ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

CLASSIFICATION AND SEPARATION OF WHATSAPP CONTENTS USING DEEP LEARNING

Ms. P. Nava Bhanu[1], B. Lakshmi Ramya[2], CH. Trujitha Lakshmi[3], K. Mounika[4]

#1M. Tech, (Ph. D)

#2, #3, #4 Students, Dept of Computer Science & Engineering, BWEC, Bapatla, Andhra Pradesh, India

ABSTRACT_ In today's digital landscape, the prevalence of messaging platforms such as WhatsApp has revolutionized the way individuals share information, including images and documents. However, the sheer volume of shared content often leads to clutter and disorganization, particularly in educational and professional contexts where study materials, official notices, and various documents are frequently exchanged. This research proposes a sophisticated solution harnessing the power of deep learning to automatically classify and organize WhatsApp content into distinct categories, including circulars, question papers, marksheets, handwritten notes, printed notes, and diverse document formats such as PDFs, DOCs, Excel sheets, CSVs, PPTs, and TXT files.

1.INTRODUCTION

Orders are methodicallly separated into gatherings and classes in light of their qualities. By using data to train computers, image classification has emerged to bridge the gap between computer vision and human vision. Picture order is accomplished by grouping pictures into foreordained classes in light of the substance of the vision. Spurred by [1], this article portrays the investigation of picture characterization utilizing profound learning. Customary picture order techniques are essential for the field of computerized reasoning (artificial intelligence), officially known as AI. AI comprises of an element extraction motor that removes significant highlights like edges and surfaces, and a grouping motor that characterizes in light of the separated highlights. The principal limit of AI is that it tends to be isolated, yet it can extricate explicit elements on the picture, not trademark highlights from the preparation dataset. This inadequacy is killed by utilizing profound learning [2]. Profound learning (DL) is a subfield of AI that can be advanced by a remarkable estimation

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

technique. Profound learning models have been acquainted with forever disintegrate data in a homogeneous design that people experience. To accomplish this, profound learning utilizes a various leveled construction of numerous calculations, addressed as a counterfeit brain framework (ANN). ANN's design is reenacted utilizing the organic brain organization of the human mind

2.LITERATURE SURVEY

Classification and Separation of WhatsApp Images Using Machine Learning

AUTHORS:

ThasneemRafathSk, V. Komalatha, G. Sandeep, K. Bhanu Rekha, S. Ravi Kishan

Abstract:

In this digital era, Internet has become an integral part of human lives. Internet and social networks have become very popular, allowing anyone to easily share pictures, text, audio and video files. Among all the applications, WhatsApp has become quite famous due to its ease of use and it has replaced almost all the other messaging apps. Apart from sending messages, images and videos over it, one more reason for the heavy usage of WhatsApp is the exchange of study notes and materials by the students during the time of the examination and end up with a lot of images to be deleted at the end of each semester. And also the notices and brochures in every semester gets mixed up with other images and these needs to be separated for easy reference. As the WhatsApp folder may have many other images, selecting the study material images, brochures, etc., one by one from the other images and then deleting them is a tedious process. Henceforth, this research work has utilized machine learning to build a model for detecting and extracting the images from the WhatsApp images folder. Further, the proposed model classifies the study notes images into printed and handwritten notes. Notices and brochures received on WhatsApp are separated into a new folder. And also, screenshots and photos are grouped into separate folders. The proposed model has been built by using a deep learning concept called the Convolutional Neural Network [CNN] and by using Python's Keras library. It takes an image and decides its category and then the action is taken accordingly.

Social Network Identification Through Image Classification With CNN

AUTHORS:

Roberto Caldelli, Chang-Tsun Li, Irene Amerini

ABSTRACT:

Identification of the source social network based on the downloaded images is an

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

important multimedia forensic task with significant cybersecurity implications in light of the sheer volume of images and videos shared across various social media platforms. Such a task has been proved possible by exploiting distinctive traces embedded in image content by Social Networks (SNs). To further advance the development of this area, we propose a novel framework, called FusionNET, that integrates two established Convolutional Neural Networks, with the former (named 1D-CNN) learning discriminative features from the histogram of DCT coefficients and the latter (named 2D-CNN) inferring unique attributes from the sensor-related noise residual of the images in question. The separately learned features are then fused by FusionNET to inform the ensuing source identification or source-oriented image classification component. A series of experiments were conducted on a number of image datasets across various social networks and instant messaging apps (IMAs) to validate the feasibility of FusionNET also in comparison with the performance of the 1D-CNN and 2D-CNN. Encouraging results were observed.

"Image phylogeny forests reconstruction,"

AUTHORS:

F. de O. Costa, M. A. Oikawa, Z. Dias, S. Goldenstein, and A. R. de Rocha

Abstract:

Today, a simple search for an image on the Web can return thousands of related images. Some results are exact copies, some are variants (or near-duplicates) of the same digital image, and others are unrelated. Although we can recognize some of these images as being semantically similar, it is not as straightforward to find which image is the original. It is not easy either to find the chain of transformations used to create each modified version. There are several approaches in the literature to identify near-duplicate images, as well as to reconstruct their relational structure. For the latter, a common representation uses the parent-child relationship, allowing us to visualize the evolution of modifications as a phylogeny tree. However, most of the approaches are restricted to the case of finding the tree of evolution of the near-duplicates, with few works dealing with sets of trees. Since one set of near-duplicates can contain an independent subsets, it is necessary to reconstruct not only one phylogeny tree, but several trees that will compose a phylogeny forest. In this paper, through the analysis of the state-of-the-art image phylogeny algorithms, we introduce a novel approach to deal with phylogeny forests, based on different combinations of these algorithms, aiming at improving their reconstruction accuracy.

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

We analyze the effectiveness of each combination and evaluate our method with more than 40 000 testing cases, using quantitative metrics.

3.PROPOSED SYSTEM

In this project we are using CNN (convolution neural networks) algorithm to classify WhatsApp contents to different categories such as Question Papers, Mark sheets, Printed papers, hand written papers, circulars, PDFs, DOCs, PPTs, XLSX, CSV. CNN algorithm will get trained on above mention categories to build a classification model. This model can be applied on test images to predict image type.

To train CNN we have used same dataset given by and below showing that dataset images.

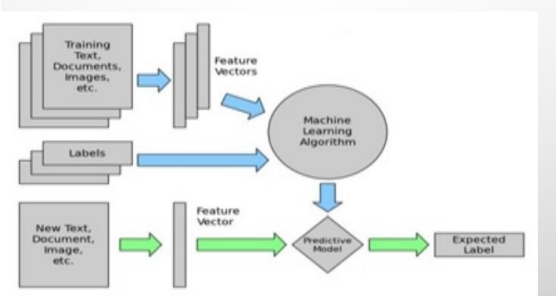


Fig 1: Proposed Architecture

4. ABOUT DATASET

3.1 IMPLEMENTATION

Upload WhatsApp Image Dataset: using this module we will upload dataset to application and then read all contents and categories from dataset

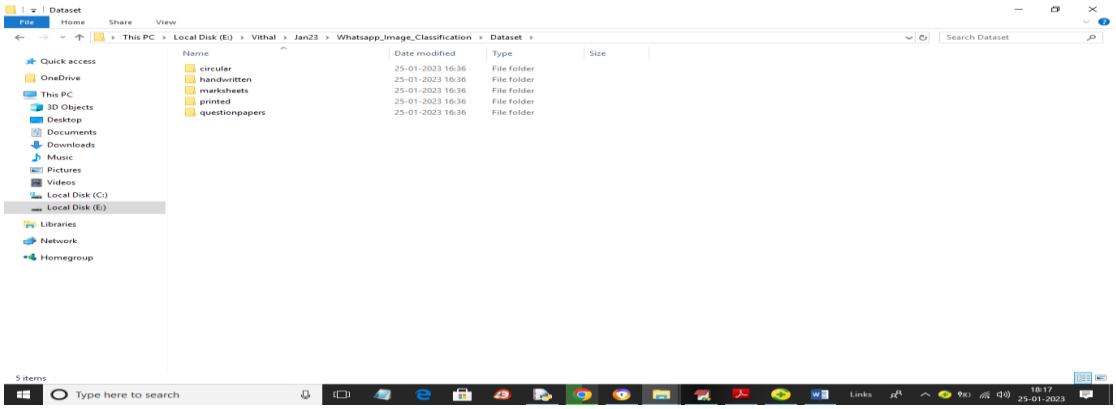
Preprocess Dataset: using this module we will resize all images to equal size and then normalize image pixel values and then shuffle the dataset. After processing all images will be split into train and test where application using 80% dataset for training and 20% for testing

Train CNN Algorithm: using this module we will input 80% dataset to CNN to trained a model and then 20% test images will be applied on trained model to calculate prediction accuracy

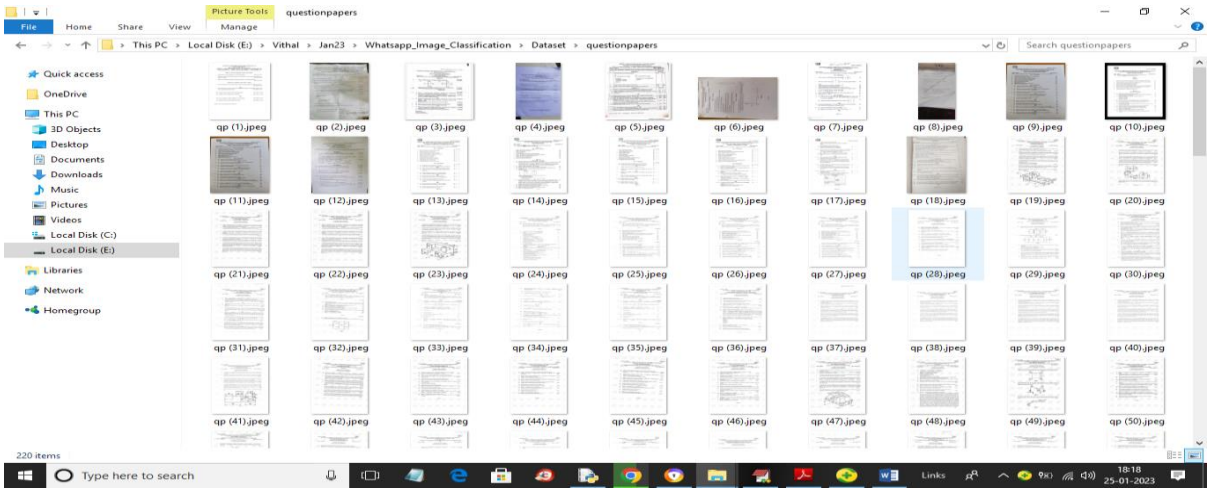
CNN Training Graph: using this module we will plot CNN training and loss graph

WhatsApp Image Classification: using this module we will upload test image and then CNN will predict or classify type of images and stores into separate specific folders.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issue.1 March 2024

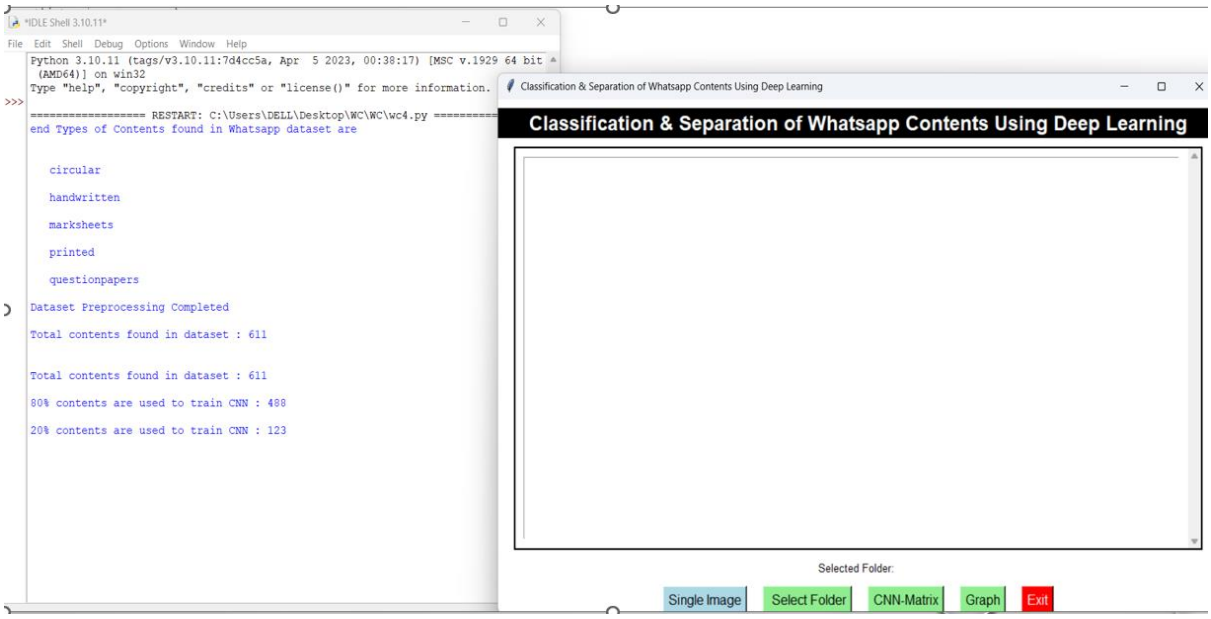


In above dataset folder we have 5 different types of images and just go inside any folder to view that type of image like below screen



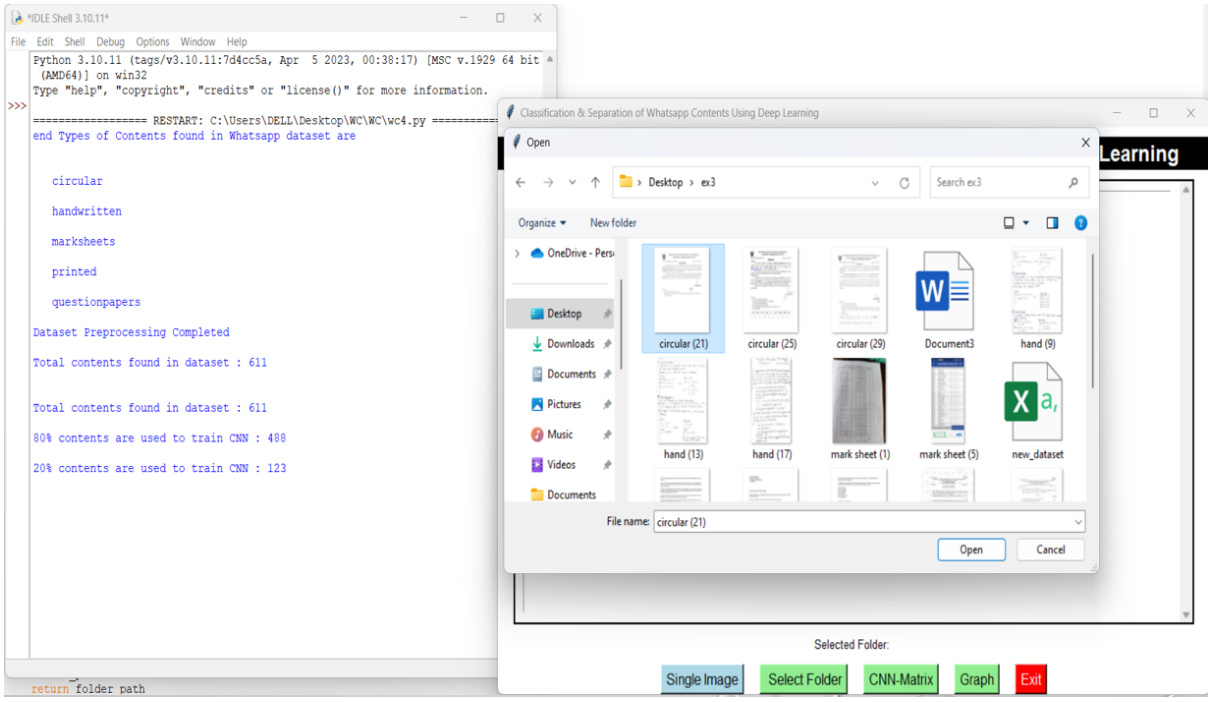
In above screen we can see some images from dataset and by using this images CNN will get trained

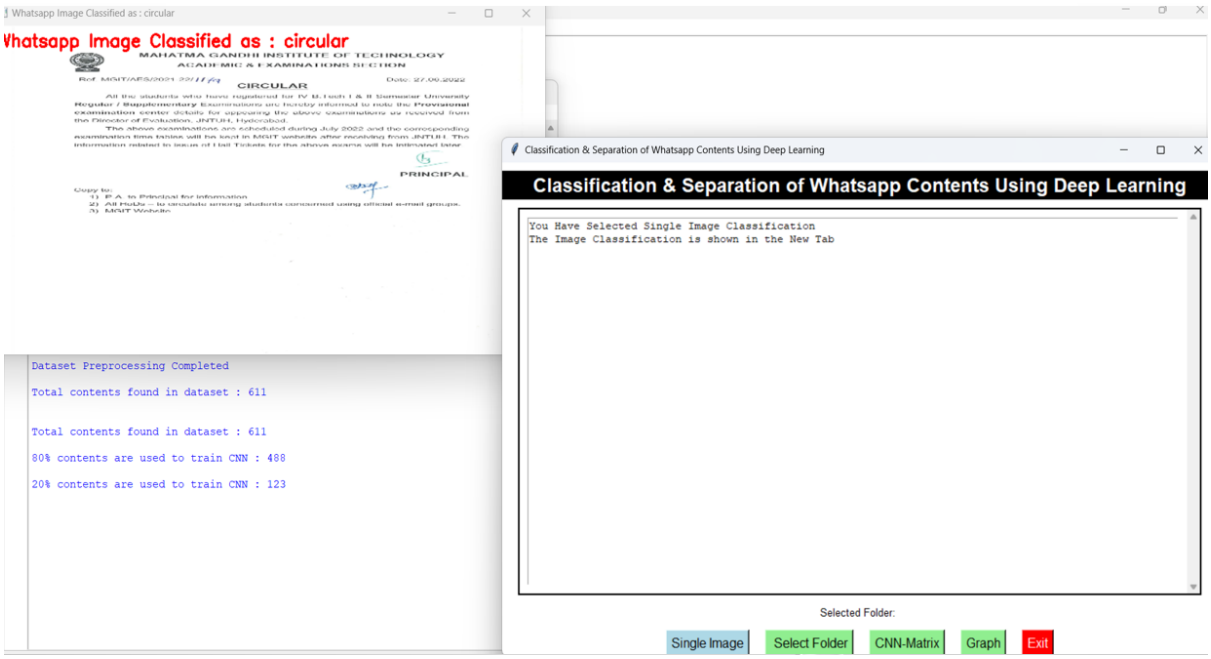
4.RESULTS AND DISCUSSION



INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issue.1 March 2024

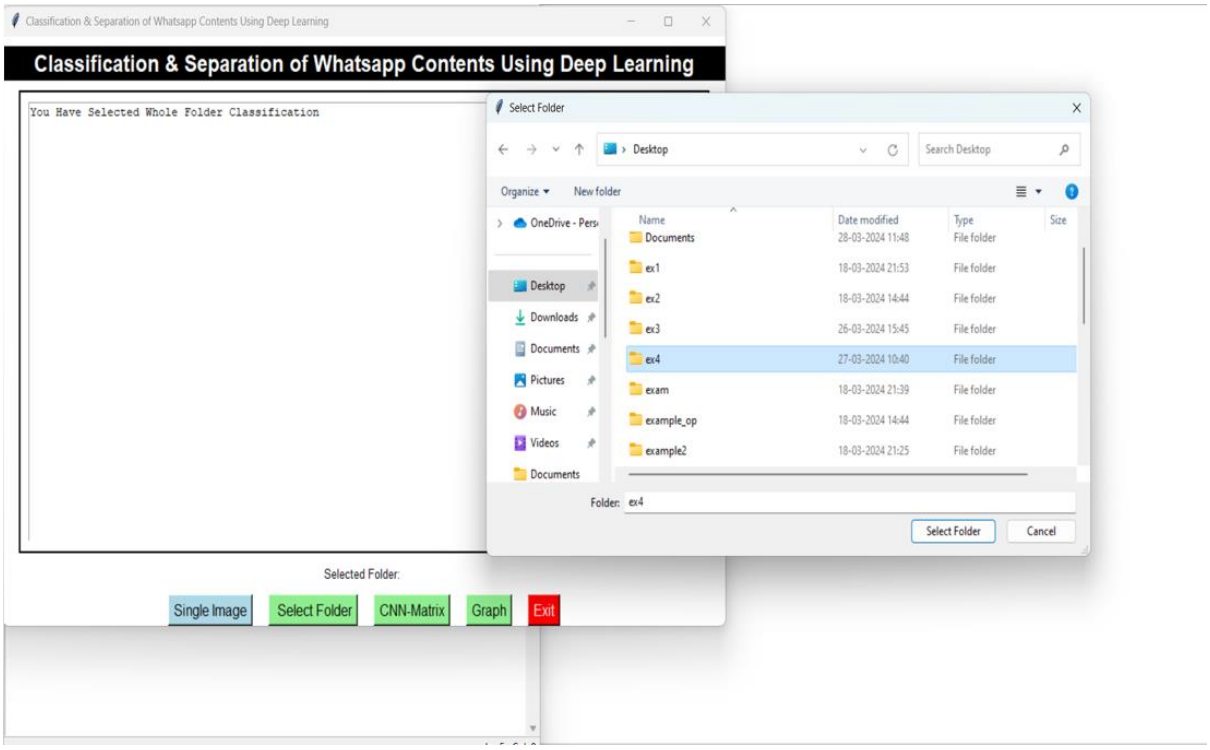
In the above screen we can see the output interface. Now click on the 'SELECT IMAGE' button. we will redirect to the files to select image. After selecting the image, it can predict what was that image. We can see the results in below page.



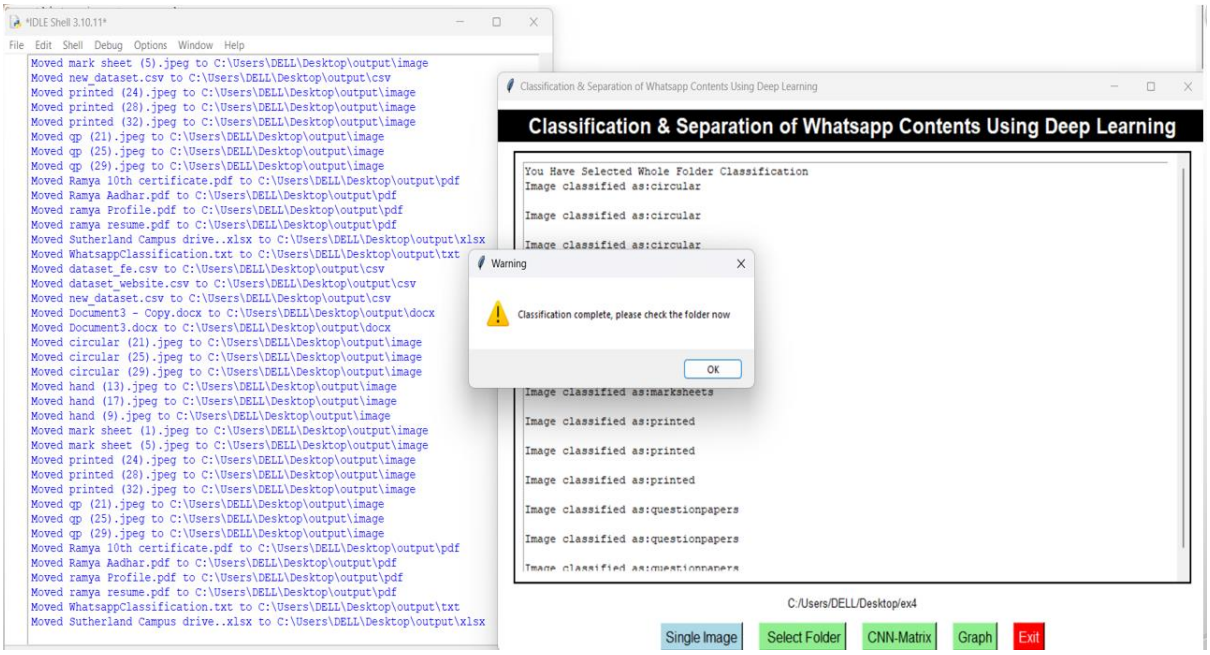


Now click on the 'SELECT FOLDER' button to get the below page.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issue.1 March 2024

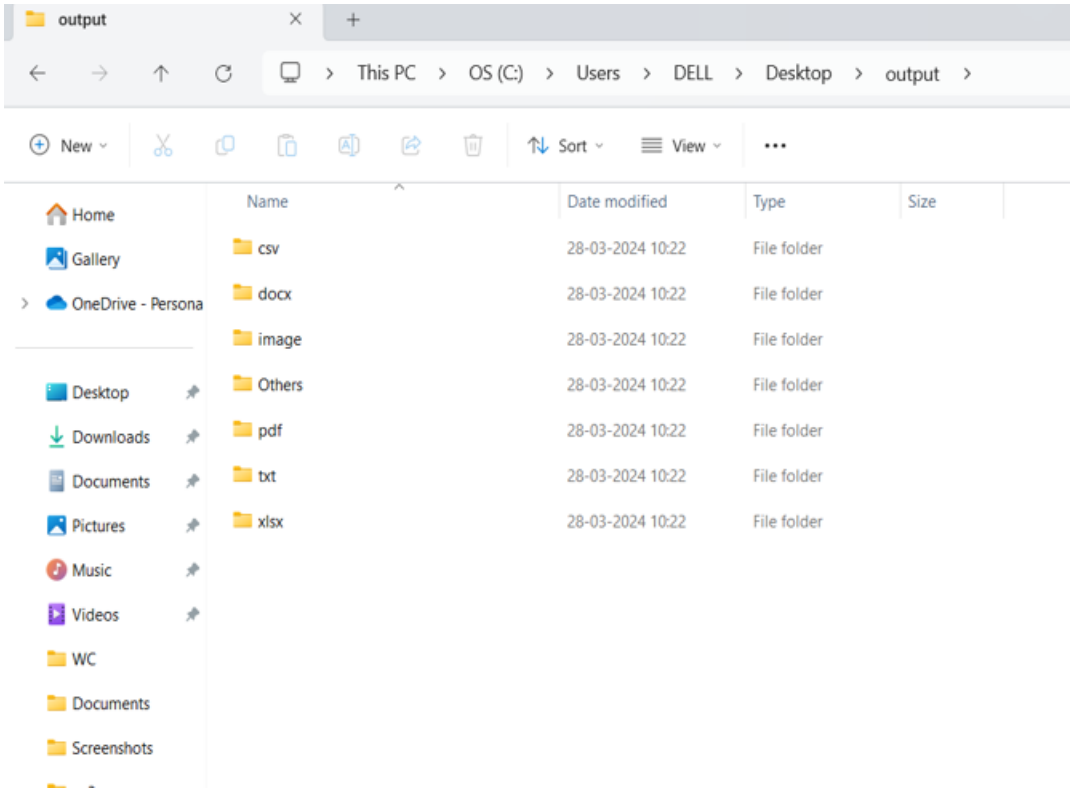


After selecting the folder which contains images, various types of documents like Pdfs, Docs, Excels, Ppts, Text files etc., all will be classified into their specified folders. We can see the Results in the below Screens.

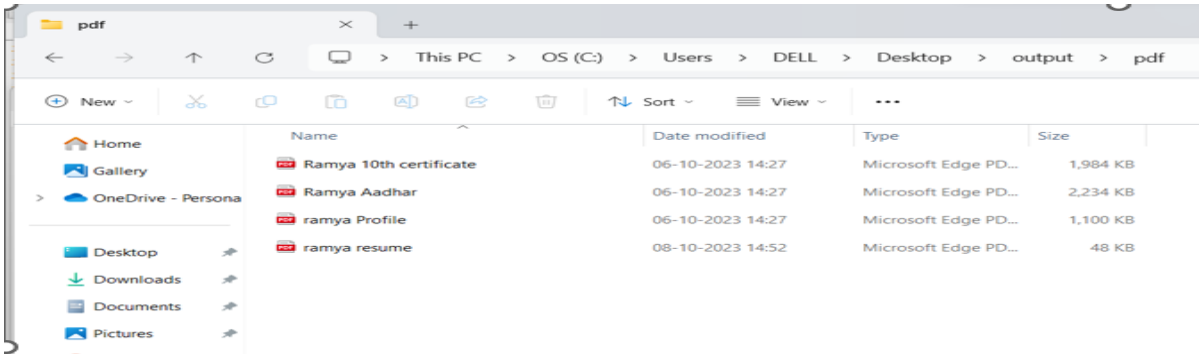


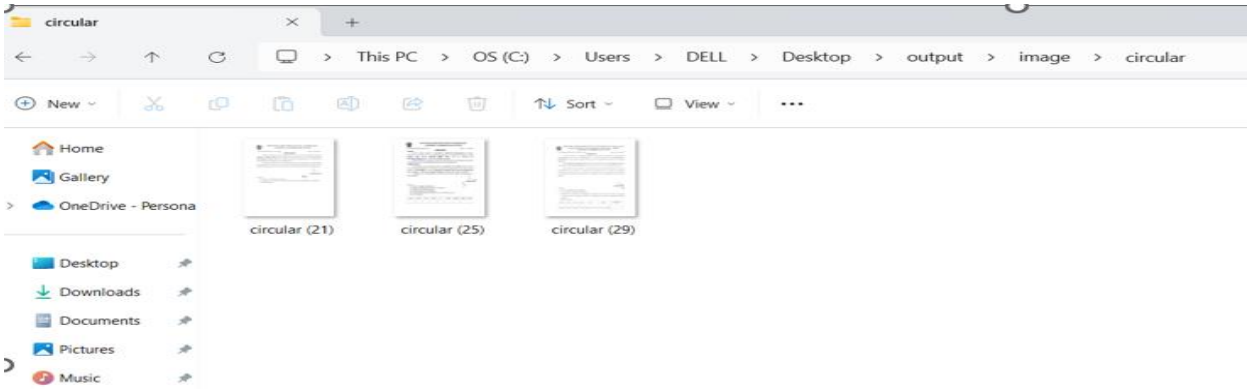
After completion of classification, we will be redirected to the output Folder.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024



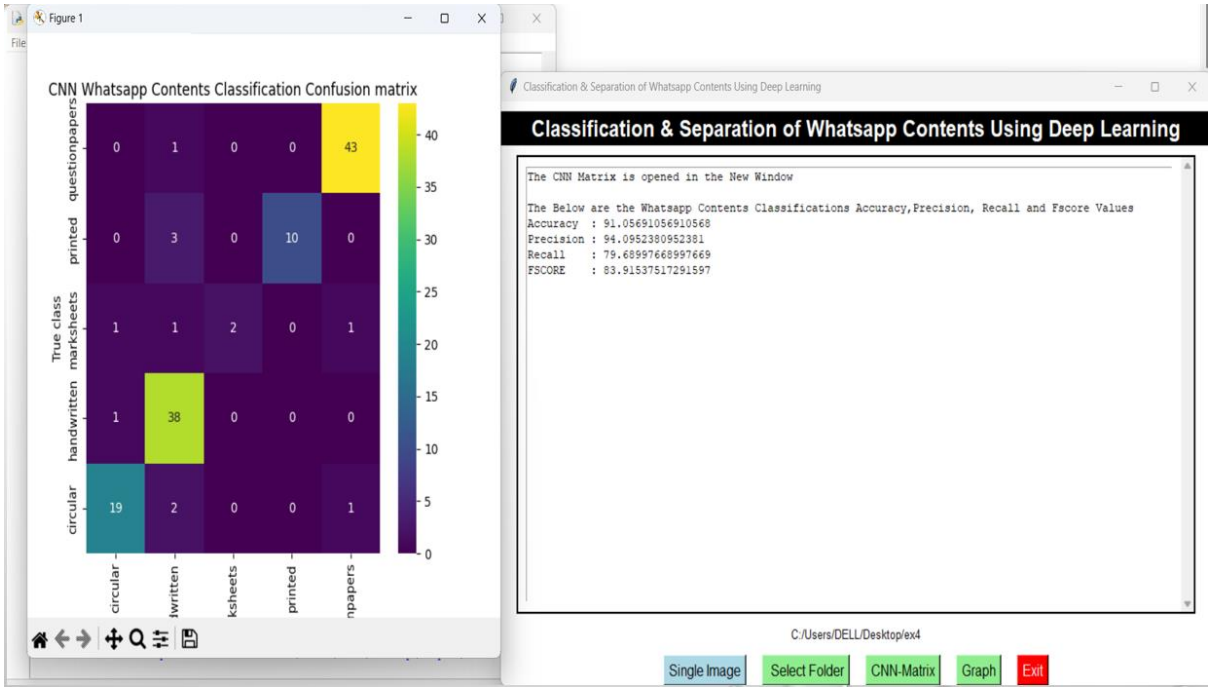
In the above screen we can see the classified folders separated with their specified extensions. If we click on the folders, we can see the documents related to the specific folder. In the image folder we are having sub folders of Circular, Hand Written, Printed, Question Papers.



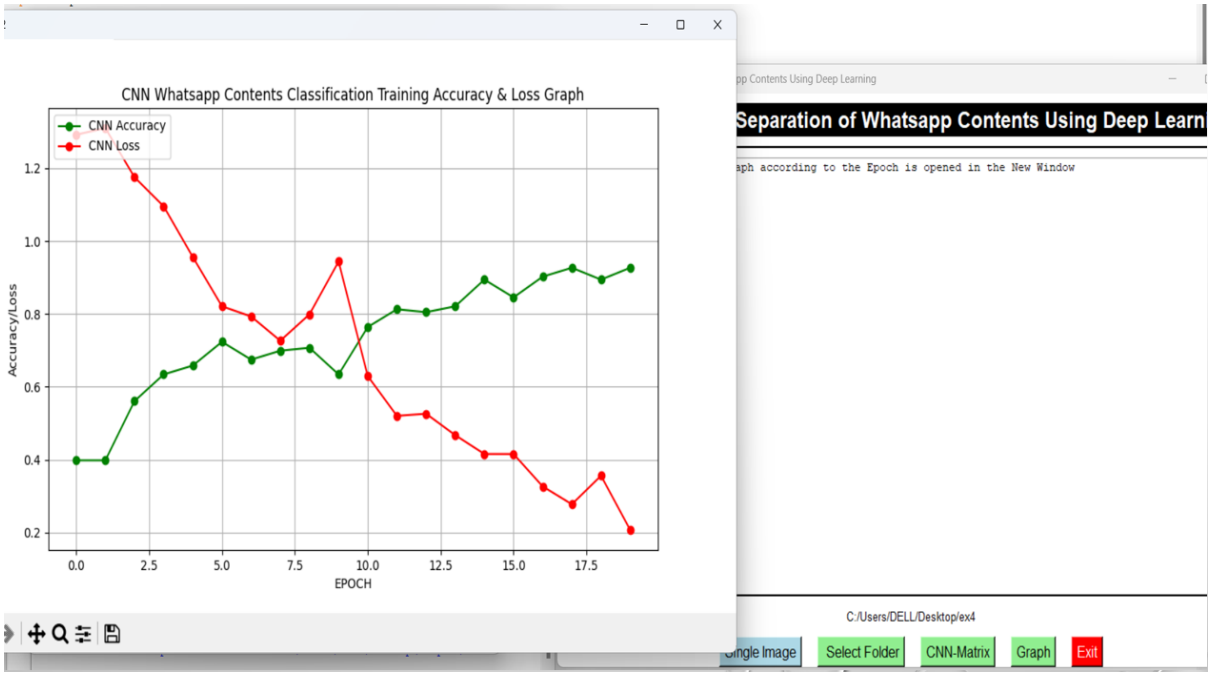


INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

Now click on the 'CNN MATRIX' to get the below Results.



In above screen with CNN we got 91% accuracy and we can see precision, recall and FSCORE metric. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and all blue colour boxes contains INCORRECT prediction count which are very few and different colour boxes contains CORRECT prediction count which are high in numbers so we got 93% accuracy. Now click on 'CNN Training Graph' button to get below page



ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

In above CNN training graph x-axis represents training epoch and y-axis represents accuracy and loss values. Green color line represents Training Accuracy and red color line represents Training LOSS and in above graph we can see with each increasing epoch accuracy got increase and reached closer to 1 and loss got decreased and reached closer to 1.

5.COCNLUSION

In the era of digital communication, WhatsApp has become a cornerstone for sharing images and documents, facilitating seamless exchange of study materials, official notices, and various files. However, managing the influx of content within the WhatsApp folder can pose challenges, requiring users to manually sift through cluttered files. This research addresses this issue by presenting a deep learning-based solution for automatically organizing and classifying WhatsApp content.

Through the utilization of Convolutional Neural Networks (CNNs), this study has developed a robust model capable of accurately categorizing WhatsApp images and documents into distinct classes, including circulars, question papers, marksheets, handwritten notes, printed notes, and various document formats such as PDFs, DOCs, Excel sheets, CSVs, PPTs, and TXT files. The model's architecture was meticulously designed and trained on a diverse dataset, achieving high accuracy and performance metrics in classification tasks.

The results of this research demonstrate the efficacy of leveraging deep learning techniques for enhancing the management of WhatsApp content. By automating the classification and organization process, users are relieved of the burden of manually sorting through their WhatsApp folders, saving time and effort. This streamlined approach not only improves user experience but also promotes efficiency and productivity, particularly for students and professionals who heavily rely on WhatsApp for communication and file sharing.

In conclusion, the proposed deep learning-based image and document classifier offers a valuable solution to the prevalent issue of content organization within the WhatsApp ecosystem. As the digital landscape continues to evolve, the integration of AI-driven solutions holds immense potential for optimizing user workflows and enhancing the overall usability of

messaging platforms. Future research endeavors may explore further refinements to the model architecture, as well as the integration of additional features and functionalities to address evolving user needs and preferences.

REFERENCES

[1] Karuppusamy, P. "Building Detection using Two-Layered Novel Convolutional Neural Networks." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 29-37.

[2] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," Journal of Machine Learning Research, vol. 11, no. Dec, pp. 3371–3408, 2010.

[3] Manoharan, J. Samuel. "Capsule Network Algorithm for Performance Optimization of Text Classification." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 1-9.

[4] G. Schaefer and M. Stich, "UCID - an uncompressed colour image database," in Proceedings of the Storage and Retrieval Methods and Applications for Multimedia, 2004, pp. 472–480. [

5] Manoharan, Samuel, and Narain Ponraj. "Analysis of Complex Non-Linear Environment Exploration in Speech Recognition by Hybrid Learning Technique." Journal of Innovative Image Processing (JIIP) 2, no. 04 (2020): 202-209.

[6] J. Yang, G. Zhu, and Y.-Q. Shi, "Analyzing the effect of jpeg compression on local variance of image intensity," Trans. Img. Proc., vol. 25, no. 6, pp. 2647–2656, Jun. 2016. [Online]. Available: http://dx.doi.org/10.1109/TIP.2016.2553521.

[7] Hamdan, Yasir Babiker. "Faultless Decision Making for False Information in Online: A systematic Approach." Journal of Soft Computing Paradigm (JSCP) 2, no. 04 (2020): 226- 235

[8] L. Van Der Maaten, "Accelerating t-sne using tree-based algorithms," J. Mach. Learn. Res., vol. 15, no. 1, pp. 3221–3245, Jan. 2014.

[9] Dhaya, R. "Hybrid Machine Learning Approach to Detect the Changes in SAR Images for Salvation of Spectral Constriction Problem." Journal of Innovative Image Processing (JIIP) 3, no. 02 (2021): 118-130.

[10] B. Tondi, "Pixel-domain adversarial examples against CNN-based manipulation detectors," Electronics Letters, vol. 54, pp. 1220–1222(2), October 2018. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/el.2018.6469