Plant Seeding Classification By IBM Watson

1. INTRODUCTION

1.1 Overview:

Agriculture is vital for human survival and remains a major driver of several economies around the world; more so in underdeveloped and developing economies. With increasing demand for food and cash crops, due to a growing global population and the challenges posed by climate change, there is a pressing need to increase farm outputs while incurring minimal costs. One major reason for reduction in crop yield is weed invasion on farmlands. Weeds generally have no useful value in terms of food, nutrition or medicine yet they have accelerated growth and parasitically compete with actual crops for nutrients and space. Inef?cient processes such as hand weeding has led to signi?cant losses and increasing costs due to manual labor. Plants continue to serve as a source of food and oxygen for all life on earth. In continents like Africa, where agriculture is predominant, proper automation of the farming process would help optimize crop yield and ensure continuous productivity and sustainability. In recent times, the state of agriculture and the amount of work people need to put in to check if plants/food is growing correctly is phenomenal, because it is 2019 and workers still need to organize and recognize the difference between different plants and weeds. People who are working in the agriculture field still have to have the ability to sort and recognize different plants and weeds, which does take a lot of time and a lot of effort in the long term.

This is where Artificial Intelligence can actually help benefit those workers, as the time and energy to identify plant seedlings will be much shortened. The ability to do so effectively can mean better crop yields and in the long term will result in better care for the environment. As by identifying the difference among the plants and weeds in a timely manner where it is highly accurate can positively impact agriculture.

1.2 Purpose:

- Know fundamental concepts and techniques of Convolutional Neural Network.
- Gain a broad understanding of image data.
- Know how to pre-process/clean the data using different data preprocessing techniques.
- Know basics of Transfer Learning and build Xception model
- know how to build a web application using the Flask framework.

? The user interacts with the UI (User Interface) to choose the image.
? The chosen image analyzed by the model which is integrated with flask application.
? Xception Model analyzes the image, then prediction is showcased on the Flask UI.

To accomplish this, we have to complete all the activities and tasks listed below

- Data Collection.
    - Create Train and Test Folders.
- Model Building
    - Importing the Model Building Libraries
    - Loading the model
    - Adding Flatten Layers

    - Adding Output Layer
    - Creating a Model object:
    - Configure the Learning Process
    - Import the ImageDataGenerator library
    - Configure ImageDataGenerator class
    - Apply ImageDataGenerator functionality to Trainset and Testset
  - Training
    - Train the Model
    - Save the Model
  - Testing
    - Test the model
  - Application Building
    - Create an HTML file
    - Build Python Code
    - Run the application
    - Final Output

2. LITERATURE SURVEY:

2.1 Existing problem:

We develop the idea of classification problem solving by starting with the common sense notion and relating it to the reasoning that occurs In heuristic programs.

1. Simple classification :

AS the name suggests, the simplest kind of classification problem is to identify some unknown object or phenomenon as a member of a known class of objects or phenomena. Typically, these classes are stereotypes that are hierarchically organized, and the process of identification is one of matching observations of an unknown entity against features of known classes. A paradigmatic example is identification of a plant or animal, using a guidebook of features, such as coloration, structure, and size. Some terminology we will find helpful: The problem is the object or phenomenon to be identified; data are observations describing this problem; possible solutions are patterns (variously called schema, frames or units); each solution has a set of features (slots or facets) that in some sense describe the concept either categorically or probabilistically; solutions are grouped into a specialization hierarchy based on their features (in general, not a single hierarchy. but multiple, directed acyclic graphs); a hypothesis is a solution that is under consideration: evtdence is data that partially matches some hypothesis; the output is some solution. The essential characteristic of a classification problem is that the problem solver selects from a set of pre-enumerated solutions. This does not mean, of course, that the "right answer" is necessarily one of these solutions, just that the problem solver will only attempt to match the data against the known solutions, rather than construct a new one. Evidence can be uncertain and matches partial, so the output might be a ranked list of hypotheses. Besides matching, there are several rules of inference for making assertions about solutions. For example, evidence for a class is indirect evidence that one of its subtypes is present. Conversely, given a closed world assumption, evidence against all of the subtypes is evidence against a class. Search operators for finding a solution also capitalize on the hierarchical structure of the solution space. These operators include: refining a hypothesis to a more specific classification; categorizing the problem by considering superclasses of partially matched hypotheses; and discriminating among hypotheses by contrasting their superclasses [31.32, 121. For simplicity, we will refer to the entire process of applying these rules of inference and operators as refinement. The specification of this process-a control strategy-is an orthogonal issue which we will consider later.

2. Data abstraction: In the simplest problems, data are solution features. so the matching and refining process is direct. For example, an unknown organism in MYCIN can be classified directly given the supplied data of gram stain and morphology. For many problems, solution features are not supplied as data, but are inferred by data abstraction. There are three basic relations for abstracting data in heuristic programs: l qualitative abstraction of quantitative data (“if the patient is an adult and white blood count is less than 2500, then the white blood count is low“); l definitional abstraction (“if the structure is one-dimensional of network construction, then its shape is a beam”); and l generalization in a subtype hierarchy (“if the client is a judge, then he is an educated person ‘I). These interpretations are usually made by the program with certainty; thresholds and qualifying contexts are chosen so the conclusion is categorical. It is common to refer to this knowledge as “descriptive, ” “factual,” or “definitional.”

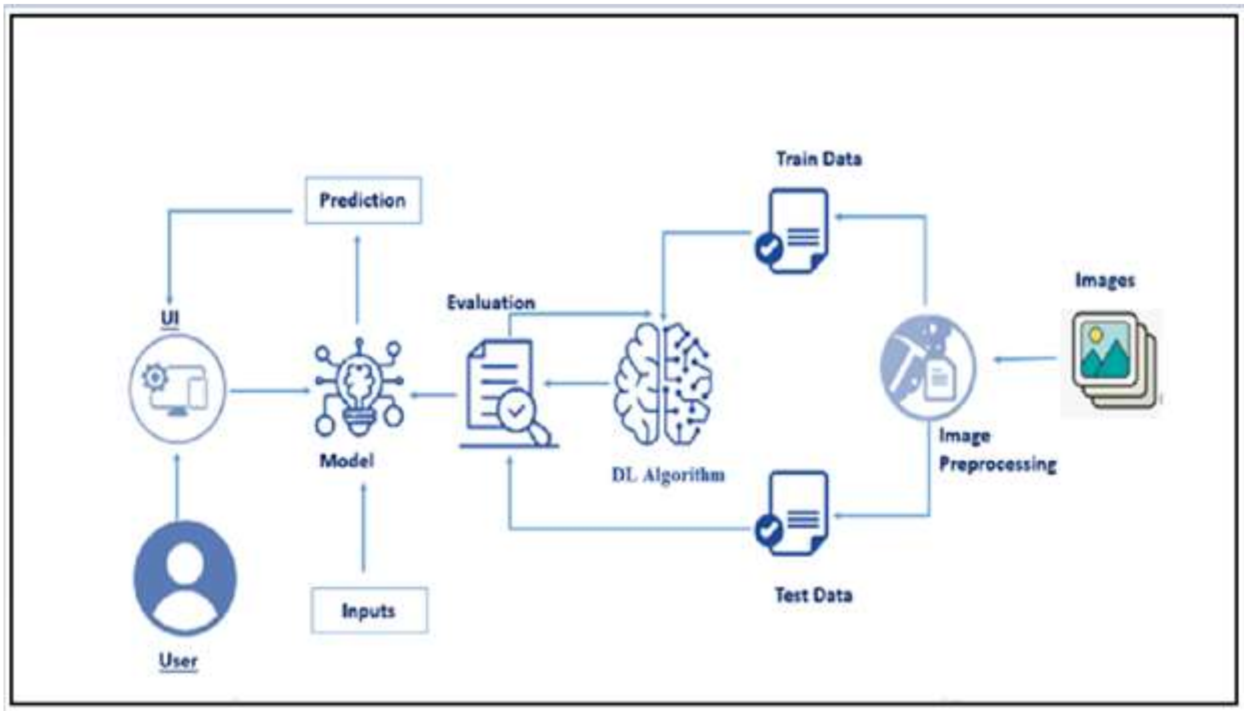
3. Heuristic classification: P In simple classification, the data may directly match the solution features or may match after being abstracted. In heuristic classification, solution features may also be matched heuristically. For example, MYCIN does more than identify an unknown organism in terms of features of a laboratory culture: It heuristically relates an abstract characterization of the patient to a classification of diseases. We show this inference structure schematically, followed by an example (Figure Z-1). Basic observations about the patient are abstracted to patient categories, which are heuristically linked to diseases and disease categories.

2.2 Proposed solution:

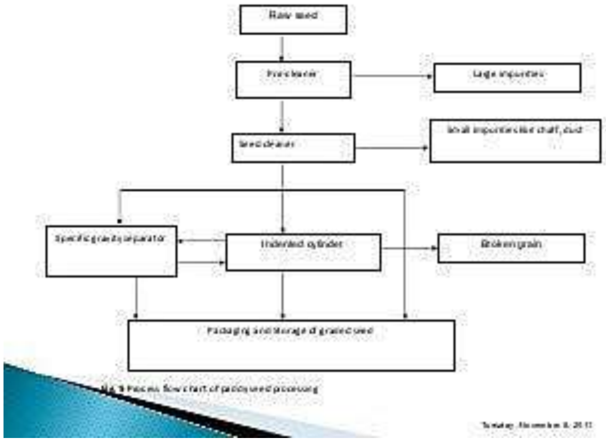
We first summarize the applications we have considered by observing that all classification problem solving involves selection of a solution. We can characterize kinds of problems by what is being selected: l diagnosis: solutions are faulty components (SOPHIE) or processes affecting the device (MYCIN); l user model: solutions are people stereotypes in terms of their goals and beliefs (first phase of GRUNDY); l catalog selection: solutions are products, services, or activities, e.g., books, personal computers, careers, travel tours, wines, investments (second phase of GRUNDY); l theoretical analysis: solutions are numeric models (first phase of SACON); l skeletal planning: solutions are plans, such as packaged sequences of programs and parameters for running them (second phase of SACON, also first phase of experiment planning in MOLGEN [Is]). Acommon misconception is that the description “classification problem” is an inherent property of a problem, opposing, for example, classification with design [37]. However, classification problem solving, as defined here, is a description of how a problem is solved by a particular problem solver. If the problem solver has a priori knowledge of solutions and can relate them to the problem description by data abstraction, heuristic association, and refinement, then the problem can be solved by classification. For example, if it were practical to enumerate all of the computer configurations Rl might select, or if the solutions were restricted to a predetermined set of designs, the program could be reconfigured to solve its problem by classification. 15 Furthermore. as illustrated by ABEL. it is incorrect to say that medical diagnosis is a “classification problem.” Only routine medical diagnosis problems can be solved by classification [32]. When there are multiple. interacting diseases. there are too many possible combinations for the problem solver to have considered them all before. Just as ABEL reasons about interacting states, the physician must construct a consistent network of interacting diseases to explain the symptoms. The problem solver formuiates a solution; he doesn't just make yes-no de&ions from a set of fixed alternatives. For this reason. Pople calls non-routine medical diagnosis an ill-structured problem [36] (though it may be more appropriate to reserve this term for the theory formation task of the physician-scientist who is defining new diseases). ln summary, a useful distinction is whether a solution is selected or constructed. To select a solution, the problem solver needs experiential ("expert") knowledge in the form of patterns of problems and solutions and heuristics relating them. To construct a solution, the problem solver applies models of structure and behavior, by which objects can be assembled, diagnosed, or employed in some plan. Whether the solution is taken off the shelf or is pieced together has important computational implications for choosing a representation. In particular, construction problem-solving methods such as constraint propagation and dependency-directed backtracking have data structure requirements that may not be easily satisfied by a given representation language. For example-returning to a question posed in the introduction-applications of EMYCIN are generally restricted to problems that can be solved by classification.

3. Theoretical Analysis

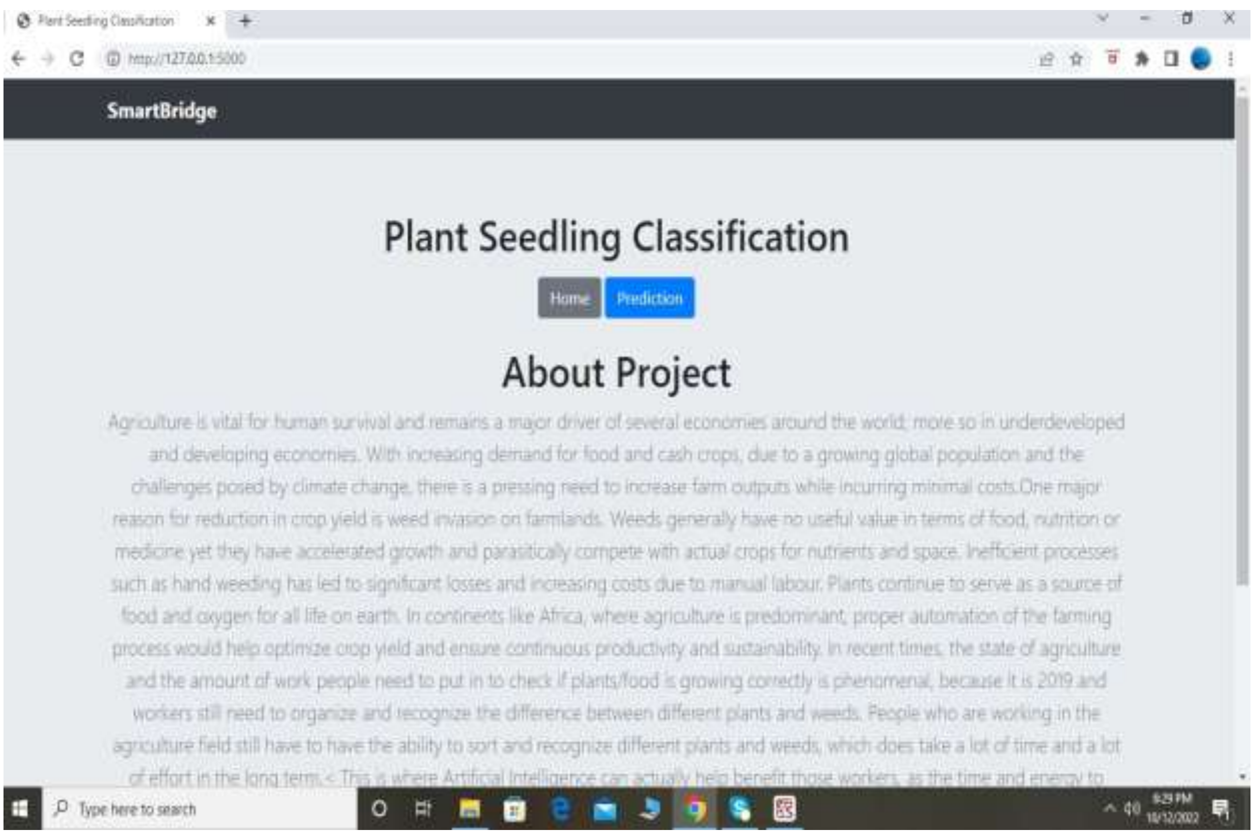
3.1 Blockdiagram:

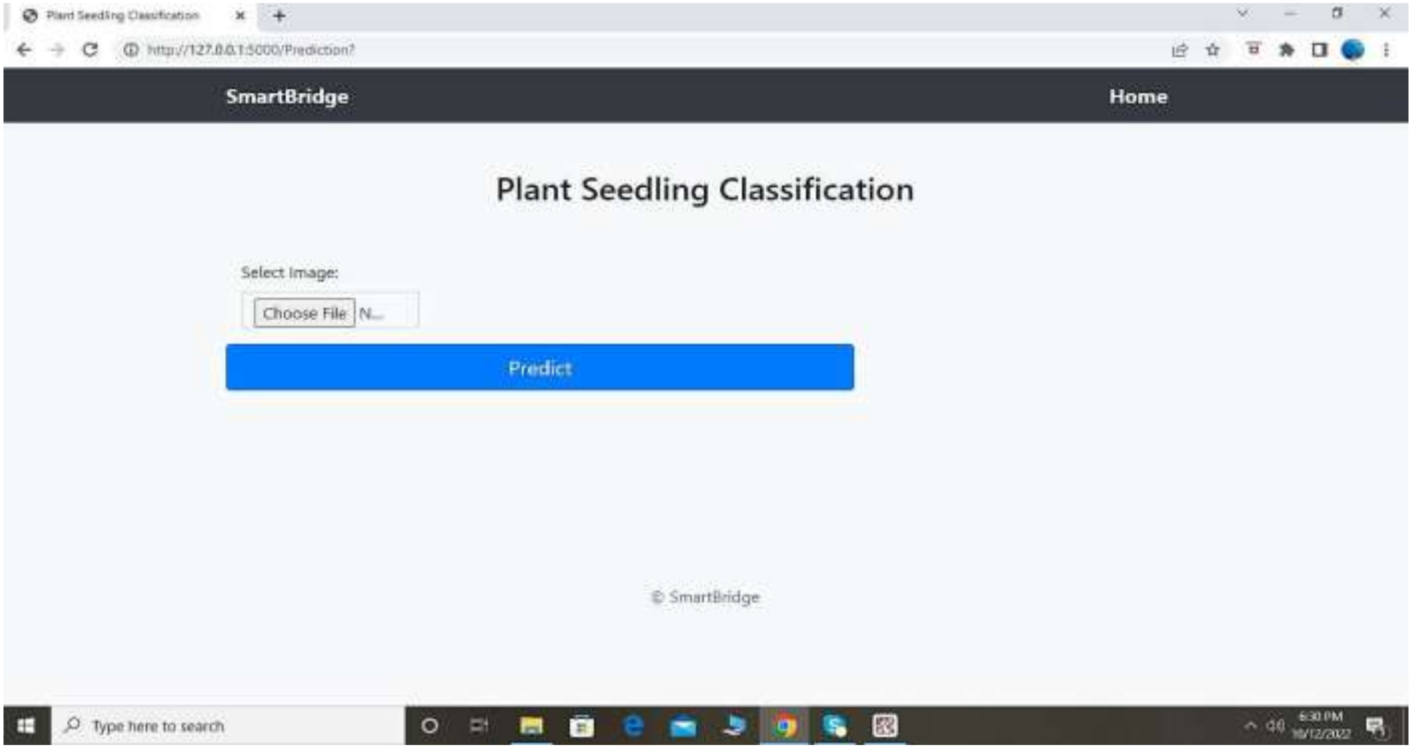


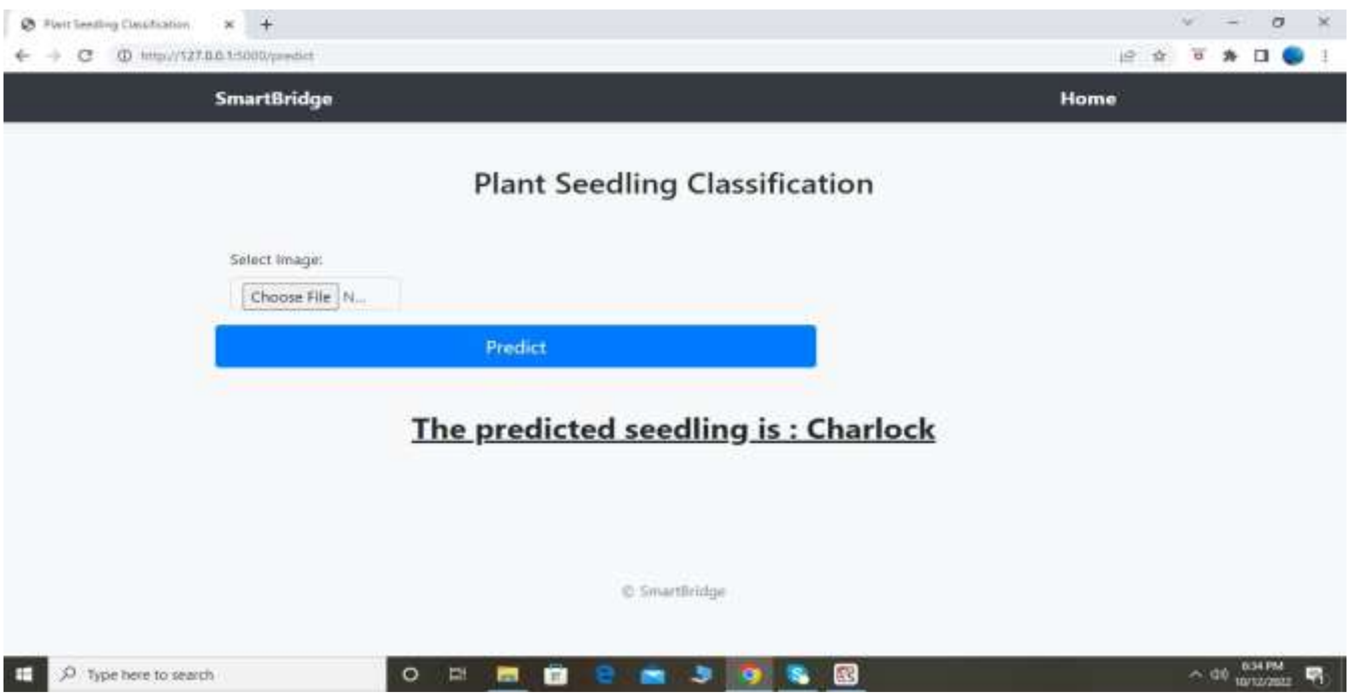
6.Flow Chart:



7. Result:




Plant Seedling Classification
http://127.0.0.1:5000/Prediction?
SmartBridge
Home
Plant Seedling Classification
Select Image:
Choose File
N...
Predict
© SmartBridge
Type here to search
6:30 PM
10/12/2022


SmartBridge
Home
Plant Seedling Classification
Select Image:
Choose File
N...
Predict
The predicted seedling is : Charlock
© SmartBridge
Type here to search

8. Advantages and Disadvantages:

Advantages:

- Uniformity due to planting of homogeneous seedlings.
- Improved final field stand and more efficient harvesting.
- Losses of expensive seed are minimised with effective use of plugs.
- Reduced risk as a result of shorter exposure in the field to pests and diseases such as rodents.
- Use of seedlings usually leads to earliness in yield.
- Germination losses in harsh climatic conditions such as extreme heat and cold are overcome.
- Reduced time in the field contributes to saving of water and other resources.
- Predictable delivery times of seedlings from the nursery helps to improve planning and crop cycling.

Disadvantages:

- Labour cost is high.
- Seedlings are exposed to possible injury during handling.
- Plants tend to grow more slowly than direct seeding because of recovery time after transplanting,

9. Applications:

1)Agriculture Sector

2)Nursary

3)Seed generating office

10. Conclusion:

CNN is used to identify, scaling, translation and other forms of images. This project is plant seedlings classification. While doing this project we come to know about various concepts like deep learning, machine learning, CNN i.e. convolution neural network etc. The code had been successfully implemented using CNN in Jupyter and Python IDEs. We also come to know about how plant seedlings are classified using various machine learning tools specially CNN. At output, we got one .csv file which comprises of image file and species which we call it as seedlings classification. In future works, we can detect disease on identified plant seedling. We can classify herbal plants. To increase understanding of details of target object, further research is needed. Another dimension is to perform new experiments when more public datasets become available.

11. Future Scope:

I) Genetic manipulation of population by increasing the frequency of desirable alleles in cross pollinated crops and introducing male sterile in self pollinated crops like wheat and Rice.

II) Intensive breeding of pulses and oil seed crops as it was done in cereals and other crops.

III) Proper breeding methods with improved crop management practises.

IV) Use of heritability methods with improved crop management practises.

V) Development of improved high yielding varieties of vegetable and seed crops.

VI) Quality Improvement in Oil seed and Vegetables.

VII) Use of transgenic plants as a medicine. E.g. Potato.

VIII) Development of varieties which are desirable for mechanical threshing and cultivation.

12. Bibliography

https://www.kaggle.com/c/plant-seedlings-classification

https://edubirdie.com/examples/the-classification-of-plant-seedling/

https://ieeexplore.ieee.org/document/9480654

Appendix:

Source Code

```
# importing the necessary dependencies
from flask import Flask,request,render_template
import numpy as np
import pandas as pd
from tensorflow.keras.preprocessing import image
from tensorflow.keras.applications.xception import Xception, preprocess_input
from tensorflow.keras.models import load_model
import pickle
import os
app = Flask(__name__)
model = load_model("seedling_exception.h5")
@app.route('/')# route to display the home page
def home():
return render_template('index.html') #rendering the home page
@app.route('/Prediction')
def prediction(): # route which will take you to the prediction page
return render_template('predict.html')
@app.route('/Home',)
def my_home():
return render_template('index.html')
@app.route('/predict',methods=["POST","GET"])# route to show the predictions in a web UI
def upload():
if request.method == 'POST':
f = request.files['img']
print("current path")
```

```
basepath = os.path.dirname("__file__")
print("current path", basepath)
filepath = os.path.join(basepath,'uploads',f.filename)
print("upload folder is ", filepath)
f.save(filepath)
img = image.load_img(filepath,target_size = (224,224))
  x = image.img_to_array(img)
  x = np.expand_dims(x,axis =0)
  x=preprocess_input(x)
   #a= x.shape

preds = np.argmax(model.predict(x), axis=1)
#print("prediction",preds)
index = ['Black-grass','Charlock','Cleavers','Common Chickweed','Common wheat','Fat Hen','Loose Silky-bent','Maize','Scentless Mayweed','Shepherds Purse','Small-flowered Cranesbill','Sugar beet']
text = "The predicted seedling is : " + str(index[preds[0]])
#return str(text)
return render_template("predict.html",z=text)
if __name__=="__main__":
app.run(debug=False)
```