VNTurtle: Harnessing Image Recognition for Conservation of Endangered Turtles in Vietnam

Phuc H. Le[1][0000−0002−6297−8640], Thuy T. Nguyen[2][0009−0007−7451−3430], Timothy E.M. McCormack[2][0009−0003−7565−2345], and Thanh-Trung Phan[1][0000−0001−7299−0785]

[1] Fulbright University Vietnam, Vietnam
{lehoangphuc1820, phanthanhtrung86}@gmail.com
[2] Asian Turtle Program of Indo-Myanmar Conservation, Vietnam
{ntthuy, tmccormack}@asianturtleprogram.org

Abstract. Vietnam hosts 26 tortoise and freshwater turtle species, including two endemics, many endangered due to illegal trade. This study explores the creation of datasets for deep-learning models in turtle conservation. We compiled data on 27 species from conservation organizations, crowdsourcing, and Cuc Phuong National Park. Using this dataset, deep-learning models achieved 70% top-1 and 96% top-5 accuracy on a curated test set, confirming its production readiness. We integrated the vision model into a web app that identifies turtles and provides ecological, conservation, and legal insights. These findings support real-world conservation efforts and highlight technology's role in biodiversity protection.

Keywords: Deep Learning · Turtle Conservation · AI for Good.

1 Introduction

Vietnam is a biodiversity hotspot for tortoise and freshwater turtle (TFT) species, playing a critical role in global turtle conservation [1]. Four of the world's 25 most endangered TFT species occur in Vietnam. However, turtles are among the most heavily trafficked animals, exploited for food, exotic pets, and religious practices [49]. This pressure has severely reduced populations: 25 of Vietnam's 26 turtle species are now listed under CITES [49], and the recently described Spotted softshell turtle (*Pelodiscus variegatus*) is proposed as Critically Endangered [2]. Despite these protections, demand for turtles as exotic pets is increasing among Vietnamese youth [50].

Despite strong legal protection in Vietnam [3], tortoise and freshwater turtle (TFT) species remain severely threatened by illegal trade, particularly via online markets [4,5,51]. Although conservation efforts and strict legal enforcement exist [48], effective prosecution depends on accurate species identification. This is challenging for non-experts due to morphological similarity among species and the influx of non-native turtles through the pet trade, which are regulated differently under Vietnamese law. Therefore, a rapid species identification tool could substantially enhance enforcement and conservation outcomes.

Turtle identification requires specialized expertise, as no single visual trait reliably distinguishes species [3]. While some species have distinctive features—such as the serrated jaws of the Yellow-headed temple turtle (*Heosemys annandalii*) or the eye-like head spots of the Four-eyed turtle (*Sacalia quadriocellata*)—species within the same genus are often difficult to differentiate. For instance, three *Cyclemys* species are nearly identical and differ only in subtle plastron and head-stripe patterns. Traditional identification relies on illustrated field guides [6,7,8,3], which are time-consuming and error-prone for non-experts. As a result, law enforcement frequently consults external experts, a process that can take hours to days and often fails due to poor image quality or obscured features. An automatic, fast, and accurate turtle-identification system would therefore substantially improve enforcement efficiency.

Difficulties in turtle identification hinder public reporting and law enforcement, creating a major bottleneck in conservation efforts. A portable mobile application with automatic turtle identification could empower both authorities and the public to respond more effectively to violations. Recent advances in computer vision, particularly CNN-based image classification [9], have enabled high accuracy in animal recognition, including aquatic and reptile species [10,11], with up to 96% accuracy reported for limited TFT datasets [12]. However, classification remains challenging for morphologically similar TFT species, and no prior work addresses a broad range of Vietnam's native TFTs. Existing identification apps perform poorly or even promote turtle trade, highlighting the absence of a reliable tool for Vietnam's turtle conservation needs.

In light of these pressing issues and findings, this paper seeks to explore the applicability of computer vision in building a production-ready automatic turtle classification system. With the above motivation, our paper aims to address the following research questions (RQ):

- **RQ1.** Considering the importance of turtles in Vietnam, what datasets could be used to develop a recognition model?
- **RQ2.** What combination of datasets and neural network architectures optimizes the accuracy of an automated classification system for identifying diverse TFT species?
- **RQ3.** What considerations are necessary for integrating this system into an application that raises public awareness and encourages proactive species protection among users?

The contributions of this research are the following:

1. We curated a diverse TFT image dataset from three sources: annotated field data from Indo-Myanmar Conservation's Asian Turtle Program (via an MOU with Indo-Myanmar Conservation), crowdsourced observations from iNaturalist, and field images collected at the Turtle Conservation Center in Cuc Phuong National Park, Vietnam.
2. We evaluated a CNN-based model (Inception-v3) and a Vision Transformer (DINOv2) across datasets with controlled, cluttered, and mixed backgrounds.

The fine-tuned DINOv2 (ViT-B/14) achieved the best performance, reaching 96% top-5 accuracy on the test set, while also revealing more cost-efficient alternatives for task-specific classification.

3. We developed a web-based application that identifies endangered TFT species and provides ecological and legal information, incorporating user feedback to improve usability. To our knowledge, this is the first such AI-powered conservation tool in Vietnam, advancing public awareness and enforcement support.

The paper is organized as follows: Section 2 presents the literature review, Section 3 describes the data collection process, followed by data preprocessing in Section 3.2. Section 4 reports the experiments with various classification models and datasets. Section 5 describes how the model is integrated into the application supporting conservation. Finally, Section 6 discusses the findings, limitations, and future directions, with the conclusion in Section 7.

2 Literature Review

2.1 Turtle Identification in Ecology

Morphological Classification. Traditionally, TFT identification relies on morphological features documented in field handbooks for law enforcement [3,6,7]. These guides describe traits such as shell shape, color, and body structure, but they are difficult for non-experts to use and often result in slow or inaccurate identification. Morphological traits can vary with age and overlap across species, making taxonomy unreliable in some cases [52], as seen in closely related species like *Cuora bourreti* and *Cuora picturata*. Consequently, molecular methods now offer more objective species identification through genetic analysis.

Genetic Analysis. Genetic analysis enables objective TFT identification by revealing evolutionary relationships and species-specific traits, and is widely used in taxonomy and wildlife forensics [13,14,15,16]. DNA sequencing complements morphology by identifying species, hybrids, and new taxa [17,18,19]. However, molecular methods are resource-intensive, sensitive to experimental conditions, and impractical for field or law enforcement use [16]. Thus, genetics is best suited for final verification, while recent computer vision approaches offer a practical alternative for automated visual identification.

2.2 Turtle Identification using Computer Vision

Computer vision uses algorithms to interpret and understand digital images, offering promising applications in turtle identification through non-intrusive, automated recognition. The two main approaches in computer vision for this purpose are Feature Engineering and Deep Learning.

Feature Engineering. Feature Engineering involves extracting significant features from raw data using domain knowledge. In turtle identification, features like shape, coloration, and plastron patterns are manually extracted and used to train machine learning (ML) models. Studies have used this method for identifying freshwater turtles, tortoises, and sea turtles. Beugeling et al. used convexity-concavity analysis of carapace contours for individual identification, highlighting the importance of plastron features but noting the need for human intervention in segmentation [20]. Luís Pina et al. found color to be the most descriptive feature in classifying sea turtles [21], and Paixao et al. confirmed that color texture features are more effective than shape features [22]. However, the reliance on handcrafted features limits this approach. Deep learning offers solutions by providing automatic feature extraction, leading to more effective downstream ML models.

Deep Learning. Deep learning automatically learns hierarchical features from raw data using multi-layer neural networks optimized via backpropagation [23,24]. Its impact on computer vision began with CNNs such as AlexNet [25], which leverage convolution for image locality and translation invariance [26]. More recently, transformer-based models like Vision Transformers (ViT) have achieved strong performance in vision tasks, though they require large datasets [27,28]. Since 2018, deep learning has been widely adopted in ecology for tasks including wildlife detection, classification, monitoring, and conservation applications [29,30,31,32]. Deep learning has been applied in turtle conservation mainly for sea turtles, including drone-based detection [33,34] and identifying illegal tortoiseshell products [35]. For freshwater TFT species, limited work exists; Liu et al. showed Inception v3 performed best on five Chinese species, though with small-scale data [12]. Overall, computer vision research largely neglects morphologically similar freshwater turtles.

In HCI, animal identification systems support nature engagement and learning [36], with applications such as Pl@ntNet and Seek using deep learning to raise public awareness [37,38]. However, these tools lack support for Vietnam's turtle species. To our knowledge, no prior work has developed a turtle identification model or application tailored to Vietnam, which constitutes our main contribution.

3 Dataset (RQ1)

We collected training data for 27 species. Although Vietnam has 26 native TFT species, we included the Chinese softshell turtle (*Pelodiscus sinensis*), a species endemic to China but widely farmed and protected in Vietnam, raising the total classes to 27. Data came from three sources: a turtle conservation organization (IMC/ATP)[3], iNaturalist, and our own collection at the Turtle Conservation Center (TCC)[4], combined into a unified *Dataset*. Unless otherwise stated, *Dataset* refers to this aggregate. Images fall into three categories: (a) *Contextual Images* show turtles in natural or human environments; (b) *Documentation Images* are field records on paper with location/time/species info;

[3]The Asian Turtle Program (ATP), under Indo-Myanmar Conservation (IMC), a UK- and Vietnam-registered NGO (No. 1126123), focuses on conserving threatened species with its main focus on freshwater turtles and tortoises through research, rescue, captive breeding, and education. IMC is based at Bac Ha C14 Building, Hanoi, Vietnam.

[4]Turtle Conservation Center is a collaboration project managed by Cuc Phuong National Park, Ninh Binh, Vietnam, and the IMC/ATP.

(c) *Cataloging Images* are standardized photos on white backgrounds with IDs, taken at TCC, shown in Figure 1.

3.1 Data Collection

Dataset Acquired from Third-party Organization (IMC/ATP). Under an MOU with our researcher, IMC/ATP provided 12,711 images exclusively for this study. These images, taken in natural habitats, homes, rescue centers, and during trade confiscations, were captured by IMC/ATP field staff using multiple devices (e.g., mobile phones, DSLRs), offering high variability in quality. Most images show a single turtle, with occasional multi-turtle shots from confiscations. This dataset includes *Contextual, Documentation, and Cataloging Images.*

Dataset Collected from Crowdsourcing Platform (iNaturalist). We curated 4,043 *Research grade* [39] images from iNaturalist, a peer-reviewed crowdsourcing platform. These *Contextual Images* depict turtles in natural environments and introduce substantial variation in lighting, background, morphology, devices, and geography, including observations outside Vietnam. Although imbalanced across species, this dataset complements IMC/ATP field images by enriching the *Dataset* with diverse, wild and geographically varied perspectives.

Dataset Collected in Cuc Phuong National Park (Self-collected). We collected 7,925 images at TCC to supplement rare species with *Contextual Images* and provide *Cataloging Images.* TCC hosts over 2,000 turtles, covering 20 of 27 target species[5]. Cataloging followed IUCN/CITES guidelines [40], capturing multiple anatomical views, while contextual images were taken in naturalistic enclosures. Images were captured on two mobile devices under IMC/ATP supervision, adhering to animal welfare standards. Although constrained by camera quality and species availability, TCC data substantially improved dataset balance and coverage across species and image types.

3.2 Data Preprocessing

Cleaning Data. The IMC/ATP and iNaturalist datasets contained noisy images due to diverse sources. We manually filtered out low-quality, irrelevant, or misleading images, including those that were too dark, depicted dead or deformed turtles, showed only partial anatomy, non-turtle subjects, grouped individuals, or chessboard-pattern backgrounds to avoid bias. After cleaning, the final *Dataset* contained 17,775 high-quality images.

Cropping The Turtles. To reduce background noise, turtles were automatically cropped using GroundingDINO [41], an open-set object detector. The detector was evaluated on 500 carefully annotated images, achieving an IoU of

[5]Seven species — *Pelochelys cantorii*, *Pelodiscus variegatus*, *Rafetus swinhoei*, *Cuora cyclornata*, *Amyda ornata*, *Manouria impressa*, and *Cyclemys atripons* — are not present at the center. The reasons for their absence range from the center's inability to replicate the species' natural habitat conditions to the specifics of wildlife confiscation logistics, where these species have been directed to other facilities better suited for their case.

(a) (b) (c)

Fig. 1: Samples of three turtle image types: (a) *Contextual Images*; (b) *Documentation Images*; and (c) *Cataloging Images*. Note that IDs, location, etc. are sensitive and sealed in black.

0.91 and correctly localizing 97% of turtles. Based on this performance, we applied it to crop the entire dataset without further fine-tuning. For each image, only the highest-confidence detection was retained, resulting in 17,734 cropped images for downstream classification, while preserving visual traits critical for distinguishing morphologically similar species.

Splitting Data. The dataset was split into three subsets based on background: *White*, *Clutter*, and *Both*. *White* includes documentation and cataloging images, while *Clutter* consists of contextual images. Backgrounds were classified in HSV space by detecting white pixels; images with a white-area ratio above 0.15 were labeled *White*, otherwise *Clutter*, with manual corrections as needed. The final subsets contain 5,303 (*White*), 12,431 (*Clutter*), and 17,734 (*Both*) images (Table 1).

(a) (b)

Fig. 2: Examples of (a) *Clutter* and (b) *White* images.

4 Experiments (RQ2)

We evaluated the robustness of the turtle classification system across model and data dimensions, focusing on model generalization and background effects. Two approaches were tested: Deep Neural Networks (DNN) with classification heads and K-Nearest Neighbors (KNN) using feature embeddings. Inception-v3

Table 1: Data Composition of Each Sub-dataset

ID	Species name	White	Clutter	Both
1	*Amyda ornata*	3	171	174
2	*Cuora amboinensis*	232	790	1,022
3	*Cuora bourreti*	1,232	382	1,614
4	*Cuora cyclornata*	0	174	174
5	*Cuora galbinifrons*	473	584	1,057
6	*Cuora mouhotii*	270	1,133	1,403
7	*Cuora picturata*	87	388	475
8	*Cyclemys atripons*	0	102	102
9	*Cyclemys oldhamii*	496	332	828
10	*Cyclemys pulchristriata*	41	626	667
11	*Geoemyda spengleri*	324	165	489
12	*Heosemys annandalii*	58	460	518
13	*Heosemys grandis*	117	321	438
14	*Indotestudo elongata*	56	1,012	1,068
15	*Malayemys subtrijuga*	131	177	308
16	*Manouria impressa*	24	415	439
17	*Mauremys annamensis*	19	480	499
18	*Mauremys mutica*	187	431	618
19	*Mauremys sinensis*	174	1,028	1,202
20	*Palea steindachneri*	30	279	309
21	*Pelochelys cantorii*	0	88	88
22	*Pelodiscus sinensis*	174	523	697
23	*Pelodiscus variegatus*	9	61	70
24	*Platysternon megacephalum*	806	1,029	1,835
25	*Rafetus swinhoei*	3	427	430
26	*Sacalia quadriocellata*	47	622	669
27	*Siebenrockiella crassicollis*	310	231	541
	Total	5,303	12,431	17,734

[42] and DINOv2 [43] served as base models, trained on a Google Cloud VM with an Nvidia Tesla T4 GPU.

4.1 Preparation of Training Data

The training data consisted of 17,734 cropped images, with a separate *Universal Test* of 212 iNaturalist images for benchmarking. These images, including a maximum of 10 per class, were excluded from the *Clutter* and *Both* sub-datasets. iNaturalist data was chosen for its high variability in contributors, backgrounds, turtle orientations, devices, and geolocation-based morphological differences, reflecting real-world data. The remaining 17,522 images were split into training, validation, and test sets (70-15-15), maintaining the class ratio across splits. The distribution is summarized in Table 2.

4.2 Evaluation Metrics

Model performance was evaluated via cross-dataset testing, where models trained on one dataset were tested on unseen data from another, simulating real-

Table 2: All datasets used in our model training

Dataset	Training Set	Validation Set	Test Set
White	3,712	796	795
Clutter	8,512	1,855	1,852
Both	12,224	2,651	2,647
Universal Test	-	-	212

world deployment. Generalization was further assessed using the *Universal Test*. Metrics included top-1 accuracy, top-5 accuracy, and Micro/Macro F1 scores, with the Macro–Micro gap indicating performance on rare species. For DNNs, all metrics were used, while KNN models were evaluated using top-5 accuracy.

4.3 Model Selection

We evaluated two models: Inception-v3 [42], a CNN achieving 78.8% top-1 accuracy on ImageNet and effective in prior turtle classification studies [12], and DINOv2 [43], a transformer-based ViT model trained on 142M images with 86.7% top-1 ImageNet accuracy. Despite differing model sizes, the comparison emphasizes real-world classification performance, with Inception-v3 serving as a baseline and DINOv2 leveraging strong self-supervised feature learning.

4.4 Model Training

Turtle Classification with Deep Neural Networks. We compared transfer learning (TL) and finetuning (FT) on DINOv2 and Inception-v3. TL froze ImageNet-pretrained weights [44] and trained only the classification head, while FT updated all parameters. Experiments were conducted on the *White*, *Clutter*, and *Both* sub-datasets. Identical classifier heads, data augmentation, and cross-entropy loss with Adam optimizer were used, with learning rates in [1e-6, 1e-5, 1e-4, 1e-3]. DINOv2 was evaluated in both TL and FT settings, while Inception-v3 was finetuned as a baseline, with training time scaling with dataset size.

Turtle Classification with K-Nearest Neighbors. We used the DNNs trained in Section 4.4 as feature extractors for KNN models, and compared them with pre-trained (PT) DINOv2 and Inception-v3-ImageNet. This evaluated feature quality before and after finetuning, and tested DINOv2's claim of strong general-purpose representations [43] in the turtle classification task.

4.5 Results

The detailed evaluation results are summarized in Table 3, Table 4, and Table 5. For clarity, "TL" refers to "Transfer learning", "FT" refers to "Fine-tuning", and "PT" refers to "Pre-trained".

Comparison of Models. We calculated the average difference between pairs of models across all train-test combinations, as shown in Tables 6 and 7. DINOv2 generally outperformed Inception-v3 in both transfer learning and fine-tuning. In the DNN approach, DINOv2 (FT) was the top performer in terms of top-1

Table 3: Deep Neural Networks - Top-1 Accuracy

		DINOv2 (FT)			DINOv2 (TL)			Inception-v3 (FT)		
Tested \\$Trained$	Metrics	White	Clutter	Both	White	Clutter	Both	White	Clutter	Both
White	Top-1	0.976	0.805	**0.988**			0.975	0.962	0.747	0.977
	$F1_{macro}$	0.858	0.661	**0.914**	-	-	0.895	0.786	0.565	0.876
	$F1_{micro}$	0.976	0.805	**0.988**	-	-	0.975	0.962	0.747	0.977
Clutter	Top-1	0.363	0.925	**0.928**			0.892	0.298	0.886	0.884
	$F1_{macro}$	0.25	0.904	**0.91**	-	-	0.871	0.204	0.861	0.871
	$F1_{micro}$	0.363	0.925	**0.928**	-	-	0.892	0.298	0.886	0.884
Both	Top-1	0.399	0.832	**0.947**			0.918	0.341	0.78	0.914
	$F1_{macro}$	0.291	0.79	**0.928**	-	-	0.903	0.249	0.71	0.897
	$F1_{micro}$	0.399	0.832	**0.947**	-	-	0.918	0.341	0.78	0.914
Universal Test	Top-1	0.264	0.67	**0.698**			0.656	0.17	0.58	0.599
	$F1_{macro}$	0.221	0.615	**0.624**	-	-	0.569	0.135	0.495	0.542
	$F1_{micro}$	0.264	0.67	**0.698**	-	-	0.656	0.17	0.58	0.599

Table 4: Deep Neural Networks - Top-5 Accuracy

	DINOv2 (FT)			DINOv2 (TL)			Inception-v3 (FT)		
Tested \\$Trained$	White	Clutter	Both	White	Clutter	Both	White	Clutter	Both
White	0.996	0.979	**1.0**	-	-	1.0	0.99	0.95	1.00
Clutter	0.651	0.989	**0.992**	-	-	0.985	0.58	0.98	0.98
Both	0.671	0.982	**0.995**	-	-	0.99	0.60	0.95	0.99
Universal Test	0.538	0.958	0.962	-	-	**0.967**	0.43	0.87	0.86

Table 5: K-Nearest Neighbors - Top-5 Accuracy

	DINOv2 (PT)			DINOv2 (FT)			Inception-v3 (PT)			Inception-v3 (FT)		
Tested \\$Trained$	White	Clutter	Both	White	Clutter	Both	White	Clutter	Both	White	Clutter	Both
White	0.992	0.888	0.995	0.996	0.929	**0.998**	0.972	0.625	0.975	0.986	0.839	0.995
Clutter	0.553	0.956	0.958	0.514	0.958	**0.969**	0.35	0.854	0.853	0.403	0.949	0.952
Both	0.581	0.905	0.97	0.545	0.936	**0.978**	0.39	0.683	0.891	0.441	0.867	0.965
Universal Test	0.458	0.83	0.83	0.368	0.816	**0.835**	0.278	0.623	0.604	0.231	0.698	0.731

accuracy and F1 scores, with a 0.04 average gap between Micro and Macro F1, indicating consistent performance. It outperformed DINOv2 (TL) and Inception-v3 (FT) by 4% and 10%, respectively, in top-1 accuracy on the *Universal Test* set. For top-5 accuracy, no major difference was found between the transfer learning and fine-tuning versions of DINOv2. However, DINOv2 (TL) achieved a 0.5% higher score in the *Universal Test*, demonstrating the effectiveness of the pre-trained features. DINOv2 (PT) had a 15% reduction in training time compared to DINOv2 (FT), making transfer learning a more cost-effective option for top-5 accuracy. DINOv2 (FT) performed well on the *Both* dataset, benefiting from a large training set, but struggled with the *White* dataset due to its limited data and ViT's preference for global relationships. In the KNN approach, models using DINOv2 features outperformed those with Inception-v3, confirming DINOv2's strong feature extraction capabilities. While KNN showed inferior results to DNN, it could be a cost-effective solution for scaling the model with new species by using small amounts of high-quality images. However, the best performance on the *Universal Test* set was achieved using the DNN approach, indicating that optimal results require downstream training with machine learning models rather than just using raw features in KNN.

Table 6: Average Differences among Deep Neural Networks

Metrics	DINOv2 (FT) vs. Inception-v3 (FT)	DINOv2 (TL) vs. Inception-v3 (FT)	DINOv2 (FT) vs. DINOv2 (TL)
Top-1	0.05	0.02	0.03
Top-5	0.04	0.03	0
F1	0.06	0.01	0.03

Table 7: Average Differences among K-Nearest Neighbors

	DINOv2 (PT) vs. DINOv2 (FT)	Inception-v3 (FT) vs. Inception-v3 (PT)	DINOv2 (PT) vs. Inception-v3 (FT)
Top-5	0.01	0.08	0.07

Impact of Dataset Background. The experiments revealed that models trained on the *Both* dataset consistently performed better across all evaluation metrics, compared to models trained on single datasets. Models trained on the *Clutter* sub-dataset also performed well on both *Clutter* and *White*, while those trained solely on *White* showed good performance only on the *White* test set. This highlights the importance of background diversity in improving model generalization to different backgrounds. Models trained on the *Clutter* sub-dataset were close to those trained on the *Both* dataset in terms of top-5 accuracy, suggesting that diverse backgrounds aid in generalizing to specific clean-background use cases. The results emphasize that background diversity is crucial for creating adaptable and robust computer vision models. Additionally, the choice of evaluation approach depends on the use case. For scenarios involving human interaction, outputting top-5 predictions can enhance accuracy, while for automated systems, top-1 accuracy is key. Factors such as accuracy tolerance, dataset type, and computational resources also influence the best approach. These findings help practitioners select the most suitable model for their specific needs.

5 Building an Automatic Turtle Identification Application for Conservation (RQ3)

This section describes how we developed a mobile application that provides reliable identifications and assists users in making prompt and informed conservation-oriented decisions during turtle encounters, especially in situations where expert opinions may not be readily available. The application incorporates the best above deep-learning models to facilitate the automatic identification of turtles. The application includes two main parts, i.e., the server side and the client side. We then discuss user feedback.

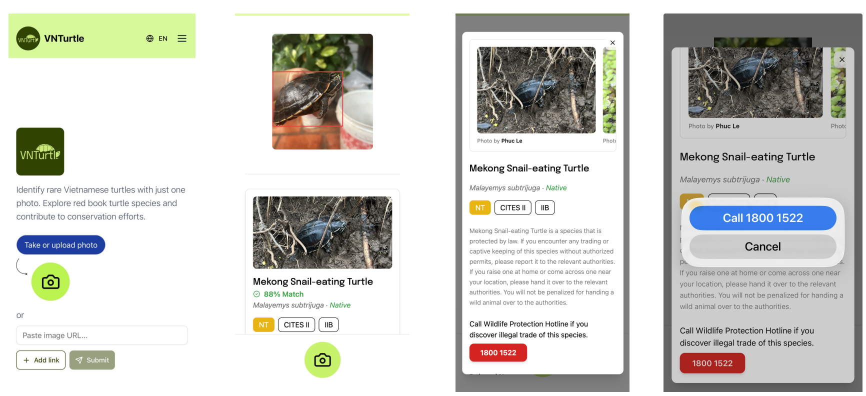
5.1 The server side

The application identifies turtles from a single image using a client–server architecture. A fine-tuned DINOv2 model performs classification, while a YOLOv8m detector [46], trained on IMC/ATP data with GroundingDINO annotations, localizes turtles before classification. Detection reduces background noise, filters non-turtle images, and handles multi-turtle inputs by selecting one instance. The detected turtle is highlighted with a bounding box for user transparency.

5.2 The client side

Technical perspective. The application was developed as a mobile-optimized web app using Next.js [47]. Designed for quick field use, the app presents the model's top-5 predictions with the turtle's *common name*, *confidence score*, and *reference images*, encouraging user verification and informed interpretation of results.

User Interface Design. The app provides a simple, bilingual (Vietnamese–English) interface for image capture, species identification, and reporting. It displays predicted species with confidence, conservation status, and legal protections in clear language, along with actionable guidance and a hotline for reporting. Figure 3 shows the main UI features.



(a) (b) (c) (d)

Fig. 3: The main features of our web app: (a) the home page; (b) identification results showing predicted species and confidence score; (c) detailed species information (conservation status and legal protection); and (d) a hotline button for reporting or assistance.

5.3 User Feedback

We conducted an online user survey with 15 participants from wildlife conservation backgrounds to evaluate usability and impact. Overall reception was positive, with average scores of 4.4 for user experience and 4.2 for performance, and high ratings (4.6) for identification and conservation information. Lower scores for encounter guidance (3.8) and qualitative feedback highlighted issues with display on some devices, difficulty identifying non-native species, and unclear placement of actionable steps. Despite the small sample size, users valued features such as common names and conservation status, and feedback pointed to clear directions for improvement in interface design, content visibility, and support for broader species coverage.

6 Discussion

Turtle Identification Conservation Impacts. Beyond the web app, the system has broad potential in turtle conservation, including automated monitoring of illegal trade on social media, support for law enforcement, and evidence-based policymaking. It can be integrated into education, citizen science, and offline fieldwork tools, and extended to other endangered species. Overall, this work demonstrates how AI can enhance conservation efforts, public engagement, and wildlife protection in Vietnam and beyond.

Ethical Considerations. Data collection followed strict ethical standards, prioritizing animal welfare and privacy. Turtle photography at Cuc Phuong National Park minimized handling and avoided harm, with sessions aligned to routine care and all metadata removed to protect sensitive locations. The application also incorporates ethical safeguards by presenting ranked predictions with confidence scores and reference images to reduce misidentification risks, especially in law enforcement. To prevent misuse, it omits location and dietary information, emphasizes legal protections, and ensures user privacy by not storing images or personal data unless explicitly contributed.

Limitation. The model currently supports only native species, limiting identification of non-native turtles. This could be mitigated by adding an "Unsupported Species" class or a separate filtering model. Recall for rare species such as *Amyda ornata* and *Rafetus swinhoei* remains a key challenge, highlighting the need to address data imbalance through techniques like over- or undersampling. Future work should also compare more comparable architectures, such as ConvNext-v2 [45], to enable fairer evaluation between CNN- and Transformer-based models.

7 Conclusion

Our contributions address two primary research questions by comparing model performance (DINOv2 vs. Inception-v3) across dataset types. Fine-tuned transformer-based DINOv2 models performed best, while pre-trained DINOv2 combined with KNN or transfer learning offered a cost-effective alternative. Diverse backgrounds in training data also improved performance. Thus, fine-tuning DINOv2 on a varied dataset is an effective strategy. Our work provides a practical guide for economically training turtle ID models. We also present an application that integrates vision models with conservation data to support TFT protection in Vietnam, highlighting key design and ethical considerations to enhance user engagement and minimize risks.

Acknowledgments. We express our sincere gratitude to TPBank for its generous funding through the TPBank STEM Scholarship, which has been instrumental in enabling this work. We also extend our special thanks to the Asian Turtle Program team, including Mr. Ha Hoang, Ms. Hang Tran, and other turtle caretakers, for their invaluable support in providing essential data and sharing expertise, which significantly enhanced the scope and depth of this study.

References

1. HUDSON, Rick. Turtles in trouble: the world's 25+ most endangered tortoises and freshwater turtles—2011. 2011.
2. FARKAS, Balázs, et al. A new species of Pelodiscus from northeastern Indochina (Testudines, Trionychidae). ZooKeys, 2019, 824: 71.
3. HENDRIE, Douglas B.; BUI Dang Phong; McCORMACK, Tim; HOANG Van Ha; VAN DIJK, Peter Paul; NGUYEN Tai Thang; NGUYEN Thanh Luan. Guidebook to Identify Freshwater Turtles and Tortoises in Vietnam. 3rd ed. Dong Nai: Dong Nai Publisher, 2022.
4. PHAN, Bich Huong; DUONG, Van Tho; TRAN, Thi Thanh Hai. *In the Maze: An Outlook of Illegal Wildlife Trade before and during the Covid-19 Pandemic in Vietnam* [online]. People and Nature Reconciliation, 2021. Available at: `https://www.nature.org.vn/en/wp-content/uploads/2022/01/In-the-Maze-13Jan22-EN.pdf`, last accessed 2022/04/21.
5. VAN, Thong Pham, et al. Longitudinal monitoring of turtle trade through Facebook in Vietnam. Herpetological Journal, 2019, 29.1.
6. *Guide to the Identification of Turtles and Tortoises Species Controlled under the Convention on International Trade in Endangered Species of Wild Fauna and Flora*. Canada: Environment Canada and PROFEPA (SEMARNAP), 1999. ISBN 0-662-64169-8. Available at: `https://cites.org/eng/virtual-college/cites-identification-guide-turtles-and-tortoises`, last accessed 2022/04/21
7. HANFEE, Fahmeeda. A WWF India Field Guide to Freshwater Turtles and Tortoises of India. TRAFFIC-India/WWF-India, 1999.
8. SINGH, Shailendra; BADOLA, Saket; FERNANDES, Merwyn. *Tortoises and Freshwater Turtles of India*. 2019.
9. CHEN, Leiyu, et al. Review of image classification algorithms based on convolutional neural networks. Remote Sensing, 2021, 13.22: 4712.
10. LIU, Xuefeng, et al. Real-time marine animal images classification by embedded system based on mobilenet and transfer learning. In: OCEANS 2019-Marseille. IEEE, 2019. p. 1-5.
11. BINTA ISLAM, Sazida, et al. Animal species recognition with deep convolutional neural networks from ecological camera trap images. Animals, 2023, 13.9: 1526.
12. LIU, Jiangchuan, et al. Classification and recognition of turtle images based on convolutional neural network. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020. p. 052044.
13. SHAFFER, H. Bradley; MEYLAN, Peter; MCKNIGHT, Mark L. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Systematic Biology, 1997, 46.2: 235-268.
14. FRITZ, Uwe; HAVAS, Peter. Checklist of chelonians of the world. Vertebrate Zoology, 2007, 57: 149-368.
15. CACCONE, Adalgisa, et al. A molecular phylogeny of four endangered Madagascar tortoises based on mtDNA sequences. Molecular Phylogenetics and Evolution, 1999, 12.1: 1-9.

16. NISHANT, K.; VRIJESH, K. Y.; AJAY, K. R. Wildlife forensic: current techniques and their limitations. J Forensic Sci Criminol, 2017, 5.4: 402.
17. STUART, Bryan L.; PARHAM, James Ford. Molecular phylogeny of the critically endangered Indochinese box turtle (Cuora galbinifrons). Molecular phylogenetics and evolution, 2004, 31.1: 164-177.
18. KUNDU, Shantanu, et al. Pet and turtle: DNA barcoding identified twelve Geoemydid species in northeast India. Mitochondrial DNA Part B, 2018, 3.2: 513-518.
19. ALACS, Erika; JANZEN, Fredric; SCRIBNER, Kim. Genetic issues in freshwater turtle and tortoise conservation. 2007.
20. BEUGELING, Trevor; BRANZAN-ALBU, Alexandra. Computer vision-based identification of individual turtles using characteristic patterns of their plastrons. In: 2014 Canadian Conference on Computer and Robot Vision. IEEE, 2014. p. 203-210.
21. PINA, Luís; RAJAMANICKAM, Leelavathi; NG, S. Feature extraction of the carapace for marine turtle species categorization. International Journal of Scientific Engineering and Technology, 2016, 5.9: 425-429.
22. DA PAIXAO, Wdnei R., et al. Texture classification of sea turtle shell based on color features: Color histograms and chromaticity moments. International Journal of Artificial Intelligence & Applications, 2018, 9.2: 55-67.
23. LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. nature, 2015, 521.7553: 436-444.
24. GOODFELLOW, A. Ian; Bengio, Yoshua; Courville. Deep Learning, 2016.
25. KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25.
26. STEVENS, Eli; ANTIGA, Luca; VIEHMANN, Thomas. Deep Learning with PyTorch: Build, train, and tune neural networks using Python tools. Manning, 2020.
27. VASWANI, Ashish, et al. Attention is all you need. Advances in neural information processing systems, 2017, 30.
28. DOSOVITSKIY, Alexey. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
29. CHRISTIN, Sylvain; HERVET, Éric; LECOMTE, Nicolas. Applications for deep learning in ecology. Methods in Ecology and Evolution, 2019, 10.10: 1632-1644.
30. BEERY, Sara; MORRIS, Dan; YANG, Siyu. Efficient pipeline for camera trap image review. arXiv preprint arXiv:1907.06772, 2019.
31. AHUMADA, Jorge A., et al. Thematic Section: Bringing Species and Ecosystems Together with Remote Sensing Tools toDevelopNewBiodiversityMetricsandIndicators Wildlife Insights: A Platform to Maximize the Potential of Camera Trap and Other Passive.
32. TABAK, Michael A., et al. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. Ecology and evolution, 2020, 10.19: 10374-10383.
33. GRAY, Patrick C., et al. A convolutional neural network for detecting sea turtles in drone imagery. Methods in Ecology and Evolution, 2019, 10.3: 345-355.
34. AI and cloud combine to help protect vulnerable marine turtle populations in northern Australia and Cape York,`https://news.microsoft.com/en-au/features/turtles-and-ai/`, last accessed 2022/04/21
35. SEE Shell App,`https://www.seeturtles.org/see-shell`, last accessed 2022/04/21
36. WEBBER, Sarah, et al. Engaging with nature through technology: A scoping review of HCI research. In: Proceedings of the 2023 CHI conference on human factors in computing systems. 2023. p. 1-18.

37. AFFOUARD, A.; JOLY, A.; LOMBARDO, J.; CHAMP, J.; GOEAU, H.; CHOUET, M.; GRESSE, H.; BOTELLA, C.; BONNET, P. *Pl@ntNet automatically identified occurrences*. Version 1.8 [online]. Pl@ntNet, 2023. Available at: https://ipt.plantnet.org/resource?r=queries&v=1.8, last accessed 2022/04/21
38. Seek by iNaturalist, https://www.inaturalist.org/pages/seek_app, last accessed 2023/12/12
39. The research grade, https://www.inaturalist.org/posts/39072-research-grade, last accessed 2024/01/01
40. Guidance To Photographing Live Tortoises and Freshwater Turtles for Identification | CITES, https://cites.org/eng/virtual-college/guidance-photographing-live-tortoises-and-freshwater-turtles-identification, last accessed 2024/01/01
41. LIU, Shilong, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In: European conference on computer vision. Cham: Springer Nature Switzerland, 2024. p. 38-55.
42. SZEGEDY, Christian, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2818-2826.
43. OQUAB, Maxime, et al. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193, 2023.
44. DENG, Jia, et al. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009. p. 248-255.
45. WOO, Sanghyun, et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023. p. 16133-16142.
46. YOLO by Ultralytics, https://github.com/ultralytics/ultralytics, last accessed 2023/06/01
47. NextJS, https://nextjs.org/, last accessed 2026/01/12
48. Selling exotic turtles on Facebook, young man was sentenced to 6.5 years in prison, https://vov.vn/phap-luat/chuyen-ban-rua-canh-tren-facebook-nam-thanh-nien-linh-an-65-nam-tu-905901.vov, last accessed 2022/04/15
49. A Snapshot Exotic Species Trade in Vietnam 2021, https://env4wildlife.org/wp-content/uploads/2021/11/A-Snapshot-of-Exotic-Species-Trade-in-Vietnam-2021-EN-November-4-2021-final.pdf, last accessed 2022/04/15
50. Turtle Market Online - Vibrant Turtle Market on Facebook and Youtube in Vietnam in 2021, https://nature.org.vn/vn/wp-content/uploads/2022/10/2022.-PanATP_Baocao_ChoRuaOnline_Final.pdf, last accessed 2022
51. Follow the illegal trading chain in the center of Sai Gon,https://thanhnien.vn/post-864430.html, last accessed 2022/01/15
52. 2.1: Species Concepts,https://bio.libretexts.org/Courses/University_of_California_Davis/BIS_2B%3A_Introduction_to_Biology_-_Ecology_and_Evolution/02%3A_Biodiversity/2.01%3A_Species_Concepts, last accessed 2022/04/21