Smart File Manager

[1]Atharv Dalvi, [2]Rishikesh Tikekar, [3]Mayur Nile,[4]Prof. Johnson Mathew

[1]B.E. Student, [2]B.E. Student, [3]B.E. Student,[4]Professor of Computer Engineering,
Department of Computer Engineering,
Datta Meghe College of Engineering, Airoli, Navi Mumbai, Maharashtra 400708, India.

Abstract : The need for developing this Android File Manager arose from the fact that there is so much information that needs to be stored and kept secret from others. Therefore this application named 'Android File Manager' will help secure all of your information. This application helps to provide a high facility to move information from one location to another location in a cell phone. The goal of this project is to provide security features and transfer data from one end to another end. The first time when we open the application it will give help on how to use it for the new user. When we enter inside the application we see various functionality regarding the use. Keeping unnecessary images or duplicate media can affect overall performance of mobile devices. Replication of multimedia is a common problem faced by mobile users. These duplicate media files can occupy a large amount of mobile storage area. The existence of such duplicate media files in storage areas indicates the presence of redundancy and lowers the mobile performance. The detection of duplicate media files is challenging due to the large set of media present in mobile storage. Duplicate media has two types. First type is exact duplicate, where pictures that match each other bit by bit, pixel by pixel. Exact duplicate pictures have 100% similar content. Second type is near duplicate, here images are not exactly duplicate, that is their binary form will be different but it can be visually similar. Near duplicate image forms when the original image goes through editing processes such as format changing, mapping, colour, scaling etc. WhatsApp has over 390.1 million active users. WhatsApp has now become an important part of our lives. It helps us to connect with the world. However, it can be a little annoying when it automatically downloads unnecessary images like memes. It makes the mobile gallery a complete mess and occupies mobile memory. Deleting the entire WhatsApp folder is a very bad option because it might contain important personal images. Manual deleting these memes requires great effort and valuable time. The proposed system aims to identify duplicate multimedia files and detect memes with few more features. Multimedia cleaner system is divided into 4 modules. These modules are exact-duplicate image, audio and pdf detection, near duplicate image detection, memes remover and WhatsApp status saver

Keywords– Smart File Manager, Meme Remover, Duplicate image remover.

I. INTRODUCTION

Android by delivering tons of great features that work well together to make your life easier in the long run. You'll see the usual features like moving, sharing and deleting the files on your phone. Android File Manager delivers a great experience with all of the features you didn't realize that you wanted. From managing files, to storage, their features all work well together delivering a simple and easy to use app. The application will connect with online database to store the user uploaded files. In this world of saving storage space and optimization technologies and various state of the art methods being deployed to overcome this issue, secondary storage of mobile phone is the primary storage area used to store information. So, if there exist duplicate media stored on your mobile phone, then no doubt we are wasting the precious space. Due to such type of duplicate media stored on mobile storage space, user may run out of memory space, which leads to shortage of storage and user may lose some of the important information. It will also lead to slowdown in the performance of a mobile phone. Hence, it is necessary to find and delete all those media, which are duplicate and contain the same information. We will be developing an easy to use application, which is used for removing duplicate media from android mobile phones, namely "Multimedia Cleaner" application which will not only find duplicate media but also remove these media from gallery, which will save a lot of memory space of mobile phones. This application will also detect Memes such as pepe meme, success kid meme, doge meme etc. With the help of this application, user can permanently delete those meme from WhatsApp folder. The application is developed using Flutter, which uses Dart language. To detect memes TensorFlow model is used which is trained on collab. TensorFlow model size is less and requires less storage space therefore it can be used to add machine learning functionality to Mobile or Embedded devices. This model is developed in python language. Dataset for this model contains pepe memes, doge memes, drake memes. This application will be able to detect exact duplicate images and near duplicate images as well.

II. PROBLEM STATEMENT

Android by delivering tons of great features that work well together to make your life easier in the long run. You'll see the usual features like moving. sharing and deleting the files on your phone. as well as less common features like compressing files. and streaming files located in cloud storage. Of course, you'll also find the ability to easily browse your files and manage your files. Android File Manager delivers a great experience with all of the features you didn't realize that you wanted. From managing files. to storage. Their features all work well together delivering a simple and easy-to-use application. This application provides security to our files and we transfer the files from 'ane to another location. In this application, we provide audio players, videos, and see images. This is a beautiful application by which we make life simpler by accessing this file manager. This application has provided access to all the files. folders etc from the mobile phone. The main features of the application are, we see all the folders that are hidden from the user by another application. So this project helps users a lot. This application shows how much memory remains either in the phone memory or other external memory.

III. LITERATURE REVIEW

3.1 A Review on Near-Duplicate detection of images using computer vision technique.
This project includes pixel based pattern and feature extraction method.this paper was published on 2 january 2020.It has two or three drawbacks.Some of them are complexity in NDD,computational complexity,optimum values.

3.2 Searching for near duplicate images in cloud databases.

This project used a method of locality sensitive hashing and BRISK method.this paper was published on 19 january 2017.But the accuracy of this method was very low.

3.3 Research on Base64 encoding algorithm and PHP implementation.

This project was implemented using Base64 extraction method..this paper was published on 06 december 2018. It was quite a helpful method.

3.4 Near Duplicate image detection.

The paper was published in 2014 showing an idea of using segmentation and clustering of images to find duplicate images.the only drawback of this method was it has too many features of extraction which leads to time penalty.

IV. PROPOSED STUDY

Base64 encoding mechanism is used to get exact-duplicate media which includes image, audio and pdf. Base64 can be implemented using 4 steps: 1. Get binary format of file. 2. Make a group of 6 bits. 3. Convert bits into decimal value. 4. Convert decimal value to a character using Base64 encoding table.

Near duplicate images can appear when a particular object is captured multiple times when the background is moving such as panorama. Near duplicate images may form when an original image is scaled, rotated, transformed or changes colour. Detecting different versions of the original image is quite difficult as it requires large computation power. This module can detect near duplicate images which are captured or downloaded from current date till past five days. Dhash is 14 also known as "Difference hash". It uses a python library to generate a hash for an image. Dhash is almost similar to Perceptual hash (phash) and Average hash (ahash) but it performs much better. As ahash works on average value and phash works on frequency patterns, dhash works on gradients.

V. IMPLEMENTATION

As per architecture diagram, the project is divided into three modules: File Manager, Duplicate Images(Near and Exact), Meme Classifier.

5.1 File Manager: This project provides a feature of File Manager which is completely open source and does not collect any data of the user, which in turn maintains user privacy to its fullest. Users may also be able to perform basic operations like creating folders, copy, paste, delete, etc.

5.2 Duplicate Images: Duplicate images can appear when a particular object is captured multiple times when the background is moving such as panorama. Near duplicate images may form when an original image is scaled, rotated, transformed or changes colour. Detecting different versions of the original image is quite difficult as it requires large computation power. This module can detect near duplicate images which are capture or downloaded from current date till past five days

5.3 Meme Classifier: It is a heavy task to detect memes from a large dataset of images. Processing all images might crash an application. TensorFlow Lite can solve this problem. TensorFlow Lite uses python language which converts pre-trained machine learning modules into TensorFlow Lite modules (.tflite). TensorFlow Lite module is developed in such a way that it gives accurate and quick results. TF Lite module is a light-weight module hence it occupies less storage space and can be integrated to Mobile applications. A model is trained to detect memes which a user receives from WhatsApp. This model can detect many types of memes like Doge, Pepe, Drakeposting, Success Kid etc.

VI. SYSTEM ARCHITECTURE

As per architecture diagram, the project is divided into six modules exact duplicate images, exact duplicate audio, exact duplicate pdf, near duplicate images, Memes finder, WhatsApp forwards, WhatsApp status images and videos respectively.
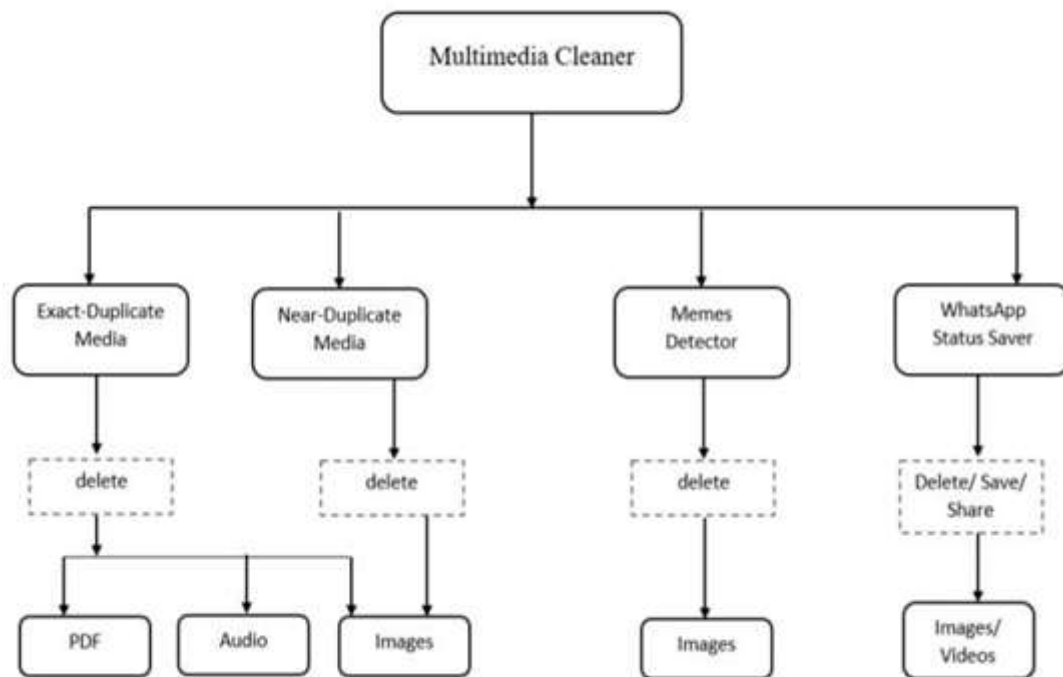
Architecture Diagram

There are two types of duplicate image, exact duplicate image and near duplicate image. Exact- Duplicate Images means images with 100% similar content. This happens when a user acciden- tally downloads the same image twice or gets the same image on whatsapp from more than one person. Near duplicate image means image with almost similar content.Near-duplicate image detection is the task of finding different versions of the same image. Images that are not exact duplicates in binary form, but can be visually identified as the same image having undergone various edit- ing steps such as color mapping, scaling, format changing, etc.

Third part of the project is detecting forwards of WhatsApp folders, forwards can be memes, thoughts,wishes etc.

Fourth part is WhatsApp status saver. This will allow users to save WhatsApp status images and videos.
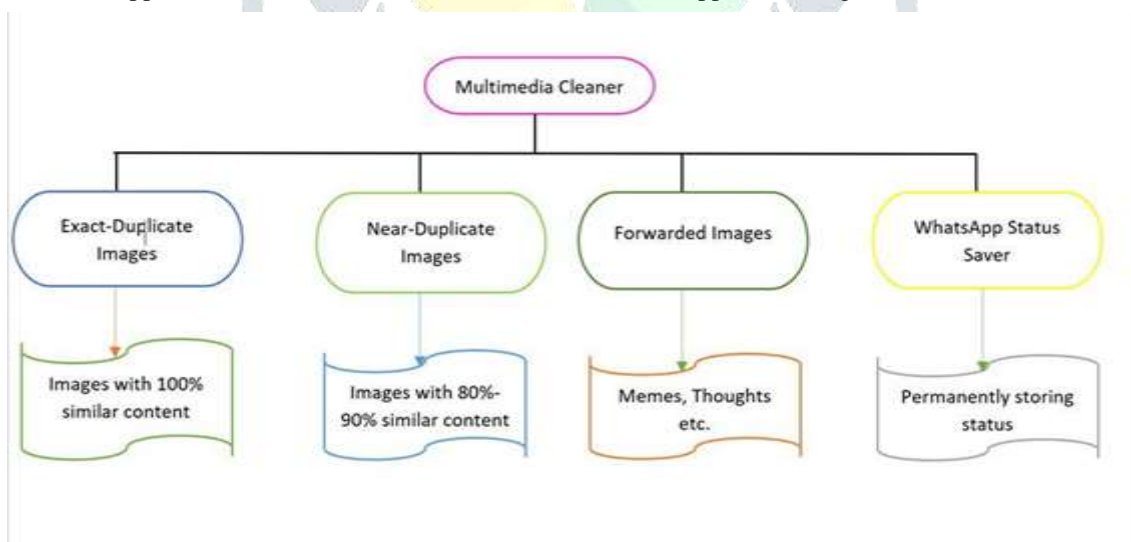

System Design

To get Exact-Duplicate media we have used Base64 encode. It is an encoding technique which helps in faster detection of duplicates.

DHash algorithm is used in near duplicate image detection.

With the help of TensorFlow Lite model we can detect memes from WhatsApp folders. Tensor- Flow model is a lightweight model that's why it requires less computation time. These features allow the TensorFlow model to easily integrate with Android/iOS applications.

As we know, WhatsApp status is available for 24hrs only and we can not save that status per- manently. Our application provides a feature to save, share or delete WhatsApp status images or videos.

VII. SCREENSHOTS
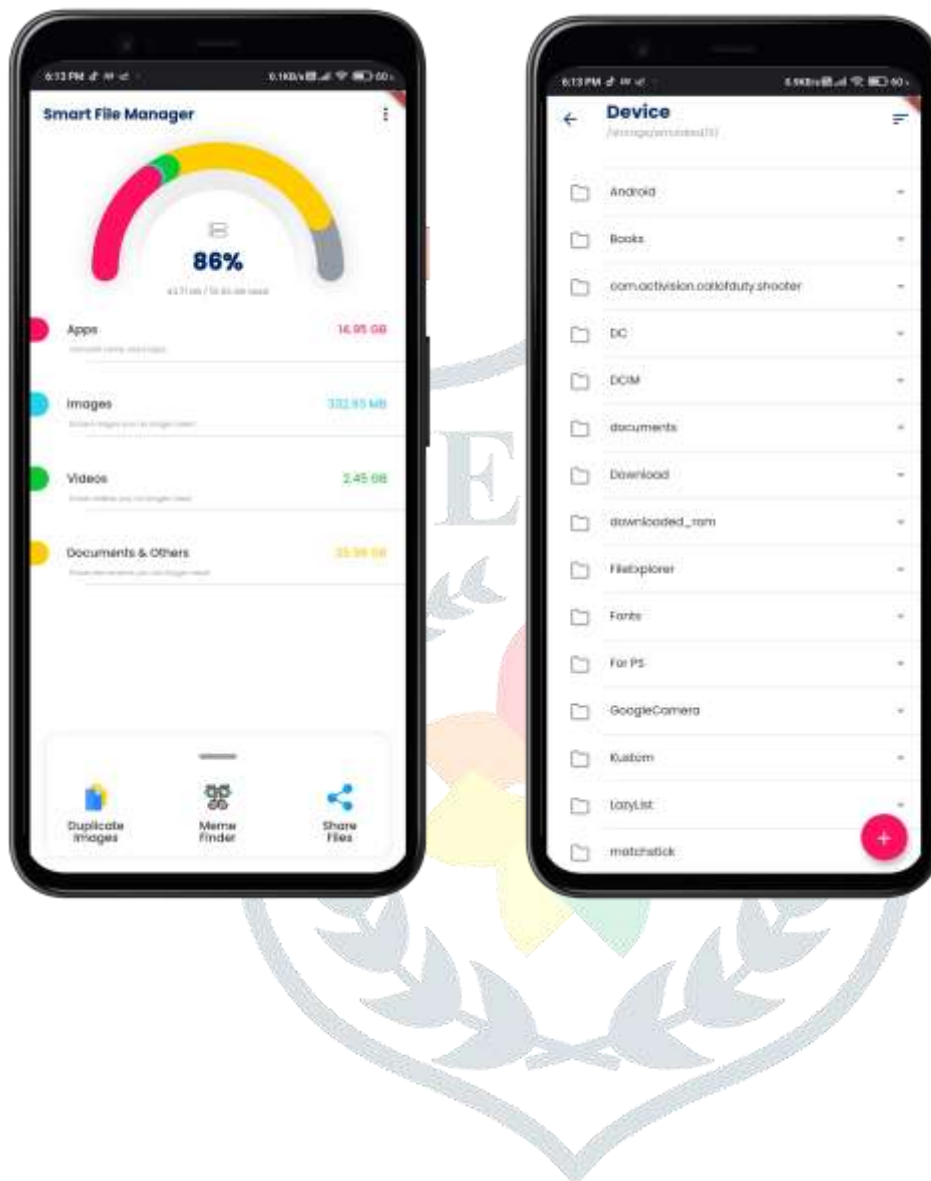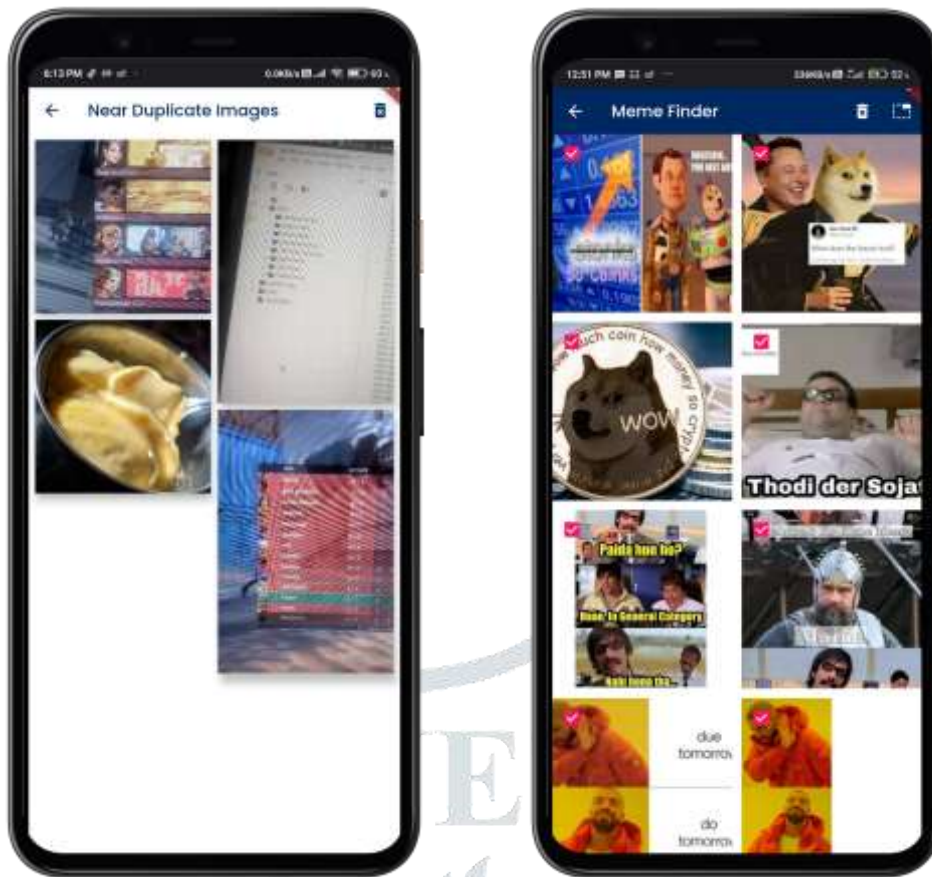
VIII.   CONCLUSION AND FUTURE ENHANCEMENT

The main aim of this project is to free-up mobile storage space for users. It focuses on deletion of duplicate media (images, pdf, audio) and memes from WhatsApp folder. This unwanted media occupies a lot of memory space. Our application can delete those unwanted images as well as duplicate media. Along with this our application has many features like users can save, share and delete WhatsApp status images and videos. Till now our application is able to detect exact media such as images, pdf and audio. Also our application can detect near duplicate images. We are able to save, share and delete WhatsApp Status images and videos. In future, we will try to reduce computation time required by an application to scan all files.

IX. REFERENCES

[1] K. K. Thyagharajan, G. Kalaiarasi R.M.D. Engineering College, Chennai, India, Sathyabama Institute of Science and Technology, Chennai, India A Review on Near-Duplicate Detection of Images using Computer Vision Techniques. 06 January 2020

[2] Maneesha, Inderveer Chana Searching of Near Exact Duplicate Images in Cloud Database. 2016 International Conference on Inventive Computation Technologies (ICICT), 2017.

[3] Somchai Wen, Wen Dang School of Information Science and Technology Yunnan Normal University Kunming, China Research on Base64 Encoding Algorithm and PHP Implementation, 2018.

[4] Rahulkrishna Yandrapally, Andrea Stocco, Ali Mesbah University of British Columbia Vancouver, BC, Canada rahulyk@ece.ubc.ca Near-Duplicate Detection in Web App Model Inference. ICSE '20, May 23–29, 2020, Seoul, Republic of Korea

[5] V.B. Nemirovskiy, A.K. Stoyanov Tomsk Polytechnic University, Tomsk, Russia Nearduplicate image recognition. 2014 International Conference on Mechanical Engineering, Automation and Control Systems (MEACS), 2014.

[6] https://www.base64decode.org/.

[7] http://www.hackerfactor.com/blog/?/archives/529-Kind-of-Like-That.html