* paintings, gods (figure 2.2a), plants, and animals (figure 2.2b) are the main subjects depicted with well-crafted lines and vibrant colors. Every painting tells a story connected to their everyday existence. Each story is based on various folktales and beliefs. Painters use dashes, lines, and dots to their paintings to show movement and depth. These designs hold great spiritual significance in addition to being purely decorative. Every one of them stands for a prayer intended to ward against ill luck and bring good fortune.

Previously produced on walls, floors, and the outside of homes, modern *Gond* art has moved onto canvas and paper and now employs both natural and artificial colors. By advancing through contemporary media while maintaining its cultural roots, this adaptation has helped *Gond* art become more widely known. *Gond* paintings' vivid workmanship and narrative quality make them a powerful chronicle of tribal mythology and an immortal medium of cultural expression.

(a) Lord Hanuman
Source: Gond Art by Venkat Shyam, Image link



(b) Abstract Tiger and Birds
Source: Gond Art by Shyam, Image link

Figure 2.2: Various themes in Gond paintings

2.3 Kalamkari

Kalamkari is a traditional Indian art form known for its elaborate designs that convey stories. It is native to Andhra Pradesh and Telangana. "Kalamkari" means "pen craft," where "kalam" is the

Hindi word for pen and ”kari” is the word for workmanship. Twenty-three processes are involved in creating this artwork, including block printing, hand painting, starching, cleaning, and more. Each step is necessary to produce the intricate and colorful results for which *Kalamkari* is known.

(a) Goddess Lakshmi
Source: Kalamkari Art by K Siva Reddy, Image link

(b) Dashavtar
Source: Kalamkari Art by K Siva Reddy, Image link

Figure 2.3: Various themes in Kalamkari paintings

Kalamkari painters primarily use aspects of nature, scenes from epics like the Mahabharata and Ramayana, and Hindu deities to create mythical and religious motifs in their works. For example Lord Vishnu’s dashavatars and Goddess lakshmi depicted in figures 2.3b, 2.3a respectively. The use of natural dyes made from plants, minerals, and roots makes this art form unique and environmentally friendly. Traditionally, an artist’s brush is a bundle of fine hair linked to the pointy end of a bamboo or date palm stick. This tool is essential to achieve the delicate lines and minute details that characterize *Kalamkari* art.

Mostly, these paintings show scenes from Krishna’s life as well as episodes from the epics like Mahabharata and Ramayana. It is a labor-intensive art form since the procedure is rather precise and includes numerous stages of dyeing, bleaching, hand painting, block printing, starching, cleaning, and more.

2.4 Kalighat

The 19th-century *Kalighat* paintings used to be hung close to the Kali Temple in Kolkata, India's *Kalighat* district. At first intended to be sacred memories for temple guests, these paintings gradually started to portray a wider range of subjects, including social settings and contemporary life. Large, broad brushstrokes that skillfully depict people and settings characterize *Kalighat* paintings.

(a) Goddess Kali's Triumph over Mahishasura
Source: Kalighat Art by Sonali Chitrakar, Image link

(b) Temptations of the Kitchen
Source: Kalighat Art by Sonali Chitrakar, Image link

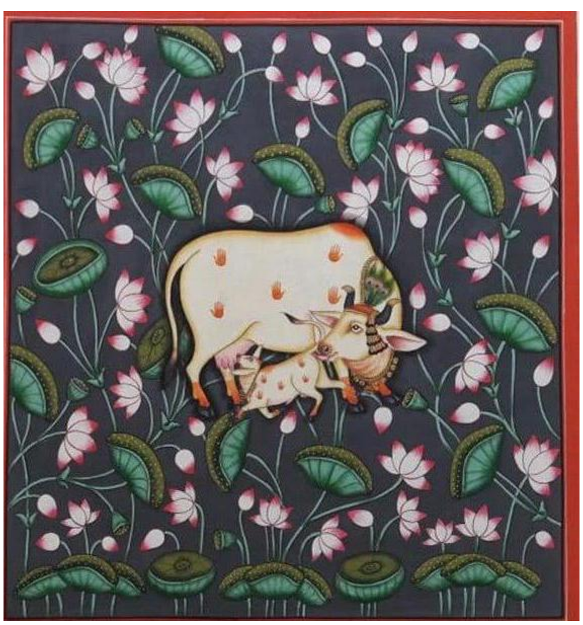
Figure 2.4: Various themes in Kalighat paintings

Paintings in the traditional *Kalighat* style were done on mill-made paper using home-made natural colors. Strong colors and simple, flowing lines make these paintings easily recognizable. Apart from depicting gods (2.4a) and goddesses, this painting technique is particularly well-known for its portrayals of regular people, such housewives (2.4b) and merchants, as well as situations from contemporary life.

The *Kalighat* paintings, a kind of social criticism, also captured the spirit and changing stories of the fast expanding community surrounding the *Kalighat* shrine. Bold simplicity and directness of the painting technique, together with a happy and lighthearted undertone, respectfully convey significant thoughts about morals and society. So defining a certain era in the history of Indian art, paintings from *Kalighat* link the old with the modern, the holy and the profane.

2.5 Pichwai

Pichwai paintings are traditional cloth-based artworks from Nathdwara, Rajasthan. They are made to be the background for the deity Shrinathji, a manifestation of Lord Krishna (2.5b). These works of spiritual and artistic value are popular for their rich hues and intricate detailing.



(a) Divine Maternal Grace
Source: Pichwai by Shehzaad Ali Sherani, Image link

(b) Cows of Devotion: Lord Srinath's companion
Source: Pichwai Art by Dinesh Soni, Image link

Figure 2.5: Various themes in Pichwai paintings

Pichwai paintings, which first appeared more than 400 years ago, represent a variety of events from Lord Krishna's life, particularly his early years and his activities around Govardhan hill. Typical scenes are of the Annakut festival, Raas Leela (dance), and Krishna among lotuses and cows, which stand for his nurturing and protecting function.

Pichwai painters use an exacting style that emphasizes fine details and a rich color scheme that combines synthetic and natural pigments. These paintings frequently use lotus pond as cows as key elements (2.5a). Gold gives the artwork a holy radiance. The styles have evolved throughout time, but they have never wavered from the old techniques; younger artists have added their own originality while yet honoring the themes' sacredness.

The ability of these paintings to arouse devotion is truly appreciated. They are seen to be a dynamic storytelling style that both spiritually offers and preserves the culture of Rajasthan by bringing lord Krishna's lore in brilliant colors and detailed compositions.

2.6 Mata Ni Pachedi

Gujarat is the birthplace of *Mata ni Pachedi*, a spiritual art form honoring the Mother Goddess. The phrase "Mata ni Pachedi" means "behind the mother goddess," and these textiles are utilized as backdrops or temple hangings in the goddess's temples. The Gujarati *Vaghari* community, who were not permitted to enter temples, invented their own transportable shrines, which gave rise to this art style.

Mata ni Pachedi is distinguished by its elaborate patterns and colorful narratives that relate the stories of the Mother Goddess in all of her guises as shown in 2.6a, 2.6b. Cotton fabric is traditionally hand-drawn, block-printed, and dyed to naturally occur colors, mostly red, black, and white. The goddess is usually depicted in the central panel, surrounded by images of her devotees, tales, and other deities, forming a complex narrative tableau.

In addition to expressing religion, the artwork serves as a cultural legacy, preserving the customs and handicrafts of the *Vaghari* people. *Mata ni Pachedi* has developed into a popular art form retaining its essential religious and cultural components.

(a) Ethereal voyage
Source: MataNiPachedi by Dilip Chiitara, Image link

(b) Meldi Maa
Source: MataNiPachedi by Kiri Jayanti Bhai, Image link

Figure 2.6: Various themes in Mata ni Pachedi paintings

2.7 Madhubani

From the *Mithila* region of Bihar, India, comes *Madhubani* paintings, a fascinating folk art form distinguished by vivid colors and elaborate patterns. Women in this area have traditionally painted in this technique. It is also referred to as Mithila art. Intially, it functioned as a kind of wall art intended to enhance houses and arouse positive emotions.

The Hindu mythological motifs that run throughout *Madhubani* paintings include Krishna, Rama, Shiva, Durga, Lakshmi, and Saraswati. Scenes from royal courts, social gatherings and beautiful natural elements like fish, peacocks, the sun, the moon and holy plants like tulsi (holy basil) are interspersed with these religious themes. Every component is given a symbolic meaning. As an example, fish stand for wealth and fertility, and peacocks 2.7a for divine and passionate love.

Since geometric patterns cover the whole canvas, leaving no portion unpainted, *Madhubani* art is unique and is thought to guard against evil spirits. For outlines, artists use double lines filled with vivid, contrasting colors. Traditionally, the paints came from natural sources: charcoal, soot for black, turmeric for yellow, rice flour for white, and indigo for blue. Handcrafted from bamboo sticks with cotton wrapped around them are the brushes.

In addition to decorating, this art form depicts the *Mithila* region's cultural and religious fabric visually, therefore highlighting the social function of women 2.7b as carriers of spirituality and culture.

(a) Duet of Brilliance
Source: Madhubani Art by Ambika Devi, Image link

(b) Half portrayal of woman wearing jewelry
Source: Madhubani Art by Vibhuti Nath, Image link

Figure 2.7: Various themes in Madhubani paintings

2.8 Pattachitra

Pattachitra, an ancient and highly admired art style from *Odisha* in Eastern India, is widely recognized for its ornate decorations and legendary tales. The term "Pattachitra" (from Sanskrit) literally translates to "cloth painting," which sums up this age-old method of painting on properly prepared canvases. In order to create this canvas, fabric is coated with a mixture of chalk and gum, polished, and left glossy for painting.

Pattachitra's style is characterized by the use of vibrant colors and intricate patterns, and it is primarily focused on events from Krishna's life (2.8b) as well as Hindu deities like Jagannath, Balabhadra, and Subhadra. The Mahabharata and Ramayana epics are also popular subjects. *Pattachitra* exclusively employs organic hues derived from minerals and vegetables. The brushes are made from the hair of domestic animals. Painstaking brushwork combined with rich color palettes creates compositions that are both highly detailed and emotionally charged.

Pattachitra paintings are characterized by their elaborate borders (can be found in 2.8a, 2.8b), which are embellished with geometric, floral, and foliate designs that both elevate and improve the religious significance of the image. A *Pattachitra* painting can be a tedious process that takes days or even weeks to complete, depending on its size and intricacies.

Pattachitra has always painted highly religious scenes, and the topics of her paintings are often accompanied by musical performances. This custom fits perfectly with Odisha's very spiritual culture and long history of producing art and handicrafts inspired by religious devotion.

(a) Crimson Divinity: The Artwork of Lord Ganesh
Source: Pattachitra Art by Apindra Swain, Image link

(b) Cosmic Transformations: Blessings of Krishna
Source: Pattachitra Art by Apindra Swain, Image link

Figure 2.8: Various themes in Pattachitra paintings

2.9 Phad

Phad painting is a style of scroll painting that originated in Rajasthan and depicts intricate religious tales of regional deities and mythical heroes. The word ”Phad,” which means ”fold” in Rajasthani, refers to the way these gigantic paintings are usually folded like books and only opened for performances. The two primary deities of Rajasthan’s *Phad*, Pabuji and *Devnarayan* (2.9a, 2.9b), are portrayed as the defenders of the inhabitants and are revered as folk heroes.

(a) Couple seated on a camel
Source: Phad Art by Kalyan Joshi, Image link

(b) Royal Garden: The Queen and her Maiden
Source: Phad Art by Kalyan Joshi, Image link

Figure 2.9: Various themes in Phad paintings

Traditionally, *Phad* paintings are made on fabric or canvas that has been treated with stone dust and rice starch to create a rigid surface that is appropriate for painting. Artists use natural colors, which they create from minerals and plants. Phad’s color scheme is incredibly vibrant, utilizing primary hues such as red, green, blue, and yellow, which gives the story lines an amazing visual impact.

Every painting is meant to function as a traveling shrine that is transported from village to village by Bhopa, or priest-singers, who narrate the folktales of the gods shown in the *Phad*. The Bhopas tell these stories through song, using the *Phad* as a visual aid. The combination of music and art improves the cultural life of the communities by preserving and passing on the stories and the art form from one generation to the next. In addition to being a form of artistic expression, the *Phad* paintings encapsulate the merger of art, religion, and history, and are a vital part of Rajasthan's living folk legacy.

2.10 Tanjore

Tanjore paintings are known for their vibrant colors, surface richness, and complex iconography. They come from Tamil Nadu, a region in South India with a rich cultural legacy. This art form, which was created in the late 16th century during the *Chola* dynasty's rule, has prospered under the support of royals, wealthy merchants, and the British. The majority of *Tanjore* paintings are religious in nature. They are distinguished by their exquisite use of semi-precious stones and gold leaf giving them alluring look (figure 2.10a,2.10b).

A *Tanjore* painting is created in several painstaking steps. The base, which is made of a piece of fabric covered in a wooden board, is prepared first (Palagai padam). To create a smooth surface, a coating of binding medium and chalk powder is applied to the cloth. On this base, a preliminary drawing of the deity is made. 'Embossing' is the following process, where parts of the painting such as decorations and clothing are given raised, three-dimensional effects by mixing limestone with a binding agent.

After the embossing is complete, the painting is embellished with colorful paint and gold foil, making the figures stand out. Precious and semi-precious stones are used for finishing touches, enhancing the painting's visual appeal. Hindu gods and goddesses are among the most often painted topics.

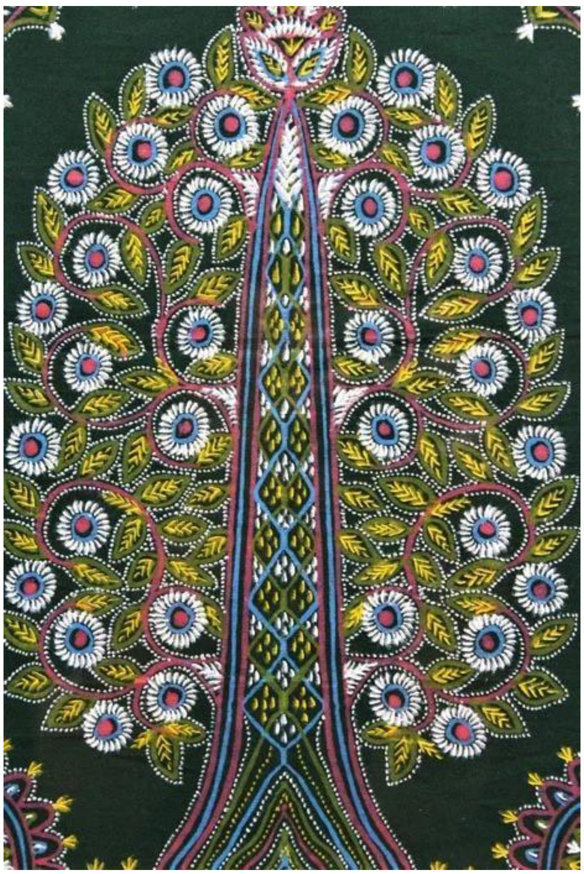
2.11 Rogan

After the embossing is complete, the painting is embellished with colorful paint and gold foil, making the figures stand out. Precious and semi-precious stones are used for finishing touches, enhancing the painting's visual appeal. Hindu gods and goddesses are among the most often painted topics.

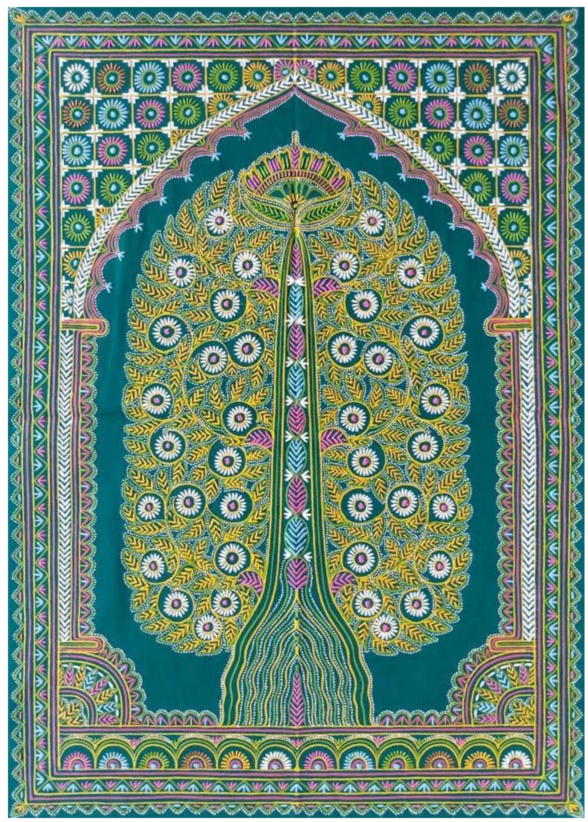
(a) Gopal: Tanjore painting
Source: Tanjore Art by Sanjay Tandekar, Image link

(b) The Peacock: Tanjore Art
Source: Tanjore Art by Sanjay Tandekar, Image link

Figure 2.10: Various themes in Tanjore paintings



(a) Kutch Rogan Art
Source: Rogan Art, Image link



(b) Tree of Life Rogan painting
Source: Rogan Art by Rizwan Khatri, Image link

Figure 2.11: Various themes in Rogan paintings

Kutch, Gujarat in India is the birthplace of the unusual and fascinating cloth painting technique known as *Rogan* art. This type of art uses a metal rod to apply thick, vibrantly colored paint—made from castor oil and natural dyes—on fabric. The Persian origin of the term ”Rogan” means ”oil-based.” A single-family in Kutch has been maintaining the technique for decades, making it a genuinely unique legacy of expertise and tradition.

The first step in the *Rogan* painting process is boiling castor oil for almost twelve hours until it becomes sticky and honey-like. The resulting mixture is then combined with natural pigments to make *Rogan* paste. The artist deftly applies the paste to one-half of the cloth using a metal rod or stylus. After that, the fabric is folded to make a mirror image, which results in elaborate, symmetrical patterns (2.11a, 2.11b). Typical *Rogan* art includes geometric patterns, peacocks, floral themes, and occasionally depictions of trees of life, which represent the abundance of nature.

Rogan painting is very patient and precise work since the paint is intricately twisted into patterns and designs without ever coming into contact with the fabric. This unique process creates extraordinarily beautiful artwork that is both aesthetically pleasing and rich in cultural significance. Despite its challenges, *Rogan* art is currently experiencing a rebirth that is attracting global art fans and providing a steady stream of cash for those who create it.

2.12 Warli

A unique type of tribal art, *Warli* painting is credited to the *Warli* tribe of Maharashtra, India. This art form has been handed down through the years, keeping its originality and cultural relevance, going back to the 10th century AD. Mostly monochromatic, *Warli* art is distinguished by the application of a white pigment, which is a combination of rice paste and water, onto a harsh, frequently reddish-brown background—traditionally the color of the mud walls of rural houses.

The everyday social and natural factors as well as human activity within the tribe serve as the inspiration for the main motifs of *Warli* paintings. *Warli* emphasizes human figures shown in a very basic form—circle heads and two inverse triangles linked at the tip forming the body, with stick-like limbs and legs—in contrast to many conventional forms that depict mythological characters or deities. This intentional minimalism reflects the core of *Warli* culture, which values balance and simplicity in daily living.

The ”Tarpa dance” is one of the most often occurring subjects in *Warli* art. The heart of *Warli* tradition is the *Tarpa*, a trumpet-like instrument performed during many joyous events. As they move around the *Tarpa* player in a circular rhythm, people holding hands represent social cohesiveness and community vitality. A major theme in *Warli* paintings as well, the circle symbolizes the cyclical aspect of existence and is a reflection of the sun and moon, which are essential to agricultural cycles and daily living.

(a) Scenic Warli Painting
Source: Warli Art by Dilip Bahotha, Image link

(b) Warli painting illustrating nature and lifestyle
Source: Warli Art by Dilip Bahotha, Image link

Figure 2.12: Various themes in Warli paintings

Warli art frequently depicts animals, birds, trees, hills, and rivers; the natural world is very important. Figures 2.12a, 2.12b are some of the examples of common themes in Warli paintings. Each one of these components represents a deeper bond between the *Warli* tribe and their surroundings and has cultural value beyond simple representation. Frequently displayed animals represent affluence and social standing in society; these include goats, horses, and cows. Especially parrots and peacocks, birds are considered lucky and happy symbols. Numerous images of trees, particularly the holy Banyan tree, represent food and the essential function of nature in tribal life.

Another frequent subject in *Warli* paintings is the daily activities of the peasants. The tribe depends heavily on farming, fishing, hunting, and food collection, as seen by the many scenes of these pursuits. Frequently shown as doing essential tasks in *Warli* home life, women are shown cooking, getting water, or grinding food. Along with recording daily life, these images honour the *Warli* tribe's sustainable and peaceful way of existence.

Together with the fine details and motifs, the natural dyes and bamboo sticks used as a brush highlight the enduring ties of *Warli* community to nature and their surroundings. In addition to being ornamental, this art form helps to tell and maintain the tribe's customs and folklore throughout many generations. *Warli* painting, then, distinguishes itself as a documentary story of the *Warli* tribe as well as an artistic medium, highlighting their close bond with the natural world and their traditional values of sustainability, simplicity, and cohesiveness of the group.

2.13 Summary of Related Works

An overview of the many Indian folk art styles is given in Table 2.1, which also summarizes important academic studies and research on each. It presents a significant study that has been conducted to document, analyze, and conserve the unique artistic, historical, and cultural components of these traditional arts. Each entry in the table represents a noteworthy development in the scientific and cultural knowledge of India's rich artistic heritage.

Table 2.1: Works done for various Indian folk art forms

Art Form	**Related Work**
Bhil	[1] Comparing Indigenous and Formal Children's Illustrations [2] Culture of Bhil in Central India [3] Tradition and evolution of Rajasthan's tribal Art
Gond	[4] Historical evolutional through Gond paintings [5] Transition from folklore to visual art [6] Semiotics in Gond Paintings
Warli	[7] Synthesis of Stick like figures using GANs [8] Evolution and Transition of Warli paintings [9] Charachteristics and themes in warli paintings
Kalamkari	[10] Structural analysis of Kalankari scrolls and temple murals [11] Kalamkari paintings on temple cloth

Art Form	Related Work
Phad	[12] Phad as visual narrative of Rajasthan [13] A review of traditions of Phad [14] Pictorial Narratives of Kaavad and Phad Traditions
Tanjore	[15] Analysis of Tanjore paintings [16] Tanjore Painting restoration using segmentation
Mata ni Pachedi	[17] Tales of Goddess (mata) on textiles [18] An overview of representation of female deities
Madhubani	[19] Historical journey of Madhubani [20] Concept of geometry in Madhubani painting [21] Classification of Madhubani paintings using transfer learning
Pichwai	[22] Reviving Pichwai Art through Contemporary Design [23] Pichwai art depicting lord Krishna's life
Rogan	[24] Pattern and motifs in Rogan paintings [25] Journey of Rogan paintings
Kalighat	[26] A review of Kalighat paintings [27] Exhibition of Kalighat Paintings at Cleveland Museum [28] Pigment analysis of nine Kalighat paintings
Pattachitra	[29] Review of motifs, tools, colors etc in Pattachitra paintings [30] Patuas' socio-economic impact through Pattachitra study

Chapter 3

Classification and Tagging

3.1 Scope

Indian folk paintings have a rich mosaic of symbols, colors, textures, and stories, making them an invaluable repository of cultural legacy. The diversity is profoundly reflected in the field of art [31], where each community presents a unique style of folk paintings depicting their day-to-day activities like agriculture, marriage, religious practices, hunting, etc. These paintings are valuable because of their distinctive use of colors, symbols, and textures, which serve as historical records of the respective folk lifestyle in addition to being an artistic expression. This historical significance makes it important to preserve them in their original form and protect their geographical tags. With this motivation, the proposed work is carried out to efficiently classify twelve types of folk paintings, namely *Warli, Bhil, Gond, Kalighat, Kalamkari, Pichwai, Rogan, Mata Ni Pachedi, Madhubani, Tanjore, Pattachitra*, and *Phad*. In addition to classification, the proposed tag generation mechanism can help in offering an enhanced search experience to the customers.

Each folk art is unique in its narrative, representation of objects/animals/humans, and the colors or material used. *Warli* art from Maharashtra stands out for its depiction of social life. Geometric patterns like triangles (mountains) and circles (sun) are used to depict harmony with nature. *Gond* art comes from central India and is characterized by its vibrant colors as well as its use of dashes, lines, and dots. *Bhil* art, also from central India, captures the folklore and their daily lifestyle using only colorful dots and motifs. *Kalamkari* art from Andhra Pradesh and Telangana portrays narratives from Hindu mythology. These can be found on temple walls and on cotton textiles. *Madhubani* art from Bihar is characterized by its complex geometric patterns, mostly depicting religious themes. *Pichwai* art originates from Rajasthan and is celebrated for its detailed depictions of Lord Krishna (a popular

Hindu god). Lotus and cow are the main elements in *Pichwai* paintings. Gujarat's *Mata Ni Pachedi* depicts cultural stories representing the Mother Goddess. *Rogan* art is another art form practiced in Gujarat in which, thick castor-oil-based paints are used to create fluid dark-colored designs. *Tanjore* paintings from Tamil Nadu can be identified by the gold foil work giving them a luminous appearance. *Phad* paintings from Rajasthan mostly depict deities and local heroes through scrolls that narrate epic tales. *Pattachitra* from Odisha is mainly known for its intricacy depicting Hindu mythology on large cloth-based scrolls. The *Kalighat* paintings from West Bengal, originally souvenirs from the *Kalighat* temple, have evolved into a distinct style featuring bold brushstrokes and reflecting societal themes. An artistic representation of the Indian folk arts can be found in footnote [1].

This chapter presents a novel approach to classifying these paintings into distinct art forms and tagging them with their unique salient features. A custom dataset named *FolkTalent*, comprising 2279 digital images of paintings across 12 different forms, has been prepared using websites that are direct outlets of Indian folk paintings. Tags covering a wide range of attributes like color, theme, artistic style, and patterns are generated using *GPT4*, and verified by an expert for each painting. Classification is performed employing the *RandomForest* ensemble technique on fine-tuned Convolutional Neural Network (CNN) models to classify Indian folk paintings, achieving an accuracy of 91.83%. Tagging is accomplished via the prominent fine-tuned CNN-based backbones with a custom classifier attached to its top to perform multi-label image classification. The generated tags offer a deeper insight into the painting, enabling an enhanced search experience based on theme and visual attributes. The proposed hybrid model sets a new benchmark in folk painting classification and tagging, significantly contributing to cataloging India's folk-art heritage.

3.2 Motivation

There have been multiple attempts [21, 32, 33, 34, 35] to create a dataset for Indian folk arts. This section covers a comprehensive overview of existing datasets, followed by an analysis of the prior works on Indian folk arts, thereby setting the ground for the motivation behind the proposed method.

Datasets: Varshney et al. [21] proposed a dataset comprising five different forms of *Madhubani* paintings, namely "Bharni, Godna, Kachni, Kohbar, and Tantrik." It contains 680 digital images collected from websites as used in ours. This dataset focuses only on one main art form. Kumar et al.

[1] https://cdn.shopify.com/s/files/1/1194/1498/files/Folk_Art_Map_of_India_2019.jpg

[32] compiled a dataset of over 2400 paintings from 8 different folk arts, namely *Mural, Pattachitra, Kalamkari, Portrait, Madhubani, Warli, Kangra,* and *Tanjore*. This was the first successful attempt at creating a dataset containing a diverse range of folk painting styles, with equal distribution across the art types. A similar attempt was made by Mane and Shrawankar [33] with around 1000 paintings across 26 different folk-art forms. The dataset size is too small to cover the huge diversity of 26 classes. A recent attempt was made to create a dataset [36] for Indian Visual arts, which includes four broad categories of art forms, namely, Sculpture, Pottery, Painting and Architecture. It consists of 4000 images, 1000 from each class, sourced from Google images. Further, Podder et al. [34] created a dataset focusing exclusively on Indian monuments, which contains images of various sculptures and paintings depicting mythological stories. Images are taken from different angles and under different illuminations.

Numerous studies have been conducted using various tasks on these datasets to retrieve desired patterns. As the paintings are analyzed, it is crucial to capture and analyze visual patterns to draw conclusions for the hypothesis. With the evolution of CNN-based models, it becomes a lot more efficient to address the nature of the proposed work, that is, image classification via transfer learning on Indian folk arts.

Classification of paintings has been well explored within the field of computer vision [37, 38, 39, 40]. Convolutional Neural Networks (CNNs) [41] serve as the fundamental framework for various downstream tasks like classification [42, 41], object detection [43, 44], image segmentation [45, 46], etc. Four prominent CNN models namely, ResNet-50 [47], EfficientNet-B0 [48], Inception-V3 [49] and VGG-16 [50] have made significant contributions in the field of image classification [37, 38, 39, 40, 51, 42] and tagging [52, 53, 54]. Based on these findings, the CNN models seem to learn and perform efficiently on image classification tasks. The foundational VGG model [50], introduced in 2014, highlighted simplicity through a deep stack of 3x3 convolutional layers, serving as a robust baseline. Inception model or *GoogLeNet* [49] includes parallel filters and an inception module to capture features at different scales. *ResNet* [47] brought a revolution in the field with its introduction in 2015, mainly for its skip connection trick that helped it overcome the vanishing gradient problem. *EfficientNet* [48] optimized both computational cost as well as model size, balancing depth, width, and resolution through the principle of compound scaling. These models represent the primary milestones in the evolution of CNNs and continue to offer diverse solutions in the field of deep learning for computer vision applications.

Existing classification approaches: A recent study [21] explored the classification of five forms of *Madhubani* paintings via transfer learning with a pre-trained CNN. This work is divided into two modules, classification using ‘handcrafted features’ and ‘automatically extracted features.’ The second approach leverages CNN-based backbones like *InceptionV3* [49] and *InceptionResNet* [55] for feature extraction and reported a classification accuracy of 98.82%. Despite having a high classification accuracy, the scalability of this dataset is limited due to its confined focus on just one folk art form. Another attempt [32] was made to classify Indian paintings using Support Vector Machine (SVM) [56] on the features extracted from CNN models like *AlexNet* [57] and *VGG* [50]. The proposed dataset was perfectly balanced, consisting of 8 different art forms on which they obtained a classification accuracy of 86.56%. Another work demonstrates an efficient query-based retrieval [33], where an accuracy of 82% was obtained on a dataset of 1000 paintings with 26 classes via transfer learning (on pre-trained CNN) approach. Even with many classes, the representation within each class is less diverse due to small class sizes. Hence, the dataset is considered to lack diversity. While there are many classes, the samples within each class might not adequately capture the variability and range of styles, techniques, or motifs typical to that folk art form. Having few samples per class leads to the memorization (overfitting) issue, thereby limiting its generalization and wide-scale application.

To summarize, the main challenges in the existing works include (1) limited diversity within each class in the dataset, (2) scope of improving classification score for a wider spectrum of Indian folk paintings, (3) skewed category-wise distribution within the datasets, and (4) the absence of validation split while training CNN models. Additionally, the existing literature focuses only on high-level details of paintings and does not delve into the granular details, including unique elements of saliency inherent in the paintings.

To address the gaps in the existing works, a new classification method is proposed involving an ensemble-based technique applied to extracted features from multiple fine-tuned CNN backbones. The dataset boasts a larger volume of 2279 images across 12 diverse forms of folk paintings. Furthermore, the proposed work is the first attempt at image tagging on Indian folk paintings and aims at devising a model to efficiently generate tags for 12 different forms of folk paintings, which can be used for efficient retrieval using keywords in the future.

With this background, the motivation behind the proposed work is classifying and tagging Indian folk paintings using CNN-based architectures in multi-class and multi-label classification settings, respectively. It can be observed from Table 3.6 that each model learnt and focused on slightly different features or aspects of the data, and no single model consistently outperformed the others across all metrics. To further fine-tune the performance and utilize the popular CNN models collectively to make classification decisions, the *RandomForest* [58] based ensemble technique is used to aggregate the performance of each strong-performing CNN model to arrive at a collective decision.

3.3 Methods

This section is aimed at discussing the proposed dataset and the methodology. Section 3.3.1 talks about the proposed dataset, FolkTalent that covers 12 different forms of Indian folk paintings and shows their compositionality per class with split proportion for training, validation, and testing. To capture better granularity, GPT4 [59] was utilized to generate tags for each image. Section 3.3.2 elaborates on classification and tagging methodology adopted on FolkTalent. The Section 3.3.3 highlights the ensemble approach performed across fine-tuned CNN models.

3.3.1 FolkTalent Dataset

The FolkTalent dataset comprises 2279 images across 12 different classes of folk paintings. The images are collected from the websites [60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78] which are the direct sellers of Indian folk paintings. For each painting, the bio of the painter has been scrutinized to ensure the authenticity of the paintings. Each painter whose work is posted on sites has inherited the art form as an accomplished artist. This was one of the main criteria for choosing images for the dataset to ensure no counterfeits were considered for the analysis.

The dataset contains paintings of Bhil (191 images), Gond (183 images), Mata Ni Pachedi (185 images), Kalighat (184 images), Kalamkari (184 images), Madhubani (187 images), Pattachitra (195 images), Phad (214 images), Pichwai (187 images), Tanjore (191 images), Rogan (185 images) and Warli (190 images). Additionally, for each painting, an array of tags/keywords are generated using GPT4 [59] followed by a manual review. The tags include colors used, theme, patterns used (like dots, dashes, etc.), art style, and the elements of saliency like flora, fauna, celestial elements (sun, moon, and stars), human figures, agricultural elements, deities, etc. Figure 3.1 is an example of a painting from *FolkTalent* with its tags listed in caption. On average, 30 tags are generated for each painting. Some common tags across each artform are shown in Table 3.1. Furthermore, a category-wise segmentation of common tags can be found in Table 3.2.

Figure 3.1: A sample painting from the dataset along with the tags

Source: A Warli Village, Warli Art by Dilip Bahotha, Image link

Tags: Warli_art, dark_brown_background, stars, moon, white_figures, dancing, cooking, cows, birds, fields, white_patterns, playing_instruments, animal_herding, riding_cart, floral_patterns, crops, ploughing, huts, village_houses, canopies, musical_instruments, celebration, harvesting, dance, geometric_designs, horses, trees, circular_dance_pattern, circular_dance, group_activities

Art Form	Common Tags
Warli Paintings	White, Red, Dance, Harvest, Marriage, Daily life, Animals, Trees, Nature, Tarpa Dance, Lines, Chowk
Bhil Paintings	Red, Yellow, Green, Nature, Gods, Festivals, Animals, Birds, Dots, Daily life, Trees, Flowers
Gond Paintings	Red, Blue, Green, Yellow, Deer, Gods, Animals, Birds, Trees, Nature, Dots, Dashes
Kalamkari Paintings	Floral, Mythology, Gods, Blue, Red, Yellow, Green, Birds, Animals, Temples, Fabric
Pichwai Paintings	Krishna, Cows, Lotus, Nature, Temples, Festivals, Blue, Green, Red, Yellow, Gold, Fabric
Madhubani Paintings	Mythology, Nature, Festivals, Geometric patterns, Animals, Birds, Plants, Gods, Rituals, Flowers, Daily life
Phad Paintings	Gods, Heroic figures, Bold lines, Green, Yellow, Red, Blue, Rituals, Musical instruments, Animals
Tanjore Paintings	Gold foil, Gods, Rich colors, Gems, Blue, Red, Green, Mythology, Temples, Jewelry
Rogan Paintings	Floral, Birds, Geometric patterns, Yellow, Red, Blue, Green, Fabric, Trees, Nature
Pattachitra Paintings	Mythology, Epics, Red, Yellow, Blue, White, Black, Gods, Nature, Animals, Birds
Mata Ni Pachedi Paintings	Goddess, Rituals, Red, Black, White, Temples, Festivals, Cloth, Animals, Birds, Nature
Kalighat Paintings	Mythology, Bold lines, Gods, Epics, Daily life, Red, Yellow, Blue, Black, Animals, Birds

Table 3.1: Common tags for each art form

The dataset is split into 3 sets: training, validation, and testing to assess the reliability and generalizability of proposed models. Training split containing 1364 images (60%), is used to iteratively adjust the model parameters with the goal of minimizing the loss function. Whereas to maintain the model's

Table 3.2: Category-wise Tags

Category	**Tags**
Color	Red, Yellow, Blue, Green, Black, White, Gold
Theme	Mythology, Nature, Festivals, Daily Life, Gods, Goddess, Animals, Birds, Plants, Dance, Tarpa Dance, Harvest, Marriage, Rituals, Epic
Style	Geometric patterns, Lines, Dots, Dashes, Bold lines
Nature	Animals, Birds, Trees, Plants, Flowers, Deer, Cow, Lotus

generalizability, the model's performance was consistently checked on the validation set which comprises of 450 images, that is 20% of the dataset (Figure 3.2) in every pass. Dataset balance is maintained to ensure similar set size per class.

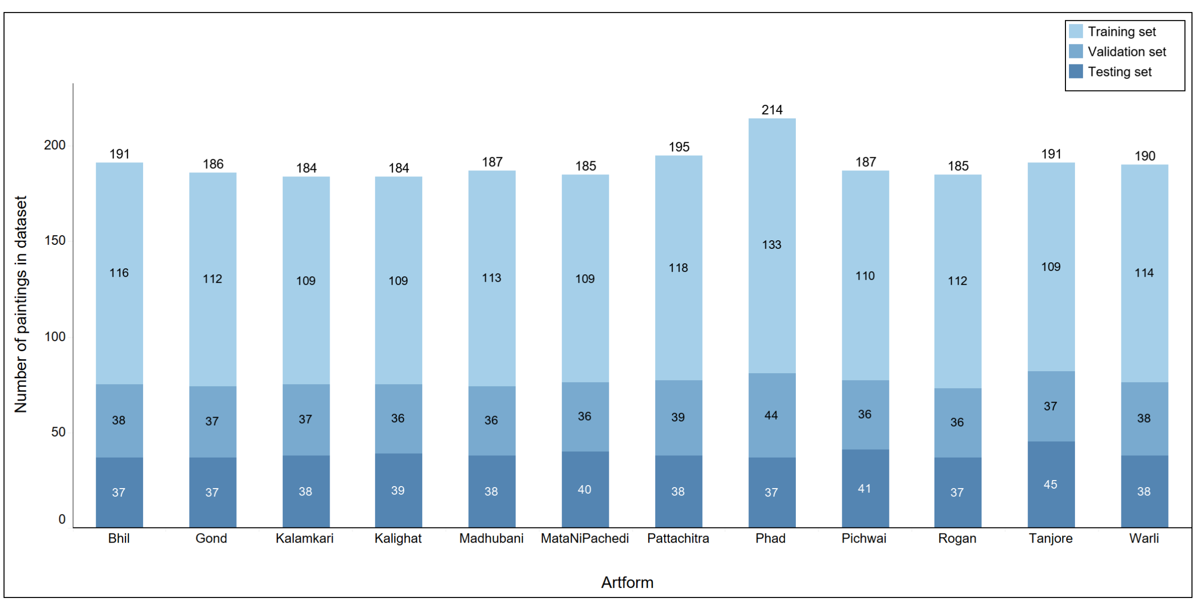


Figure 3.2: Image distribution across classes and partitions.

The validation set is essential for assessing model's performance during training, enabling fine-tuning, and preventing over-fitting by evaluating the model on unseen data. Incidentally, validation split has not been used so far in the existing works of classification of Indian folk paintings. Models are typically check pointed based on the best validation score, raising concerns about a potential over-fitting to the validation set. To ensure an unbiased evaluation of the model's performance on unseen data, the test dataset is introduced. In the proposed work, the test set comprises 20% of images. This set mimics the real-world scenario by exposing the model to completely unseen data. Notably, all three partitions are mutually exclusive and exhaustive.

3.3.2 Pre-processing

It is a critical initial step in any machine learning or data analysis work aimed at preparing raw data for further analysis or modeling, which involves a series of operations to clean, transform, and organize the data to make it suitable for the specific task at hand. The process begins with the frame and the background removal from the painting to avoid conflict with the actual content in the painting. The border within the painting was kept intact as it contributes to saliency and has unique significance based on the folk community and the themes. For example, *Warli* paintings depicting a wedding scene mostly contain a floral border as its hallmark feature. For tagging, the captioning potential of *GPT4* was leveraged to generate keywords and tags, describing the key aspects of the painting. Semantically similar words were aggregated and "synonym replacement" was done to represent them via a single tag, e.g., *GPT4* generated tags like 'celebrated', 'celebrating', 'feast', 'celebration', 'festivity', were all represented with a single keyword "celebration". Next, a word vocabulary of Next, a word vocabulary of 1500 tags, is generated consisting of all such tags generated.

Using this vocabulary, a multi-label encoder transformed labels (tags) into a binary vector to indicate the presence of each tag in the vocabulary. This prepares the data for multi-label classification task. A demonstration of the tag-to-binary vector transformation is provided in Tables 3.3 and 3.4.

Data Augmentation: As the name suggests, it is used to expand the dataset by applying various transformations and modifications to the existing dataset thus increasing its diversity and size. As variation invokes regularization, this in turn improves robustness of the model. Unlike usual image augmentation, color-based augmentation has been skipped in the proposed method, to preserve the significance of various colors in each art form and maintain their authenticity. Further, transformations like horizon-

Table 3.3: Image tags for various samples

Image	**Tags**
Image 1	Sun, Stars, Cow
Image 2	Lotus, Cow
Image 3	Stars, Lizard
Image 4	Sun, Stars

Table 3.4: Binary vectors for four sample images

Image	Sun	Stars	Lotus	Cow	Lizard
Image 1	1	1	0	1	0
Image 2	0	0	1	1	0
Image 3	0	1	0	0	1
Image 4	1	1	0	0	0

tal and vertical flips, as well as scaling that are standard practice to apply, have been carried out. Finally, all the images were resized to 224x224 pixels across three channels, RGB (Red Green Blue).

The final dataset following pre-processing, and augmentation comprises 2279 images, each sized at 224x224 pixels. Additionally, it includes binary vectors associated with tagging information for each image.

3.3.3 Fine-tuning CNN models for classification

Problem Statement: Given a folk painting image as input, the objective is to accurately classify it into one of twelve predefined categories, rendering a multi-class classification task. Further, to enhance our understanding of the image content, the associated tags (unique salient elements specific to each category) are predicted for each image, thereby making it a multi-label classification task. The generated tags serve to provide interpret-ability and insight into the model's decision-making process.

Transfer learning with pre-trained CNN & their pre-processor: With the proposed dataset, the task is to build an algorithm that efficiently classifies Indian folk paintings. Based on the promising performance on *ImageNet* classification dataset [79], four pre-trained CNN architectures, namely, *VGG16*, *ResNet50*, *EfficientNetB0* and *InceptionV3* were employed. For the classification of Indian folk painting, every CNN architecture offer unique benefits. The simplicity and depth of *VGG16* helps it to effectively capture complex creative details unique to these artworks. Skip connections in *ResNet50* enable it to manage challenging textures and patterns seen in many folk art forms. *EfficientNetB0* balances computational cost and accuracy to effectively analyze high-resolution images typical of art datasets. Multi-scale feature extraction is where *InceptionV3* outshines, which is essential for spotting both fine details and more general style aspects particular to different folk painting genres. These designs taken together improve the model's resilience and capacity for generalizing across several forms of Indian folk painting. This enables the model to capture both local and global features effectively. Combining these viewpoints helps the model make better predictions and become more robust and generalizable. The CNN backbones mentioned require the images in a specific format due to the way they are trained. Hence, all the images are resized to 224x224 pixel resolution (except for *InceptionV3* with 299x299) across all 3 RGB channels. Further, they are normalized across channels using corresponding pre-computed statistics from *ImageNet* [79].

Fine Tuning CNN: In the proposed analysis, the model is trained with a batch size of 128 for 100 epochs. Early stopping with a patience of 15 is applied to stop training the if no further improvement is observed in validation accuracy for 15 consecutive epochs. The model is check pointed to save the last best weights based on validation accuracy. Additionally, learning rate is optimized using "Reduce Learning rate on Plateau (ReduceLROnPlateau)" [80] to enable maximum learning and a reduction in loss. *ReduceLROnPlateau* [80] effectively handles slow convergence by overcoming plateaus in the loss function surface. To achieve this, the patience was set to 8, which specifies the LR controller to drop the learning rate (reduction factor) to 50% of current if validation accuracy (monitoring criteria) does not improve for 8 consecutive epochs. The final set of hyper-parameter values is finalized based on best validation accuracy.

Model Optimization: For optimization, Adam (Adaptive moment estimation) [81] is used that combines the merits of two popular optimizer called *RMSprop* [82] and Momentum [83] for effectively updating the model's weight during training. Overall, it adapts learning rate to increase optimization with accumulated gradients from the previous passes that helps in smoother updates. The convergence observed is much faster than traditional algorithms like Stochastic Gradient Descent (SGD) [84] especially in case of varying curvature irregular loss [81]. As loss function, "Categorical cross-entropy (CE)" was used which is most preferable in multi-class classification tasks. It computes dissimilarity (entropy) between the predicted probability scores (per-class probability) and the true label (one-hot vector). With the objective of minimizing this loss function, models try to assign a higher probability score for correct class and lower for the incorrect ones. In the case of tagging, "Binary cross-entropy (BCE)" loss function is used, since it involves multi-label classification. The working principle of BCE is similar to CE, with the only difference being, all the classes in multi-label setting are considered to be independent.

Classification head: This is a module (head) of the simple feed-forward network (FFN) [85] comprising 12 neurons. It is added atop of the headless pre-trained CNN backbones to tailor the entire architecture to the specific task i.e., classifying 12 classes instead of the default 1000 [55]. The backbone (excluding the FFN) acts as a feature descriptor and is thus unfrozen during training to fine-tune (adapt) the proposed dataset. Parallelly, FFN (the classifier) too gets trained to correctly identify the labels. Similarly, for tagging, this FFN architecture is adapted to accommodate different number of classes (tags). In the design, FFN is added after Global Average Pooling (GAP) layer which produces 1024-dimensional features capturing the overall semantics of an image. These features serve as an input for the FFN to generate desired labels. Architecturally, FFN comprises a fully connected linear layer with 1024 output neurons followed by, *ReLU* [86] (Rectified Linear Unit) activation function, and another linear layer with 12 output neurons, followed by Softmax activation [87] to derive class probabilities. For tagging, the last linear layer is modified to have 1500 output neurons with Sigmoid activation function [88]. Unlike Softmax, Sigmoid treats all the labels independently.

Inference: In this stage, the entire pipeline explained previously is executed, starting from the image-processing, and continuing until obtaining predictions from the classification head, all while maintaining the model's weights frozen. For multi-class classification, the class index is determined by taking the argmax of the predicted vector, which is then looked up against a class-index dictionary to obtain the corresponding class name. For tagging, the soft prediction scores are binarized using a threshold of 0.5, and all tag names corresponding to the vector elements that exceed the threshold are selected.

Ensembling Fined-tuned CNN architectures with Random Forest Ensembling offers a powerful approach to harnessing the collective decisions of various models, enhancing overall robustness, and reducing bias. Here, a similar strategy was adopted by incorporating a Random Forest [35] (a decision-tree-based ensembling technique) into the workflow. The probabilistic predictions from the top three fine-tuned CNN models (*ResNet50*, *InceptionV3*, and *EfficientNetB0*) were combined and fed into the Random Forest algorithm, as shown in figure 3.3. This helped in generating a more informed decision regarding the correct label by leveraging the diverse insights offered by each model.

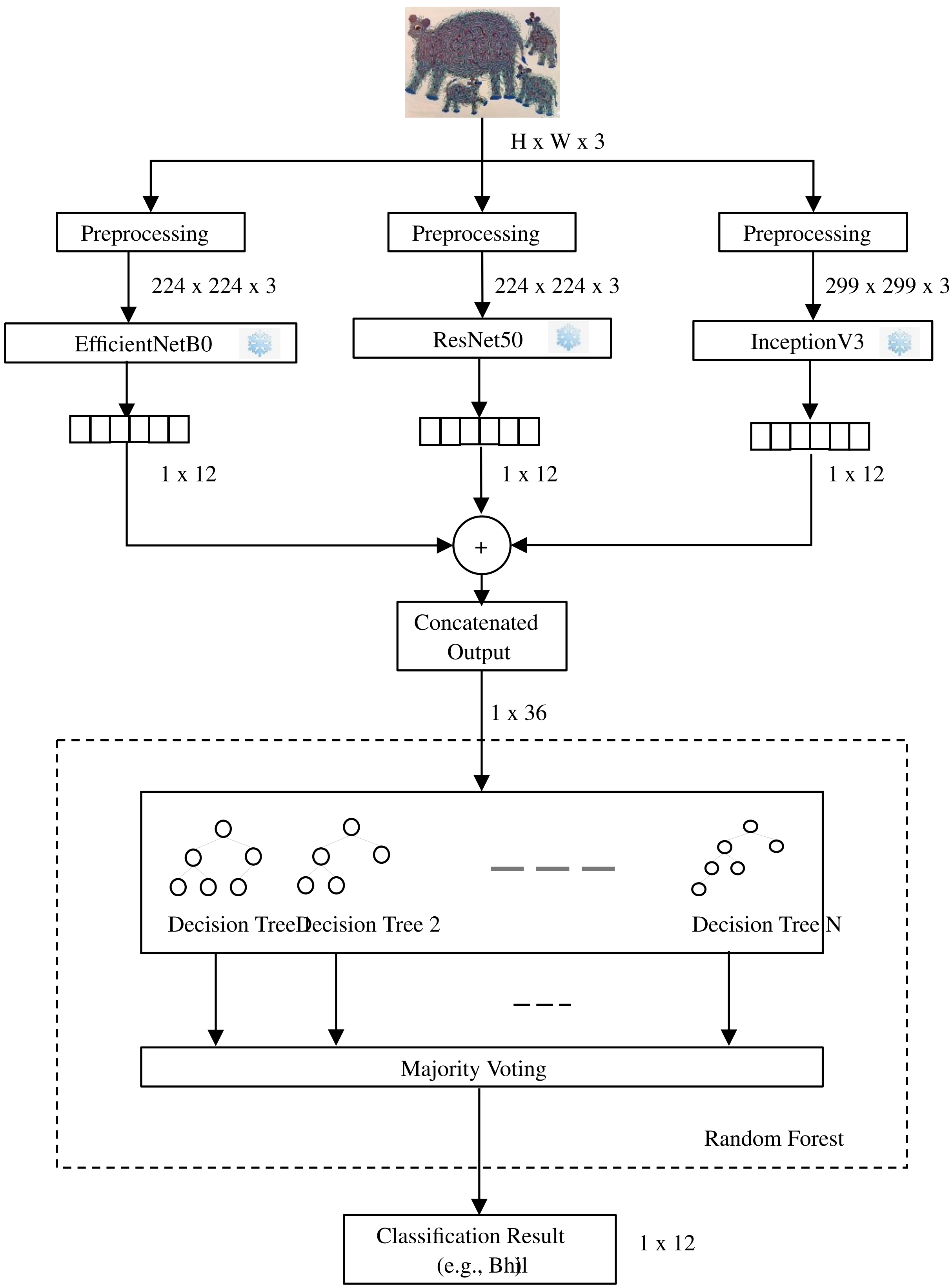


Figure 3.3: Ensemble of fine-tuned CNN models using RandomForest

Table 3.5: Performance of fine-tuned CNN models and their ensemble on FolkTalent for classification

CNN Model	Training Accuracy	Validation Accuracy	Testing Accuracy
VGG16	75.28	73.83	62.8
InceptionV3	99.93	90.44	88.82
ResNet50	99.78	94	90.53
EfficientNetB0	99.93	91.61	90.97
Random Forest (Ensemble)	100	94.00	91.83

3.4 Results

In this section, the results are presented that are obtained from CNN fine-tuning as well as ensembling as shown in Table 3.5.

As evident from the table, *EfficientNetB0* has achieved the highest test accuracy, scoring 90.97%, while *ResNet50* achieved highest validation accuracy of 94% with test accuracy being 90.53%. Further, *VGG16* exhibited the lowest test accuracy of 62.8% as well as a low training and validation accuracy of 75.28% and 73.83%, respectively. One of the possible reasons is the vanishing gradient problem due to the lack of skip connections, which *ResNet* has incorporated. This limits its ability to focus on intricate details which might be the defining feature of the image. Also, unlike the pre-training of *VGG*, the dataset possesses a lot of elements (objects), leading to a lot more confusion. This aspect reflects its inability towards efficient classification for the use case and, hence, is not considered in further analysis. *InceptionV3* on contrary showed better performance with an accuracy of 88.82% on test and 90.44% on validation split. Next, we combine (ensemble) decisions from Inception, *ResNet*, and *EfficientNet* via Random Forest to achieve the best score on both validation and test, i.e., 94% and 91.83%, respectively. To have a detailed understanding of class-wise results, confusion matrix is generated for each CNN counterpart as well as for ensemble. For *ResNet50*, *EffcientNetB0* and *InceptionV3*, confusion matrices are shown in figures 3.4, 3.5, and 3.6, respectively. Each matrix shows a significant classification score with minimal errors. Further, Table 3.6. shows the combined results from the three fine-tuned CNN backbones.

To evaluate the performance for multi-label classification for tagging, a new metric called 'mean Average Precision (mAP)' is used because of its efficient handling of imbalanced classes, and threshold independence. Further, *mAP* considers both Precision and Recall into a single metric offering insights

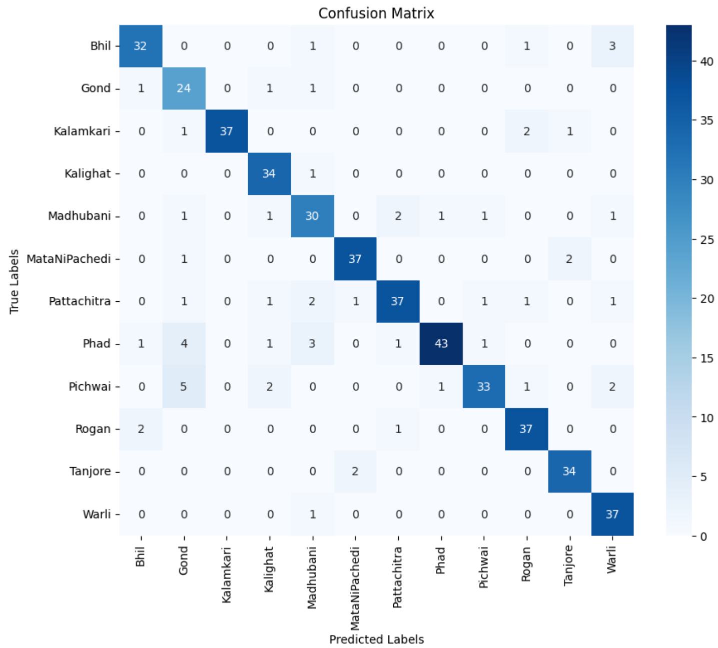


Figure 3.4: Confusion Matrix for ResNet50

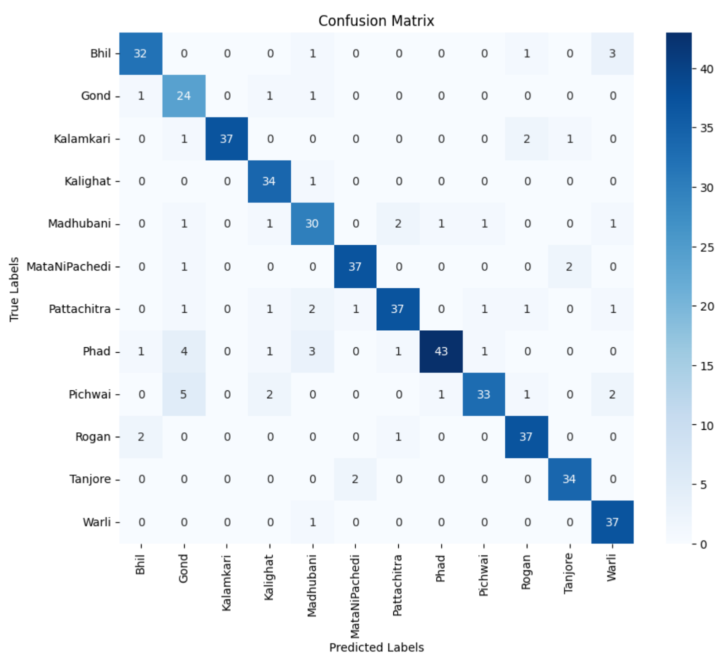


Figure 3.5: Confusion Matrix for EfficietNetB0

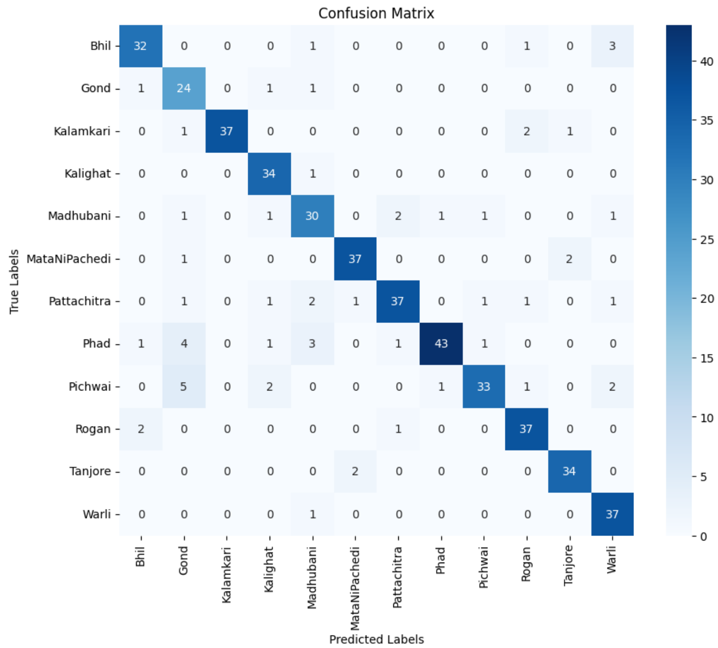


Figure 3.6: Confusion Matrix for InceptionV3

into the precision-recall trade-off for each class independently. Table 3.7 shows the *mAP* scores obtained corresponding to all three CNN-based models.

Hyper parameter tuning: To optimize the performance of the models in both multi-class and multi-label settings, effective hyper parameter tuning is performed, which plays a crucial role in fine-tuning model parameters to achieve optimal results. In the experimentation, various crucial hyper parameters pertinent to training deep neural networks have been explored, including learning rate, batch size, and learning rate reduction factor. Specifically, learning rates were sampled from the available options of 0.001 and 0.0001, while batch size varied across 32, 64, and 128. For the learning rate reduction factor, values of 0.2 and 0.5 were used. This enhances convergence and stability during training. Subsequently, for *RandomForest*, the experimentation was done with hyper parameters, including the number of estimators (n_estimators) and the maximum depth of each tree (max_depth), with values (100, 200, 400, 800, 1000) and (10, 25, 35, 50), respectively. Additionally, the effect of the minimum number of samples required to split an internal node (min_sample_split) was explored across 2, 4, 8, and 16. Upon a comprehensive evaluation, a learning rate of 0.001, batch size of 32, and a reduction factor of 0.5 yielded optimal performance of 96.22 on the validation set and 92.04 on the test set for *EfficinetNetB0*. For ensembling, the combination of 100 estimators, a maximum depth of 25, and a minimum sample

Table 3.6: Combined results (Accuracy) for ResNet50, EfficientNetB0, InceptionV3

Art Form	**InceptionV3**	**ResNet50**	**EffcientNetB0**
Bhil	84.21	87.5	89.47
Gond	64.86	84.21	81.08
Kalamkari	100	81.08	100
Kalighat	90.24	100	95.12
Madhubani	87.17	89.74	92.31
Warli	92.5	89.47	94.73
Phad	95.55	95.55	88.88
Pichwai	92.10	89.47	94.73
MataNiPachedi	97.37	86.49	100
Pattachitra	81.08	90.24	89.19
Rogan	86.84	100	97.36
Tanjore	91.89	91.89	94.59

Table 3.7: Performance metrics on Tagging

CNN Model	**Training mAP**	**Validation mAP**
InceptionV3	86.82	83.05
ResNet50	97.53	81.13
EfficientNetB0	90.33	84.15

split of 8, exhibited superior performance. Further details and results can be referred to the ablation tables below.

To better understand the impact of various hyper parameters on the performance of our *ImageNet* models and the ensembling process using *RandomForest*, an ablation study is conducted. Table 3.8 and 3.9 show the experimentation done with *ResNet50*, *InceptionV3*, and *EfficientB0*, whereas Table 3.10 details their ensemble (*RandomForest*) setting. For each hyper parameter combination, their respective validation and test scores are reported, compared, and discussed.

The experimentation is done with the three promising CNN backbones, namely *ResNet50*, *InceptionV3*, and *EfficientNetB0*. As observed in Table 3.8, *EfficientNetB0* has an overall superior perfor-

Table 3.8: Experiments with learning rate=0.001

Model	Batch Size	Learning Rate Reduction Factor	Validation Accuracy	Test Accuracy
EfficientNetB0	32	0.2	94.89	92.69
	32	0.5	96.22	92.04
	64	0.2	94.89	91.83
	64	0.5	94.89	92.26
	128	0.2	95.33	91.18
	128	0.5	94.89	90.75
ResNet50	32	0.2	92.67	88.82
	32	0.5	92.67	88.39
	64	0.2	92.44	89.03
	64	0.5	93.11	90.32
	128	0.2	93.11	88.82
	128	0.5	93.11	87.31
InceptionV3	32	0.2	92.0	87.96
	32	0.5	92.0	87.74
	64	0.2	90.89	88.6
	64	0.5	92.22	88.17
	128	0.2	90.22	87.1
	128	0.5	91.33	89.46

mance. The best hyper parameter combination is the learning rate of 0.001, batch size of 32, and a reduction factor of 0.5, which yielded an optimal performance of 96.22 on the validation set and 92.04 on the test set for *EfficinetNetB0*.

Similarly, for Table 3.9, with a learning rate of 0.0001, *ResNet50* is the winner with the best parameter combination: learning rate of 0.0001, batch size of 32, and a reduction factor of 0.2, which yielded optimal performance of 94.89 on the validation set and 91.61 on the test set. However, after comparing both tables, *EfficientNetB0* is found to be the best-performing model.

As it can be observed from Table 3.10, increasing N estimators have improved the performance initially and then plateaued. This can be attributed to the fact that *RandomForest* is an ensemble of

Table 3.9: Experiments with learning rate=0.0001

Model	Batch Size	Learning Rate Reduction Factor	Validation Accuracy	Test Accuracy
EfficientNetB0	32	0.2	94.67	92.69
	32	0.5	94.22	90.54
	64	0.2	94.22	92.26
	64	0.5	91.56	90.97
	128	0.2	92.89	89.46
	128	0.5	91.56	91.4
ResNet50	32	0.2	94.89	91.61
	32	0.5	94.22	92.26
	64	0.2	93.78	91.18
	64	0.5	92.89	91.18
	128	0.2	92.67	90.54
	128	0.5	92.89	90.54
InceptionV3	32	0.2	92.22	86.67
	32	0.5	92.0	88.17
	64	0.2	88.67	91.61
	64	0.5	89.78	90.75
	128	0.2	88.22	91.61
	128	0.5	88.44	85.38

strong learners (decision trees) and has a lot of them ($\tilde{8}00$ or 1000) where the order is comparable to the proposed dataset size, making it complex and exhausting its learning capacity. In the process, it becomes invariant to other hyper parameters and produces low to no-variance decisions.

For ensembling, the combination of 100 estimators, a maximum depth of 25, and a minimum sample split of 8 exhibited superior performance.

3.5 Discussion

This work proposes a dataset, *FolkTalent*, that consists of 2273 paintings (digital images) across 12 different folk-art forms. To facilitate comprehensive understanding, a comprehensive array of tags

Table 3.10: Hyperparameter tuning for RandomForest

N Estimators	Max Depth	Min Sample Split	Validation Accuracy	Test Accuracy
100	10	2, 4, 8, 16	94	92.47
	25, 35, 50	8	94	92.9
	25, 35, 50	2, 4, 16	94	92.69
200	10, 25, 35, 50	2, 4, 8, 16	93.11	92.69
400	10	2, 4, 8, 16	93.78	92.9
	25, 35, 50	2, 4, 8, 16	93.78	93.12
800	10, 25, 35, 50	2, 4, 8, 16	93.78	93.12
1000	10	2, 4, 8, 16	93.78	92.9
	25, 35, 50	2, 4, 8, 16	93.78	93.12

describing colors, attributes, objects, patterns, and their unique elements of saliency, are generated using GPT-4 and subsequently verified by an expert.

A novel ensemble approach has been proposed for the effective classification of Indian folk paintings on the *FolkTalent* dataset. The ensemble approach showed a notable change in validation and test accuracy on the dataset with a score of 94.00 and 91.83 percent respectively. Furthermore, with tagging, the proposed method showed a remarkable mean Average Precision (mAP) score of 84.15% on the validation dataset.The methodology employed generates an average of 30 image tags for each of the 12 painting styles, which have potential applications within the domain.

The proposed work is currently limited to twelve types of Indian folk paintings only, namely *Warli, Bhil, Gond, Kalamkari, Madhubani, Phad, Rogan, Tanjore, Mata Ni Pachedi, Kalighat, Pattachitra,* and *Pichwai*. In the proposed dataset, each class contains an average of 190 paintings, which may affect developing standalone CNN based deep-neural models for tasks like image classification or tagging. However, given the size, fine-tuning (domain adaptation) of a pre-trained model is feasible as proposed.

The work can be extended to other Indian folk-art forms like *Pithora, Chittara, Cheriyal, Sohrai, Manjusha, Thangka*, etc. with an increased per class image count. In addition to tagging, the work can also be adapted for object detection from these paintings which can later facilitate the creation of detailed narratives associated with each painting. The generated tags can be incorporated together with textual description/caption in online folk painting catalogs that can significantly improve the user experience by improving efficiency and accuracy of search and retrieval functions. Additionally, the tags can be used to detect the authentic paintings based on the language/narrative (in the form of symbols, objects etc.) used for each art form, thereby contributing to the preservation of Indian folk arts.

Chapter 4

Object detection from Warli paintings

4.1 Scope

Warli paintings, an Indian folk art form, depict distinctive cultural elements with their own symbols, items, and colors. Apart from being an artistic expression, these paintings tell stories that are fundamental to the *Warli* community and function as a great cultural guide. Despite its deep-rooted cultural heritage, to the best of our knowledge, no corpora with object-wise coordinate annotation and corresponding description exist to date. Hence, we present *WarliScan*, a novel dataset comprising 250 digital scans of *Warli* paintings and their annotations. These annotations provide precise bounding box coordinates for each unique item in the paintings, along with corresponding labels for each item. These would be useful for developing models for automating the authenticity validation of *Warli* paintings. Furthermore, this dataset can also be used to enable automated comprehension of cultural artifacts, thus facilitating the preservation of Indian culture. To present the efficacy of this dataset, an object detection baseline is developed by fine-tuning the *YoloV8n* model, which results in a Mean Average Precision score of 0.585.

Indian folk art represents the cultural diversity of the people and the connection between humans and nature. These ancient art forms not only represent the country's vibrant past but also, importantly, reflect the social and religious customs of different regions. From cloth to tapestry weaving, sculpturing, instruments, or painting, the art is a symbol of people from a certain region of the country. In addition to its artistic value, the materials used, for example, natural dyes extracted from flowers for cloth coloring, are based on scientific knowledge.

Traditional paintings, for example whether they originate from the villages of Bihar's *Madhubani* or the arid regions of Rajasthan's *Phad*, showcases unique patterns and subjects that narrate tales of deities, natural elements, celebrations, and ceremonial practices. These art forms are preserved through generations and are being passed down as a skill from ancestors to their descendants. *Warli* paintings are one such style that stands out for its highly minimalist yet expressive approach, offering a glimpse into the lives of the *Warli* community. *Warli* is one of the largest tribes in India, originating back in the Neolithic period.

Warli art[1] is more than just an aesthetic form; it is a means of recording the tribe's folk history, invoking powers, and depicting everyday life. Traditionally, these paintings were done on the rough walls of village huts with a mixture of branches, earth, and cow dung, creating a reddish-brown background with white as the only color used, derived from a mixture of rice paste and water and gum as a binder. The paintings incorporate geometric motifs, like circles, triangles, and squares, to represent the tribe's deep understanding of symmetry in nature, as shown in Figure 4.1.

As an example, the circle denotes the celestial bodies of the sun and the moon, while the triangle is derived from mountains. The pointed trees reflect the profound reverence of the *Warli* community towards the natural world. Additionally, the square symbolizes a sacred parcel of land or a cultural enclosure. On critical analysis of the elements of *Warli* paintings, it can be observed that the dance forms like *Tarpa*, religious rituals, and traditional weddings are frequently shown. These themes serve as both creative expressions and the essential components of the social structure within *Warli* society.

Originating from the *Warli* tribe of Maharashtra and Madhya Pradesh, states of India, the painting serves as a narrative of tribal folklore and their lifestyle. This section delves deeper into several elements, themes, and objects commonly depicted in *Warli* paintings. These paintings not only decorate their mud walls but also preserve their traditions. Figure 4.2 shows various common themes in *Warli* paintings.

Daily Life and Activities: One of the most common themes in *Warli* paintings is the depiction of everyday life of the community. Scenes of farming, harvesting, fishing, and hunting are common, reflecting the tribe's deep connection with nature and their dependence on these activities for sustenance. Men and women are often shown engaged in these tasks, which emphasizes the community's values of labor and cooperation.

[1] https://www.jstor.org/stable/40463595

Figure 4.1: An example of Warli Painting

Source: Warli Art by Nikita Mundekar, Image link

Nature and Animals: Nature plays a crucial role in *Warli* art, showcasing the tribe's belongingness to their natural surroundings. Trees, hills, rivers, and animals are not just background elements for the community but are often central to the narratives depicted. Animals like tigers, elephants, horses, and birds are frequently illustrated, symbolizing various spiritual and practical aspects of *Warli* life. The portrayal of flora and fauna in a harmonious manner highlights the tribe's philosophical view of coexistence with nature.

Rituals and Festivals: *Warli* paintings vividly illustrate the various rituals and festivals celebrated by the tribe. These include scenes from weddings, dances, and harvest festivals, particularly the Tarpa dance, where people join hands, forming chains around a central figure playing the Tarpa (a traditional trumpet-like instrument). This dance is a metaphor for community life and is a recurrent motif in *Warli* art, symbolizing unity and joy.

Sacred Elements: Geometric shapes like circles, triangles, and squares are imbibed with significant meanings. The circle represents the sun and the moon, which is a symbol of timelessness and the eternal cycle of life. Triangles in *Warli* paintings are believed to depict mountains and pointed trees but also represent different deities of the tribe. The square appears to denote sacred enclosures or a piece of land, often seen in the depiction of “Chowk” or ritual spaces.

Symbolic Representation of Social Structures: The layout of *Warli* paintings represents the *Warli* social structure. The combination of scenes laid out in a circular pattern mimics their circular huts. The placement of objects and figures within the painting denotes hierarchical relationships or social roles, providing deeper insights into the social order and governance of the tribe.

Philosophical and Mythological Elements: *Warli* art form also delves into philosophical and mythological themes, depicting deities and spirits that protect the village and its people. These figures are symbolized in abstract designs rather than being depicted in human form. They are often represented by natural elements like the sun, moon, and earth, which hold religious significance to the *Warli* community.

Through these themes, in addition to adornments, *Warli* paintings document the community’s beliefs, folklore, and history, serving them as an invaluable cultural catalog that carries the essence of their heritage. While exploring the significance of these elements, it becomes clear that each stroke and shape in *Warli* art has an associated narrative crafted with profound meanings and communal values.

Warli paintings, which are abundant in symbols and themes, contain story lines that are extremely precious to cultural heritage. In the field of computer vision, object detection models such as *YOLO* [89] have the ability to analyze and classify these components on a large scale, providing a comprehensive analysis. The technological approach not only helps in preservation but also makes them accessible for analytical purposes. The upcoming section outlines the scope of performing object detection in the proposed dataset.

(a) Warli painting illustrating daily life
Source: Warli Art by Dilip Bahotha, Image link

(b) Warli painting illustrating nature and animals
Source: Warli Art by Jivya Soma Mashe, Image link

(c) Tarpa dance in Warli Painting
Source: Warli Art by Nikita Mundekar, Image link

(d) Dev chowk in Warli painting
Source: Warli Art, Image link

Figure 4.2: Various themes in Warli paintings

4.2 Motivation

The progression of object detection techniques in computer vision ranges from basic to sophisticated approaches, gradually increasing detection speed and accuracy. Originally, characteristics were extracted from a picture by scanning it at several scales using methods such as sliding windows paired with a histogram of oriented gradients (HOG) [90]. A major breakthrough happened with the development of region proposal-based techniques, such as R-CNN [91], which used selective search to predict object placements with more accuracy than sliding window techniques.

Fast R-CNN [92] and Faster R-CNN [93] were introduced later and included deep learning to improve the region suggestions and perform object categorization more quickly. The breakthrough was made possible by YOLO (You Only Look Once) [89] and SSD (Single Shot MultiBox Detector) [94], which used a single neural network to the whole image and predicted several bounding boxes and class probabilities in a single assessment, therefore dramatically increasing the detection speed.

YOLO: The baseline model of *YOLO* significantly reduced the complexity of object detection. It was done by predicting bounding frames and class probabilities in a single pass of the neural network, thereby streamlining the detection process. Further iterations incorporated multiple enhancements.

The accuracy and efficiency of *YOLOv2* [95] (YOLO9000) were enhanced through the implementation of batch normalization along with high-resolution classifiers and anchor boxes. *YOLOv3* [96] made additional improvements to the model with the addition of multi-scale predictions and a deeper feature extractor, which was built upon the Darknet-53 architecture. In order to further improve the velocity and precision of real-time applications, *YOLOv4* [97] incorporated functionalities like Cross-Stage Partial connections, Mish activation, and data augmentation methods. *YOLOv5* [98], despite being unofficial within the *YOLO* series, introduced streamlined processes and enhanced performance during the deployment stages by incorporating diverse export formats across multiple platforms. The recent iterations, *Scaled-yolov4* [99] and *YOLOv7* [100], incorporate additional enhancements and optimizations to maintain YOLO's competitiveness in relation to the cutting edge of real-time object detection. Every subsequent iteration enhanced and incorporated functionalities that optimize performance across a multitude of object detection benchmarks.

With these developments, labor-intensive feature engineering is giving way to end-to-end trainable systems offering high accuracy and real-time performance by combining feature extraction, bounding-box regression, and classification into a single model.

Datasets: In the field of object detection, several key datasets have been elementary to the development and evaluation of algorithms. Of these, noteworthy are:

1. **Pascal VOC:** It [101] is among the original datasets offering annotated photos for classification, detection, and segmentation. For systems of early object detection, it has proved essential.

2. **MS COCO (Microsoft Common Objects in Context):** [102]: It provides a richly annotated, wider scale of photos and item types. It has extensive application in cutting-edge object detection research.

3. **ImageNet:** It [103] is mainly used for classification. It also includes a sizable collection of annotated photos for object detection that are very helpful for training deep-learning models.

4. **Cityscapes:** Concentrated on urban settings, it is [104] necessary for uses like automated (driverless) cars since it offers photographs that are specially tagged to help with an understanding of urban street scenes.

Size, intricacy, annotations, and usual use cases of these datasets differ greatly, which affects the direction and effectiveness of object detection systems. Noteworthy progress has been made in the field of object detection from paintings [105, 106, 107], wherein attempts have been undertaken to achieve object recognition from visual arts.

The objective of the research by Westlake et. al. [105] was to identify individuals in a range of photographs or cartoons. The researchers utilized R-CNN for detection and achieved a maximum AP (Average Precision) value of 0.59. This was accomplished by employing fast R-CNN with *VGG16* as the backbone, which was pre-trained on the *ImageNet* dataset and then fine-tuned using the proposed People-Art dataset.

Another attempt was made by Smirnov [106] to detect objects from digital images in real time. A feature fusion approach was employed by combining objects and styles trained using CNN backbones and further passing the output through Support Vector Machine (SVM) [108]. An improved *mAP* was observed after augmentation, as compared to the baselines.

Another comprehensive work was done by Bengamra et al. [107], which includes a detailed survey on object detection in visual arts. This comprises an elaborate comparison and analysis of various studies across different art forms.

Given the scarcity of research on object detection in Indian folk paintings, the motivation behind the proposed work is to efficiently detect objects from *Warli* paintings across different themes.

4.3 Methods

The proposed work started with dataset creation. The dataset consists of 250 *Warli* paintings along with their annotations to denote the coordinates of each object present within the image. The paintings are taken from websites [60, 61, 62, 63, 64, 65, 66] which are direct sellers of Indian folk paintings. This is done to ensure authenticity and avoid picking deep fakes.

Annotation Method: Annotations were performed using the Python library *LabelImg* [109], which provides an interface to manually create bounding boxes. *LabelImg* is preferred due to its simplicity and effectiveness in annotation tasks, allowing users to easily label objects by drawing rectangles around them, which is important for training object detection models. The annotation process employed an iterative approach. Initially, 50 images were fully annotated manually. Subsequently, these annotations were used to fine-tune a *Yolov8n* model [110]. Using this trained model, the next set of 50 paintings was annotated, followed by a manual review. This cycle of fine-tuning the model with fully annotated images and generating annotations for a new set was repeated until all 250 images were annotated.

Train-Test-Validation Split: As shown in Figure 4.3, upon annotation, the dataset is divided into three partitions, namely, training, validation, and test. The training comprises 60% of the dataset, while the validation and test set each contained 20 per cent. The distribution of labels in each class across the three splits namely train, validation and test is shown is Table 4.1.

Model Training: To optimize performance and precision, the experimental architecture for training the object detection model was utilized. The training configuration employed a desktop-based configuration file called *objectDetect.yaml*, in which the dataset parameters, including URLs to training data and class definitions, were exhaustively mentioned. By implementing this setup, it was guaranteed that all essential data inputs (images, annotations, and labels) were accurately designated and readily available throughout the training phase. The computation of the model was performed on two GPUs, as

Table 4.1: Counts per class for each Partition

Classes	**Training set**	**Validation set**	**Test set**
bird	2250	840	233
pot	272	60	200
cooking	11	3	2
woman	1200	300	283
man	1128	250	550
tree	311	66	50
plant	517	70	67
basket	161	120	83
hut	200	120	100
cow	228	70	73
peacock	72	30	33
tarpaDance	28	10	8
sun	67	20	11
drum	17	11	3
moon	28	4	1
chowk	11	1	2
bullockcart	39	10	1
cloud	32	8	6
well	17	3	1
ant	333	14	19

indicated by the device parameter being set to 'mps', which signifies the use of Apple's Metal Performance Shaders to improve processing capabilities.

A significant number of epochs, precisely 500, were incorporated for training to provide the model with sufficient time to acquire knowledge and adjust to the characteristics of the dataset. The images were set at a resolution of 640x640 pixels to effectively manage computational demands while ensuring detailed feature recognition was maintained. The training procedure was coordinated by utilizing the *train* method of the model, which methodically oversaw the data movement within the model

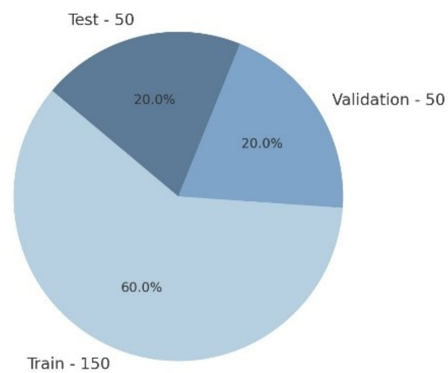


Figure 4.3: Data Distribution across each partition

Table 4.2: mAP Values for YOLO V8n Pre and Post Fine-Tuning

Model	mAP@0.5
Yolov8 pre-trained	0.195
Yolov8 fine-tuned	0.585

and guaranteed optimal utilization of GPU resources. The primary objective of this all-encompassing configuration was to attain elevated degrees of precision in object detection while preserving feasible computational burdens.

4.4 Results

Mean Average Precision (mAP) values were evaluated by the study of several *YOLO* variants. The investigation started out with baseline *mAP* values established using the original *YOLO* design. The same model was next refined using training data from our dataset. Performance was significantly improved by this fine-tuning procedure. The Average Precision values for each class for every iteration is showing in Table 4.3. The last iteration *YoloV8* with an *mAP* of 0.58 showed the greatest improvement. The details of the ablation can be found in Table 4.2 below.

The precision-recall curves for the several classes found in the dataset are shown in the graph in Figure 4.4; each class is indicated by a different color. The curves demonstrate the trade-off for each class between recall (x-axis) and precision (y-axis) as the decision threshold is changed. Greater performance is indicated by higher numbers on both axes.

Table 4.3: Average Precision values for top 15 labels Across Iterations

Class	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
bird	0.614	0.655	0.787	0.837	0.848
pot	0.000	0.000	0.172	0.577	0.445
cooking	0.000	0.000	0.000	0.000	0.000
woman	0.495	0.464	0.634	0.678	0.702
man	0.274	0.339	0.456	0.453	0.609
tree	0.799	0.653	0.654	0.776	0.822
plant	0.443	0.672	0.734	0.810	0.782
basket	0.000	0.000	0.423	0.506	0.586
hut	0.563	0.554	0.600	0.508	0.547
cow	0.045	0.132	0.542	0.774	0.810
peacock	0.000	0.000	0.000	0.000	0.052
tarpaDance	0.000	0.000	0.995	0.995	0.995
sun	0.000	0.000	0.028	0.498	0.500
bullockcart	0.000	0.000	0.000	0.995	0.995
cloud	0.000	0.000	0.000	0.000	0.076
all classes	0.216	0.231	0.402	0.561	0.585

Observations: The best performances are shown by the ‘tarpaDance’, ‘bullockCart’, and ‘bird’ classes, whose precision-recall curves approach the upper-right corner, indicating great detection accuracy. The ‘Cooking’ class was not successfully detected at all, as seen by its precision of zero across all recall levels. Additionally, performing poorly with extremely low precision is the ’cloud’ class. The ’all classes’ curve provides the average performance across all classes; an *mAP* of 0.585 indicates a moderate overall detecting capability.

After fine-tuning the backbones, the results from the model with the best performance are shown in figures 4.5 and 4.6. Detailed ablations can be found in the supplementary. From Figure 4.5, it can be seen that the model could easily recognize birds and plants. It failed to detect a few birds sitting on the tree, forming part of the tree branches.

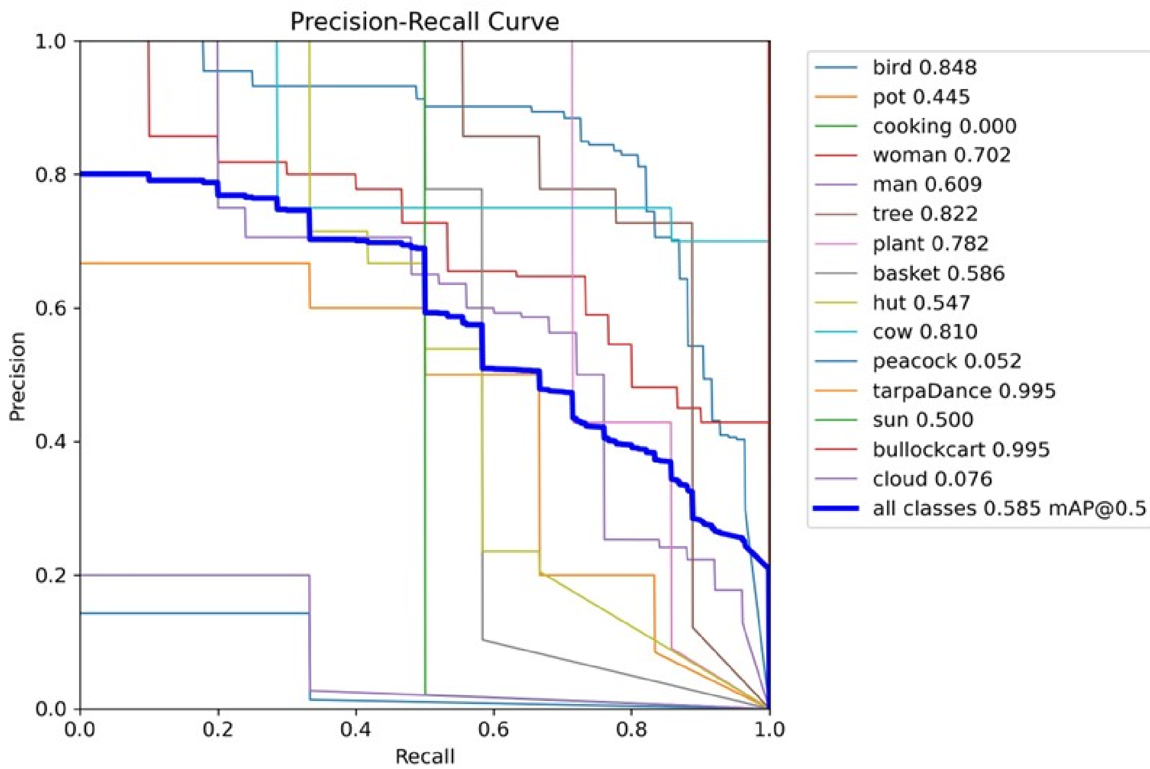


Figure 4.4: Precision-Recall curve for various classes

On the contrary, in Figure 4.6, we see that the model could effectively detect even minute elements like pots and baskets. It could also efficiently detect cows, plants, trees, birds, men, women, huts, pots, and baskets.

4.5 Discussion

The ablation study shows that the pre-trained object detection models trained in the real-world images failed to detect the hand-drawn characters in *Warli* paintings, thereby creating a motivation for manual annotation of *Warli* paintings. The best performance on our *WarliScan* dataset is obtained using the *YoloV8* model. An mAP of 0.585 is observed. The results demonstrate the application of *YOLO* models for object detection in *Warli* paintings, achieving significant improvements in mean average precision through iterative training and fine-tuning processes as shown in 4.3. The use of a custom dataset, *WarliScan*, has proven effective in enhancing the model's ability to recognize culturally significant objects within the paintings, ultimately contributing to the digital preservation and understanding of *Warli* art. By training models on *WarliScan* dataset, the models' capacity to recognize and comprehend culturally important patterns is enhanced, which is important for both safeguarding the artistic legacy and facilitating authenticity assessments in the growing art industry. The use of improved object detection in digital preservation is a significant step towards protecting and preserving the cultural tales found in *Warli* paintings in their pristine form.

The proposed work can be extended in multiple ways by future studies on deep fake detection. The robustness and generalizability of the object detection algorithms may be further enhanced by increasing the count and the diversity of *Warli* paintings in the database. Investigating the incorporation of further sophisticated machine learning methods, such as generative adversarial networks (GANs) or semantic segmentation, may provide a further understanding of the themes and stylistic components of the paintings. Utilizing the proposed models to additional Indian folk-art genres may offer a thorough technical solution for a wider range of cultural heritage preservation and research.

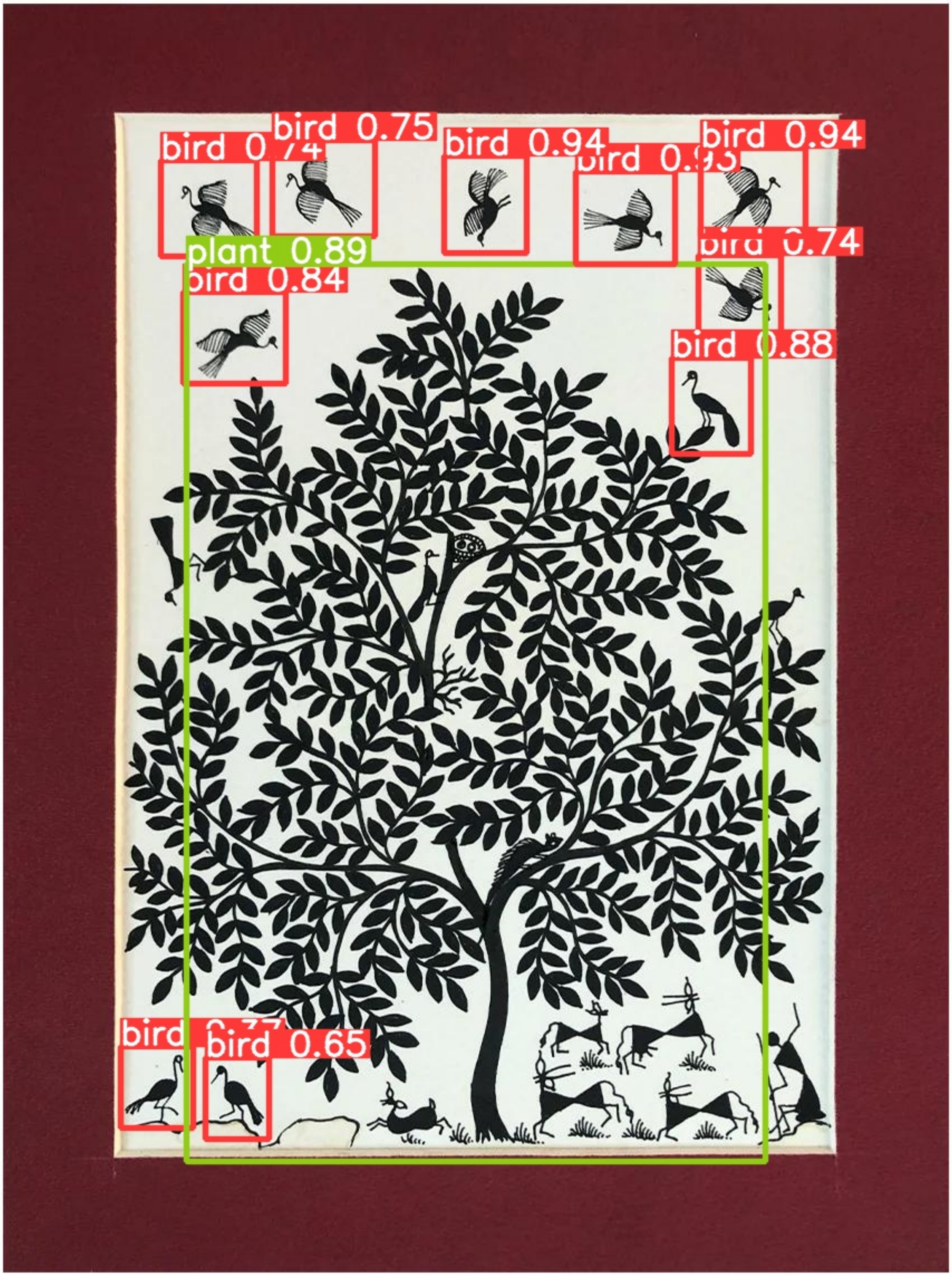


Figure 4.5: Object Detection in a simple Warli painting

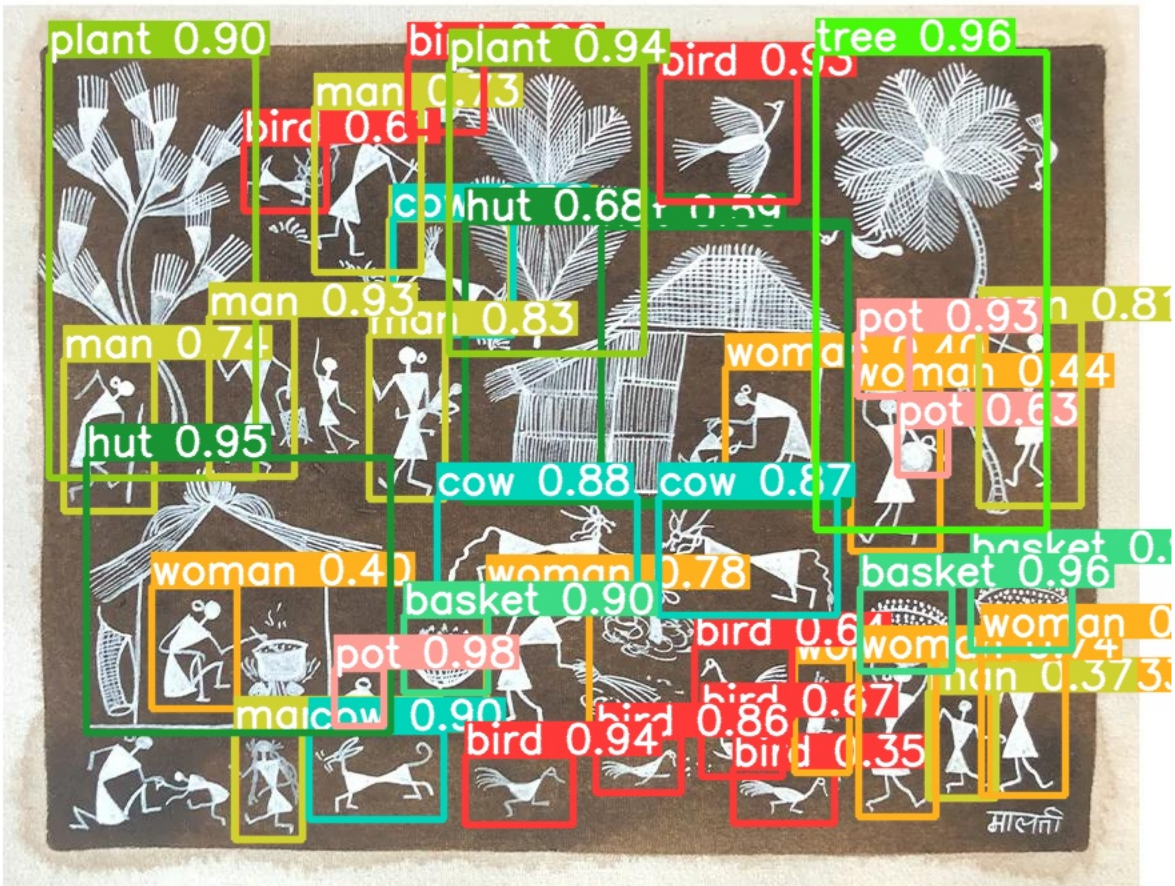


Figure 4.6: Object detection in a complex and intricate Warli painting

Chapter 5

Conclusion

The proposed work represents a groundbreaking stride in both cultural studies and technology. The two studies under discussion, “FolkTalent: Enhancing Classification and Tagging of Indian Folk Paintings” and “WarliScan: Object Detection from Warli Paintings,” demonstrate how state-of-the-art machine-learning techniques can be employed to preserve and comprehend the intricate subtleties of Indian folk paintings.

To begin with, the “WarliScan” program closes a significant gap in the cultural preservation of Warli paintings by introducing a unique dataset designed for object detection within these artworks. This effort advances our understanding of Warli iconography and paves the way for developing increasingly sophisticated algorithms that can identify and categorize intricate elements in folk art worldwide. The YoloV8n model’s mean average precision score of 0.58 indicates potential improvements in these technologies’ dependability and accuracy in the context of art analysis. However, “FolkTalent” expands the scope of image classification and labeling to include a broader range of Indian folk art styles. This hybrid model, which combines the best CNN architectures with ensemble RandomForest, achieves an impressive 91.83

Despite these advancements, there are some limitations to the proposed work. The current focus is limited to Warli paintings for object detection, and the model has shown occasional failures in detecting various components. Common failures include peacocks, cooking scenes, and other elements like birds sitting on the branches of tree that blend in with the surrounding environment. Additionally, the study focuses on only twelve Indian folk painting types: Warli, Bhil, Gond, Kalamkari, Madhubani, Phad, Rogan, Tanjore, Mata Ni Pachedi, Kalighat, Pattachitra, and Pichwai, with an average of 190 paintings per class. This limits the development of standalone CNN models for classification or tagging. The usage of multi-label classification for tagging instead of more effective contrastive learning models like CLIP or BLIP is another constraint.

There is a great possibility going ahead to expand the dataset to include other folk-art forms such as Pithora, Chittara, Cheriyal, Sohrai, Manjusha, Thangka, and others. Search and retrieval efficiency could be much enhanced by modifying the object identification framework to provide detailed narratives and combining produced tags with textual descriptions in online catalogs. Using contrastive learning models such as CLIP or BLIP might help to improve tagging performance. The model's generalizability and resilience would be improved by increasing the variety and count of Warli paintings in the WarliScan collection. Furthermore, using more advanced machine learning techniques, such semantic segmentation or generative adversarial networks (GANs) could offer deeper understanding of the concepts and artistic elements of Warli paintings.

By using innovative technology to guarantee accessibility and preservation, both projects show a great regard for old cultural legacy. Given the many difficulties involved in physically maintaining modern art, the study on efforts at digital preservation provides important new perspectives. The methods used set a standard for related uses in other fields of cultural preservation, implying that the combination of artificial intelligence and art could result in major progress in our knowledge and capacity to protect our common cultural legacy. These systems also show the advantages of multidisciplinary co-operation by combining fields including computer vision, cultural studies, and artificial intelligence, so advancing a better knowledge of artistic expressions from many civilizations. These technologies seem to have endless future uses; they could improve museum visits with augmented reality or generate more dynamic, interactive internet archives allowing art to be accessed worldwide.

In conclusion, techniques for researching and preserving Indian folk art, such as deep learning and machine learning, contribute to preserving cultural narratives and improve our capacity to educate people worldwide about these rich traditions. These technologies have the potential to lead to even more advanced techniques and tools for cultural preservation, ensuring that folk art endures and inspires future generations.

Chapter 6

Publications

[1] N. Hada, A.K. Singh, K. Vemuri, “FolkTalent: Enhancing Classification and Tagging of Indian Folk Paintings, ” 2023, International Conference on FOSS Approaches towards Computational Intelligence and Language Technology (ICFOSS-CIL24), March 2024.

[2] N. Hada, K. Vemuri, “WarliScan: Object Detection fromWarli Paintings” submitted at Conference on AI, Ethics, and Society (AIES), 2024. Conference in October 2024, notification in July 2024.

Bibliography

[1] D. Bandyopadhyay and N. Sabnani, "Interpretation of narrative illustrations: A comparative study of illustrations by the children from indigenous and formal art education backgrounds," in *ICERI2015 Proceedings*. IATED, 2015, pp. 1925–1933. 23

[2] B. Joseph, "The bhil culture in central india," *Prajñā Vihāra: Journal of Philosophy and Religion*, vol. 14, no. 1-2, 2013. 23

[3] T. Bagchi and P. Chaudhuri, "Tribal paintings of rajasthan: A micro-study of continuity and change," *Man in India*, vol. 95, no. 3, pp. 463–482, 2015. 23

[4] G. Malik and S. Sethi, "Gond paintings: a historical evolution," *International Journal of Applied Home Science*, vol. 5, no. 4, pp. 832–836, 2018. 23

[5] S. R. Padhi and M. P. Goswami, "Tribal folklore to aesthetic and religious painting: Transition of oral narratives to visual art," *Journal of religion and health*, vol. 63, no. 2, pp. 877–888, 2024. 23

[6] M. P. Goswami and P. Yadav, "Dots and lines: Semiotics of the motifs in gond painting," *Journal of Media and Communication*, vol. 3, no. 2, pp. 35–50, 2020. 23

[7] N. Das, S. Kundu, and S. Deb, "Image synthesis of warli tribal stick figures using generative adversarial networks," in *2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 2021, pp. 266–271. 23

[8] P. Masram and N. Vishwakarma, "Warli painting in maharashtra: Evolution and transition," *Think India Journal*, vol. 22, no. 14, pp. 9613–9626, 2019. 23

[9] K. M. Rao, "Warli painting: Aesthetics of indigenous art form of india," *ShodhKosh: Journal of Visual and Performing Arts*, vol. 3, no. 2, pp. 207–215, 2022. 23

[10] J. Phogat and M. Raja, “A study of kalamkari scroll paintings in particular and temple murals in andhra pradesh-a structural analysis,” *Journal of Pharmaceutical Negative Results*, pp. 1823–1831, 2022. 23

[11] M. Divakala, M. Vasantha, N. Campus, and O. H. tech City, “Kalamkari-the painted temple cloths.” 23

[12] S. K. Shekhawat, “Phad: The visual oral narrative of rajasthan,” 1819. 24

[13] S. Sharma, U. K. Dharmalingam, and S. Karmakar, “A scoping review on lesser-known tradition of rajasthan—‘phad’,” in *International Conference on Research into Design*. Springer, 2023, pp. 227–239. 24

[14] D. Joshi, “Translating orality: Pictorial narrative traditions with reference to kaavad and phad,” *Cankaya University Journal of Humanities and Social Sciences*, vol. 12, no. 1-2, pp. 85–98, 2018. 24

[15] S. VIMALA, “Analysis of the tanjore marathas’paintings,” *International Journal of Educational Review, Law And Social Sciences (IJERLAS)*, vol. 2, no. 5, pp. 1409–1415, 2022. 24

[16] P. S, B. S, S. M, and V. T, “Restoration of tanjore paintings using segmentation and in-painting techniques,” *Heritage Science*, vol. 10, no. 1, p. 41, Mar 2022. 24

[17] N. Singh and V. Gupta, “Mata ni pachedi: Tales of the goddess on textiles from gujrat, india,” *Craft Research*, vol. 15, no. 1, pp. 109–126, 2024. 24

[18] P. Pathak, “Representation of female deities in mata-ne-pachhedi and contemporary changes in craft.” 24

[19] S. Ghosh, “Madhubani painting—vibrant folk art of mithila,” *Art and Design Review*, vol. 8, no. 2, pp. 61–78, 2020. 24

[20] P. Lata, P. Rani, and R. Kumari, “Geometrical conception in folk art: Madhubani painting of bihar,” *Boletin de Literatura Oral-The Literary Journal*, vol. 10, no. 1, pp. 3265–3269, 2023. 24

[21] S. Varshney, C. V. Lakshmi, and C. Patvardhan, “Madhubani art classification using transfer learning with deep feature fusion and decision fusion based techniques,” *Engineering Applications of Artificial Intelligence*, vol. 119, p. 105734, 2023. 24, 26, 27

[22] P. Yadav, R. Guru, and U. Saini, “Pichwai: The tapestry art of nathdwara.” 24

[23] ——, “Pichwai: The tapestry art of nathdwara.” 24

[24] T. Rathi, “Study of traditional motifs and patterns vocabulary of rogan painting from gujarat,” 2019. 24

[25] B. Meena and A. Chitra, “A glorious journey of roghan painting from kutch, india.” 24

[26] P. Sanyal, “Kalighat paintings: A review,” *Chitrolekha International Magazine on Art and Design: Volume 3, Number 1, 2013*, p. 2, 2013. 24

[27] D. Sarma and B. Sullivan, “Under the gaze of kali: Exhibitionism in the kalighat painting exhibition at the cleveland museum of art,” *Bruce M. Sullivan*, pp. 26–34, 2015. 24

[28] M. Wheeler, L. Burgio, and M. Shulman, “Materials and techniques of kalighat paintings: pigment analysis of nine paintings from the collections of the victoria and albert museum,” *Journal of the Institute of Conservation*, vol. 34, no. 2, pp. 173–185, 2011. 24

[29] A. Pareek and S. Pant, “” pattachitra”-a traditional hand painted textile of orissa.” *International Journal of Recent Research Aspects*, vol. 2, no. 4, 2015. 24

[30] P. Dey, N. Ray, C. K. Dash, D. K. Das, and P. K. Das, “Mapping of socio-economic condition and occupational structure of patuas in the context of pattachitra (scroll painting): A study from bharatpur village in bankura, west bengal, india.” 24

[31] E. Sharma, “Tribal folk arts of india,” *Journal of International Academic Research for Multidisciplinary*, vol. 3, no. 5, pp. 300–308, 2015. 25

[32] S. Kumar, A. Tyagi, T. Sahu, P. Shukla, and A. Mittal, “Indian art form recognition using convolutional neural networks,” in *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2018, pp. 800–804. 26, 27, 28

[33] K. Mane and U. Shrawankar, “An improved indexing technique for tribal art retrieval system,” in *2023 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE, 2023, pp. 1–5. 26, 27, 28

[34] D. Podder, M. Shashaank, J. Mukherjee, and S. Sural, “Ihird: A data set for indian heritage image retrieval,” *Digital Techniques for Heritage Presentation and Preservation*, pp. 51–73, 2021. 26, 27

[35] A. Sharma and R. Jadon, “Indian visual arts classification using neural network algorithms,” in *International Conference On Innovative Computing And Communication*. Springer, 2023, pp. 483–493. 26, 36

[36] ——, “Indian visual arts classification using neural network algorithms,” in *International Conference On Innovative Computing And Communication*. Springer, 2023, pp. 483–493. 27

[37] K. Mondal and H. Anita, “Categorization of artwork images based on painters using cnn,” in *Journal of Physics: Conference Series*, vol. 1818, no. 1. IOP Publishing, 2021, p. 012223. 27

[38] S. Kumar, A. Tyagi, T. Sahu, P. Shukla, and A. Mittal, “Indian art form recognition using convolutional neural networks,” in *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2018, pp. 800–804. 27

[39] F. Milani and P. Fraternali, “A dataset and a convolutional model for iconography classification in paintings,” *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 14, no. 4, pp. 1–18, 2021. 27

[40] H. Yang and K. Min, “Classification of basic artistic media based on a deep convolutional approach,” *The Visual Computer*, vol. 36, no. 3, pp. 559–578, 2020. 27

[41] A. Munir, J. Kong, and M. A. Qureshi, “Overview of convolutional neural networks,” 2024. 27

[42] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk, and C. Von Kalle, “Skin cancer classification using convolutional neural networks: systematic review,” *Journal of medical Internet research*, vol. 20, no. 10, p. e11936, 2018. 27

[43] S. Bhardwaj, “Object detection using convolutional neural networks: A review,” *International Journal of Innovative Research in Computer Science & Technology*, vol. 9, no. 6, pp. 290–294, 2021. 27

[44] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object detection using convolutional neural networks," in *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018, pp. 2023–2027. 27

[45] F. Sultana, A. Sufian, and P. Dutta, "Evolution of image segmentation using deep convolutional neural network: A survey," *Knowledge-Based Systems*, vol. 201, p. 106062, 2020. 27

[46] D. N. Gonçalves, V. A. de Moares Weber, J. G. B. Pistori, R. da Costa Gomes, A. V. de Araujo, M. F. Pereira, W. N. Gonçalves, and H. Pistori, "Carcass image segmentation using cnn-based methods," *Information Processing in Agriculture*, vol. 8, no. 4, pp. 560–572, 2021. 27

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 27

[48] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114. 27

[49] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826. 27, 28

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 27, 28

[51] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 34–42. 27

[52] M. Jocic, D. Obradovic, V. Malbasa, and Z. Konjo, "Image tagging with an ensemble of deep convolutional neural networks," in *Proceedings of the 2017 International Conference on Information Society and Technology, ICIST Workshops, Beijing, China*, 2017, pp. 17–20. 27

[53] H. Jang and S. Cho, "Automatic tagging for social images using convolution neural networks," *Journal of KIISE*, vol. 43, no. 1, pp. 47–53, 2016. 27

[54] T. Kim, J. Lee, and J. Nam, "Sample-level cnn architectures for music auto-tagging using raw waveforms," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 366–370. 27

[55] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017. 28, 36

[56] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000. 28

[57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. 28

[58] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017. 29

[59] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. 29

[60] ArtZolo, "Homepage," https://www.artzolo.com/, n.d., accessed: date. 29, 53

[61] Memeraki, "Paintings," https://www.memeraki.com/collections/paintings, n.d., accessed: date. 29, 53

[62] U. Tribes, "Homepage," https://universaltribes.com/, n.d., accessed: date. 29, 53

[63] Gaatha, "Homepage," https://shop.gaatha.com/, n.d., accessed: date. 29, 53

[64] I. F. Art, "Homepage," https://indianfolkart.org/, n.d., accessed: date. 29, 53

[65] Artisera, "Homepage," https://www.artisera.com/, n.d., accessed: date. 29, 53

[66] E. I. Art, "Homepage," https://www.exoticindiaart.com/, n.d., accessed: date. 29, 53

[67] D. Ethnic, "Desically Ethnic – Homepage," https://desicallyethnic.com/, accessed: 2024-05-14. 29

[68] T. Swstik, "The Swstik – Homepage," https://theswstik.com/, accessed: 2024-05-14. 29

[69] A. Kejriwal, "Ayush Kejriwal – Homepage," https://ayushkejriwal.com/, accessed: 2024-05-14. 29

[70] Khoj Crafts, "Khoj Crafts – Homepage," https://khojcrafts.com/, accessed: 2024-05-14. 29

[71] Tvami, "Tvami – Homepage," https://tvami.com/, accessed: 2024-05-14. 29

[72] P. Mirra, "Prosperity Mirra – Homepage," https://www.prosperitymirra.com/, accessed: 2024-05-14. 29

[73] Mangala Tanjore Paintings, "Mangala Tanjore Paintings – Homepage," https://mangalatanjorepaintings.com/, accessed: 2024-05-14. 29

[74] MyAngadi, "MyAngadi – Homepage," https://www.myangadi.com/, accessed: 2024-05-14. 29

[75] GiTagged, "GiTagged – Homepage," https://www.gitagged.com/, accessed: 2024-05-14. 29

[76] Dastkari Haat Samiti, "Kalighat Paintings," https://dastkarihaat.com/collections/kalighat-paintings, accessed: 2024-05-14. 29

[77] Tamil Tanjore Art Gallery, "Shop," https://ttagallery.in/shop, accessed: 2024-05-14. 29

[78] Pepperfry, "Pepperfry – Homepage," https://www.pepperfry.com/, accessed: 2024-05-14. 29

[79] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015. 34, 35

[80] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012. 35

[81] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 35

[82] M. C. Mukkamala and M. Hein, "Variants of rmsprop and adagrad with logarithmic regret bounds," in *International conference on machine learning*. PMLR, 2017, pp. 2545–2553. 35

[83] H. Attouch and J. Peypouquet, "The rate of convergence of nesterov's accelerated forward-backward method is actually faster than 1/k^2," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1824–1834, 2016. 35

[84] L. Bottou, "Online algorithms and stochastic approximations," *Online learning in neural networks*, 1998. 35

[85] A. Zell, *Simulation neuronaler netze*. Addison-Wesley Bonn, 1994, vol. 1, no. 5.3. 36

[86] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018. 36

[87] I. Goodfellow, Y. Bengio, and A. Courville, "6.2. 2.3 softmax units for multinoulli output distributions," *Deep learning*, vol. 180, 2016. 36

[88] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *International workshop on artificial neural networks*. Springer, 1995, pp. 195–201. 36

[89] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788. 49, 51

[90] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893. 51

[91] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587. 51

[92] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448. 51

[93] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015. 51

[94] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37. 51

[95] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271. 51

[96] ——, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. 51

[97] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020. 51

[98] G. Jocher, A. Stoken, J. Borovec, A. Chaurasia, and L. Changyu, "ultralytics/yolov5. github repository," *YOLOv5*, 2020. 51

[99] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," in *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 2021, pp. 13 029–13 038. 51

[100] ——, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464–7475. 51

[101] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303–338, 2010. 52

[102] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755. 52

[103] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255. 52

[104] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223. 52

[105] N. Westlake, H. Cai, and P. Hall, "Detecting people in artwork with cnns," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 825–841. 52

[106] S. Smirnov and A. Eguizabal, "Deep learning for object detection in fine-art paintings," in *2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo)*. IEEE, 2018, pp. 45–49. 52

[107] S. Bengamra, O. Mzoughi, A. Bigand, and E. Zagrouba, "A comprehensive survey on object detection in visual art: taxonomy and challenge," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 14 637–14 670, 2024. 52, 53

[108] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995. 52

[109] D. Tzutalin, "tzutalin/labelimg," *Github*, 2015. 53

[110] G. Jocher, A. Chaurasia, and J. Qiu, "Yolo by ultralytics (version 8.0. 0)[computer software]," *YOLO by Ultralytics (Version 8.0. 0)[Computer software]*, 2023. 53