**HUMAN IMAGE CLASSIFICATION ON MACHINE LEARNING WITH PYTHON**

*Summer Internship Report Submitted in partial fulfillment*

*of the requirement for undergraduate degree of*

**Bachelor of Technology**

**In**

**COMPUTER SCIENCE ENGINEERING**

**By**

**CHINTALA SHALINI (221810306012)**

*Under the Guidance of*

**Mrs. S. Mounika**

****

**Department of Computer Science and Engineering**

**GITAM School of Technology**

**GITAM Deemed to be University**

**Hyderabad Campus -502329**

# **DECLARATION**

# 

# I submit this Summer Internship entitled "HUMAN IMAGE CLASSIFICATION” to GITAM School of Technology, GITAM Deemed to be University, Hyderabad campus in partial fulfillment of the requirements for the award of the degree of “Bachelor of Technology” in **“Computer Science Engineering”**

# I declare that it was carried out independently by me under the guidance of

# Mrs. S Mounika, Assistant Professor-GITAM School of Technology, GITAM Deemed to be University, Hyderabad, India.

# 

# The results embodied in this report have not been submitted to any other University or Institute for the award of any degree or diploma.

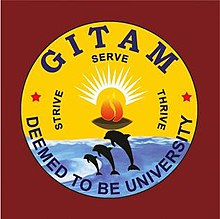
# 

# PLACE: HYDERABAD

# DATE: 25-10-2021

CHINTALA SHALINI

221810306012



**Department of Computer Science and Engineering**

**GITAM (Deemed to be University)**

**Rudraram Mandal, Sanga Reddy district, Patancheruvu, Hyderabad, Telangana 502329**

**CERTIFICATE**

This is to certify that the Summer Internship Report entitled “HUMAN IMAGE CLASSIFICATION” by using python with machine learning is being submitted by shalini chintala in partial fulfillment of the requirement for the award of Bachelor of Technology in “COMPUTER SCIENCE ENGINEERING” at GITAM University, Hyderabad.

It is faithful record work carried out by him at the Computer Science Engineering Department, GITAM School of Technology, GITAM Deemed to be University, Hyderabad Campus under my guidance and supervision.

**Mrs. S Mounika**   **Prof. S. Phani Kumar**

Assistant Professor Professor and HOD

Department of CSE Department of CSE

**CERTIFICATE OF COMPLETION OF SUMMER INTERNSHIP**



**INTERNSHIP CERTIIFICATE**



# 

# **ACKNOWLEDGEMENT**

# 

# First, I would like to thank Mr. Vungarala V Subrahmanyam, the CEO of Verzeo Edu Tech, The Technical Program Manager for giving me the opportunity to do an internship within the organization.

# I also would like to thank all the people that worked along with me with their patience and openness. They created an enjoyable working environment. It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

# I am highly indebted to Principal N. Seetha Ramaiah, for the facilities provided to accomplish this internship. I would like to thank the Head of the Department Phani Kumar S for his constructive criticism throughout my internship.

I would like to thank Dr. B. Rajesh, College internship coordinator Sri L.V Satyanarayana internship coordinator Department of CSE for their support and advice to get and complete internship in the above said organization. I hereby wish to express a deep sense of gratitude to Mrs. S Mounika, Assistant Professor, Department of Computer Science and Engineering, for their esteemed guidance, moral support and invaluable advice provided by them for the success of this internship.

I am extremely grateful to my department staff members and friends who helped me in successful completion of this internship.

# .

# SHALINI CHINTALA 221810306012

# **ABSTRACT**

# 

# **TABLE OF CONTENTS**

# **Introduction.**............................................................ 6

# 1.1. Context of Project............................................ 7

# 1.2. Background Study............................................ 8

# 1.3. Project Planning............................................... 9

# 1.4. Purpose……………………………………… 10

# 2. **System Design**

# 2.1. Design.............................................................. 11

# 2.2. User Characteristics......................................... 12

# 2.3 System Information…………………………. 12

# 2.4 System Analysis………………………………12

# 2.5 Feasibility Analysis……………………………13

3. **Hardware and Software Requirement**

3.1. Hardware Required.............................................14

# 3.2. Software Required..............................................14

# 4. **Implementing Tools for the Project**

4.1. Tools………………………………………….15

# 5. **Project Database**

# 5.1 Database Design..................................................18

# 6. **Project Model View**...................................................18

# 7. **Software Testing**

# 7.1 Why software testing is needed………………. 23

7.2 Testing Strategy………………………………. 24

7.3 White box testing………………………………24

7.4 Black box testing ……………………………. 24

# 8. **Conclusion**

# 8.1 Conclusion ......................................................... 25

8.2 Future aspect……………………………………25

# **INTRODUCTION**

Image Classification is one of the most fundamental tasks in computer vision.

And for a reason

Image classification has revolutionized and propelled technological advancements in the AI field, from the automobile industry to medical analysis and automated perception in robots.

But how does image classification actually work, and what are its benefits and limitations?

Image classification (or Image recognition) is a subdomain of computer vision in which an algorithm looks at an image and assigns it a tag from a collection of predefined tags or categories that it has been trained on.

Vision is responsible for 80-85 percent of our perception of the world, and we, as human beings, trivially perform classification daily on whatever data we come across.

Therefore, emulating a classification task with the help of neural networks is one of the first uses of computer vision that researchers thought about. Let's explore it in more detail.

Types of image classification techniques:

A variety of algorithms can solve image classification as a task. Broadly, we can classify them into supervised and unsupervised algorithms.

Supervised classification

In supervised classification, the classification algorithm is trained on a set of images along with their corresponding labels.

This helps the algorithm predict the correct tag for images that it has not yet seen with the help of information it has extracted from labeled sample data.

During training, the algorithm extracts feature from the image matrix as data that is important enough to be processed. These features represent the image in a lower-dimensional feature space and allow the classifier to classify images based on them.

During the evaluation, features of test images are collected and classified again with the help of the network, which now knows the typical features of every class it has been trained with.

Popular supervised methods of classification based on machine learning algorithms include:

Support Vector Machines

Decision Trees

K Nearest Neighbors

Popular neural networks used for Supervised Image Classification include Alex Net, ResNet, Dense Net, and Inception.

**1.1 CONTEXT OF PROJECT**

The ‘Online E-commerce Web application’ Services department strives to provide solutions to develop and transfer easy and efficient ways in the digital age and to help reduce the human pressure and time. To help support shop collections, digital initiatives, and external partner institution digital projects, It provides services that include the digitization of analog objects, metadata management, digital preservation, and discovery and access of digital collections.

“Shop Management System '' is a web application written for all operating systems, designed to help users maintain and organize shops virtually. This software is easy to use for both beginners and advanced users. It features a familiar and well thought out, an attractive user interface, combined with strong searching Insertion and reporting capabilities. The report generation facility of the shop system helps to get a good idea of which are the various items brought by the members, making it possible to get the product easily.

The ‘Online E-commerce Web application’ Services department strives to provide solutions to develop and transfer easy and efficient ways in the digital age and to help reduce the human pressure and time. To help support shop collections, digital initiatives, and external partner institution digital projects, It provides services that include the digitization of analog objects, metadata management, digital preservation, and discovery and access of digital collections.

“Shop Management System” is a web application written for all operating systems, designed to help users maintain and organize shops virtually. This software is easy to use for both beginners and advanced users. It features a familiar and well thought out, an attractive user interface, combined with strong searching Insertion and reporting capabilities. The report generation facility of the shop system helps to get a good idea of which are the various items brought by the members, making it possible to get the product easily.

**1.2 BACKGROUND STUDY**

We assess the tendency of state-of-the-art object recognition models to depend on signals from image backgrounds. We create a toolkit for disentangling foreground and background signal on ImageNet images, and find that (a) models can achieve non-trivial accuracy by relying on the background alone, (b) models often misclassify images even in the presence of correctly classified foregrounds--up to 87.5% of the time with adversarial chosen backgrounds, and (c) more accurate models tend to depend on backgrounds less. Our analysis of backgrounds brings us closer to understanding which correlations machine learning models use, and how they determine models' out of distribution performance.

**1.3 PROJECT PLANNING**

Project planning is part of project management, which relates to the use of schedules such as Gantt charts to plan and subsequently report progress within the project environment. Initially, the project scope is defined and the appropriate methods for completing the project are determined. Following this step, the durations for the various tasks necessary to complete the work are listed and grouped into a work breakdown structure.

The logical dependencies between tasks are defined using an activity network diagram that enables identification of the critical path. Float or slack time in the schedule can be calculated using project management software. Then the necessary resources can be estimated and costs for each activity can be allocated to each resource, giving the total project cost. At this stage, the project plan may be optimized to achieve the appropriate balance between resource usage and project duration to comply with the project objectives. Once established and agreed, the plan becomes what is known as the baseline. Progress will be measured against the baseline throughout the life of the project

**1.4 PURPOSES**

# The objective of image classification is **to identify and portray**, as a unique gray level (or color), the features occurring in an image in terms of the object or type of land cover these features actually represent on the ground. Image classification is perhaps the most important part of digital image analysis.

1. **SYSTEM DESIGN**

**2.1 DESIGN**

Prior to 2017, most renditions of neural network models were coded in a batch scripting style. As AI researchers and experienced software engineers became increasingly involved in research and design, we started to see a shift in the coding of models that reflected software engineering principles for reuse and design patterns.

The introduction of design patterns also helped advance convolutional neural networks (as well as other network architectures) by aiding other researchers in understanding and reproducing a model's architecture.

A procedural style for reuse was one of the earliest versions of using design patterns for neural network models. Understanding the architecture of the procedural reuse design pattern is crucial if you are going to apply it to any model you are building. Once you see how the parts work, individually and together, you can start working with the code that builds these parts -- code that is available free for downloading. I show how the procedural reuse design pattern makes it easier to reproduce model components by applying it to several formerly state-of-the-art models including VGG, ResNet, ResNeXt, Inception, DenseNet and SqueezeNet. This offers both a deeper understanding of how these models work, as well as practical experience reproducing them.



**2.2 USER CHARACTERISTICS**

Admin The administrator has all the rights to access the system. He is the one who has all rights to view the members and product details, modify those details. He can add various products based on the category. He can also set the available quantity of a product and its reasonable price. Also, he can also set discounts on various occasions. Admin can also view the details of a member. The admin has the power to generate the scratch card so that users can also use the recharge card to buy various products.

Users The user can log in to the system by using his specific email and password. Users can view the products and order the products according to their own needs. He can view his profile and update his details. He can update his personal information by logging into the system. Users can find various products by using the search option easily. update his details. He can update his personal information by logging into the system. Users can find various products by using the search option easily.

**2.3 SYSTEM INFORMATION**

This system is an automated Shop Management System. Through the software, users can add members, add products, search products, update information, edit information, and buy the product in quick time. The system has the following advantages: User friendly interface Fast access to database Search facility Look and Feel Environment

**2.4 SYSTEM ANALYSIS**

System Analysis refers to the process of examining a situation with the intent of improving it through better procedures and methods. System Analysis is the process of planning a new system to either replace or complement an existing system. But before any planning is done the old system must be thoroughly understood and the requirements determined. System analysis is therefore the process of gathering and interpreting facts, diagnosing problems and using the information to recommend improvements in the system. System analysis is conducted with the following objectives in mind: Evaluate the system concept for feasibility. Perform economic and technical analysis. Allocate functions to hardware, software people, database and other system elements. Establish cost and schedule constraints. Create a system definition that forms the foundation for all the subsequent engineering work.

**2.5 FEASIBILITY ANALYSIS**

Whatever we think need not be feasible. It is wise to think about the feasibility of any problem we undertake. Feasibility is the study of impact, which happens in the organization by the development of a system. The impact can be either positive or negative. When the positives nominate the negatives, then the system is considered feasible. Here the feasibility study can be performed in two ways such as technical feasibility and Economical Feasibility.

**2.5.1 TECHNICAL FEASIBILITY**

It is technically feasible, since there will not be much difficulty in getting required resources for the development and maintaining the system as well. All the resources needed for the development of the software as well as the maintenance.

Data Collection –

1. Scrape or Collect Human Images containing images of both male and female

2. Data Cleaning - Remove unwanted data from the dataset.

3. Data Preprocessing - Process the data as per the model requirement

4. Model Building - Build a model over the dataset that will classify the image of Indian origin Human from others and classify and save 1000 Indian origin Human Images. The model should work for all Skin tones (Fair, Mild, and Dark) and for both males and females

**2.5.2 ECONOMICAL FEASIBILITY**

Development of this application is highly economically feasible. The organization needed not spend much money for the development of the system already available. The only thing to be done is making an environment for development with effective supervision. If we are doing so, we can attain the maximum usability of the corresponding resources Even after the development, the organization will not be in a condition to invest more in the organization. Therefore, the system is economically feasible.

**3. HARDWARE AND SOFTWARE REQUIREMENTS**

**3.1 HARDWARE REQUIRED**

**Processor**: Pentium IV or Above

**RAM**: 2GB or above

**Hard Disk**: 50GB or above

**Input Devices**: Keyboard, Mouse

**Output Devices**: Monitor

**3.2 SOFTWARE REQUIRED**

**Operating System**: Linux, Ubuntu, Mac, Windows XP, 7, 8, 8.1, 10

**Frontend**: HTML, CSS, Bootstrap, JavaScript, React

**Backend**: PYTHON and others.

**Database: SIPI Image Database**

**Localhost**: python, Anaconda.

1. **IMPLEMENTING TOOLS FOR THE PROJECT**
   1. **TOOLS**

* [**TensorFlow**](https://www.tensorflow.org/): An open-source platform for the implementation, training, and deployment of machine learning models.
* [**Kera’s**](https://keras.io/): An open-source library used for the implementation of neural network architectures that run on both CPUs and GPUs.
* [**Pandas**](https://pandas.pydata.org/): Data analysis and modification library.
* [**Matplotlib**](https://matplotlib.org/): Tool utilized to create visualization plots in Python such as charts, graphs and more
* [**NumPy**](https://numpy.org/): Enables several mathematical computations and operations of array data structures.

## **TENSORFLOW**

TensorFlow is an end-to-end open-source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. Easy model building.

Robust ML production anywhere.

Powerful experimentation for research.

**KERAS**

## Iterate at the speed of thought. Kera’s is the most used deep learning framework among top-5 winning teams on [Kaggle](https://www.kaggle.com/). Because Kera’s makes it easier to run new experiments, it empowers you to try more ideas than your competition, faster. And this is how you win.

**PANDAS**

pandas is a fiscally sponsored project of [NumFOCUS.](https://numfocus.org/)

# **MATPLOTLIB**

# Visualization with Python.

# Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in **python.**

**NUMPY**

Nearly every scientist working in Python draws on the power of NumPy.

NumPy brings the computational power of languages like C and Fortran to Python, a language much easier to learn and use. With this power comes simplicity: a solution in NumPy is often clear and elegant.

1. **PROJECT DATABASE**

**5.1 DATABASE DESIGN**

Database is critical for all businesses. A good database does not allow any form of anomalies and stores only relevant information in an ordered manner. If a database has anomalies, it is affecting the efficiency and data integrity. For example, delete anomalies arise upon the deletion of a row which also forces other useful data to be lost. As such, the tables need to be normalized. This fulfils the last objective of ensuring data are accurate and retrieved correctly. Database files are the key source of information into the system. It is the process of designing database files, which are the key source of information to the system. The files should be properly designed and planned for collection, accumulation, editing and retrieving the required information.

1. **PROJECT VIEW**





1. **SOFTWARE TESTING**

**7.1 Why Software Testing is Needed? Tool-bars work properly?**

Are all menu functions and pull-down sub functions properly listed? Is it possible to invoke each menu function using a logical assumption that if all parts of the system are correct, the goal will be successfully achieved? In adequate testing or non-testing will lead to errors that may appear a few months later. Testing represents an interesting anomaly for the software engineer. During earlier software engineering activities, the engineer attempts to build software from an abstract concept to a tangible product. Now comes testing. The engineer creates a series of test cases that are intended to “demolish” the software that has been built. In fact, testing is the one step in the software process that could be viewed (psychologically, at least) as destructive rather than constructive. Testing requires that the developer discard preconceived notions of the “correctness” of software just developed and overcome a conflict of interest that occurs when errors are uncovered. If testing is conducted successfully (according to the objectives stated previously) it will uncover errors in the software. As a secondary benefit, testing demonstrates that software functions appear to be working according to specification, that behavioral and performance requirements appear to have been met. In addition, data collected as testing is conducted provides a good indication of software reliability and some indication of software quality as a whole. But testing cannot show the absence of errors and defects, it can show only that software errors and defects are present. It is important to keep this (rather gloomy) statement in mind as testing is being conducted.

**7.2 TESTING STRATEGY**

There are types of testing that we implement. They are as follows: While deciding on the focus of testing activities, study project priorities. For example, for an online system, pay more attention to response time. Spend more time on the features used frequently. Decide on the effort required for testing based on the usage of the system. If the system is to be used by a large number of users, evaluate the impact on users due to a system failure before deciding on the effort. This creates two problems: Time delay between the cause and appearance of the problem. The effect of the system errors on files and records within the system. The purpose of the system testing is to consider all the likely variations to which it will be suggested and push the systems to limits. The testing process focuses on the logical intervals of the software ensuring that all statements have been tested and on the functional interval is conducting tests to uncover errors and ensure that defined input will produce actual results that agree with the required results. Program level testing, modules level testing integrated and carried out. There are two major types of testing: White Box Testing, Black Box Testing.

**7.3 WHITE BOX TESTING**

White box sometimes called “Glass box testing” is a test case design that uses the control structure of the procedural design to drive the test case. Using white box testing methods, the following tests were made on the system.

a) All independent paths within a module have been exercised once. In our system, ensuring that case was selected and executed checked all case structures. The bugs that were prevailing in some parts of the code were fixed.

b) All logical decisions were checked for the truth and falsity of the values.

**7.4 BLACK BOX TETSING**

Black box testing focuses on the functional requirements of the software. This is black box testing enables software engineering to derive a set of input conditions that will fully exercise all functional requirements for a program. Black box testing is not an alternative to white box testing; rather it is a complementary approach that is likely to uncover a different class of errors that white box methods like. Interface errors. Performance in data structure. Performance errors. Initializing and termination errors.

**RAW DATA CLEANING CODE:**

{

"cells": [

{

"cell\_type": "code",

"execution\_count": 15,

"metadata": {},

"outputs": [],

"source": [

"import cv2\n",

"import pandas as pd\n",

"import numpy as np\n",

"import matplotlib.pyplot as plt\n",

"%matplotlib inline"

]

},

{

"cell\_type": "code",

"execution\_count": 16,

"metadata": {},

"outputs": [],

"source": [

"face\_cascade=cv2.CascadeClassifier('./haarcascades/haarcascade\_frontalface\_default.xml')\n",

"eye\_cascade = cv2.CascadeClassifier('./haarcascades/haarcascade\_eye.xml')"

]

},

{

"cell\_type": "code",

"execution\_count": 17,

"metadata": {},

"outputs": [],

"source": [

"def get\_cropped\_image\_if\_2\_eyes(image\_path):\n",

" img = cv2.imread(image\_path)\n",

" gray = cv2.cvtColor(img,cv2.COLOR\_BGR2GRAY)\n",

" faces = face\_cascade.detectMultiScale(gray,1.3,5)\n",

" for (x,y,w,h) in faces:\n",

" roi\_gray = gray[y:y+h, x:x+w]\n",

" roi\_color = img[y:y+h, x:x+w]\n",

" eyes = eye\_cascade.detectMultiScale(roi\_gray)\n",

" if len(eyes)>=2:\n",

" return roi\_color"

]

},

{

"cell\_type": "code",

"execution\_count": 22,

"metadata": {},

"outputs": [],

"source": [

"path\_to\_data='./dataset/'\n",

"path\_to\_cr\_data='./cleaned/'"

]

},

{

"cell\_type": "code",

"execution\_count": 23,

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"['./dataset/Indian', './dataset/Non Indian', './dataset/test']"

]

},

"execution\_count": 23,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"import os\n",

"img\_dirs=[]\n",

"for entry in os.scandir(path\_to\_data):\n",

" if entry.is\_dir():\n",

" img\_dirs.append(entry.path)\n",

"img\_dirs"

]

},

{

"cell\_type": "code",

"execution\_count": 24,

"metadata": {},

"outputs": [],

"source": [

"import shutil\n",

"if os.path.exists(path\_to\_cr\_data):\n",

" shutil.rmtree(path\_to\_cr\_data)\n",

"os.mkdir(path\_to\_cr\_data)"

]

},

{

"cell\_type": "code",

"execution\_count": 25,

"metadata": {},

"outputs": [

{

"name": "stdout",

"output\_type": "stream",

"text": [

"Indian\n",

"Non Indian\n",

"test\n"

]

}

],

"source": [

"cropped\_image\_dirs=[]\n",

"file\_name\_dirs={}\n",

"\n",

"\n",

"for img\_dir in img\_dirs:\n",

" count = 1\n",

" file\_name=img\_dir.split('/')[-1]\n",

" print(file\_name)\n",

" \n",

" for entry in os.scandir(img\_dir):\n",

" \n",

" roi\_color = get\_cropped\_image\_if\_2\_eyes(entry.path)\n",

" if roi\_color is not None:\n",

" cropped\_folder = path\_to\_cr\_data+file\_name\n",

" \n",

" if not os.path.exists(cropped\_folder):\n",

" os.makedirs(cropped\_folder)\n",

" cropped\_image\_dirs.append(cropped\_folder)\n",

" \n",

" cropped\_file\_name=file\_name+str(count)+'.jpg'\n",

" cropped\_file\_path=cropped\_folder+'/'+cropped\_file\_name\n",

" \n",

" cv2.imwrite(cropped\_file\_path, roi\_color)\n",

" #file\_name\_dirs[file\_name].append(cropped\_file\_path)\n",

" count+=1"

]

},

{

"cell\_type": "code",

"execution\_count": null,

"metadata": {},

"outputs": [],

"source": []

}

],

"metadata": {

"kernelspec": {

"display\_name": "Python 3",

"language": "python",

"name": "python3"

},

"language\_info": {

"codemirror\_mode": {

"name": "ipython",

"version": 3

},

"file\_extension": ".py",

"mimetype": "text/x-python",

"name": "python",

"nbconvert\_exporter": "python",

"pygments\_lexer": "ipython3",

"version": "3.9.2"

}

},

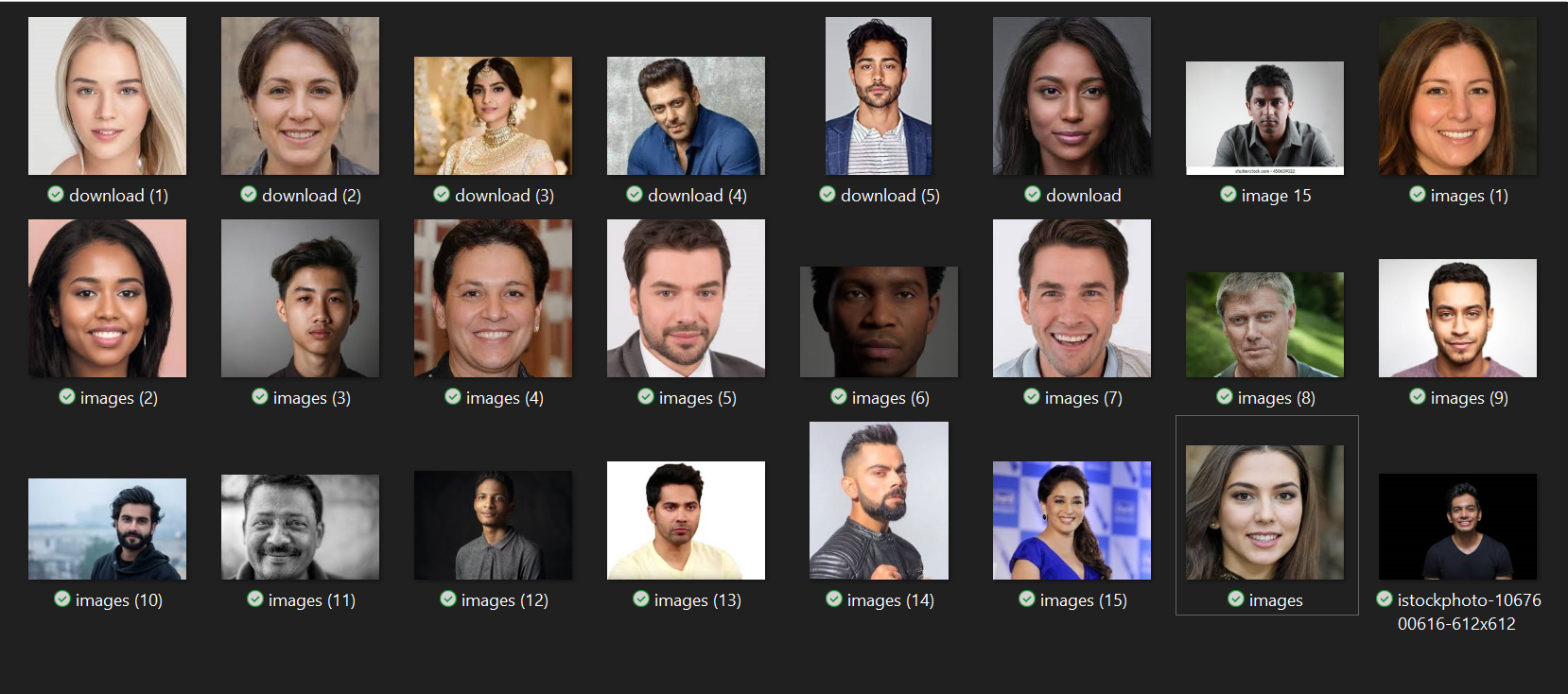
"nbformat": 4,

"nbformat\_minor": 4

}

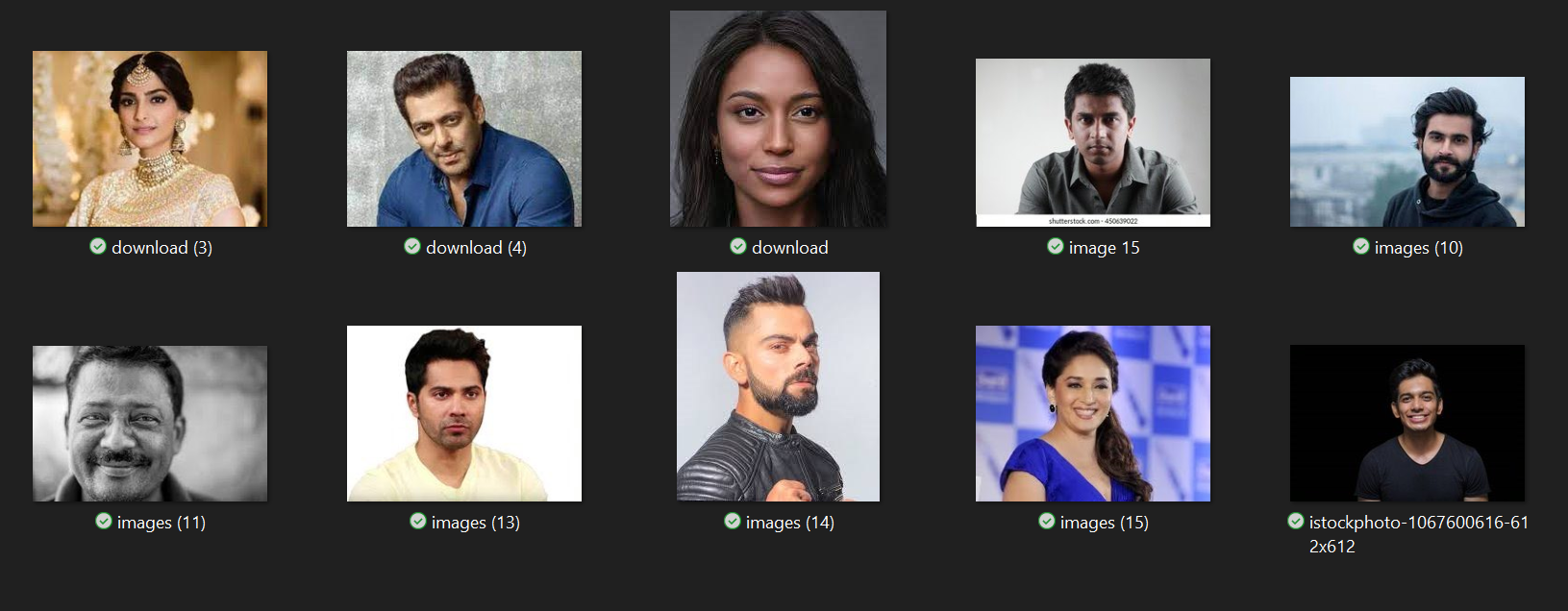
**DATA SET: INPUT**

With Indian and non-Indian faces:



**OUTPUT:**

Only Indian faces



1. **CONCLUSION & FUTURE ENHACEMENT**

**8.1 CONCLUSION**

The task of classifying news manually requires in-depth knowledge of the domain and expertise to identify anomalies in the text. In this research, we discussed the problem of classifying fake news articles using machine learning models. The data we used in our work is collected from the World Wide Web and contains news articles from various domains to cover most of the news rather than specifically classifying political news

**8.2 FUTURE ASPECT**

The project has a very vast scope in future. The project can be implemented on intranet in future. Project can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion. With the proposed software of database Space Manager ready and fully functional the client is now able to manage and hence run the entire work in a much better, accurate and error free manner. The following are the future scope for the project. Should be added payment gateway Can be added inventory management system Can be added multiple branches Can be added multilingual to this site and many features can be added this project to make it more robust.