BIT PAPER ASSESSMENT CONSUMING PROFOUND KNOWLEDGE

[1]Name of 1st Ushaben Barad
[1]Designation of 1st Assistant Professor
[1]Name of Department of 1st Faculty of Engineering
[1]Name of organization of 1st Gokul Global University, Sidhpur, Patan, Gujarat – India

Abstract— In today's era of automation, there is a growing need for automating the evaluation of hand-written MCQs (Multiple Choice Questions) answer sheets. The traditional method of manually checking these answer sheets is time-consuming and requires a significant amount of manpower. This automated approach can help reduce the workload of teachers and make the evaluation process more efficient. With the increasing popularity of online learning and the need for fast and accurate evaluations, this application can be a valuable tool for educational institutions.

Keywords— The assessment of multiple-choice questions, the extraction of handwritten text, and the utilization of deep learning algorithms.

INTRODUCTION

In the modern world, evaluating MCQs answer sheets is a crucial part of many examinations, and it is usually done manually by teachers. However, this can be a daunting task, particularly if the number of students is significant. Automating the MCQs answer checking process can make it more efficient, transparent, and fair, eliminating any chances of biasedness from the teacher's side. While OMR sheets are commonly used for MCQs exams, they require expensive evaluation machines and can be challenging to bubble correctly. Furthermore, there are no existing tools to evaluate handwritten MCQs.

This project aims to introduce an application that utilizes deep learning to evaluate MCQ answer-type exams. The application can be implemented in various educational institutions to streamline the evaluation process. When the application is launched, the user is presented with two options: Evaluate a Single Image or a Folder of Images. If the user selects the "Implement with Image" option, they are prompted to upload the Image, key, marks for the correct answer, and marks for the wrong answer. After filling in these fields, the user can press the "Result" button, and the marks obtained are displayed on the output screen. Alternatively, if the user selects the "Implement with Folder of Images" option, they are prompted to upload a zip file of images and an excel sheet with roll numbers, key, marks for the correct answer, and marks for the wrong answer. Upon filling in these fields and pressing the "Result" button, an excel sheet with marks is downloaded as output.

LITERATURE SURVEY (RELATED WORK)

• The paper by Jamshed Memon, Maira Sami, Rizwan Ahmed Khan, and Mueen Uddin in 2017 discusses the significance of Optical Character Recognition (OCR) in converting handwritten documents into analyzable, editable, and searchable data. They highlight the use of artificial intelligence and machine learning tools in this field.

• Shubham Sanjay Mor, Shivam Solanki, and Saransh Gupta's 2017 paper presents a new approach to handwriting recognition using Deep learning and computer vision. They utilize the Emnist dataset and Convolutional Neural Networks (CNN) to create an Android application for handwritten text detection and conversion into digital form.

• they discuss the increasing use of digital technology for capturing images containing textual data that needs to be stored or edited. They propose the use of Tesseract OCR Engine for Optical Character Recognition.

• They explore the evolution of human-machine communication and its impact on natural communication modalities such as speech, gestures, sound, and vision.

SYSTEM IMPLEMENTATION

Development Enviornment:

The analysis of student's answer sheets in our application is performed using Deep Learning techniques, which is executed in Google Colab Notebook. This Notebook is a widely used open-source tool for conducting data science projects and experiments, and it is based on the web. Many machine learning applications utilize this notebook. To develop the frontend of our application, we used the Anvil framework. Anvil is a Python framework that makes it easy to create a frontend for our Python application. We can also deploy our backend code to Anvil server and access our application from anywhere in the world using the web.

Dataset Preparation

We are using two datasets in our project. One of the dataset being used in our project is "Handwriting Recognition," which was downloaded from Kaggle. It includes over 400,000 handwritten names collected through charity projects. Character recognition technology uses image processing techniques to transform characters from scanned documents into digital formats. While it works well with machine-printed fonts, recognizing handwritten characters is still challenging for machines due to the vast differences in individual writing styles. The dataset contains a total of 206,799 first names and 207,024 surnames, which have been split into a training set (331,059), testing set (41,382), and validation set (41,382).The another is "HandWritten_Character" dataset is a publicly available dataset on Kaggle. It contains 20,000 images of handwritten characters from 0 to 9 and A to Z in uppercase and lowercase letters. The dataset is suitable for training and evaluating machine learning models for character recognition tasks. Each image is of size 32 x 32 pixels and is in grayscale format. The dataset can be downloaded for free from the Kaggle website and can be used for research and educational purposes.

Preprocessing:

To prepare the "Handwriting Recognition" dataset images for CNN model training, the following preprocessing steps are necessary.

• Initially, converting the images to grayscale is essential since character recognition does not require color information. This action simplifies feature extraction and reduces image dimensions.

• The images' pixel values may be normalized next to a standard range to make the training process more efficient. Normalizing is accomplished by dividing the pixel values by the maximum pixel value of 255 in grayscale images. Doing this prevents neural network activation function saturation.

• The images are then resized to a fixed size to ensure all images have the same dimensions for CNN input. Data augmentation techniques can be employed to create more variations in the dataset. This technique involves random image transformations such as rotation, scaling, and shearing to create new data variations.

• Finally, dividing the dataset into training, validation, and testing subsets, typically in an 80:10:10 ratio, is required. The training set is used to train the CNN model, while the validation set is used to fine-tune hyper parameters and prevent overfitting. Lastly, the testing set is used to evaluate the model's final performance on unseen data.

Model Building and Training

• The convolutional neural network (CNN) designed to classify images. The model consists of several layers that are arranged in a sequence to extract features from images and classify them into one of 35 possible classes.

• The first layer of the model is a 2D convolutional layer with 32 filters and a kernel size of (3,3). The layer applies ReLU activation function, which is commonly used in CNNs. The layer's padding is set to "same", ensuring the output size is identical to the input size. The layer receives a grayscale image of 32x32 pixels as input.

• Next, a max pooling layer is added with a pool size of (2,2). This layer reduces the output's spatial size from the previous layer, which helps avoid overfitting.

• Then, another 2D convolutional layer is added with 64 filters and a kernel size of (3,3). The layer applies the ReLU activation function, and it is followed by another max pooling layer with a pool size of (2,2).

• The third layer is a 2D convolutional layer with 128 filters and a kernel size of (3,3). This layer also uses the ReLU activation function, followed by another max pooling layer with a pool size of (2,2)

• The output of the previous layers is flattened into a onedimensional array and fed into a fully connected dense layer with 128 neurons and ReLU activation function. A dropout layer is added to prevent overfitting, with a dropout rate of 0.2.

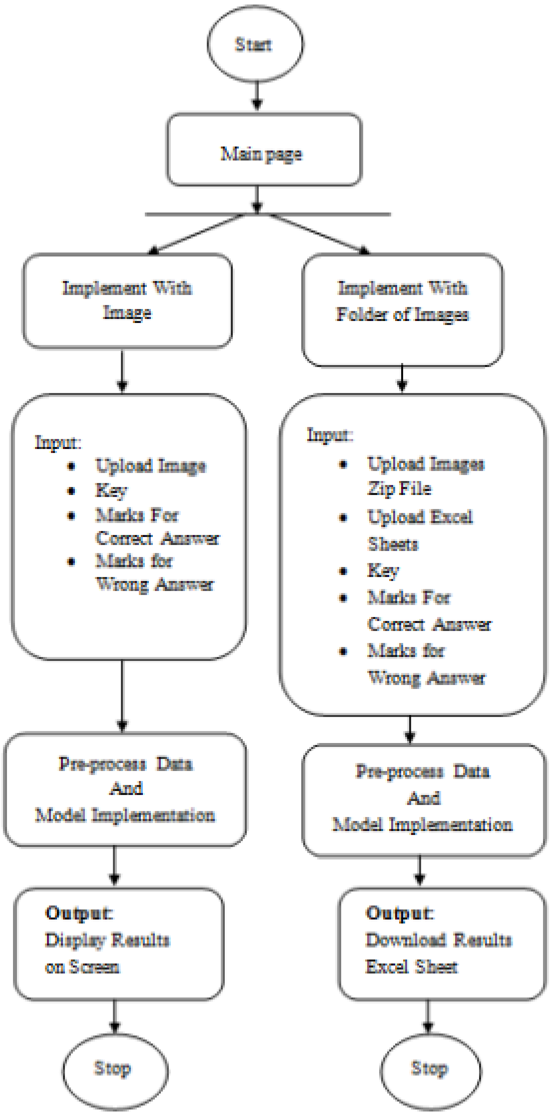
• Finally, the output is passed to a dense layer with 35 neurons and softmax activation function. The softmax function produces a probability distribution over the 35 possible classes, with each value indicating the probability of the input belonging to that class.

• In conclusion, this CNN model follows a standard architecture for image classification tasks. The successive convolutional layers extract increasingly complex features from the input image, while the max pooling layers help prevent overfitting.

The dropout layers also contribute to preventing overfitting by randomly dropping out neurons during training. The final dense layer with soft max activation function produces a probability distribution over the possible classes, allowing the model to classify input images with confidence.

• We have trained our model by up to 50 epoch .In each epoch we have gave random 2385 images as sample.

HELPFUL HINTS AND FIGURES



Flow chart

EXPERIMENTS & RESULTS

The image is taken as input and preprocessed to remove unwanted noise. Then, the image is fed into the MCQ-CNN to extract the text.

The extracted data is collected in a list and the options written by the student are extracted from within the brackets and stored in the same order in the list. After that, the list of student options is compared with the key. If the values at the same index match, marks are assigned. For a folder of images, the names are first converted to numbers, and then the

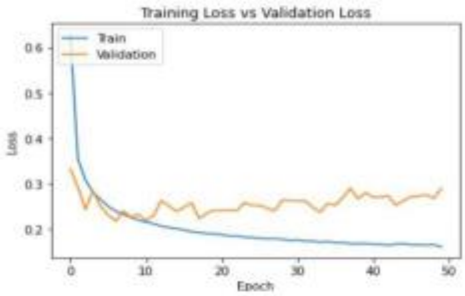


Figure 3.Loss Graph

Figure 4.Accuracy Graph

images are sorted numerically. The same procedure is followed for each image in the folder and the results are stored in an Excel sheet.. Our model is getting accuracy of 94%.

CONCLUSION

The application was developed with meticulous attention to possible errors, making it highly efficient and dependable. By automating the evaluation process, the system not only saves time but also reduces the workload of teachers, making the process more efficient. Our model is getting accuracy of 94%.

FUTURE WORK

The application is highly robust, which presents numerous opportunities for future improvisation. Once the application has been approved and authenticated, it will be implemented. Further work will involve creating an algorithm that can accurately scan OMR sheets and fill in the blanks. We will investigate this algorithm to ensure high performance and accuracy.

REFERENCES

[1] JAMSHED MEMON , MAIRA SAMI , RIZWAN AHMED KHAN AND MUEEN UDDIN Handwritten Optical Character Recognition(OCR):A Comprehensive Systematic Literature Review (SLR)Department of Computer Science, Shaheed Zulqar Ali Bhutto Institute of Science and Technology, Karachi , Pakistan . accepted July 16, 2017, date of publication July 28, 2017, date of current version August 14, 2017.

[2] SHUBHAM SANJAY MOR, SHIVAM SOLANKI, SARANSH GUPTA, SAYAM DHINGRA, MONIKA JAIN, RAHUL SAXENA "recognition handwritten text: with Deep Learning and Android" International Journal of Engineering and Advanced Technology (IJEAT) , February 2017

[3] NADA FARHANI, NAIM TERBEH AND MOUNIR ZRIGUI "Image to Text Conversion: State of the Art and Extended Work" LaTICE Laboratory, Monastir-Tunisia 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications.

[4] NISHA PAWAR, ZAINAB SHAIKH, POONAM SHINDE, PROF. Y.P. WARKE4 "Image to Text Conversion Using Tesseract" Dept. of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology, Maharashtra, India. International Research Journal of Engineering and Technology (IRJET)

2017 www.irjet.com Key Words: "Artificial Intelligence, Optical Character Recognition, Tesseract, text-to-speech".

[5] HAZEM M EL-BAKRY, ABDULAZIZ SHEHAB "Handwritten Text Recognition System based on Neural Network" prof of Mansoura University.author profiles for this publication at: https://www.researchgate.net/publication/298808334 . Article · January 2016

[6] RAMESH M. KAGALKAR "A Review On Conversion Of Image To Text As Well As Speech Using Edge Detection And Image Segmentation" K.L.E. College of Engineering & Technology. https://www.researchgate.net/publication/281281758 Article · November 2014.

[7] A S N Chakravarthy , Penmetsa V Krishna Raja , prof. P s avadhani "handwritten text image authentication using back propagation" Associate Professor Department of CSE& IT, Sri Aditya Engineering College,Surampalem, E.G.Dist , Andhra Pradesh, India. Research Scholar, JNTUK, Kakinada, E.G.Dist , Andhra Pradesh, India. Professor, Dept. of CS & SE, Andhra University, Visakhapatnam Dist, Andhra Pradesh, India.International Journal of Network Security & Its Applications (IJNSA), Sep 2011

[8] Vishwas Tanwar Manipal University Jaipur" Machine Learning based Automatic Answer Checker Imitating Human Way of Answer Checking"International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV10IS120063 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : www.ijert.org Vol. 10 Issue 12, December-2017

[9] M.F. Al-Jouie, A.M. Azmi, "Automated Evaluation of School Children Essays in Arabic," 3rd International Conference on Arabic Computational Linguistics, 2017, vol.117, pp.19-22.

[10] A.Kashi, S.Shastri and A. R.Deshpande, "A Score Recommendation System Towards Automating Assessment In Professional Courses," 2016 IEEE Eighth International Conference on Technology for Education,2016,pp.140-143.

[11] T. Ishioka and M. Kameda, "Automated Japanese essay scoring system: Jess," in Proc. of the 15th Int'l Workshop on Database and Expert Systems Applications, 2004, pp. 4-8.

[12] K.Meena and R.Lawrance, "Evaluation of the Descriptive type answers using Hyperspace Analog to Language and Selforganizing Map", Proc. IEEE International Conference on Computational Intelligence and Computing Research, 2014, pp.558-562.

[13] Vkumaran and A Sankar," Towards an automated system for short-answer assessment using ontology mapping," International Arab Journal of e-Technology, Vol. 4, No. 1, January 2015.

[14] R. Siddiqi and C. J. Harrison, "A systematic approach to the automated marking of short-answer questions," Proceedings of the 12th IEEE International Multi topic Conference (IEEE INMIC 2008), Karachi, Pakistan, pp. 329- 332, 2008.

[15] M. J. A. Aziz, F. D. Ahmad, A. A. A. Ghani, and R. Mahmod, "Automated Marking System for Short Answer examination (AMSSAE)," in Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on, 2009, pp. 47-51.

[16] R. Li, Y.Zhu and Z.Wu,"A new algorithm to the automated assessment of the Chinese subjective answer," IEEE International Conference on Information Technology and Applications, pp.228 – 231, Chengdu, China, 16-17 Nov. 2013.

[17] X.Yaowen, L.Zhiping, L.Saidong and T.Guohua," The Design and Implementation of Subjective Questions Automatic Scoring Algorithm in Intelligent Tutoring System," 2nd International Symposium on Computer, Communication, Control and Automation, Vols. 347-350, pp. 2647-2650, 2013.