VISVESVARAYA TECHNOLOGICAL UNIVERSITY
"JNANA SANGAMA", BELAGAVI - 590 018

MINI PROJECT REPORT

on

"AutoFis-Automated Identification of Fish Species"

Submitted by

Alok Illur	4SF21CD003
Chirasmita	4SF21CD005
Hrithik J Shetty	4SF21CD008
Shetty Adarsh Suresh	4SF21CD027

In partial fulfillment of the requirements for the V semester

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE)

Under the Guidance of

Dr. Mustafa Basthikodi

Professor and HOD, Department of CSE

at

SAHYADRI

College of Engineering & Management

An Autonomous Institution

MANGALURU

2023 - 24

SAHYADRI

College of Engineering & Management

Adyar, Mangaluru - 575 007

Department of Computer Science & Engineering (Data Science)

CERTIFICATE

This is to certify that the **Mini Project** entitled **"AutoFis - Automated Identification of Fish Species"** has been carried out by **Alok Illur (4SF21CD003), Chirasmita (4SF21CD005), Hrithik J Shetty (4SF21CD008) and Shetty Adarsh Suresh (4SF21CD027)**, the bonafide students of Sahyadri College of Engineering and Management in partial fulfillment of the requirements for the V semester of Bachelor of Engineering in Computer Science & Engineering (Data Science) and Engineering of Visvesvaraya Technological University, Belagavi during the year 2023-24. It is certified that all suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The mini project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

Project Guide	Project Coordinator	HOD
Dr.Mustafa Basthikodi	**Dr. Navaneeth Bhaskar**	**Dr. Rithesh Pakkala P**
Professor & HOD	Associate Professor	Associate Professor & HOD
Dept. of CSE	Dept. of ISE	Dept. of ISE

Evaluation:

Examiner's Name	Signature with Date
1.	....
2.	....

SAHYADRI

College of Engineering & Management

Adyar, Mangaluru - 575 007

Department of Computer Science & Engineering (Data Science)

DECLARATION

We hereby declare that the entire work embodied in this Mini Project Report titled **"AutoFis - Automated Identification of Fish Species"** has been carried out by us at Sahyadri College of Engineering and Management, Mangaluru under the supervision of **Dr. Mustafa Basthikodi**, in partial fulfillment of the requirements for the V semester of **Bachelor of Engineering** in **Computer Science and Engineering (Data Science)**. This report has not been submitted to this or any other University for the award of any other degree.

Alok Illur	**(4SF21CD003)**
Chirasmita	**(4SF21CD005)**
Hrithik J Shetty	**(4SF21CD008)**
Shetty Adarsh Suresh	**(4SF21CD027)**

Abstract

The marine and terrestrial ecosystems are essential components of the planet's natural balance which provide crucial resources and support to diverse forms of life. However, the sustainability of these ecosystems is under threat due to various anthropogenic activities, including over fishing and habitat destruction. Effective management strategies are paramount to mitigate these threats and preserve biodiversity.One significant challenge faced by fishermen, researchers and conservationists is the accurate identification of species. Correct species identification is vital for adhering regulations, promoting sustainable fishing practices and safeguarding the integrity of marine ecosystems. Traditional methods of species identification which is relying on manual observation and morphological characteristics are often time-consuming, labor-intensive and are prone to errors.To address this challenge, there is a pressing need for the development of an Intelligent Species Identification System which aims to harness cutting-edge technologies, such as artificial intelligence (AI), machine learning (ML) and computer vision to streamline the process of species identification. By leveraging vast databases of species information and sophisticated algorithms we can accurately identify the species from various sources such as photographs.The primary objective is to empower fishermen, marine biologists and policymakers with a user-friendly and reliable tool for species identification. Moreover it will contribute to the conservation of marine biodiversity by enhancing our understanding of species distributions, population dynamics and ecological interactions. The development and implementation of an Identification System represent a significant advancement in the field of marine conservation and sustainable management. By harnessing the power of technology it empowers stakeholders with the tools and knowledge needed to protect and preserve the invaluable biodiversity of our oceans.

Acknowledgement

It is with great satisfaction we are submitting the Mini Project Report on **"AutoFis - Automated Identification of Fish Species"**. We have completed it as a part of the curriculum of Visvesvaraya Technological University, Belagavi in partial fulfillment of the requirements for the V semester of Bachelor of Engineering in Computer Science and Engineering (Data Science).

We are profoundly indebted to our guide, **Dr. Mustafa Basthikodi** , Professor and HOD, Department of Computer Science and Engineering for timely advice, encouragement and we sincerely express our gratitude.

We also thank **Dr. Navaneeth Bhaskar**, Associate Dean (R&D) and **Ms. Ashritha K P**, Project Coordinator, Department of Information Science and Engineering & CSE (Data Science) for their constant encouragement and support extended throughout.

We express our sincere gratitude to **Dr. Rithesh Pakkala P**, Associate Professor and HOD, Department of Information Science and Engineering & CSE (Data Science) for his invaluable support and guidance.

We sincerely thank **Dr. S S Injaganeri**, Principal, Sahyadri College of Engineering and Management, who has always been a great source of inspiration.

Finally, we express our heartfelt thanks to our family and friends for their wishes and encouragement throughout the work.

Alok Illur	**(4SF21CD003)**
Chirasmita	**(4SF21CD005)**
Hrithik J Shetty	**(4SF21CD008)**
Shetty Adarsh Suresh	**(4SF21CD027)**

Table of Contents

Abstract **i**

Acknowledgement **ii**

Table of Contents **iii**

List of Figures **iv**

1 Introduction **1**

2 Literature Survey **3**

3 Problem Statement and Objectives **5**

3.1 Objectives 5

4 Methodology **6**

4.1 Architecture Diagram 6

4.1.1 Brief Explanation of Dataset: 8

4.2 Components Requirements 9

4.2.1 Tools Used : 9

4.2.2 Tech stack : 9

5 Results and Discussion **10**

5.1 Snapshots of Results Obtained : 10

5.1.1 Brief Explanation of Confusion Matrix: 12

5.1.2 Brief Explanation of Accuracy Table: 13

6 Outcome and Future Scope Work **14**

6.1 Outcomes: 14

6.2 Future Scopes: 15

7 Conclusion **16**

References **17**

List of Figures

4.1 Block Diagram of Proposed Model 7
4.2 Code snippet of Location Tracker 7
4.3 Dataset of Fish species 8

5.1 Frontend 10
5.2 Image Analyser with Sample Fish Data 11
5.3 Fish Detection with Sample Fish Data 11
5.4 Location Detector with Sample Data 12
5.5 Confusion Matrix 12
5.6 Accuracy Table of the Fish Dataset 13

Chapter 1

Introduction

In the multifaceted landscape of fisheries management, fishermen encounter a formidable obstacle that is the precise identification of a diverse array of fish species during their daily operations [1]. This challenge not only hampers the efficiency of fishing activities but also poses significant hurdles to the advancement of sustainable fishing practices. Traditional methods of species identification that is relying on manual observation and morphological characteristics are often time-consuming, labor-intensive and prone to errors [2]. Such limitations underscore the urgent need for innovative solutions that leverage cutting-edge technology to streamline the identification process.

The absence of a dependable system for species identification exacerbates this issue, resulting in a lack of crucial data essential for upholding the delicate ecological balance of marine ecosystems [1]. Recognizing this pressing need our proposed Intelligent Fish Species Identification System takes a comprehensive approach to tackle this multifaceted issue.

At the heart of this challenge lies the absence of a robust and accurate method for identifying diverse fish species. This deficiency not only impacts the livelihoods of fishermen but also poses a significant threat to the sustainability of marine ecosystems [2]. In response, our Intelligent Fish Species Identification System aims to harness the power of advanced machine learning techniques to deliver a reliable, accurate and scalable solution [3].

The initial phase of our system involves the compilation of a diverse dataset of fish species images drawn from various online repositories and databases [3]. This meticulously curated dataset serves as the cornerstone for training a machine learning model specifically a Convolutional Neural Network (CNN) [3]. The CNN is chosen for its exceptional ability to discern intricate patterns and features in images, making it an ideal candidate for the nuanced task of fish species identification [3].

By leveraging the power of pre-trained CNN's and transfer learning techniques, our system aspires to overcome the challenges associated with automated fish species recognition in the fish images [1]. Through the integration of advanced machine learning algorithms and imagery our system aims to provide real-time, accurate and scalable fish species identification capabilities [2].

Moreover, our system is designed to be user-friendly and accessible to a wide range of stakeholders, including fishermen, researchers, and policymakers [1]. By empowering stakeholders with the tools and knowledge needed for accurate species identification, our system facilitates compliance with regulations enables informed decision-making and promotes sustainable fishing practices [1]. Additionally, our system contributes to broader goals such as marine biodiversity conservation and the integration of technology into environmental conservation efforts [1].

In conclusion, the development and implementation of an Intelligent Fish Species Identification System represent a significant advancement in the field of fisheries management and marine conservation. By harnessing the power of technology, our system has the potential to revolutionize fish species identification practices, promote sustainability in fisheries, and contribute to the preservation of marine biodiversity.

Chapter 2

Literature Survey

Zhang et al. [1] explores about the various application of deep learning techniques but more importantly about convolutional neural networks (CNNs) for the identification of fish species and recognition in underwater environments. They address about the various challenges of accurately identifying fish species from autonomous underwater vehicle (AUV) imagery, which is crucial for various applications such as marine biodiversity monitoring,habitat assessment, and ecosystem management. By applying CNNs, the authors demonstrate significant improvements in the efficiency and accuracy of fish species identification compared to traditional methods. The reserach paper showcases the potential of deep learning-based approaches for underwater species identification and contribute to the conservation and management of marine ecosystems.

Mirimin et al. [2] provide a comprehensive review of the current methods and challenges in automated fish species recognition in underwater videos. They basically analyzed the various techniques and algorithms used for fish species recognition which include the traditional methods and deep learning approaches. The authors critically evaluate the performance of existing automated fish species recognition systems, highlighting their strengths, limitations and the areas for improvement. By findings from multiple studies they identifies key research gaps and future directions for advancing automated fish species recognition technology. The study serves as a valuable resource for researchers and practitioners working in the field of marine biodiversity who are monitoring and managing the ecosystem.

Beijbom et al. [3] focus on the application of machine learning techniques for automated monitoring of coral reef fish collection using underwater imagery. They basically develop and evaluate machine learning algorithms for identifying and quantifying fish species in coral reef environments aiming to enhance the efficiency and accuracy of fish species identification in underwater ecosystems. The study emphasizes the importance of automated monitoring systems in assessing coral reef health and biodiversity, and it demonstrates the potential of machine learning for advancing marine conservation efforts. By leveraging machine learning technology, the authors contribute to the development of innovative tools and methodologies for studying and protecting coral reef ecosystems.

Shoaib et al. [4] propose an automatic fish species classification system for underwater videos using pre-trained convolutional neural networks (CNNs) and transfer learning techniques. They basically check the effectiveness of the pre-trained CNN models for feature extraction and classification of the fish species in underwater videos aiming to develop efficient and accurate fish species classification systems for underwater environments. The authors demonstrate the ease of the pre-trained CNNs and transfering the learning to overcome challenges associated with limited labeled data and varying environmental conditions. By using deep learning and transfer learning, the study contributes to the development of robust and scalable solutions for automated fish species classification in underwater ecosystems.

Chapter 3

Problem Statement and Objectives

In fisheries management, the challenge of accurately identifying the diverse fish species and their precise locations presents a significant obstacle to the sustainable practices and ecological balance. Additionally, the classification of these fish species is also crucial. To address these issues, an Intelligent Fish Species Identification System is proposed. This involves compiling a comprehensive dataset of fish images to train a convolutional neural network(CNN). A user-friendly image capturing device which is tailored for the fishermen is developed, enabling a real-time species identification, classification and accurate location tracking. By integrating traditional practices with well-known technology this project aims to advance the sustainable fisheries management and contribute to the preservation of marine ecosystems.

3.1 Objectives

- To develop and train an automated system on a dataset for accurately identifying fish species from the imagery.
- To empower fishermen with a system that promotes sustainable fishing practices by ensuring
- To improve the efficiency and accuracy of fish species classification processes.
- To provide users with information on the nearest locations where the desired fish species has been observed.
- Contribute to marine biodiversity conservation efforts through data generation and analysis.

Chapter 4

Methodology

The methodology encompasses two key stages that is dataset collection and curation, and model selection and training. Dataset collection involves sourcing fish species images from diverse online sources and databases which is followed by meticulous curation and labeling to enhance model efficacy. This systematic approach aims to develop a precise and reliable system capable of accurately identifying and classifying various fish species based on their attributes.

4.1 Architecture Diagram

Dataset Collection and Curation:

- Gather a diverse dataset of fish species images from various online sources and databases.
- Emphasize meticulous curation and labeling for improved model efficacy.
- Ensure representation of a wide range of fish species to enhance the system's robustness.

Model Selection and Training:

- Choose a suitable machine learning model, specifically a Convolutional Neural Network (CNN).
- Train the model using the curated dataset to accurately identify different fish species.
- Implement techniques such as data augmentation to improve the model's generalization.

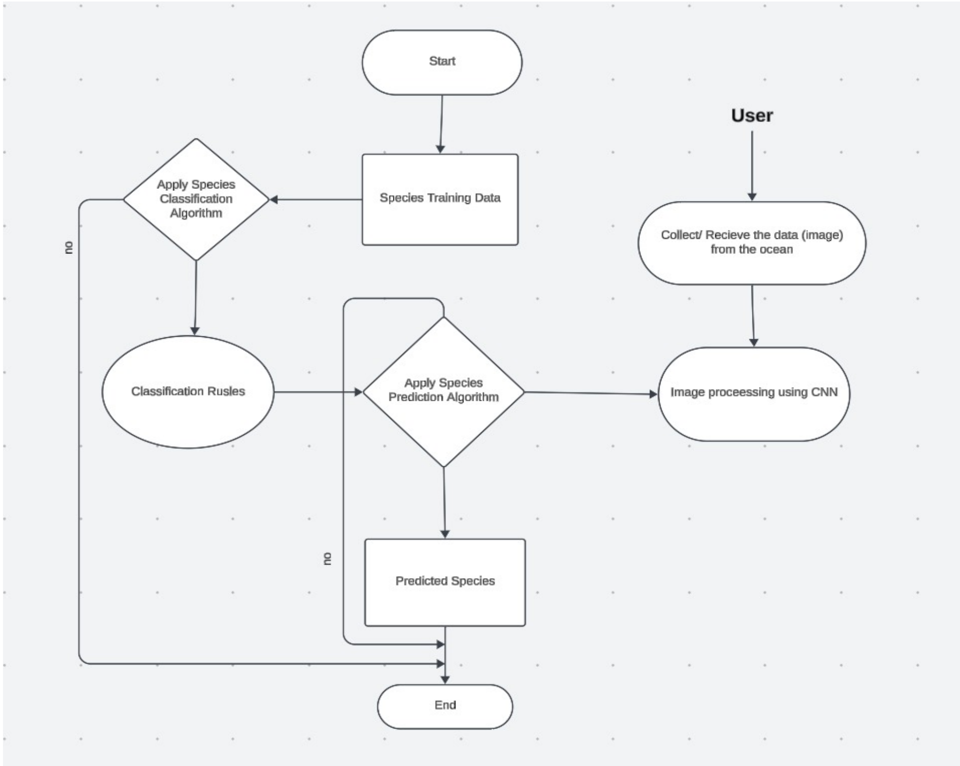

Figure 4.1: Block Diagram of Proposed Model

```python
def haversine(lat1, lon1, lat2, lon2):
    lat1, lon1, lat2, lon2 = map(radians, [lat1, lon1, lat2, lon2])
    dlat = lat2 - lat1
    dlon = lon2 - lon1
    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
    c = 2 * atan2(sqrt(a), sqrt(1 - a))
    distance = 6371 * c
    return distance

def find_nearest_places(reference_place, places, fish_species):
    reference_lat, reference_lon = reference_place['Latitude'], reference_place['Longitude']
    places_with_distances = []
    visited_locations = set()

    fish_species = fish_species.lower()

    for place in places:
        if fish_species in place['Common Fish Species'].lower():
            distance = haversine(reference_lat, reference_lon, place['Latitude'], place['Longitude'])
            location_name = place['Location']

            if location_name not in visited_locations:
                places_with_distances.append((location_name, distance))
                visited_locations.add(location_name)

    places_with_distances.sort(key=lambda x: x[1])

    return places_with_distances[1:7]
```

Figure 4.2: Code snippet of Location Tracker

4.1.1 Brief Explanation of Dataset:

We acquired a dataset sourced from Kaggle, comprising around 10,000 images. Employing standard data splitting practices, we divided the dataset into training and testing subsets using an 80:20 ratio, allocating 80% of the data for training and 20% for testing. Additionally, within the testing subset, we further partitioned it into validation and testing sets, each containing 10% of the original dataset. This rigorous partitioning scheme ensures robust model evaluation and performance validation. By splitting the dataset in this manner, we aim to train our model effectively on a diverse range of examples while ensuring fair evaluation through separate validation and testing phases. This approach helps mitigate overfitting and ensures the generalization capability of our machine learning model.

```
fData['Label'].value_counts()

Grass Carp               1212
Goby                      607
Glass Perchlet            407
Silver Barb               329
Knifefish                 319
Catfish                   314
Gourami                   311
Tilapia                   294
Perch                     293
Janitor Fish              286
Silver Perch              283
Tenpounder                277
Freshwater Eel            271
Indian Carp               262
Long-Snouted Pipefish     256
Mosquito Fish             254
Silver Carp               238
Snakehead                 232
Jaguar Gapote             229
Gold Fish                 206
Big Head Carp             201
Black Spotted Barb        200
Pangasius                 199
Fourfinger Threadfin      191
Mudfish                   189
Indo-Pacific Tarpon       186
Mullet                    177
Bangus                    171
Scat Fish                 154
Climbing Perch            152
Green Spotted Puffer      110
Name: Label, dtype: int64
```

Figure 4.3: Dataset of Fish species

4.2 Components Requirements

4.2.1 Tools Used :

- Google Collab
- VS code

4.2.2 Tech stack :

- Tensorflow
- CV
- Scikit learn
- Next Js
- Typescript
- React
- Html
- Flask

Chapter 5

Results and Discussion

The result obtained for this project is the successful development and deployment of the Intelligent Fish Species Identification System , a cutting-edge tool that revolutionises fish species identification in marine environments. Through the integration of advanced machine learning algorithms, including convolutional neural networks (CNNs) and transfer learning techniques it accurately identifies fish species from image , streamlining fisheries management practices and supporting marine biodiversity conservation efforts. The system's user-friendly interface and geolocation feature enhance its practical utility, enabling stakeholders to access timely and accurate information on fish species distributions and habitats. As a result, it contributes to improved decision-making processes, promotes responsible fishing practices and facilitates targeted conservation strategies aimed at preserving marine ecosystems and species diversity for future generations.

5.1 Snapshots of Results Obtained :

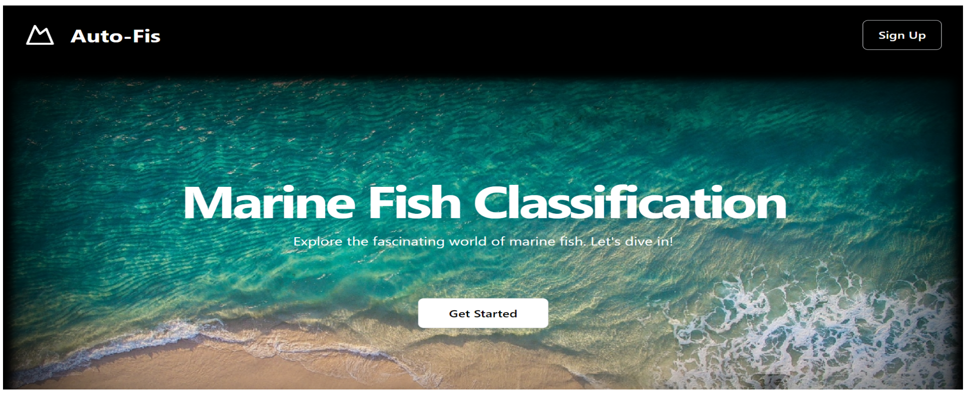

Figure 5.1: Frontend

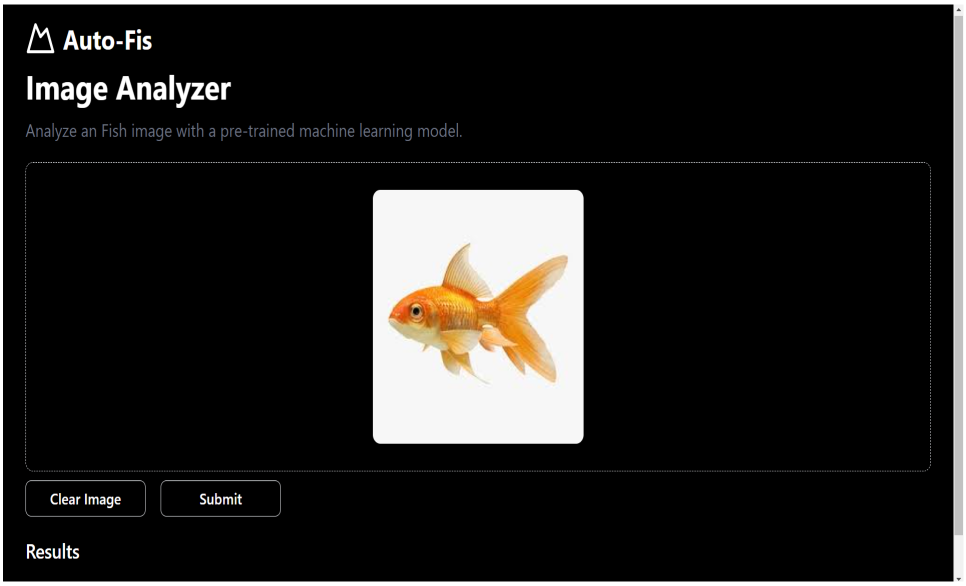

Figure 5.2: Image Analyser with Sample Fish Data

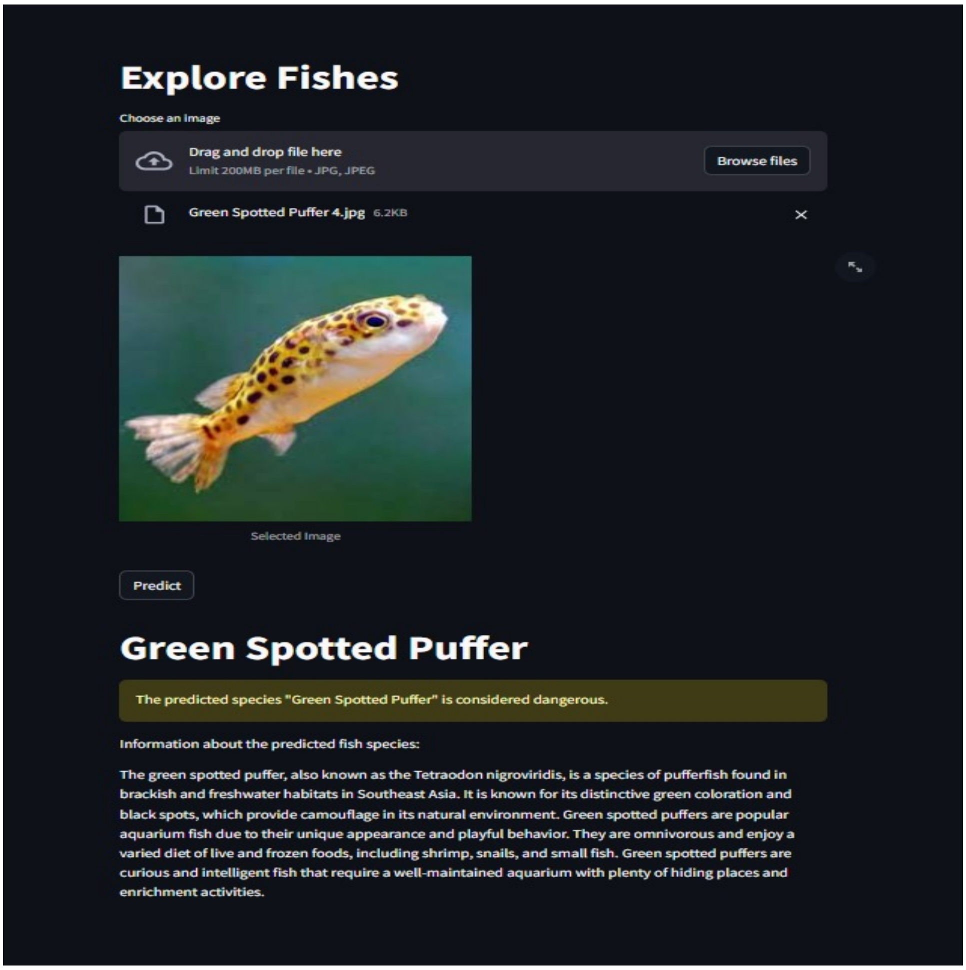

Figure 5.3: Fish Detection with Sample Fish Data

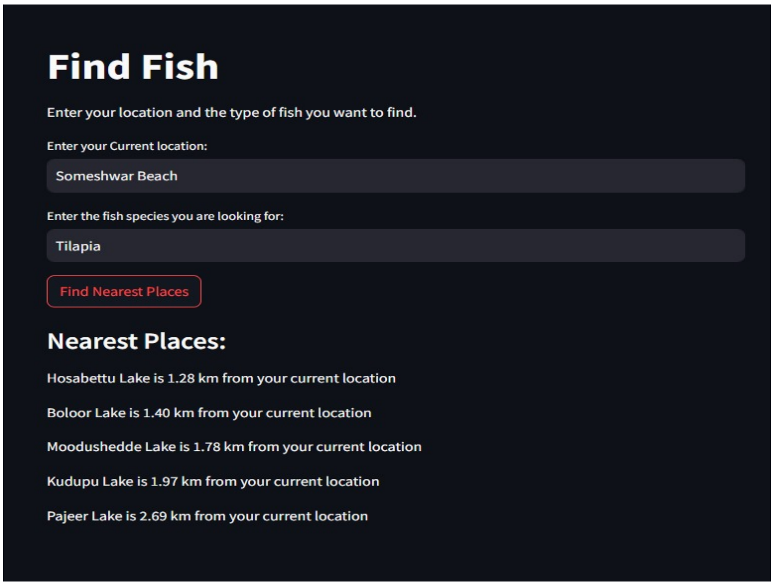

Figure 5.4: Location Detector with Sample Data

5.1.1 Brief Explanation of Confusion Matrix:

A confusion matrix is a tool used in machine learning to assess the performance of classification models by comparing predicted and actual class labels.Typically, the confusion matrix consists of four categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).In our case, due to a binary classification problem with an imbalance in predictions, we have no false positives. The remaining instances are divided between true negatives and false negatives, with approximately 373 instances classified correctly as negative and 124 instances misclassified as negative when they are actually positive.

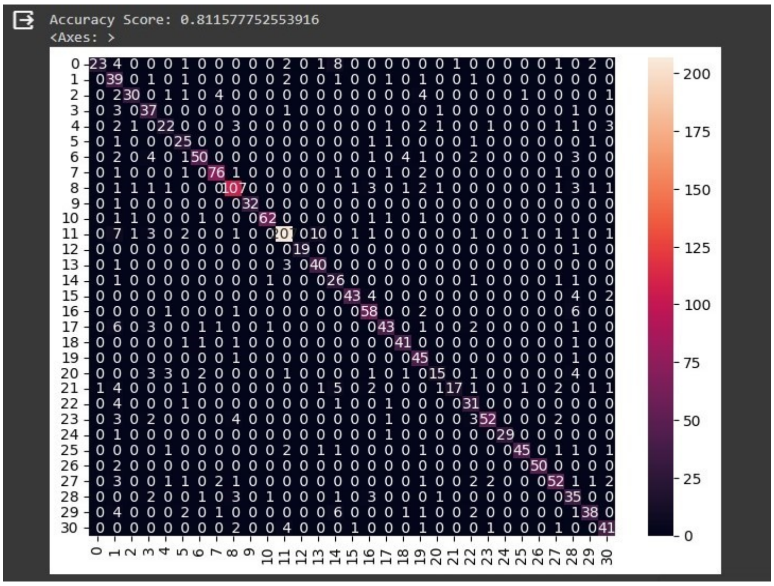

Figure 5.5: Confusion Matrix

5.1.2 Brief Explanation of Accuracy Table:

The classification report presents a comprehensive evaluation of the model's performance across 31 classes. High precision scores, such as 0.90 for "Glass Perchlet," 0.86 for "Goby," and 1.00 for "Silver Carp," indicate accurate predictions. Similarly, strong recall scores, like 0.96 for "Mosquito Fish," 0.87 for "Grass Carp," and 0.96 for "Silver Carp," show the model's ability to capture instances of these classes effectively.

F1-scores, balancing precision and recall, are also high for classes like "Gourami" (0.93), "Grass Carp" (0.90), and "Silver Carp" (0.98), indicating balanced performance. However, disparities exist for some classes, such as "Bangus" (precision: 0.96, recall: 0.53, F1-score: 0.69), "Big Head Carp" (precision: 0.42, recall: 0.83, F1-score: 0.56), and "Mudfish" (precision: 0.75, recall: 0.48, F1-score: 0.59), indicating challenges in accurate classification.

Overall accuracy stands at 81%, demonstrating the model's effectiveness. Addressing discrepancies in precision, recall, and F1-scores for specific classes is essential to enhance overall performance.

	precision	recall	f1-score	support
Bangus	0.96	0.53	0.69	43
Big Head Carp	0.42	0.83	0.56	47
Black Spotted Barb	0.88	0.68	0.77	44
Catfish	0.66	0.86	0.75	43
Climbing Perch	0.73	0.58	0.65	38
Fourfinger Threadfin	0.66	0.83	0.74	30
Freshwater Eel	0.89	0.74	0.81	68
Glass Perchlet	0.90	0.90	0.90	84
Goby	0.86	0.86	0.86	125
Gold Fish	1.00	0.97	0.98	33
Gourami	0.95	0.91	0.93	68
Grass Carp	0.93	0.87	0.90	238
Green Spotted Puffer	1.00	0.95	0.97	20
Indian Carp	0.75	0.91	0.82	44
Indo-Pacific Tarpon	0.52	0.84	0.64	31
Jaguar Gapote	0.93	0.81	0.87	53
Janitor Fish	0.77	0.85	0.81	68
Knifefish	0.84	0.73	0.78	59
Long-Snouted Pipefish	0.85	0.91	0.88	45
Mosquito Fish	0.69	0.96	0.80	47
Mudfish	0.75	0.48	0.59	31
Mullet	0.94	0.45	0.61	38
Pangasius	0.65	0.82	0.72	38
Perch	0.93	0.78	0.85	67
Scat Fish	1.00	0.94	0.97	31
Silver Barb	0.94	0.83	0.88	54
Silver Carp	1.00	0.96	0.98	52
Silver Perch	0.81	0.75	0.78	69
Snakehead	0.53	0.74	0.62	47
Tenpounder	0.86	0.68	0.76	56
Tilapia	0.77	0.80	0.79	51
accuracy			0.81	1762
macro avg	0.82	0.80	0.80	1762
weighted avg	0.84	0.81	0.82	1762

Figure 5.6: Accuracy Table of the Fish Dataset

Chapter 6

Outcome and Future Scope Work

6.1 Outcomes:

- Creation of a robust and accurate system for automated fish species identification.
- Using advanced machine learning techniques, including convolutional neural networks (CNNs) and transfer learning, to identify fish imagery and classify fish species with high precision.
- Reduced reliance on manual observation and morphological characteristics for fish species identification.
- Provision of a user-friendly tool for stakeholders such as fishermen, researchers and localites for smooth fish species identification processes and classification of target and non-target species.
- Contribution to marine biodiversity conservation efforts by generating valuable data on fish species distributions and classification.
- Addition of a location feature to enhance practical utility, providing users with information on the nearest locations where desired fish species have been observed.
- Enablement of users to locate potential habitats for specific fish species efficiently, facilitating targeted research and conservation efforts and responsible fishing practices by facilitating accurate species identification.

.

6.2 Future Scopes:

The future scope of our work could incorporate real-time data streams from sensors and underwater cameras deployed in marine environments.This integration would enable us for continuous monitoring of fish populations and environmental conditions, providing us the valuable insights for adaptive management strategies.By monitoring the environmental parameters such as temperature, salinity and water quality alongside with the behavior of we can gain unprecedented insights into the dynamics of marine ecosystems.Also to enhance the system's capabilities which can be directed towards expanding the species database to include a broader range of fish species from diverse geographic region which would inhance the coverage of species identification which leads to catering to the needs of users worldwide. Future collaborations with marine biologists and government agencies can facilitate the integration of this into larger-scale research and conservation projects. These collaborative efforts may involve the development of standardized protocols for data collection and analysis, as well as the integration of the system into broader conservation initiatives. Through these we aim to contribute significantly to marine conservation and sustainable management practices.

Chapter 7

Conclusion

In conclusion, the development of the Intelligent Species Identification System offers a promising solution to the critical challenge faced by fishermen in accurately identifying diverse fish species and classifying them during their activities. Additionally it gives you the information on the nearest locations where the desired fish species has been observed and tells you the distance from your current location.The absence of a reliable identification system not only hampers fishing efficiency but also contributes to a lack of essential data for maintaining marine ecosystem balance. Leveraging technological advancements, particularly Convolutional Neural Networks (CNNs), this system aims to streamline the identification , classification and accurate spot . Through meticulous data curation and labeling, it aspires to revolutionize fisheries management, promoting sustainable practices and contributing to the preservation of marine biodiversity. By providing a reliable means of species identification,classification and accurate location ,the system aligns with broader goals of sustainability, offering a transformative approach to harmonize human activities with the delicate ecosystems of our oceans.

References

[1] Er, M.J., Chen, J., Zhang, Y. and Gao, W., 2023. Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. Sensors, 23(4), p.1990.

[2] Mirimin, L., Desmet, S., Romero, D.L., Fernandez, S.F., Miller, D.L., Mynott, S., Brincau, A.G., Stefanni, S., Berry, A., Gaughan, P. and Aguzzi, J., 2021. Don't catch me if you can–Using cabled observatories as multidisciplinary platforms for marine fish community monitoring: an in situ case study combining Underwater Video and environmental DNA data. Science of the Total Environment, 773, p.145351.

[3] Shoaib Ahmed Siddiqui, Ahmad Salman, Muhammad Imran Malik, Faisal Shafait, Ajmal Mian, Mark R Shortis, Euan S Harvey, Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data, ICES Journal of Marine Science, Volume 75, Issue 1, January/February 2018, Pages 374–389, https://doi.org/10.1093/icesjms/fsx109

[4] Gonzalez-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, D.E., Ganase, A., Gonzalez-Marrero, Y., Herrera-Reveles, A., Kennedy, E.V., Kim, C.J., Lopez-Marcano, S. and Markey, K., 2020. Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. Remote Sensing, 12(3), p.489.