PROJECT REPORT

Project Name : Digital Naturalist-AI Enabled Tool For Biodiversity Researchers

Team Id :PNT2022TMID51703

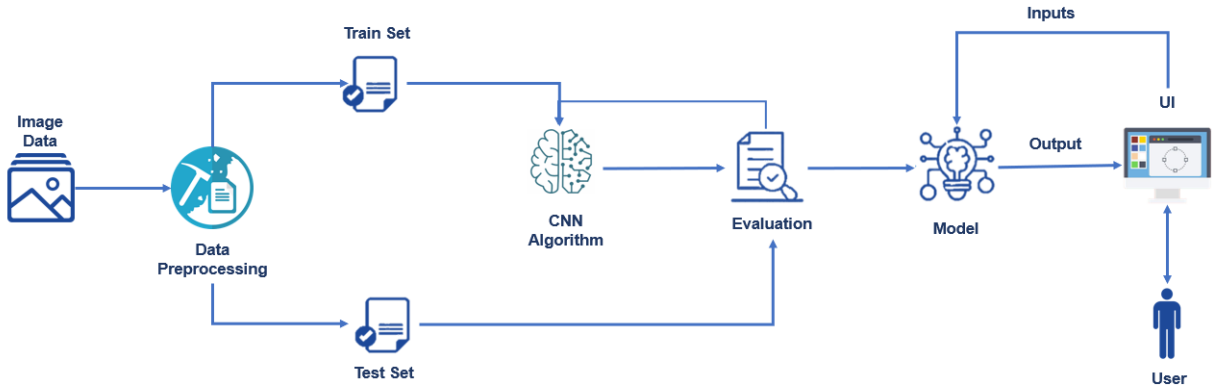
Team : JIJY A.V (Leader)
JEMI .K (Mem 1)
GRACIA ROSE .J (Mem 2)
BIBISHA M.S (Mem 3)

1. INTRODUCTION

1.1Project Overview

A naturalist is someone who studies the patterns of nature, identifies a different kind of flora and fauna in nature. Being able to identify the flora and fauna around us often leads to an interest in protecting wild spaces, and collecting and sharing information about the species we see on our travels is very useful for conservation groups like NCC.

Field naturalists can only use this web app from anywhere to identify the birds , flowers , mammals and other species they see on their hikes, canoe trips and other excursions.



In this project, we are creating a web application which uses a deep learning model, trained on different species of birds, flowers and mammals.

1.2 Purpose

To biological recording have to date typically focused on active sampling,

that is, images collected specifically for the purpose of recording wildlife (e.g.,
wildlife recording apps or camera traps). However, this has neglected large
amounts of image data that are not collected for the purposes of biological
recording, but which nonetheless may contain useful information about
biodiversity.
This includes social media imagery(e.g., Flickr and Instagram), CCTV, and
imagery collected along linear infrastructure (e.g., Google StreetView). These
unexploited image data could be rapidly analyzed using “AI naturalists” designed to
locate potential images of biodiversity and classify what they see.

2.LITERATURE SURVEY

2.2 Existing problem

- Digital Nature Photography is the definitive how – to book on photographing nature with a digital camera.
- Digital Photography is a process that uses an electronic device called a digital camera to capture an image.
- Landscape photography is one of the last bastions of traditional film, but this is slowly changing, as more and more leading photographers adopt digital technology.
- Heather is a versatile wildlife photographer whose images

combine scientific accuracy with pictorialappeal.

2.2 References

1. M. Schroeck, R. Shockley, J. Smart, D. Romero-Morales, P. Tufano
Analytics: The Real-World Use of Big Data
IBM Institute for Business Value (2012)

2.A.Y. Sun, B.R. Scanlon
How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions
Environ. Res. Lett., 14 (2019), p. 073001

3.M.A. Tabak, M.S. Norouzzadeh, D.W. Wolfson, S.J. Sweeney, K.C. Vercauteren, N.P. Snow, J.M. Halseth, P.A.D. Salvo, J.S. Lewis, M.D. White, *et al.*
Machine learning to classify animal species in camera trap images: applications in ecology
Methods Ecol. Evol., 10 (2019), pp. 585-590

4.R. Gibb, E. Browning, P. Glover-Kapfer, K.E. Jones
Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring
Methods Ecol. Evol., 10 (2019), pp. 169-185

5.B. Efron, T. Hastie
Computer Age Statistical Inference
Cambridge University Press (2016)

6.C.J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M.J. Raddick, R.C. Nichol, A. Szalay, D. Andreescu, *et al.*
Galaxy zoo: morphologies derived from visual inspection of galaxies

from the Sloan digital sky survey
Mon. Not. R. Astron. Soc., 389 (2008),

2.3 Problem Statement Definition

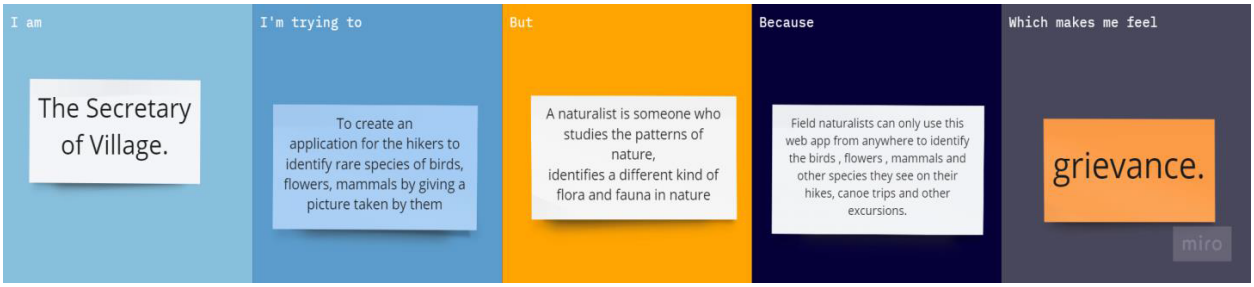
Create a problem statement is a detailed description of an issue that needsto be addressed by a problem-solving team.It is written to focus the team at the beginning, keep the team on track during the project, and to confirm that the team delivered an appropriate solution that addresses a customer need at the end of the project.
A customer problem statement outlines problems that the customer faces. It helps to figure out how the product or service will solve the problem for them. The statement helps to understand the experience to be offered to the customers.

3. IDEATION & PROPOSED SOLUTION

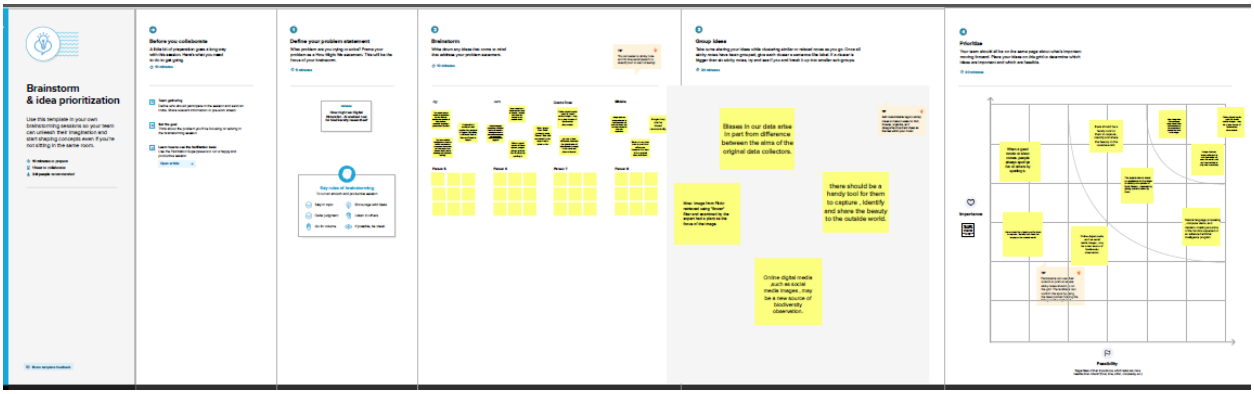
3.1 Empathy Map Canvas

3.2 Ideation and Brainstorming

Ideation:-



Brainstorming :-



3.3 Proposed Solution

S.NO	Parameter	Description
1.	Problem Statement(Problem to be solved)	speciesThe project aims to create an application for the hikers to identify rare of birds, flowers,

		mammals by giving a picture taken by them.
2.	Idea / Solution Description	A naturalist is someone who studies the patterns of nature, identifies a different kind of flora and fauna in nature. Being able to identify the flora and fauna around us often leads to an interest in protecting wild spaces, and collecting and sharing information about the species we see on our travels is very useful for conservation groups like NCC.
3.	Novelty / Uniqueness	Medical practitioners have

		been putting heavy emphasis on the enhancement of precision medicine andinventing a novel and effective cures for complex diseases. AI, on
4.	Social Impact / Customer Satisfaction The	The interest in customer experience has increased at a phenomenal rate. However, research to capture the true meaning of the concept is limited. Therefore, this study aims to address the question of what are the underlying dimensions that constitute the construct of customer

		experience.
5.	Business Model (Revenue Model)	A digital business model might be defined as a model that leverages digital technologies to improve several aspects of an organization . From how the company acquires customers, to what product/service it provides. A digital business model is such when digital technology helps enhance its value proposition.

3.4 Problem solution

Project Title: Digital Naturalist- AI enable tool for biodiversity research
Team ID: PNT2022TMID51703

Project Design Phase-I - Solution Fit Template

4.REQUIREMENT ANALYSIS

4.1 Functional Requirements:-

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Login	Login into application
FR-4	User Dashboard	Learn to access the application
FR-5	Online digital media	Such as social media images, may be a new source of biodiversity observations, but they are far too numerous for a human to practically review.
FR-6	Images	Images were predominantly

		biodiversity focused, showing single species. Non-functional Requirements: Following are the non-functional requirements of the proposed solution. FR No.

4.2 Non Functional Requirements :-

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	To create an application for the hikers to identify rare species of birds, flowers, mammals by giving a picture taken by them.
NFR-2	Security	The ever-growing number of digital sensors in the environment has led

		to an increase in tthe amount of digital data being generated.
NFR-3	Reliability	By combining social media APIs with AI classifiers, we were able to build an AI naturalist capable of creating biodiversity datasets from previouslyunexploited data sources.
NFR-4	Performance	Being able to identify the flora and fauna around us often leads to an interest in protecting wild spaces, and collecting and sharing information about the species we see on our travels is very useful for conservation

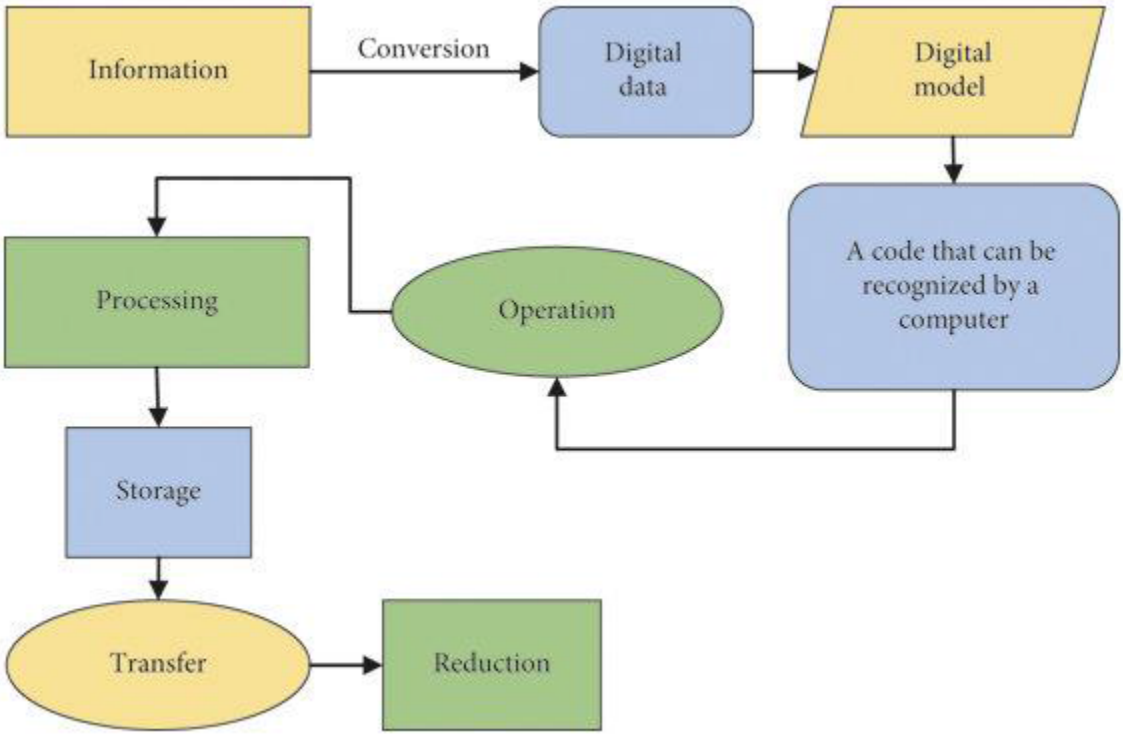
		groups like NCC.
NFR-5	Availability	When venturing into the woods, field naturalists usually rely on common approaches like always carrying a guidebook around everywhere or seeking help from experienced ornithologists. There should be a handy tool for them to capture, identify and share the beauty to the outside world.
NFR-6	Scalability	Field naturalist can only use from anywhere to identify the birds, flowers, mammals and other species they see on their hikes, canoe trips and

		Other excursions.

5.PROJECT DESIGN

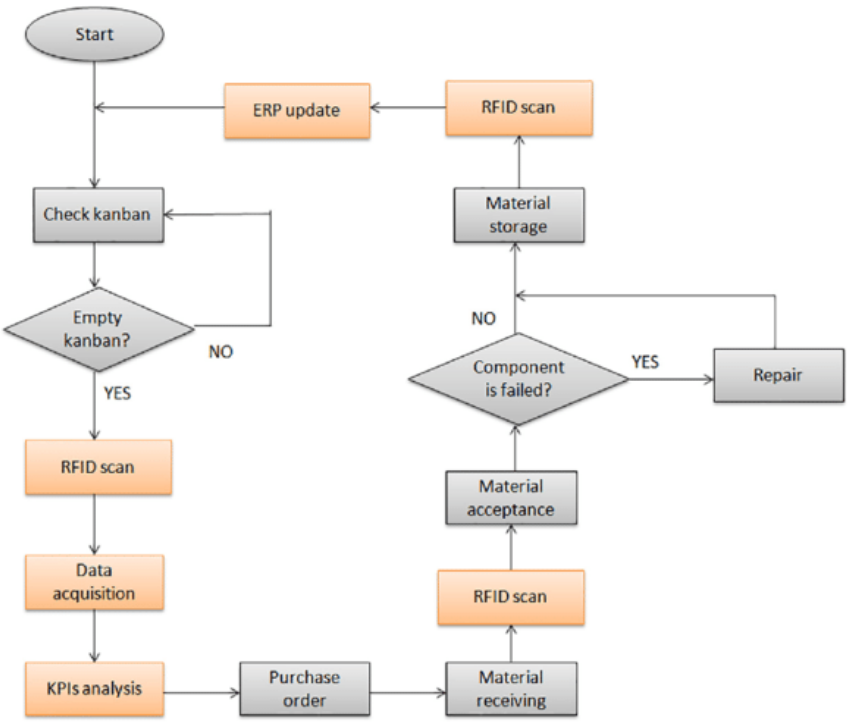
5.1 Data Flow Diagram

A Data Flow Diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DEFs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one. That’s why DFDs remain so popular after all these years. While they work well for data flow software and systems, they are less applicable now a days to visualizing interactive real-time or data base-oriented software were systems



5.2 Solution & Technical Architecture

A Solution architecture is an architectural description of a specific solution. SAs combine guidance from different enterprise architecture viewpoints as well as from the enterprise architecture viewpoints as well as from the enterprise solution architecture. The solution architecture helps ensure that a new system will fit the existing enterprise environment. A good solutions architect looks at the existing environment and analyses what technologies are available and what software product must be developed to provide the best solution for the problem that needs to be solved.



6.PROJECT PLANNING AND SCHEDULING

ctio Req me pic)	UserStoryNumber	UserStory/Task	StoryPoints	Priority	TeamMember
istr n(M e r)	UNS-1	As a user, I can register for the application by entering my email ,password,	**3**	High	Jijy A V (Team Leader)

		and confirming my password.			
	UNS-	As a user, I will receive confirmation email once I have registered For the application	**2**	High	Jemi K (Member 1)
	UNS-2	As a user, I can register for the application through Facebook	**3**	**High**	Gracia Rose J (Member 2)
	UNS-3	As a user, I can log in to the application by entering email and password	**3**	**High**	Bibisha M S (Member3)
	UNS-2	As a registered	**3**	**High**	Jijy A V (Team

		user, I need to easily login login to my registered account via the web page in minimum time			Leader)
	UNS-4	As a user, I need to have a friendly user interface to easily view and access there sources	**3**	**Medium**	Jemi K (Member 1)
	UNS-1	As a new user, I want to first register using my organization email and create a	**3**	**High**	Gracia Rose J (Member 2)

		password for The account.			
	UNS-4	**nt** As a registered user, I need to easily log in using the registered accouvia the web page	**3**	**High**	Bibisha M S (Member3)
	UNS-1	As a user, I want to first register using my email and create a password For the account	1	High	Jijy A V (Team Leader)

7. CODING & SOLUTIONING

7.1 Feature Code 1

```
# -*- coding: utf-8 -*-
"""
Created on Sat Nov 12 13:05:05 2022
@author: Alice
"""
#Importing Libraries

#Locating and loading datasets
import pathlib
from pathlib import Path
import os, gc, glob, random
from PIL import Image

#DataManagement and matrix calculations
import pandas as pd
import numpy as np

#Model Building
import tensorflow as tf
import keras
import keras.backend as K
from keras.optimizers import SGD, Adam, Adagrad, RMSprop
from keras.applications import *
from keras.preprocessing import *
from keras.preprocessing.image import ImageDataGenerator
from keras.callbacks import EarlyStopping, ModelCheckpoint
from keras.models import Sequential

from keras.layers import Dense, Conv2D, MaxPool2D, Flatten, Activation,
BatchNormalization,Dropout
from keras.models import Model
from keras.utils.np_utils import to_categorical
from sklearn.model_selection import train_test_split

# Data Visualization
import matplotlib.pyplot as plt

#Loading and testing models
from keras.models import load_model
from keras.models import model_from_json

# Directory operations
```

```
import os
from os import listdir

#
================================================================================
================================================== #
#
================================================================================
================================================== #
# ========================================DEFINING THE REQUIRED
FUNCTIONS============================================================ #
#
================================================================================
================================================== #
#
================================================================================
================================================== #
def generateListofFiles(dirName):
    """This function returns a list with exact paths of files inside the given
directory """
    listOfFile = os.listdir(dirName)
    allFiles = list()
    for fol_name in listOfFile:
        fullPath = os.path.join(dirName, fol_name)
        allFiles.append(fullPath)

    return allFiles

def Configure_CNN_Model(output_size):
    """This function defines the cnn model structure and configures the layers"""
    K.clear_session()
    model = Sequential()

 model.add(Dropout(0.4,input_shape=(224, 224, 3)))

    model.add(Conv2D(256, (5, 5),input_shape=(224, 224, 3),activation='relu'))
    model.add(MaxPool2D(pool_size=(2, 2)))
    #model.add(BatchNormalization())

    model.add(Conv2D(128, (3, 3), activation='relu'))
    model.add(MaxPool2D(pool_size=(2, 2)))
    #model.add(BatchNormalization())
```

```
    model.add(Conv2D(64, (3, 3), activation='relu'))
    model.add(MaxPool2D(pool_size=(2, 2)))
    #model.add(BatchNormalization())

    model.add(Flatten())
    model.add(Dense(512, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(256, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(128, activation='relu'))
    model.add(Dropout(0.3))

    model.add(Dense(output_size, activation='softmax'))

    return model

def PrepreocessData(subfolders):
    """Pre precess the image data in the provided category list"""
    X_data,Y_data,found = [],[],[]
    id_no=0
    #itering in all folders under Boats folder
    for paths in subfolders:
        #setting folder path for each boat type
        files = glob.glob (paths + "/*.jpg")
        found.append((paths.split('\\')[-2],paths.split('\\')[-1]))

        #itering all files under the folder one by one
        for myFile in files:
            img = Image.open(myFile)
            #img.thumbnail((width, height), Image.ANTIALIAS) # resizes image
in-place keeps ratio
            img = img.resize((224,224), Image.ANTIALIAS) # resizes image
without ratio
            #convert the images to numpy arrays
            img = np.array(img)
            if img.shape == ( 224, 224, 3):
                # Add the numpy image to matrix with all data
                X_data.append (img)
                Y_data.append (id_no)
        id_no+=1

    #converting lists to np arrays again
    X = np.array(X_data)
```

```
    Y = np.array(Y_data)

 # Print shapes to see if they are correct
    print("x-shape",X.shape,"y shape", Y.shape)

    X = X.astype('float32')/255.0
    y_cat = to_categorical(Y_data, len(subfolders))

    print("X shape",X,"y_cat shape", y_cat)
    print("X shape",X.shape,"y_cat shape", y_cat.shape)

    return X_data,Y_data,X,y_cat,found;
def splitData():
    X_train, X_test, y_train, y_test = train_test_split(X, y_cat,
test_size=0.2)
    print("The model has " + str(len(X_train)) + " inputs")
    return X_train, X_test, y_train, y_test
#
=============================================================================
========================================================= #
#
=============================================================================
========================================================= #
# ============================================LOADING THE DATA AND PRE-
PROCESSING DATA========================================================= #
#
=============================================================================
========================================================= #
#
=============================================================================
========================================================= #
# Augument the datasets with AugumentData.py.
# The AugumentData.py will generate many images with the original dataset to
increase the accuracy of the model.


# Loading the augumented data form local storage
aug_data_location = "C:/Users/0xluk/OneDrive/Documents/Digital
Naturalist/augumented data"
Folders = generateListofFiles(aug_data_location)
subfolders = []
for num in range(len(Folders)):
    sub_fols = generateListofFiles(Folders[num])
```

```
    subfolders+=sub_fols
```

8. TESTING

8.1Test case

```
# -*- coding: utf-8 -*-
"""
Created on Sat Nov 12 13:05:05 2022
@author: Alice
"""
#Importing Libraries


#Locating and loading datasets
import pathlib
from pathlib import Path
import os, gc, glob, random
from PIL import Image


#DataManagement and matrix calculations
import pandas as pd
import numpy as np


#Model Building
import tensorflow as tf
import keras
import keras.backend as K
from keras.optimizers import SGD, Adam, Adagrad, RMSprop
from keras.applications import *
from keras.preprocessing import *
from keras.preprocessing.image import ImageDataGenerator
from keras.callbacks import EarlyStopping, ModelCheckpoint
from keras.models import Sequential
from keras.layers import Dense, Conv2D, MaxPool2D, Flatten, Activation,
BatchNormalization,Dropout
from keras.models import Model
from keras.utils.np_utils import to_categorical
from sklearn.model_selection import train_test_split


# Data Visualization
```

```
import matplotlib.pyplot as plt


#Loading and testing models
from keras.models import load_model
from keras.models import model_from_json


# Directory operations
import os
from os import listdir


#
=============================================================================
========================================================= #
#
=============================================================================
========================================================= #
# ==========================================DEFINING THE REQUIRED
FUNCTIONS============================================================ #
#
=============================================================================
========================================================= #
#
=============================================================================
========================================================= #
def generateListofFiles(dirName):
    """This function returns a list with exact paths of files inside the given
directory """
    listOfFile = os.listdir(dirName)
    allFiles = list()
    for fol_name in listOfFile:
        fullPath = os.path.join(dirName, fol_name)
        allFiles.append(fullPath)

    return allFiles


def Configure_CNN_Model(output_size):
    """This function defines the cnn model structure and configures the layers"""
    K.clear_session()
    model = Sequential()
    model.add(Dropout(0.4,input_shape=(224, 224, 3)))

    model.add(Conv2D(256, (5, 5),input_shape=(224, 224, 3),activation='relu'))
```

```
    model.add(MaxPool2D(pool_size=(2, 2)))
    #model.add(BatchNormalization())


    model.add(Conv2D(128, (3, 3), activation='relu'))
    model.add(MaxPool2D(pool_size=(2, 2)))
    #model.add(BatchNormalization())


    model.add(Conv2D(64, (3, 3), activation='relu'))
    model.add(MaxPool2D(pool_size=(2, 2)))
    #model.add(BatchNormalization())


    model.add(Flatten())
    model.add(Dense(512, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(256, activation='relu'))
    model.add(Dropout(0.3))
    model.add(Dense(128, activation='relu'))
    model.add(Dropout(0.3))


    model.add(Dense(output_size, activation='softmax'))

    return model


def PrepreocessData(subfolders):
    """Pre precess the image data in the provided category list"""
    X_data,Y_data,found = [],[],[]
    id_no=0
    #itering in all folders under Boats folder
    for paths in subfolders:
        #setting folder path for each boat type
        files = glob.glob (paths + "/*.jpg")
        found.append((paths.split('\\')[-2],paths.split('\\')[-1]))

        #itering all files under the folder one by one
        for myFile in files:
            img = Image.open(myFile)
            #img.thumbnail((width, height), Image.ANTIALIAS) # resizes image
in-place keeps ratio
            img = img.resize((224,224), Image.ANTIALIAS) # resizes image
without ratio
            #convert the images to numpy arrays
            img = np.array(img)
```

```
            if img.shape == ( 224, 224, 3):
                # Add the numpy image to matrix with all data
                X_data.append (img)
                Y_data.append (id_no)
        id_no+=1


    #converting lists to np arrays again
    X = np.array(X_data)
    Y = np.array(Y_data)


    # Print shapes to see if they are correct
    print("x-shape",X.shape,"y shape", Y.shape)


    X = X.astype('float32')/255.0
    y_cat = to_categorical(Y_data, len(subfolders))


    print("X shape",X,"y_cat shape", y_cat)
    print("X shape",X.shape,"y_cat shape", y_cat.shape)


    return X_data,Y_data,X,y_cat,found;
def splitData():
    X_train, X_test, y_train, y_test = train_test_split(X, y_cat,
test_size=0.2)
    print("The model has " + str(len(X_train)) + " inputs")
    return X_train, X_test, y_train, y_test
#
================================================================================
================================================== #
#
================================================================================
================================================== #
# ========================================LOADING THE DATA AND PRE-
PROCESSING DATA================================================== #
#
================================================================================
================================================== #
#
================================================================================
================================================== #
# Augument the datasets with AugumentData.py.
# The AugumentData.py will generate many images with the original dataset to
increase the accuracy of the model.
```

```
# Loading the augumented data form local storage
aug_data_location = "C:/Users/0xluk/OneDrive/Documents/Digital
Naturalist/augumented data"
Folders = generateListofFiles(aug_data_location)
subfolders = []
for num in range(len(Folders)):
    sub_fols = generateListofFiles(Folders[num])
    subfolders+=sub_fols


X_data,Y_data,X,y_cat,found= PrepreocessData(subfolders)
# Splitting the data to Test and Train
X_train, X_test, y_train, y_test = splitData()


#
============================================================================
========================================================= #
#
============================================================================
========================================================= #
# ========================================================BUILDING THE CNN
MODEL============================================================ #
#
============================================================================
========================================================= #
#
============================================================================
========================================================= #
early_stop_loss = EarlyStopping(monitor='loss', patience=3, verbose=1)
early_stop_val_acc = EarlyStopping(monitor='val_accuracy', patience=3,
verbose=1)
model_callbacks=[early_stop_loss, early_stop_val_acc]


model = Configure_CNN_Model(6)
model.compile(loss='categorical_crossentropy',optimizer=Adam(lr=0.001),metrics=
['accuracy'])
weights = model.get_weights()
model.set_weights(weights)


#
```

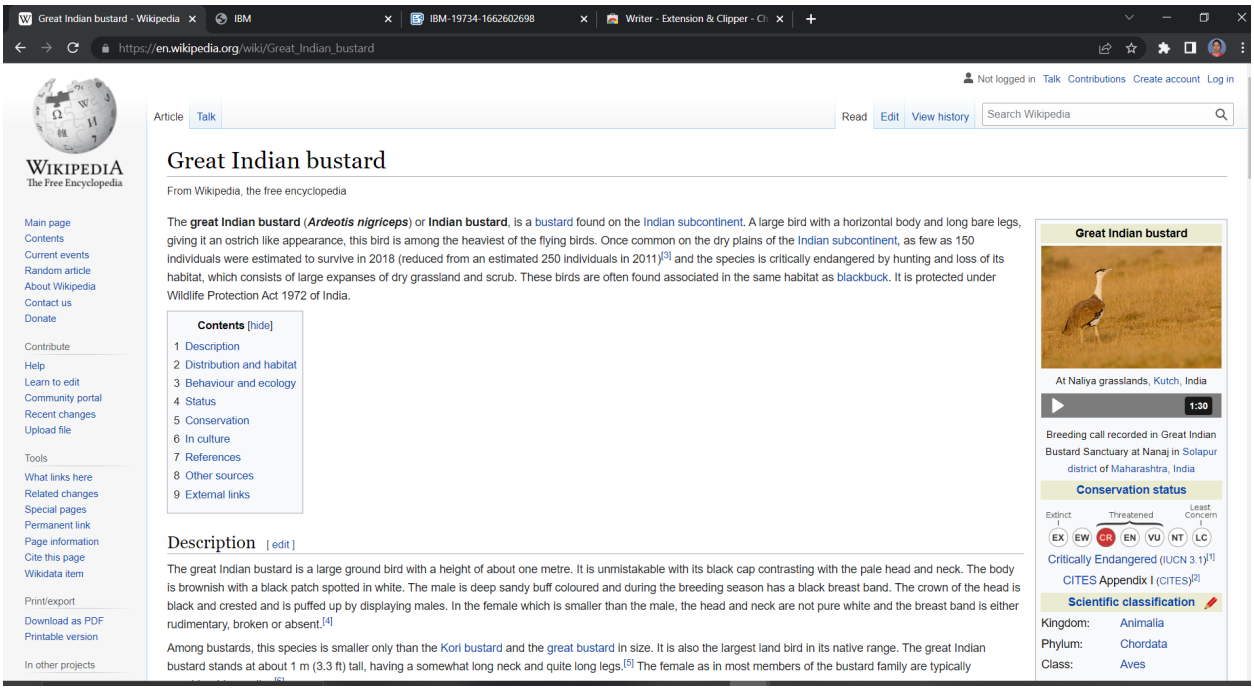
```
# ==========================================================================
# ========== #
#
# ==========================================================================
# ========== #
# ==========PREDECTING IMAGE
CLASSES========== #
#
# ==========================================================================
# ========== #
#
# ==========================================================================
# ========== #
image_number = random.randint(0,len(X_test))
predictions = model.predict([X_test[image_number].reshape(1, 224,224,3)])


for idx, result, x in zip(range(0,6), found, predictions[0]):
   print("Label: {}, Type : {}, Species : {} , Score : {}%".format(idx,
result[0],result[1], round(x*100,3)))



#predicting the class with max probability
ClassIndex=np.argmax(model.predict([X_test[image_number].reshape(1,
224,224,3)]),axis=1)
print(found[ClassIndex[0]])
#
# ==========================================================================
# ========== #
#
# ==========================================================================
# ========== #
# ==========SAVING THE MODEL
LOCALLY========== #
#
# ==========================================================================
# ========== #
#
# ==========================================================================
# ========== #
model_json = model.to_json() #indent=2
with open("DigitalNaturalist.json", "w") as json_file:
    json_file.write(model_json)
```

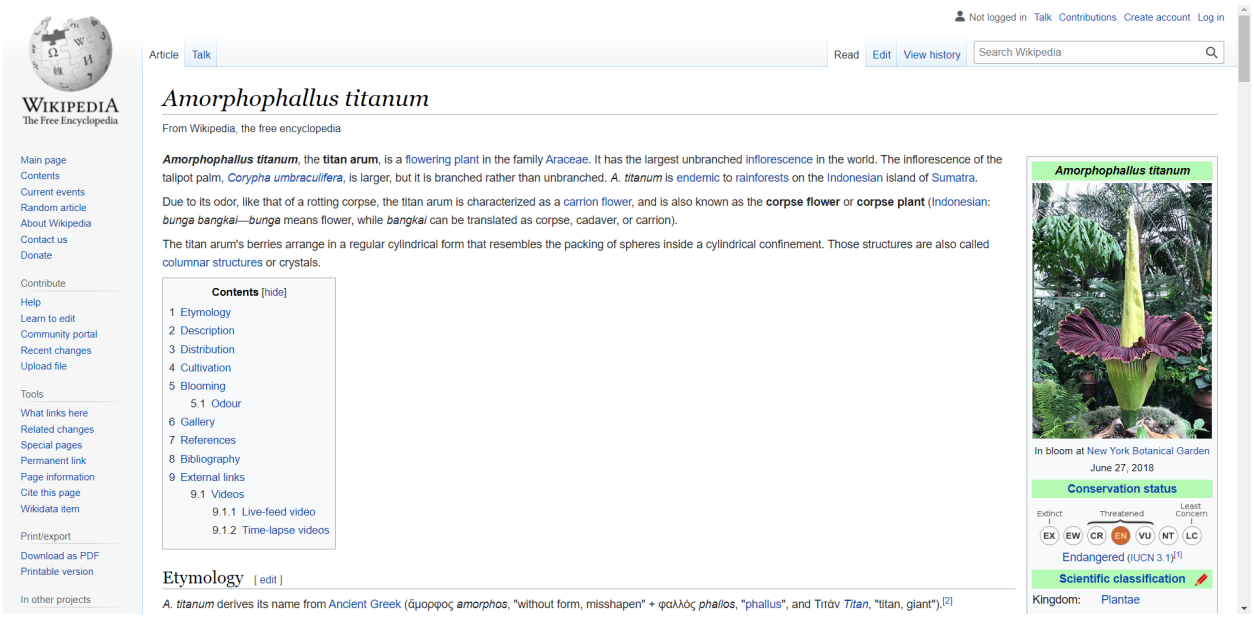
```
# serialize weights to H5
```

```
model.save_weights("DigitalNaturalist.h5")
print("Saved model to disk")
```

9.RESULTS

9.1 Performance Metrics





10.CONCLUSION

A naturalist is someone who studies the patterns of nature, identifies a different kind of flora and fauna in nature. Being able to identify the flora and fauna around us often leads to an interest in protecting wild spaces, and collecting and sharing information about the species we see on our travels is very useful for conservation groups like NCC.

Git hub link :-https://github.com/IBM-EPBL/IBM-Project-43439-1660716904

Demo video link :-https://drive.google.com/file/d/1V0vWwW-kG-KFyRIvEKbXp5VZqGQO1DyP/view?usp=drivesdk