WhatsApp Backup's Images Segregation

Prarthana Gupta[1], Aditya Bohade[2], Om Tulaskar[3], Suvarna Aranjo[4]
[1,2,3,4]*Department of Information Technology, Xavier Institute of Engineering, Mahim, Mumbai, India*

***Abstract*— One of the most downloaded apps of all time, WhatsApp allows users to send and receive instant messages from their smartphones. It uses online resources to communicate with individuals or groups using a variety of media types. Its growth has attracted the curiosity of academics who are curious about the impact of WhatsApp on the users' personal and social lives. The proposed model was developed using the Python language's Keras module and the deep learning concept of Convolutional Neural Networks (CNNs). After identifying an image's classification, the corresponding procedure is carried out. CNN has been very successful in the field of image identification. It's a tedious process to go through each WhatsApp folder and manually delete each image that symbolises a study aid, pamphlet, etc. Since then, the research team has developed a machine learning model to locate and retrieve the images from WhatsApp's image folder. Initially, the idea of instructing a computer or gadget to perform a task such as image classification piqued our interest. In addition, the concept can be used in a wide variety of practical contexts. As a result of these considerations, we have settled on the topic of image classification for our research.**

***Keywords*— WhatsApp, Deep Learning, Machine Learning, Convolutional Neural Net- work, Python, Keras Library, TensorFlow.**

I.INTRODUCTION

The advent of mobile technology has resulted in profound social and political shifts. You can hardly go anywhere these days without encountering a smartphone app. Most people have nothing but praise for the plethora of apps they use on their smartphones on a regular basis. There are a lot of people who couldn't do their daily jobs without these apps.

During exam time, students frequently use WhatsApp to send each other photos of their study materials, which can quickly fill up a phone's storage. Eliminating the study photos from the WhatsApp folder requires tediously selecting each one. This is because there could be a large number of other images in the WhatsApp folder, such as promotional materials, screenshots, etc.

Having over 2 billion users in over 180 countries, WhatsApp is a popular cross-platform messaging and calling service. According to the company, only the sender and the intended recipient can decipher the message sent through WhatsApp because it is end-to-end encrypted (E2EE). When communications are terminated at a third-party server and meddling-in-the-middle is trivial, the need for E2EE arose to prevent infrastructure and service providers from reading or tampering with user messages. [7]

Under the End-to-End Encrypted Backups initiative, WhatsApp developed and deployed an encrypted backup solution to prevent unauthorised parties, including the company itself, from accessing the contents of WhatsApp messages backed up to their infrastructure. Here, we'll go over how this project resulted in an encrypted backup solution for WhatsApp. It begins with a high-level explanation of its security architecture, then moves on to a detailed examination of its most salient protections for personal information. [7] The goal is to extract usable information from the WhatsApp backup while filtering out irrelevant messages and photos using natural language processing. With this, all necessary information is gathered. Now I can quickly and effortlessly sift through WhatsApp's archives.

II.RELATED WORK

Deep Learning Methods for Image Recognition is an article that focuses on the application of these methods to the field of image recognition. It demonstrates that image recognition is currently a popular research topic in the field of computer vision. Words and letters written by hand, facial features, images with medical significance, etc. Many of the algorithms used in deep learning are based on the ability to recognise images. It employs three crucial algorithms, including: 1. Two additional neural

networks are the CNN and the RNN. Deep neural networks are used for forecasting and image classification. In order to extract features and categorise images, convolutional neural networks are employed, while recurrent neural networks are used to process event sequences, time series, language, models, etc. Due to their multilayered nature and the fact that they must specifically deal with 2D input data, CNNs are integrated into deep learning via artificial neural networks. Multiple input and output hidden layers are possible in a CNN. Common types of hidden layers include convolutional, grouping, normalisation, and fully connected layers. When compared to other methods, CNN's tremendous success can be attributed to the fact that each layer of neurons is connected to each other via a fully connected layer. Multiple types of classification are possible, including binary, multiclass, and multilabel.

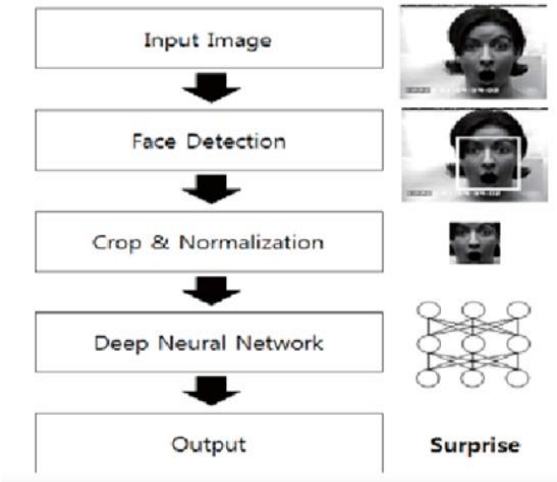

FIG. 01. DEEP LEARNING FOR IMAGE CLASSIFICATION

III.SOFTWARE REQUIREMENTS

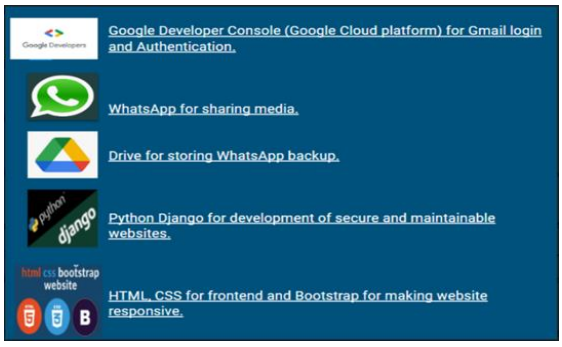

FIG.02: TOOLS.

IV. METHODOLOGY

1. Create A CNN model: A CNN model is created using the python Keras library and using the activation functions of ReLU and Softmax.
2. Data Augmentation and Training: Data Augmentation Technique is applied to the dataset to overcome the problem of fewer data and then using this augmented data, the model is trained.
3. Mount Internal Storage: Internal storage of the mobile can be mounted into the system(laptop) using the WebDAV server using which the exact path of the WhatsApp Images can be retrieved

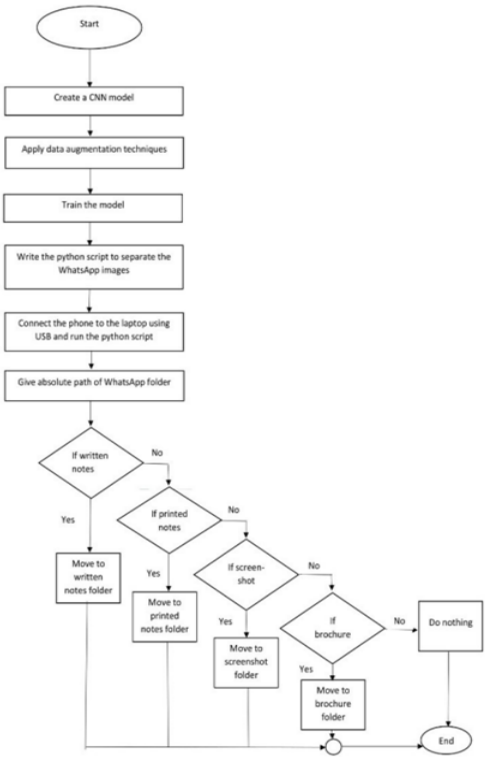

FIG. 03. METHODOLOGY [3]

Python Script: A python script is written which takes the absolute path of the WhatsApp folder in our phone and uses the CNN model to classify the images into different categories. A mobile phone is connected to the laptop and the python script is run. Then four new folders called “handwritten notes”, “printed notes”, “screenshots” and “brochures” are created in our mobile phone and all the images in the WhatsApp folder are moved to the folders according to their classification. Other images are left untouched in the WhatsApp folder itself. [3]

Data collection: The data collected consisted of five types of images, namely handwritten note images, printed note images, screenshots, brochures and other categories, including all other photographs (images without notes, etc.). To train the model, an image dataset of research materials was used. The dataset contains about 900 1000 images and its size is about 1.5 GB. Photos taken from Format.jpg Each image varies in size from 20kb to 5Mb.

FIG Non-Notes Images

FIG: Brochure Image

FIG: Screenshot Image

FIG: Handwritten-Notes Image

FIG: Printed Notes Image

FIG. 04. SAMPLE DATA SET [3]

Constructing the model: The model is built with an input layer, an output layer and three hidden layers. Relu activation function is used for input as well as hidden layers, while the output layers use the SoftMax activation function.

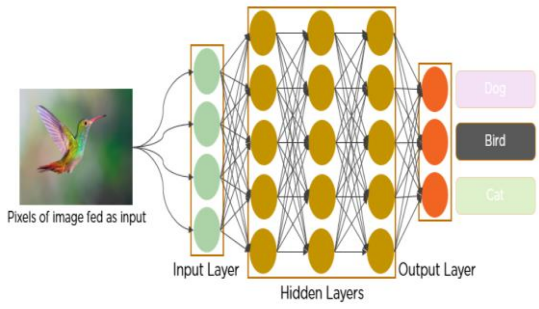

Fig. 05: CNN Model [12]

Training the Model: The dataset is first populated and then loaded. It is then trained for a given period of time and group size. Run python script: The python script containing the trained model is run when the phone is connected to the laptop. WhatsApp path as input, then split the images into different folders.

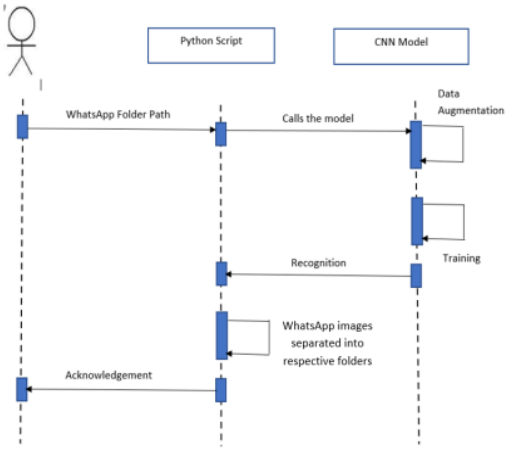

Fig. 06. Process Flow [3]

When the python script is run, it requests the input for the absolute path of the WhatsApp folder, then with the help of the trained CNN model, the images get classified and moved into the different folders.[3]

ALGORITHM:

1.Start the process

2. The phone is connected to the laptop using USB and the internal storage of the phone is mounted to the laptop for the access of the WhatsApp path.

3.The model which is trained using the dataset containing five classes of images is called inside the python script and the script is run in the CMD prompt.

4. When python script is run, it asks for the absolute path of the WhatsApp Folder which is to be provided.

5. As the script runs, the images gets classified one by one.

6. The model will separate the images according to which class they belong to, by using the multiclass classification.

7. Four new folders will get created in the phone that are “handwritten-notes", "printed-notes", "brochures” and “screenshots”.

8. As the images will get classified, they will be moved into the respective folders one by one.

9. Non-notes images do not get moved anywhere and remain in the same WhatsApp images folder itself.

10. End the process.

V. RESULTS

Before Classification:

Before Classification, all the images in the WhatsApp
folders are present in the same place and are mixed up.

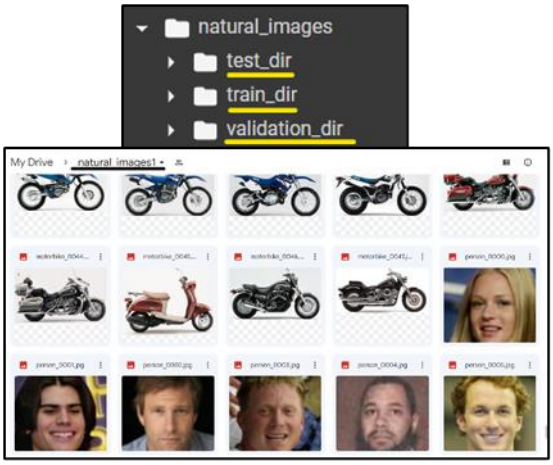

Fig. 07. Random Images Before classification

Running Script and Classifying: Then the python script which has the CNN model in it is run by giving the absolute path of the WhatsApp folder. Then the classification of the images takes place.

```
import matplotlib.pyplot as plt
from keras.preprocessing.image import ImageDataGenerator
from keras.optimizers import Adam
from tensorflow.keras.utils import img_to_array
from keras.utils import to_categorical
from sklearn.model_selection import train_test_split
import numpy as np
import pandas as pd
import random
import cv2
import os
from PIL import Image

from keras.layers import *
from keras.models import *
import keras
import tensorflow as tf
```

```
train_dir = r'/content/drive/MyDrive/natural_images/train_dir'  # Location of training images
validation_dir = r'/content/drive/MyDrive/natural_images/validation_dir' #Location of test images
test_dir =r'/content/drive/MyDrive/natural_images/test_dir' #Location of test images
```

Fig. 08. Libraries in python.

After Classification: After Classification, we can see that the images get classified into their respective folders i.e., “Good morning messages with flowers”, “pet animals”, ”transport” etc, remain in the WhatsApp folder itself.

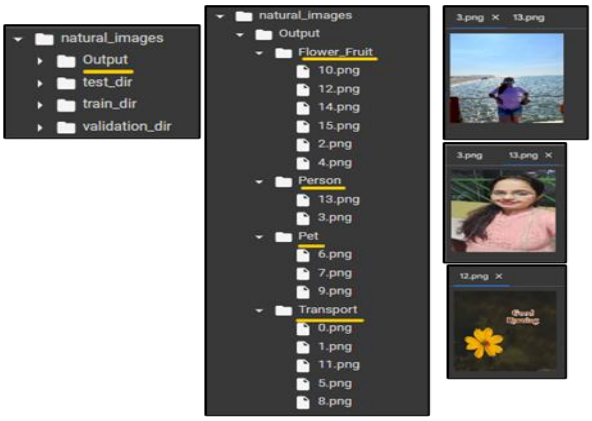

Fig.09. Images Segregated into respective folders.

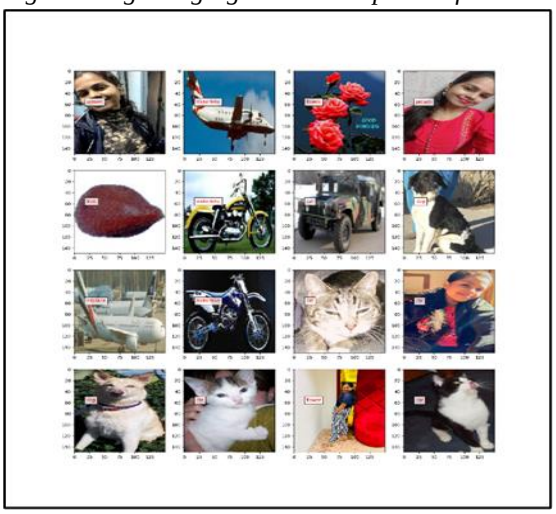

Fig. 10. Prediction done by the model

We can see here that the images get classified into separate folders even in the mobile phone.

Accuracy: The accuracy achieved for the trained model is 83.35%

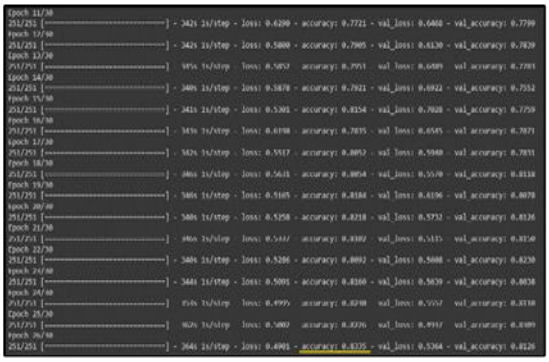

Fig. 11. Accuracy found is 83.35%

After classification, there are new folders such as “Good morning messages with flowers”, “pet animals”,” transport” etc, formed and images are transferred into their respective folders.

CONCLUSIONS

The implemented model helps us to separate the notes(meme) images and non-notes(non-meme) images and also the brochures which get filled up in our WhatsApp folder without a proper arrangement. Not only it will save a lot of time by providing easy identification of notes images but also helps in proper management of the WhatsApp media in an organized manner. By this arrangement, some of the important deadlines in the brochures can be attended.

The accuracy of the model can be increased so as to get better classification and this model can further be extended to be developed as a mobile app so that the separation can happen with the help of just the app and no USB cable and laptop will be needed.

ACKNOWLEDGMENT

We would like to thank Fr. Dr. John Rose S.J (Director of XIE) for providing us with such an environment so as to achieve goals of our project and supporting us constantly.

We express our sincere gratitude to our Honorable Principal Mr. Y.D.Venkatesh for encouragement and facilities provided to us. We would like to place on record our deep sense of gratitude to Prof. Meena Ugale , Head of Department Of Information Technology, Xavier Institute of Engineering, Mahim, Mumbai, for his generous guidance, help and useful suggestions. With deep sense of gratitude we acknowledge the guidance of our project guide Prof. Suvarna Aranjo. The time-to-time assistance and encouragement by her has played an important role in the development of our project. Helping and making sure about the project and coordinating with us at every step made the academic year smooth and efficient. We would also like to thank our entire Information Technology staff who have willingly cooperated with us in resolving our queries and providing us all the required facilities on time.

REFERENCE

[1] Survey Analysis on the usage and Impact of Whatsapp https://ieeexplore.ieee.org/document/9591866

[2] Smys, S., and Wang Haoxiang. "Na¨ıve Bayes and Entropy based Analysis and Classification of Humans and Chat Bots." Journal of ISMAC 3, no.01 (2021): 40-49 https://irojournals.com/iroismac/article/view/3/1/4

[3] CLASSIFICATION AND SEPARATION OF WHATSAPP IMAGES USING MACHINE LEARNING.https://ieeexplore.ieee.org/document /9591866

[4] Image Segmentation and its Various Techniques in Image Processing https://www.ijsr.net/get abstract.php?paper

[5] Techniques of Deep Learning for Image Recognition https://ieeexplore-ieee-org.svkm.mapmyaccess.com/document/9033628

[6] Image Classification using SVM and CNN https://ieeexplore.ieee.org/document/9132851

[7] End-to-End Encrypted Backups Security Assessment https://research.nccgroup.com/wp-content/uploads/2021/10/NCC Group WhatsApp E001000M Report 2021-10-27 v1.2.pdf

[8] Image Recognition https://searchenterpriseai. techtarget.com/definition/image-recognition.

[9] Tripathi, Milan." Analysis of Convolutional Neural Network based Image Classification Techniques." Journal of Innovative Image Processing (JIIP) 3, no. 02 (2021): 100-117

[10] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward net- works are universal approximators. Neural Networks, 2(5):359–366, 1989.

[11] Transfer learning-based Object Detection by using Convolutional Neural Networks. B N Krishna Sai; T. Sasikala. Object Detection and Count of Objects in Image using TensorFlow Object Detection API. . 2019 International Conference on Smart Systems and Inventive Technology (IC- SSIT).

[12] Introduction to Convolutional Neural Networks (CNN) https://www.analyticsvidhya. com/blog /2021/05/convolutional-neural-networks-cnn/

[13] Kaggle Natural Images Dataset. A compiled dataset of 6899 images from 8 distinct classes.https://www.kaggle.com/datasets/prasunroy/natural-images

[14] Github: aswintechguy/Deep-Learning-Projects https://github.com/aswintechguy/Deep-Learning-Projects/blob/main/Gender%20and%20Age%20Prediction%20-%20Image%20Classification%20%26%20Regres

sion/Gender%20and%20Age%20Prediction%20-
%20CNN.ipynb