Cureus Journal of Computer Science

Open Access Original Article

“Stripecode Guardians”: Individual Tiger Identification Using Machine Learning

Reetichi Pattanaik [1] , Sankar Kumar Mridha [1]

1. Department of Computer Science and Information Technology, Siksha ‘O’ Anusandhan Deemed to be University, Bhubaneswar, IND

Corresponding author: Sankar Kumar Mridha, sankarmridha@soa.ac.in

Received 11/16/2024
Review began 11/17/2024
Review ended 07/11/2025
Published 07/11/2025

DOI: https://doi.org/10.7759/s44389-024-02092-4

Abstract

The identification of individual animals and species is crucial for effective ecology and wildlife conservation. By employing advanced image processing and pattern recognition techniques, animals like tigers, zebras, and others can be distinguished based on their unique coat patterns or stripes. This article focuses on the use of tiger stripes as distinct identifiers, much like human fingerprints, allowing for precise and reliable identification to support conservation efforts. However, manual classification of these patterns requires substantial human effort and specialized knowledge. Recent technological advancements, particularly in machine learning, provide innovative solutions to streamline this process. In this paper, we propose a One-Class Support Vector Machine based on advanced machine learning techniques to identify the individual tiger through their stripes. These methods analyze various aspects of tiger images, such as color, shape, and texture of the unique stripe pattern of individual tigers. We used a curated dataset consisting of 8,370 labeled images from 9 individual tigers. Each tiger had 30 original images captured through camera traps, and for each of these, augmented versions were generated using various image transformation techniques. We validated the model using one-vs-all cross-validation strategy and observed an average accuracy of 88% in distinguishing target individuals from others. Future work includes extending the model to identify other big cats such as leopards and lions, as well as endangered species like the striped hyena and crocodiles, including Gharial, Mugger, and saltwater crocodiles. Additionally, our goal is to develop a user-friendly application for real-time identification of these species.

Categories: AI applications, Image Processing and Analysis, Machine Learning (ML)
Keywords: tiger identification, machine learning, ocsvm, pca, endangered species, wildlife protection, monitoring

Introduction

Wildlife encompasses all organisms that live freely in nature without human interference [1]. In the intricate fabric of ecology, wildlife plays a dual role as both weaver and thread, intricately bound to balance and stability. Through their existence, wildlife orchestrates the complex symphony of the food chain, with each species contributing a unique note that creates a harmonious balance. Acting as nature's architects, wildlife shapes landscapes and habitats. Their absence disrupts this delicate equilibrium. In the modern era, environmental challenges demand sophisticated solutions for conservation efforts, sustainable management, and habitat protection. The primary aim is to maintain the ecosystem’s delicate balance while enabling wildlife to coexist alongside human activities. Safeguarding natural habitats, addressing illegal wildlife trade, promoting sustainable land management practices, and fostering community engagement are critical components of achieving this harmony. Ultimately, the future of our planet and its biodiversity rests in our collective hands, urging us to act with urgency and compassion to preserve Earth's rich life forms for generations to come. Wildlife crimes, however, extend far beyond environmental degradation. These illicit activities fuel corruption, organized crime, and instability, undermining the rule of law and stunting socioeconomic growth in affected regions. They also raise profound ethical questions, challenging our moral duty to coexist with other species on Earth. Wildlife crimes include illegal trade, poaching, trafficking, and exploitation of wildlife for profit. The destruction of habitats and the killing of endangered species for financial gain pose severe threats to biodiversity and ecological balance.

The survival of endangered species is at risk, teetering on the brink of extinction due to illegal wildlife trafficking [2]. The trade of skin, meat, ivory, scales, bones, and body parts has devastated various species. A tragic example is the pangolin, which is hunted extensively for its scales and meat, making it the world’s most trafficked animal. Misconceptions about its medicinal properties and status symbolism only exacerbate the crisis. This situation extends to other species, such as tigers, whose bones and pelts are highly prized, and elephants, who are slaughtered for their ivory. The illegal wildlife trade not only threatens individual species but also global conservation efforts, fueling corruption, destabilizing ecosystems, and jeopardizing human health.

Addressing this crisis requires a unified global effort that combines stringent law enforcement, cross-

How to cite this article
Pattanaik R, Mridha S (July 11, 2025) “Stripecode Guardians”: Individual Tiger Identification Using Machine Learning. Cureus J Comput Sci 2 : es44389-024-02092-4. DOI https://doi.org/10.7759/s44389-024-02092-4

border cooperation, and campaigns to reduce demand. By dismantling trafficking networks, disrupting supply chains, and raising awareness about endangered species, we can help stem the tide of extinction and protect our planet's biodiversity. In India, tiger poaching remains a significant threat, as the illegal trade in tiger parts undermines conservation efforts to protect these iconic big cats [3]. Authorities are combating this threat with increased patrols, intelligence-led operations, and enhanced collaboration with local communities, while addressing the root causes of poaching through sustainable development programs. Raising awareness alone, however, is not enough. Conservation efforts must also address the rampant over exploitation of forest resources. In northeastern and central India, vast "empty forests" have emerged as wildlife is hunted, sold, or driven away. Wildlife trafficking has reached the scale of the drug trade, underscoring the need for robust knowledge of ecosystems, animal behavior, and geography. To combat wildlife crimes, conservation initiatives such as national parks, wildlife sanctuaries, botanical gardens, and bio-reserves have been introduced.

Tracking animal populations through censuses is a vital aspect of wildlife conservation. The International Union for Conservation of Nature (IUCN) maintains the IUCN Red List, offering comprehensive global species conservation reports [4]. National wildlife agencies gather population data through field surveys, camera trapping, satellite tracking, and other advanced techniques. These censuses provide critical insights into population trends and habitat health, enabling conservationists to devise informed strategies to mitigate threats and promote the survival of endangered species. Together, national parks, wildlife sanctuaries, botanical gardens, biosphere reserves, and census initiatives serve as strongholds in the fight against wildlife crime. They provide hope for the future, reinforcing our collective endeavor to preserve Earth's rich tapestry of life for generations to come. Male tigers typically reach sexual maturity around 4 to 5 years of age, while females mature slightly earlier, between 3 and 4 years. Unlike some species with fixed mating seasons, tigers can mate year-round. Litter sizes vary widely, typically ranging from one to seven cubs, though two to four cubs are most common. Once born, tiger cubs depend entirely on their mother for care and protection. She nurses them and teaches essential hunting and survival skills until they can live independently, around 18 to 24 months. This highlights the maternal investment required for survival in the competitive wild.

India's rich biodiversity allows for the identification of individual animals through unique characteristics. Figure *1* illustrates that animals can be distinguished by features such as coat patterns or coloration, facial features, footprints or tracks, scale patterns, and natural markings, all of which provide valuable tools for conservation efforts [5,6].

1. Coat Patterns or Coloration: Many animals exhibit distinct coat patterns or colors. For example, tigers' intricate stripes or leopards' speckled spots offer unique identifiers.

2. Facial Features: Some animals possess distinctive facial markings that aid identification, such as the unique scars on bears or facial patterns on cheetahs.

3. Footprints or Tracks: The size, shape, and arrangement of animal tracks provide clues to an animal's identity.

4. Scale Patterns: Certain reptiles and amphibians, like snakes, have unique scale patterns that aid in identification.

5. Natural Markings: Distinctive natural markings, scars, or injuries, such as ear notches or missing limbs, can also help in recognizing individual animals.

FIGURE 1: Image of facial features, coat pattern or coloration, scale pattern, footprints or track, and natural markings, respectively.

Just as fingerprints are unique identifiers for humans, animals' coat patterns, such as those found on tigers, zebras, and giraffes, act as their natural fingerprints [7]. In tigers, for example, the patterns on each side of the body can differ as shown in Figure *2*. Although these characteristics are not identical to human fingerprints, they are invaluable for researchers tracking individual animals, particularly when conventional methods fall short.

FIGURE 2: Fingerprints of two different persons and coat patterns of two different tigers, respectively.

Tiger stripes are the result of complex genetic interactions, shaped by the expression of multiple genes. This genetic variation produces the wide range of stripe patterns observed in tigers, which can differ across subspecies, such as the Bengal and Amur tigers (in Figure 3). Stripes not only indicate an individual's health and maturity but also serve as a social language, allowing tigers to recognize clan members within their dense jungle habitats. These unique patterns play a critical role in fostering bonds and hierarchies within tiger communities, aiding survival.

FIGURE 3: Different species of tigers found in the world.

Source: [8]

The dwindling tiger population in India calls for urgent conservation measures. Monitoring individual tigers' behavior and habitats is crucial for addressing the threats they face, but this task is challenging and requires keen observational skills, precision, and dedication. Traditional methods, like manually grouping and labeling camera-trapped images, are time-consuming and labor-intensive.

1. As the tiger population declines, monitoring behavior and habitats is essential for effective conservation.

2. Protecting tigers involves navigating remote environments, demanding significant skill and dedication from researchers.

3. Traditional methods, although critical, are resource-intensive, highlighting the need for innovative approaches.

4. Accurate identification and population counts are essential for conservation strategies; even minor errors can have serious implications.

5. Conservation efforts require collaboration among governments, NGOs, local communities, and researchers.

6. Protecting tigers reflects a commitment to biodiversity and a moral duty to preserve wildlife for future generations.

This research proposes an automated framework for individual tiger identification using a hybrid machine learning strategy that incorporates a Principal Component Analysis (PCA) [9,10] for feature extraction and a One-Class Support Vector Machine (OCSVM) [11,12] for anomaly-based classification. The core idea is to model the stripe pattern of a single tiger as the “normal class” and detect all other inputs as anomalies, effectively identifying that individual among others. The methodology addresses a critical gap in wildlife conservation, where manual identification is labor-intensive, time-consuming, and requires expert knowledge.

The dataset used in this study comprises nine individually identified tigers, with 30 original images per tiger collected from real-world camera trap data. To improve generalization and address data scarcity, each original image was augmented 30 times using geometric and photometric transformations, resulting in a total of 930 images per individual and 8,370 images overall. Our main methodological contribution lies in the integration of PCA with OCSVM to enhance feature representation and reduce noise. PCA was observed to yield varying impacts across model variants: for instance, in the Bahubali-Subhashree model, PCA improved accuracy slightly from 85.22% to 85.87%, while in the Spandan-Subhashree model, it decreased performance from 86.73% to 80.40%, indicating sensitivity to feature loss. Further analysis using varying sample sizes (100 to 500) demonstrated that model accuracy improved with more data, converging near 88% accuracy for both models.

Related work

The Smithsonian's National Zoo and Conservation Biology Institute provided an informative piece about tigers, covering various aspects, including their physical description, habitat, behavior, diet, reproduction, and conservation status [13]. It emphasizes their significance as apex predators and their current status as endangered due to habitat loss and poaching. The text also includes news updates and taxonomic information about tigers.

In 2013, INPO, an Independent Non-Profit Organization, was initiated by Russian President Vladimir Putin, which works for research and conservation of wildlife stock [14]. In 2017, Papillon Systems developed a system to automate the census of Amur tigers. With the Amur tiger population dwindling, the system was proposed to aid their conservation efforts. Two methods were proposed for tiger identification: one based on unique skin patterns using Papillon's AFIS algorithms and another based on full-face images using their POLYFACE system. The research showed promising results, suggesting that these algorithms could effectively be used for automated census and identification of tiger populations. This system has the potential to significantly assist scientists and specialists in studying Amur tigers, processing large amounts of data efficiently with minimal human effort. Li et al. [15] introduced the Amur Tiger Re-identification in the Wild (ATRW) dataset in their paper, which aims to address the challenge of monitoring endangered species like the Amur tiger using computer vision methods. The dataset comprises over 8,000 video clips of 92 Amur tigers, annotated with bounding boxes, pose key points, and tiger identities. Unlike previous re-identification datasets, ATRW captures tigers in diverse poses and lighting conditions. The study demonstrates the dataset's difficulty for re-ID tasks. It proposes a novel method that incorporates precise pose parts modeling in deep neural networks to improve performance significantly compared to existing methods.

A novel method for the identification of tigers for census purposes using iris pattern matching and recognition was introduced by Ghoshal et al. [16]. They proposed a study that talks about the utilization of tiger iris patterns as a unique biometric identifier. The method involves segmentation of the tiger iris using circular Hough transform, followed by normalization using Daugman's rubber sheet model. Feature extraction and encoding are performed using Gabor filters, and matching is achieved through the calculation of Hamming Distance. The study demonstrates the potential of iris recognition technology for wildlife conservation and proposes future directions for research in this area. Kathait et al. [17] proposed a methodology for classifying images of tigers into individual classes using deep learning models. While motion-activated cameras have enabled the collection of vast amounts of wildlife images, uniquely identifying individual animals remains challenging. The proposed pipeline employs the YOLOv8 model and EfficientNetB3 model with transfer learning to classify tiger images, focusing on 98 unique tigers out of 192 with a threshold of 15 images per tiger. This approach addresses the gap in existing techniques for individual animal identification within species. Kupyn and Pranchuk [18] present TigerNet, a lightweight and efficient object detection model tailored for real-time Amur tiger detection in the wild. Their approach integrates a Feature Pyramid Network with Depth Wise Separable Convolutions and a lightweight FD-MobileNet backbone, achieving a model with 600k parameters and 0.071 GFLOPs per image, suitable for edge device deployment. They also introduce a two-stage semi-supervised learning method via pseudo-labeling to distill knowledge from larger networks, enhancing performance even with limited annotated data. The model outperforms other methods in the ATRW-ICCV 2019 tiger detection sub-challenge, demonstrating superior efficiency and accuracy. The paper provides insights into the architecture, training pipeline, dataset, and experimental evaluation, showcasing its potential for wildlife conservation applications. In April 2020, Shi et al. [19] cited an approach using deep convolutional neural networks (CNNs) for automatic individual identification of Amur tigers based on their stripe patterns. The research, published in Integrative Zoology in April 2020, demonstrates the effectiveness of the CNN algorithm in accurately identifying individual tigers, with recognition accuracy rates of 90.48% for the left side and 93.5% for the right side of the body. The study emphasizes the importance of such technology for population monitoring and conservation strategies. Shi et al. [19] suggest that their approach can significantly contribute to the field of automatic individual identification technology for Amur tigers.

National Geographic, published on March 11, 2009 [20], discusses the development of photo-recognition software designed to identify tigers based on their unique stripe patterns. The software, created by Lex Hilby from Conservation Research, generates a three-dimensional model of a tiger's skin and compares different images of the same individual taken at various times and angles. This automated process simplifies the laborious task of matching tiger photos manually, offering conservationists a valuable tool for tracking tiger populations and identifying poached skins. The software's success was demonstrated through testing by Ullas Karanth from the Wildlife Conservation Society, who found it highly accurate even when matching images taken years apart or at different angles. Hilby's software has potential applications beyond tigers, possibly tracking other patterned species, such as leopards, zebras, and salamanders. In 2010, it was published in Genetics by Eizirik et al. [21] to investigate the genetic basis of mammalian coat patterns, specifically in domestic cats. Using the domestic cat as a model organism, the researchers establish three pedigrees to map the four recognized coat pattern variants. They demonstrate that at least three different loci control coat markings in domestic cats, including the Abyssinian form, spotted, mackerel, and blotched patterns. Through genetic mapping, they identify specific genomic regions associated with each pattern variant [22-24]. The study sheds light on the genetic and developmental mechanisms underlying mammalian coat patterns, providing valuable insights into the evolution of these phenotypes.

Recently, Jones and Farrow [25] proposed an OCSVM model to detect genetic drift from breast cancer images. They observed a strong correlation between increasing noise and detected inliers. One of the major issues nowadays is very unpredictable forest fires. So, Xu et al. [26] proposed a deep OCSVM model for detecting forest fires. Their proposed model is compared with state-of-the-art models with a higher fire detection rate (95.29%, 8.53% greater than state-of-the-art) and a lower error warning rate. Liu et al. [27] presented a one-class classification technique for detecting unnatural skin tissues. A CNN model was employed to drag the key features from the tissues, and the OCSVM demonstrated high accuracy on the dataset. CNN-based tiger classification was proposed by Shah and Sinha in 2025 [28]. The author applied Synthetic Minority Oversampling Technique to balance the dataset, improve the model's accuracy, and deal with the overfitting issue.

Materials And Methods

The declining population of tigers in India, exacerbated by habitat loss, poaching, and human-wildlife conflict, presents a critical conservation challenge. Despite concerted efforts to monitor and protect these iconic species, existing methods for tracking individual behavior and population trends are labor-intensive, time-consuming, and prone to error. The reliance on traditional techniques, such as manual grouping of camera-captured images, hampers the efficiency and precision of tiger population monitoring efforts. This necessitates the development of innovative and technologically driven solutions to streamline data collection, enhance accuracy, and improve the effectiveness of tiger conservation initiatives. The urgency of the situation underscores the need for a comprehensive approach that leverages advanced technologies and interdisciplinary collaboration to address the multifaceted challenges facing tiger conservation in India.

To address the challenges of tiger conservation in India, leveraging machine learning (ML) presents a promising avenue for innovation and improvement in monitoring and protection efforts [29]. ML, a subset of artificial intelligence, involves the development of algorithms that enable computers to learn from data and make predictions or decisions without being explicitly programmed. ML algorithms can analyze large volumes of data, including camera trap images, satellite imagery, and environmental variables, to identify patterns and trends that may not be apparent to human observers. This allows for more comprehensive and nuanced insights into tiger behavior, habitat preferences, and population dynamics. ML techniques significantly reduce the time and effort required for manual categorization and labeling by accurately identifying and classifying tigers and other wildlife species in images. By analyzing historical data and environmental variables, these models can inform decision-making and prioritize conservation efforts in areas where they are likely to have the greatest impact. Figure *4* illustrates the various ML techniques.

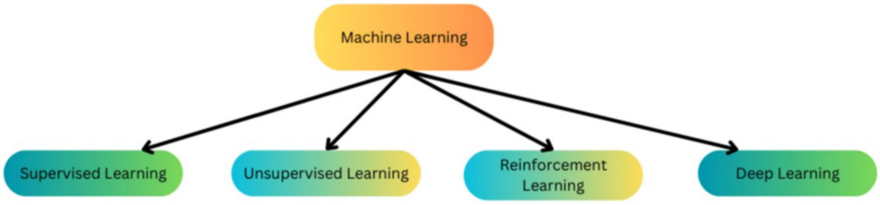

FIGURE 4: Types of machine learning approaches.

Types of machine learning

Supervised Learning

Definition: Supervised learning involves training a model on a labeled dataset, where each input is paired with the corresponding correct output [30].

Examples:

Email Spam Classification: Given a dataset of emails labeled as spam or not spam, a supervised learning algorithm learns to classify new emails based on their features.

Handwritten Digit Recognition: The model is trained on images of handwritten digits with labels (0-9), learning to recognize and classify digits accurately.

Unsupervised Learning

Definition: Unsupervised learning involves training a model on an unlabeled dataset, where the model identifies patterns and structures in the data without explicit guidance [31].

Examples:

Clustering for Customer Segmentation: An unsupervised learning algorithm segments customers based on purchasing behavior, identifying groups with similar patterns and helping businesses target specific groups with tailored strategies.

Anomaly Detection: The model learns to identify unusual data points within a dataset, useful for detecting fraud in financial transactions or spotting anomalies in network traffic.

Reinforcement Learning

Definition: Reinforcement learning trains an agent to interact with an environment to maximize cumulative rewards through trial and error, guided by feedback in the form of rewards or penalties [32].

Examples:

Autonomous Driving: Reinforcement learning can train a vehicle to navigate traffic, make decisions on accelerating, braking, or changing lanes based on sensor feedback.

Game Playing: These algorithms master complex games, such as Go or chess, by making optimal moves through self-play or against opponents, receiving rewards for successful strategies and penalties for unsuccessful ones.

In this article, we use an OCSVM model (see in Figure *5*). OCSVM is like a virtual fence, distinguishing between normal and abnormal data points within a dataset. Unlike traditional classification algorithms that require labeled data from multiple classes, OCSVM constructs a boundary or "hyperplane" around the normal data points, aiming to encapsulate them while maximizing the margin between the boundary and the data points. This hyperplane effectively defines the region of normalcy within the dataset. What sets OCSVM apart is its ability to detect anomalies without prior knowledge of what constitutes normal behavior. This makes it particularly useful in scenarios where anomalies are rare or difficult to define explicitly, making OCSVM a robust and versatile tool for anomaly detection.

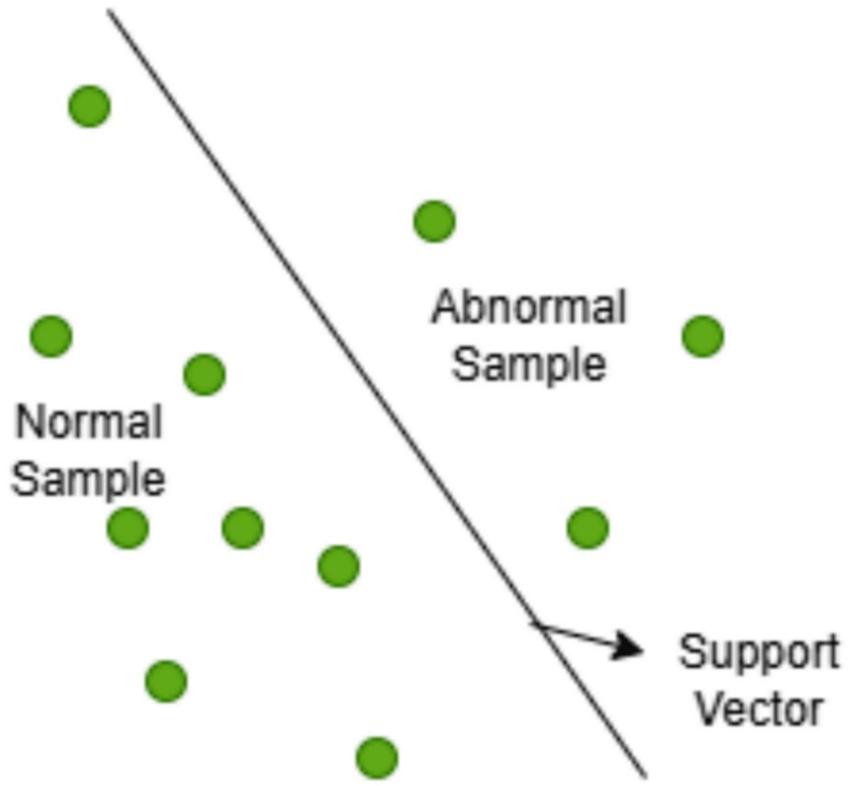

FIGURE 5: Architecture of OCSVM model.

Source: [33]

OCSVM, One-Class Support Vector Machine

PCA is a statistical method used for dimensionality reduction, which simplifies large datasets by transforming them into a set of uncorrelated variables called principal components. The primary goal of PCA is to reduce the number of variables in a dataset while retaining as much variance as possible, which often helps in removing noise and redundancy from data, improving the performance of machine learning models (see in Figure *6*).

OCSVM was chosen over deep learning models like CNNs or Transformers primarily due to the nature of the dataset and the specific goals of the study. The task involved identifying individual tigers using a relatively small and imbalanced dataset, where each tiger had a limited number of labeled images. Unlike CNNs or Transformers that require large datasets and extensive training, OCSVM excels in scenarios with limited data, especially when the focus is on learning the characteristics of a single class and detecting deviations as outliers. OCSVM is computationally efficient and integrates well with PCA for dimensionality reduction and feature extraction. This makes it suitable for identifying subtle and high-dimensional patterns, such as tiger stripes, without overfitting. Additionally, it avoids the need for deep architectures and high-end hardware, making it ideal for quick prototyping and conservation-focused applications where data is scarce but precision is critical.

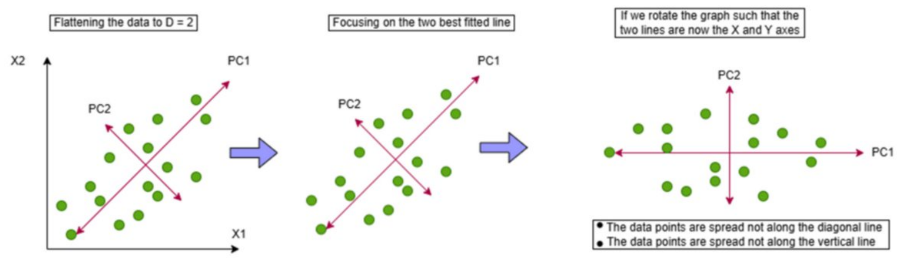

FIGURE 6: Architecture of principal component analysis.

Source: [34]

PC, Principal Component

In the tiger's social ecosystem, a dynamic model unfolds, embodying unique behavioral characteristics. Figure 7 illustrates that camera-captured images or user-provided images are fed into the model, which evaluates, preprocesses, and clusters them based on characteristics and features learned during training. The model then labels each cluster with an identifier corresponding to the tiger it represents, providing a powerful tool for wildlife researchers and conservationists. This tool enables researchers to better understand tiger behavior, population dynamics, and habitat usage, thereby contributing valuable insights to conservation efforts.

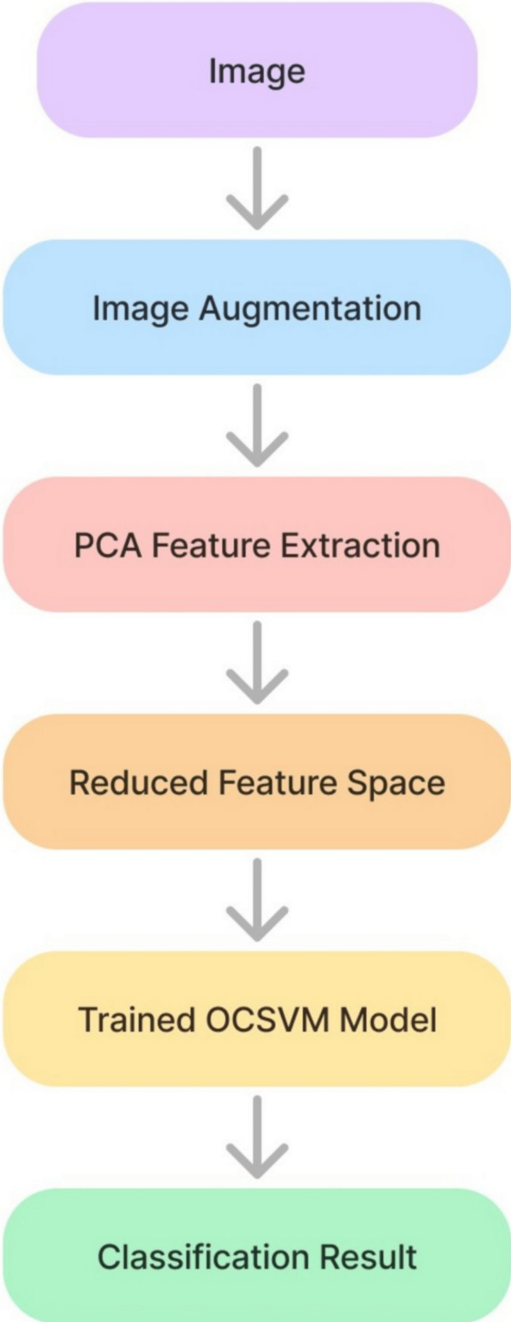

FIGURE 7: Flow chart of our proposed model.

OCSVM, One-Class Support Vector Machine; PCA, Principal Component Analysis

Dataset details

To conduct evaluation and further research in this area, a dataset of Bengal tiger images was essential. However, no open-source dataset was available. Therefore, images were manually captured on camera at Nandankanan Zoological Park. For full access to the dataset used in this study, see the supplementary materials (Dataset: *https://drive.google.com/drive/folders/1kY_H7bFVUiCidSGnS0LO9PwPcolUnyVX*, and Source Code: *https://github.com/REETICHI/StripeCode-*). This allows for the reproducibility of the experiments outlined in this paper. Images of nine tigers were collected, with 20-30 images of each tiger taken from various angles and perspectives. These included detailed views of different body parts: front face, front body pattern, dorsal body pattern, left and right side body patterns, front and back views of the ears, tail pattern, leg patterns, and neck pattern, as shown in Figure 8.

FIGURE 8: Different body parts (Stripecode) of tigers.

The nine tigers, named Spandan, Subhashree, Bahubali, Megha, Sahil, Basu, Rakesh, Abhay, and Samrat, each exhibit unique characteristics. Due to noise and background disturbances, each image was masked, and to build a more robust model, the images were further augmented, as can be seen in Figure 9.

FIGURE 9: The first two columns represent the original image and masked image, respectively, while the third and fourth columns showcase augmented images.

Results

In the ongoing evaluation of the OCSVM model for the desired task, it is essential to assess its strengths and suitability for the given dataset. OCSVM is particularly effective for anomaly detection tasks, making it suitable for identifying outliers or abnormalities in the data. Selecting the most appropriate model and configuration requires careful consideration of the specific requirements and characteristics of the task at hand. In our proposed tiger identification model, PCA plays a key component used both for dimensionality reduction and feature extraction. Tiger stripe images are high-dimensional and often include redundant or noisy features, which can degrade model performance. PCA addresses this by transforming the original features into a set of orthogonal components that retain the most significant variance. This transformation not only reduces computational complexity but also enhances feature quality by isolating the most informative patterns in the data, such as stripe geometry and texture. By emphasizing these core visual cues, PCA improves OCSVM's ability to detect outliers and differentiate between individual tigers. In our experiments, PCA increased the accuracy of the Bahubali-Subhashree model from 85.22% to 85.87%, suggesting that it helped reduce noise and improved class separability. However, in the Spandan-Subhashree model, accuracy dropped from 86.73% to 80.40%, indicating that excessive dimensionality reduction can sometimes lead to information loss. Despite this, PCA greatly contributed to more robust and efficient feature representation, resulting in improved model performance where applicable. Overall, PCA significantly improved the model's generalization, accuracy, and scalability, making it a critical component of our workflow for real-world animal conservation applications.

In this proposed work, data augmentation was used to overcome the limitation of a small dataset. Each of the nine tigers represented by 30 manually captured images under varied conditions. To ensure precision, background noise was manually masked for each image. Subsequently, extensive data augmentation was performed using geometric and photometric transformations such as flipping, blurring, rotation, scaling, cropping, brightness/contrast adjustments, translation, and Gaussian noise. Each original image was augmented 30 times, resulting in 930 images per tiger and a total of 8,370 images. This approach significantly increased dataset diversity, helping the model generalize better and improving its ability to distinguish individual tigers under varying environmental and visual conditions.

Results for OCSVM without PCA

Result for OCSVM - Bahubali as Outliers (0) and Subhashree as Inliers (1).

Number of total samples is 460.

Final Model Accuracy: 85.22%.

Classification Report

Precision, recall, and F1-scores for class 0 (outliers) and class 1 (inliers), along with weighted and macro averages, indicate strong accuracy for class 0 but low recall for class 1 (see in Figure *10*).

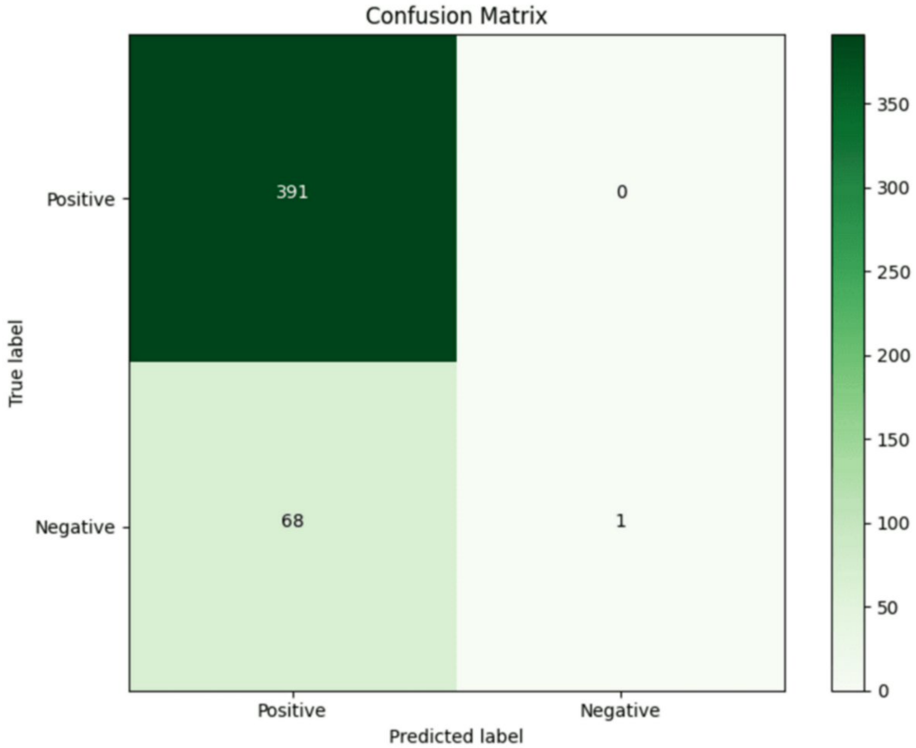

FIGURE 10: Confusion matrix for Bahubali.

Result for OCSVM - Spandan as Outliers (0) and Subhashree as Inliers (1).

Number of total samples is 505.

Final Model Accuracy: 86.73%.

Classification Report

Precision, recall, and F1-scores show strong performance for outlier detection (class 0) but relatively low recall for inliers (class 1) (see in Figure *11*).

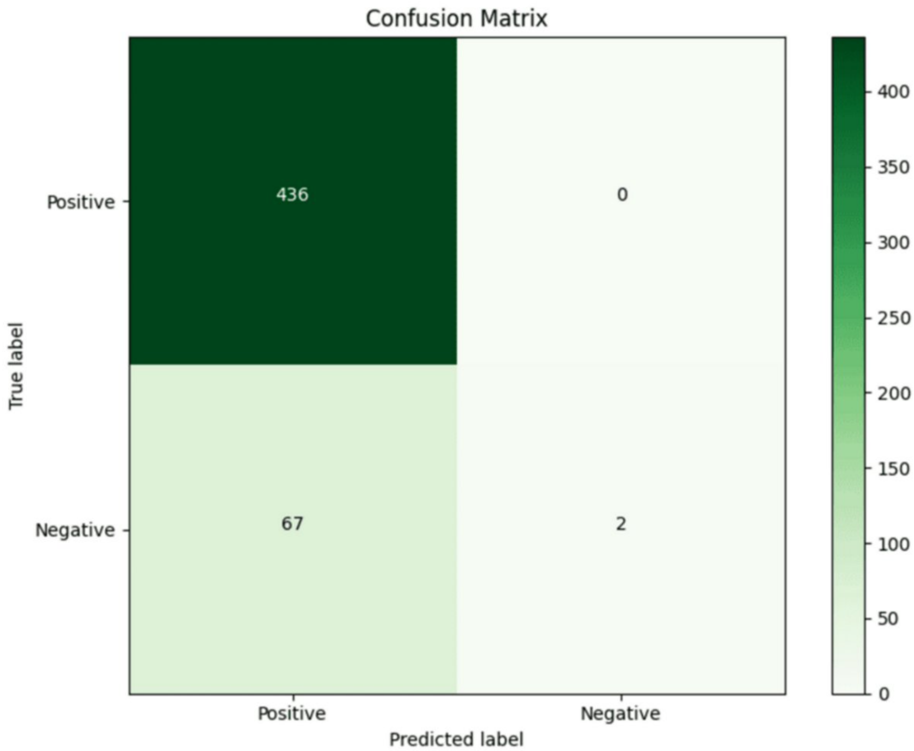

FIGURE 11: Confusion matrix for Spandan.

Results for OCSVM with PCA

Result for OCSVM - Bahubali as Outliers (0) and Subhashree as Inliers (1).

Number of total samples is 460.

Final Model Accuracy: 85.87%.

Classification Report

Slight improvement in class 1 recall, with minor changes in weighted and macro averages, suggesting a marginal benefit from PCA in this configuration (see in Figure 12).

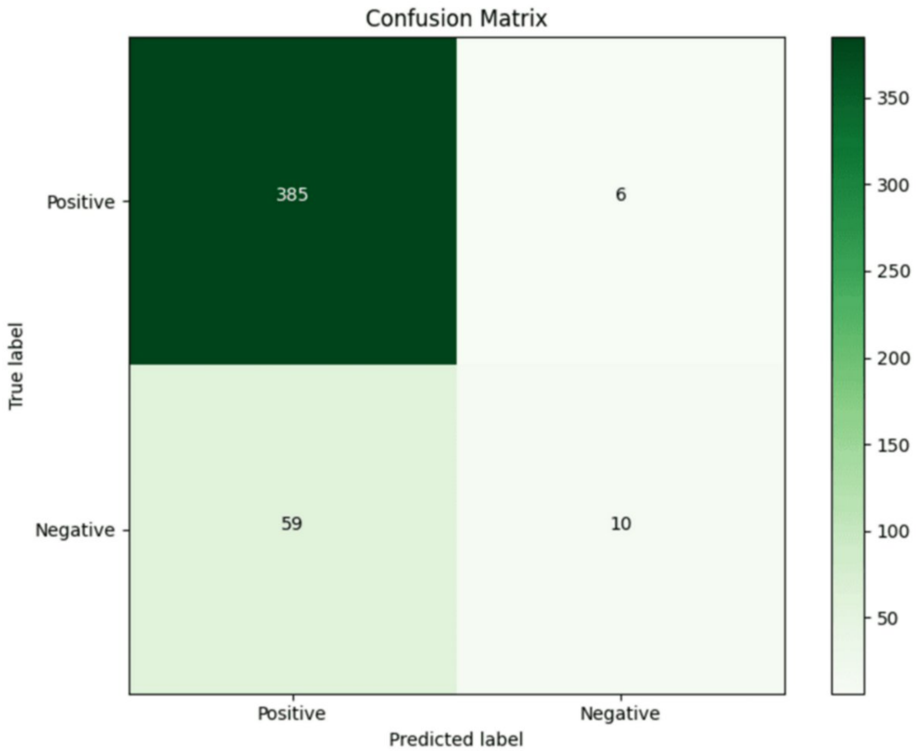

FIGURE 12: Confusion matrix for Bhahubali.

Result for OCSVM - Spandan as Outliers (0) and Subhashree as Inliers (1).

Number of total samples is 505.

Final Model Accuracy: 80.40%.

Classification Report

Performance decreases slightly after PCA application, showing that dimensionality reduction may not always benefit the model and is highly dependent on the dataset characteristics and configuration (see in Figure *13*).

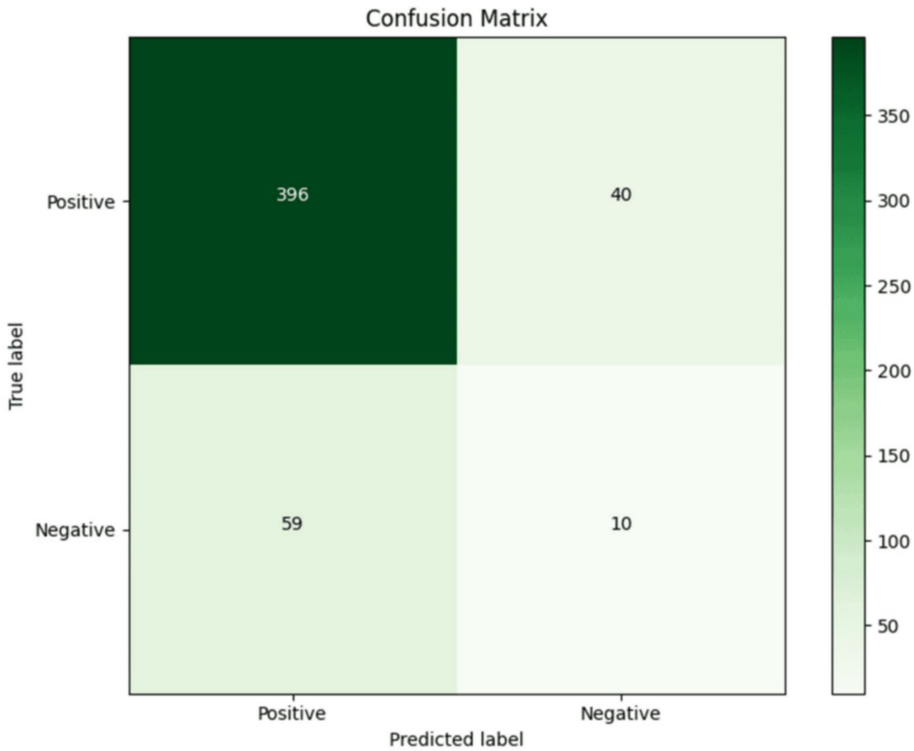

FIGURE 13: Confusion matrix for Spandan.

Result for OCSVM - PCA components

In this evaluation, three tigers (Bahubali, Subhashree, and Spandan) were considered to examine the performance of the OCSVM model.

In the first model (see in Figure *14*), Bahubali was labeled as outliers (0) and Subhashree as inliers (1). The model achieved an accuracy of 85.21% without dimensionality reduction (PCA). Upon applying PCA, the accuracy slightly improved to 85.86%, suggesting a marginal benefit from dimensionality reduction in this scenario. In the second model (see in Figure *15*), Spandan was labeled as outliers (0) and Subhashree as inliers (1). Without PCA, the model achieved an accuracy of 86.73%. However, after applying PCA, accuracy dropped to 80.39%, indicating a decrease in performance.

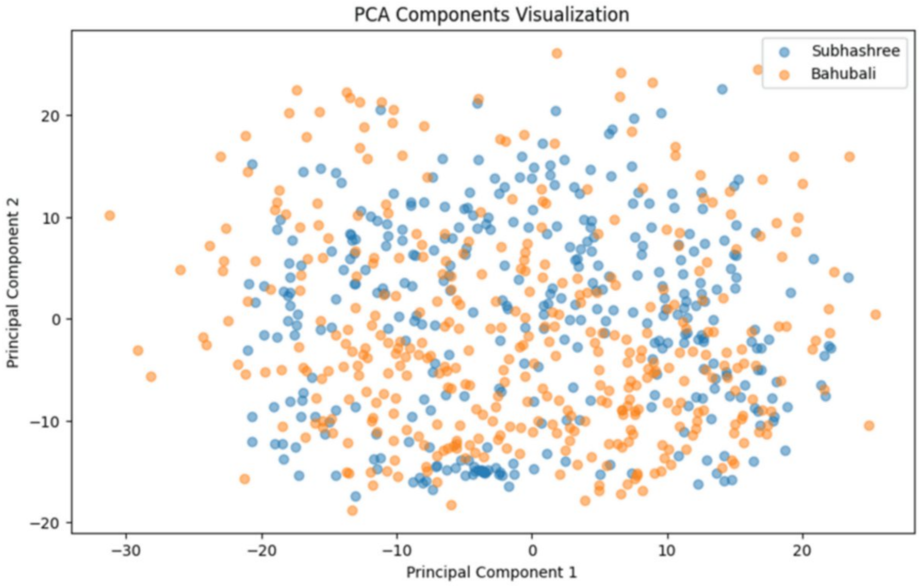

FIGURE 14: Result for OCSVM - Bahubali as outliers (0) and Subhashree as inliers (1).

OCSVM, One-Class Support Vector Machine; PCA, Principal Component Analysis

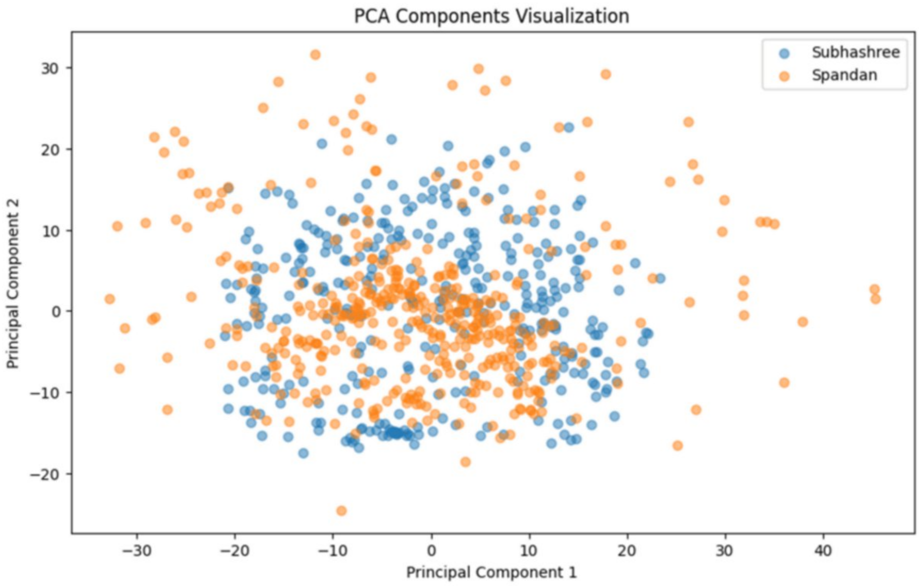

FIGURE 15: Result for OCSVM - Spandan as outliers (0) and Subhashree as inliers (1).

OCSVM, One-Class Support Vector Machine; PCA, Principal Component Analysis

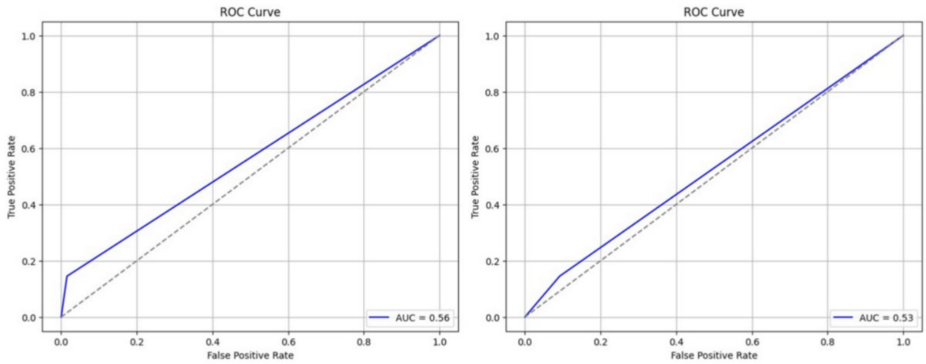

FIGURE 16: ROC curve for different tiger classes (left: Bahubali-Subhashree class, and right: Subhashree-Spandan class).

AUC, Area Under the Curve; ROC, Receiver Operating Characteristic

The analysis also includes Receiver Operating Characteristic (ROC) curves, which are used to evaluate the model's ability to distinguish between the target class (in this case, the "tiger" class) and outliers. The corresponding Area Under the Curve (AUC) values provide a quantitative measure of the model's performance. In Figure *16*, we observed that the AUC curve follows a consistent upward trend, indicating that as the model becomes better at identifying the tiger class and distinguishing it from outlier samples, its overall ability to generalize to unseen data also improves. Notably, the AUC curve appears to stabilize around 500 samples, suggesting that the model's performance reaches its optimal point after a sufficient amount of training data.

Discussion

The evaluation of the OCSVM model in this study provides several important insights into its performance for tiger classification tasks based on their stripecode. We also examined the role of PCA in dimensionality reduction. The key findings highlight both the strengths of the OCSVM model and the impact of PCA on model performance.

Cross-validation technique

In this study, a one-vs-all cross-validation technique was employed to evaluate the OCSVM model for individual tiger identification. For each of the nine tigers, 80% of their augmented images were used for training, while the remaining 20% were reserved for testing. To assess the model’s ability to distinguish between inliers and outliers, the test set was combined with images from another tiger, treated as outliers. This approach was iteratively applied to all tiger pairs, ensuring each individual served as the target class once. This technique effectively simulates where the model must recognize a known individual from unknowns, validating the OCSVM’s performance in class separation using metrics such as accuracy, confusion matrix, and classification report.

Performance of OCSVM without PCA

When evaluating OCSVM without PCA, the model demonstrated solid accuracy in classifying inliers and outliers. Specifically, for the first model (Bahubali as outliers and Subhashree as inliers), the model achieved an accuracy of 85.22%. For the second model (Spandan as outliers and Subhashree as inliers), the model performed slightly better with an accuracy of 86.73%. These results highlight OCSVM’s strength in identifying anomalies, as evidenced by strong classification of outliers (class 0), with precision and recall scores indicating good performance in outlier detection. However, recall for class 1 (inliers) was relatively low in both models, indicating that while the model was effective at identifying outliers, it struggled to consistently identify inliers, a common challenge in one-class classification tasks.

Impact of PCA on OCSVM performance

When PCA was applied to the models, there was a noticeable impact on performance, although the effects varied between the two datasets. For the Bahubali and Subhashree classification (first model), the application of PCA resulted in a slight improvement in accuracy, increasing from 85.22% to 85.87%. This modest improvement suggests that dimensionality reduction helped the model by reducing noise and possibly removing irrelevant features, thereby improving the separation between the classes. Additionally, the recall for class 1 (inliers) showed a slight increase, which indicates that PCA may have helped improve the model's sensitivity to inliers in this particular case.

In contrast, for the Spandan and Subhashree classification (second model), the use of PCA led to a noticeable decrease in performance, with accuracy dropping from 86.73% to 80.40%. This significant decline in accuracy suggests that in this case, the dimensionality reduction introduced by PCA may have led to a loss of important features or patterns in the data that were critical for distinguishing between the inliers and outliers.

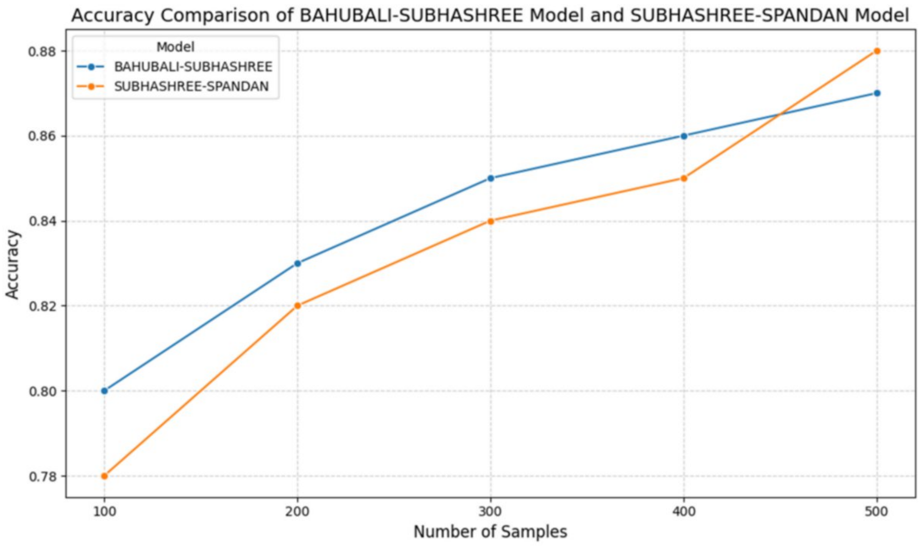

FIGURE 17: Accuracy performance based on sample size.

The observed decrease in accuracy can be attributed to the limited size of the dataset used in this study. To further assess the performance of the proposed model, we conducted an analysis of accuracy with respect to changing sample sizes. Initially, we considered a dataset consisting of 100 randomly selected samples from both categories and evaluated the model's accuracy. In subsequent iterations, we incrementally added 100 additional random samples to the dataset and observed the corresponding changes in accuracy at each step. This process allowed us to examine the impact of sample size on the model's performance. Figure *17* compares the accuracy of two models, Bahubali-Subhashree and Subhashree-Spandan, as a function of sample size, ranging from 100 to 500. The Bahubali-Subhashree model consistently outperforms the Subhashree-Spandan model, with its accuracy increasing as the sample size grows. Both models improve in accuracy with larger datasets. They converge near 500 samples, where they both reach an accuracy of about 0.88. This suggests that larger sample sizes significantly enhance the performance of both models.

The combination of OCSVM and PCA results in a highly efficient workflow. The model demonstrated strong accuracy and adaptability despite data scarcity and complex image features. Our analysis highlights the potential of OCSVM, enhanced by PCA, to significantly support wildlife monitoring and conservation efforts. The model automates the identification and tracking of tigers, offering a practical solution for protecting endangered species in their natural habitats.

The observed inconsistency in PCA impacts on model accuracy improving it for one tiger pair but reducing it for another which highlights the complexity of dimensionality reduction in this context. PCA aims to reduce noise and redundant features by transforming data into principal components, which can enhance class separability. However, if important discriminative features are lost during this reduction, model performance may decline. In the Bahubali-Subhashree model, PCA helped by emphasizing relevant patterns, slightly increasing accuracy. Conversely, in the Spandan-Subhashree model, excessive dimensionality reduction likely discarded critical information, causing a drop in accuracy. This inconsistency suggests that PCA's effectiveness depends heavily on the dataset characteristics and indicates a need for careful tuning to ensure consistent improvements across diverse datasets.

The proposed work has several limitations. First, the dataset size is relatively small, which can limit the model's ability to generalize well to new, unseen tiger images, especially given the natural variability in tiger appearances and environmental conditions. The use of PCA for dimensionality reduction showed mixed results, improving accuracy in some cases but reducing it in others, indicating that the method may not consistently preserve all critical features. In future work, we plan to expand the dataset significantly to include a larger and more diverse set of samples, which is expected to improve the performance and robustness of the model. By applying the model to a more extensive dataset, we aim to better capture the underlying patterns and enhance its generalization capability.

Conclusions

The article presents a robust and progressive approach that harnesses advanced machine learning techniques for precise and reliable tiger identification. By addressing core challenges in wildlife monitoring, anti-poaching, and biodiversity conservation, this work offers a scalable, automated system capable of identifying individual tigers using their unique stripe patterns. Through a combination of meticulous data collection, innovative image preprocessing, and rigorous model validation, the Stripecode Guardians initiative delivers a practical solution for real-world conservation needs. Its success lies not only in improving accuracy and detection but also in reducing the reliance on manual expertise. While this study primarily focuses on tigers, the underlying methodology is designed for adaptability, paving the way for broader applications across species. In this study, we explored the effectiveness of the OCSVM model for tiger classification based on their unique stripecodes. The proposed approach integrates OCSVM with PCA, which plays a crucial role in both dimensionality reduction and feature extraction. Tiger stripe images are inherently high-dimensional and often contain redundant or noisy features that can hinder model performance. PCA addresses this challenge by transforming the original features into a set of orthogonal components that capture the most significant variance in the data. This not only reduces computational complexity but also enhances the quality of input features by isolating the most informative aspects of tiger stripe patterns, such as stripe geometry, direction, and texture. By highlighting these essential visual characteristics, PCA boosts the OCSVM's ability to construct accurate decision boundaries, enabling it to better differentiate between individual tigers and detect outliers. The synergy between PCA and OCSVM results in a more robust classification pipeline, particularly suited for wildlife datasets where inter-class variation is subtle, and data availability is limited. Additionally, the performance of the OCSVM model was rigorously evaluated in conjunction with PCA to assess its effectiveness in outlier detection and individual tiger differentiation. The OCSVM model, combined with PCA, was evaluated to determine its effectiveness in tiger identification. PCA played a crucial role in enhancing the model by reducing dimensionality and extracting key features, which helped suppress noise and improve class separability. In the Bahubali-Subhashree model, PCA improved the accuracy from 85.22% to 85.87%, demonstrating a modest gain through more refined feature representation. However, in the Spandan-Subhashree model, accuracy dropped from 86.73% to 80.40%, suggesting that overly aggressive dimensionality reduction may lead to the loss of important distinguishing features. Despite this, both models showed improved performance with larger datasets, with accuracy peaking at 88% and AUC scores rising accordingly. These findings highlight the importance of dataset size and composition, and confirm the robustness of the OCSVM-PCA pipeline when properly tuned, making it an effective solution for high-dimensional wildlife image classification tasks.

The proposed approach has certain limitations. One major constraint is the limited size of the dataset, which may affect the model's generalization capability when applied to unseen tiger images. This is particularly challenging due to the inherent diversity in tiger appearances and varying environmental backgrounds. Additionally, while PCA was employed for dimensionality reduction, its impact on performance was inconsistent - enhancing accuracy in some scenarios but leading to a decline in others. This suggests that PCA may not reliably retain all essential features required for accurate identification. In summary, the "Stripecode Guardians" initiative represents a significant advancement in wildlife conservation, heralding a new era of collaboration and innovation that offers promising avenues for leveraging advanced technologies to protect and preserve our planet's diverse ecosystems.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sankar Kumar Mridha, Reetichi Pattanaik

Acquisition, analysis, or interpretation of data: Sankar Kumar Mridha, Reetichi Pattanaik

Drafting of the manuscript: Sankar Kumar Mridha, Reetichi Pattanaik

Critical review of the manuscript for important intellectual content: Sankar Kumar Mridha, Reetichi Pattanaik

Supervision: Sankar Kumar Mridha

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.
Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.
Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from

any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Wright RT, Boorse DF: Environmental Science: Toward a Sustainable Future. Pearson, Boston; 2017.
2. Choudhary S: The Wildlife (Protection) Act, 1972: A comprehensive study on wildlife conservation in India. 2024, 10.2139/ssrn.4848859
3. Uddin N, Enoch S, Harihar A, Pickles RS, Hughes AC: Tigers at a crossroads: Shedding light on the role of Bangladesh in the illegal trade of this iconic big cat. Conservation Science and Practice. 2023, 5:e12952. 10.1111/csp2.12952
4. Baillie JEM, Hilton-Taylor C, Stuart SN: 2004 IUCN Red List of Threatened Species. A Global Species Assessment. IUCN, Gland, Switzerland and Cambridge, UK.; 2004.
5. Kumar S, Singh SK, Singh R, Singh AK: Animal Biometrics: Techniques and Applications. Springer, Singapore; 2017. 10.1007/978-981-10-7956-6
6. Petso T, Rodrigo SJ Jr, Dimane M: Review on methods used for wildlife species and individual identification. European Journal of Wildlife Research. 2022, 68:3. 10.1007/s10344-021-01549-4
7. Awad AI: From classical methods to animal biometrics: A review on cattle identification and tracking. Computers and Electronics in Agriculture. 2016, 123:423-435. 10.1016/j.compag.2016.03.014
8. Pinterest. (2024). Accessed: November 1, 20224: https://in.pinterest.com/pin/477663104224892870/.
9. Ma J, Yuan Y: Dimension reduction of image deep feature using PCA. Journal of Visual Communication and Image Representation. 2019, 63:102578. 10.1016/j.jvcir.2019.102578
10. Jolliffe IT: Principal component analysis for special types of data. Principal Component Analysis. Springer, New York; 2002. 338-372. 10.1007/0-387-22440-8_13
11. Chen Y, Zhou XS, Huang TS: One-class SVM for learning in image retrieval. Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205), Thessaloniki, Greece. 2001, 1:34-37. 10.1109/ICIP.2001.958946
12. Alam S, Sonbhadra SK, Agarwal S, Nagabhushan P: One-class support vector classifiers: A survey. Knowledge-Based Systems. 2020, 196:105754. 10.1016/j.knosys.2020.105754
13. Walston J, Robinson JG, Bennett EL, et al.: Bringing the tiger back from the brink—the six percent solution. PLoS Biology. 2010, 8:e1000485. 10.1371/journal.pbio.1000485
14. Flanigan ST: Nonprofit service provision by insurgent organizations: the cases of Hizballah and the Tamil Tigers. Studies in Conflict & Terrorism. 2008, 31:499-519. 10.1080/10576100802065103
15. Li S, Li J, Tang H, Qian R, Lin W: ATRW: a benchmark for Amur tiger re-identification in the wild. 10.48550/arXiv.1906.05586
16. Ghoshal D, De P, Saha B: Identification of tigers for census by the method of tiger iris pattern matching and recognition. International Journal of Computer Applications. 2012, 49(2):19-23. 10.5120/7600-0316
17. Kathait SS, Singh V, Kumar A: Individual tiger identification using transfer learning. International Journal of Computer Applications. 2024, 186:46-50. 10.5120/ijca2024923469
18. Kupyn O, Pranchuk D: Fast and efficient model for real-time tiger detection in the wild. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South). 2019, 310-314. 10.1109/ICCVW.2019.00041
19. Shi C, Liu D, Cui Y, Xie J, Roberts NJ, Jiang G: Amur tiger stripes: Individual identification based on deep convolutional neural network. Integrative zoology. 2020, 15:461-470. 10.1111/1749-4877.12453
20. Hiby L, Lovell P, Patil N, Kumar NS, Gopalaswamy AM, Karanth KU: A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins. Biology Letters. 2009, 5:383-386. 10.1098/rsbl.2009.0028
21. Eizirik E, David VA, Buckley-Beason V, et al.: Defining and mapping mammalian coat pattern genes: multiple genomic regions implicated in domestic cat stripes and spots. Genetics. 2010, 184:267-275. 10.1534/genetics.109.109629
22. Das A, Sowmya S, Sinha S, Chandru S: Identification of a zebra based on its stripes through pattern recognition. 2021 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India. 2021, 120-122. 10.1109/ICDI3C53598.2021.00033
23. Mohapatra RK, Panda S, Acharya UR: Study on activity pattern and incidence of stereotypic behavior in captive tigers. Journal of Veterinary Behavior. 2014, 9:172-176. 10.1016/j.jveb.2014.04.003
24. Reddy KPK, Aravind R: Measurement of asymmetry of stripe patterns in animals. 2012 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India. 2012, 1-5. 10.1109/SPCOM.2012.6290240
25. Jones WS, Farrow DJ: One-class support vector machines for detecting population drift in deployed machine learning medical diagnostics. Scientific Reports. 2025, 15:12157. 10.1038/s41598-025-94427-x
26. Xu Y, Ma Y, Ye Q, Fu L, Yang X: Interpretable deep one-class model for forest fire detection. Expert Systems with Applications. 2025, 281:127657. 10.1016/j.eswa.2025.127657
27. Liu X, Ouellette S, Jamgochian M, Liu Y, Rao B: One-class machine learning classification of skin tissue based on manually scanned optical coherence tomography imaging. Scientific Reports. 2023, 13:867. 10.1038/s41598-023-28155-5
28. Shah M, Sinha H: Automated image identification and classification of tiger beetles based on deep learning model. IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC). 2025, 01125-01132. 10.1109/CCWC62904.2025.10903740
29. Kotsiantis SB, Zaharakis ID, Pintelas PE: Machine learning: a review of classification and combining techniques. Artificial Intelligence Review. 2006, 26:159-190. 10.1007/s10462-007-9052-3
30. Osisanwo FY, Akinsola JET, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J: Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT). 2017, 48:128-138. 10.14445/22312803/IJCTT-V48P126

31. Olaode A, Naghdy G, Todd C: Unsupervised classification of images: a review. International Journal of Image Processing. 2014, 8:325-342.
32. Mou L, Saha S, Hua Y, Bovolo F, Bruzzone L, Zhu XX: Deep reinforcement learning for band selection in hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing. 2022, 60:1-14. 10.1109/TGRS.2021.3067096
33. Researchgate. (2024). Accessed: November 1, 20224: https://www.researchgate.net/figure/A-conceptual-sketch-of-the-OCSVM-classifier-1-target-class-1-outlier-class_fig3_3....
34. Linkedin. (2024). Accessed: November 1, 20224: https://www.linkedin.com/pulse/gentle-introduction-principal-components-analysis-michael-wynn/.