IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13472

A Deep Learning Framework for Breaking Text-Based CAPTCHAs

Dr. SRINIVAS BABU P [1], DEEPTHI P [2], HARSHITHA R [3], LAVANYA Y [4], NANDITHA S [5]

Professor, Department of Electronics and Communication, East West Institute of Technology, Bengaluru, India[1]

Department of Electronics and Communication, East West Institute of Technology, Bengaluru, India[2]

Department of Electronics and Communication, East West Institute of Technology, Bengaluru, India[3]

Department of Electronics and Communication, East West Institute of Technology, Bengaluru, India[4]

Department of Electronics and Communication, East West Institute of Technology, Bengaluru, India[5]

Abstract: Websites can enhance their security and protect against malicious Internet attacks by implementing CAPTCHA verification to distinguish between human users and automated bots. Text-based CAPTCHAs are commonly used as they are easy for humans to solve but challenging for machines to decipher. This research introduces a CNN model that utilizes binary images to recognize CAPTCHAs efficiently. The project involves creating an advanced Captcha Recognition System using deep learning on a Raspberry Pi. In real-time, the Raspberry Pi processes images with the help of OpenCV, applying the trained model to authenticate captchas. This innovative approach demonstrates the practical use of deep learning on edge devices, strengthening security through automated captcha verification and showcasing the potential for IoT security solutions in real-world applications.

Key terms: Convolutional neural network; OpenCV; Automated CAPTCHA verification.

1.INTRODUCTION

In today's digital world, ensuring internet security and verifying users are crucial to safeguard online platforms from malicious activities like bots, spam, and fraud. CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) have been widely adopted to address these issues. CAPTCHAs test a user's ability to solve puzzles or recognize distorted characters, distinguishing humans from automated scripts. However, with advancing technology, traditional CAPTCHAs are becoming less effective against sophisticated automated tools. This has created a demand for more robust solutions, leading to the integration of deep learning techniques as a promising approach. This project focuses on developing a CAPTCHA recognition system using deep learning. Deep learning, a subset of artificial intelligence, has shown impressive capabilities in tasks like image and text recognition. By utilizing deep neural networks, the project aims to address the challenge of CAPTCHA recognition by training a model to accurately solve various types of CAPTCHAs.

A. PROBLEM STATEMENT

The CAPTCHA recognition system involves in developing a system that can accurately and efficiently recognize and figure out CAPTCHAs to distinguish between human users and automated bots. The challenge lies in creating a robust algorithm that can effectively process and interpret the distorted characters that are commonly used in CAPTCHAs. The system must be able to handle various types of CAPTCHAs with different levels of complexity and noise, ensuring high accuracy in identifying and extracting the correct characters. Additionally, the system should be capable of real-time

DOI: 10.17148/IJARCCE.2024.13472

processing to provide timely verification and authentication for online users. The goal is to design a reliable and scalable solution that enhances internet security by preventing unauthorized access and fraudulent activities on digital platforms.

B. OBJECTIVES

Develop a system that can accurately recognize and figure out CAPTCHAs with a high level of precision to ensure reliable verification of human users. Develop a compact and portable system that can be easily deployed in various environments and scenarios. Design a system that can efficiently process CAPTCHAs using the computational power of Raspberry Pi while maintaining high accuracy.

2. METHODOLOGY

The system utilizes deep learning, particularly the Convolutional Neural Network (CNN), to develop a model for recognizing captchas. The Raspberry Pi is used to handle the dataset processing. Once the model is trained, it is incorporated into the Raspberry Pi setup for automated captcha recognition and verification. This approach merges cutting-edge technologies to build a smart and efficient captcha recognition system.

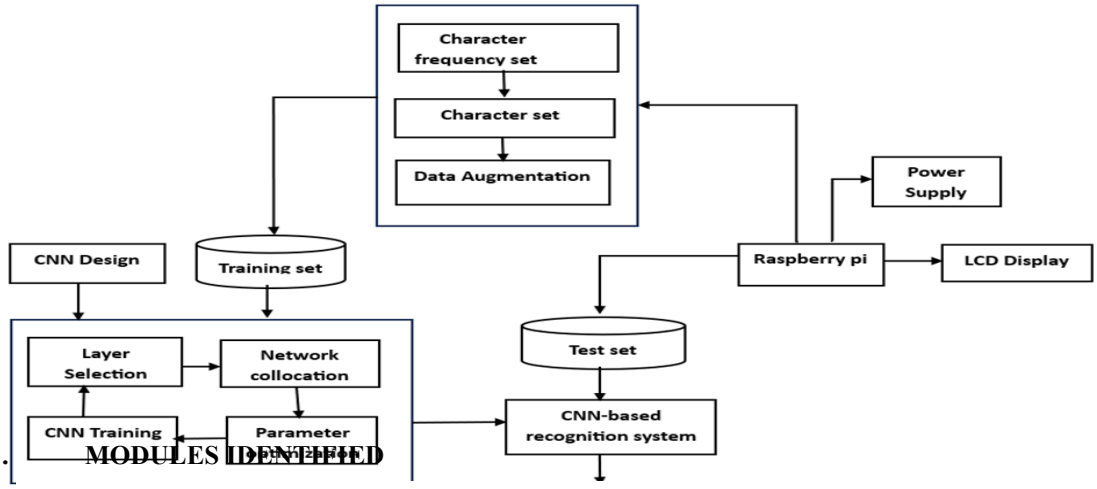

Fig 1 : BLOCK DIAGRAM

1. MODULES IDENTIFIED

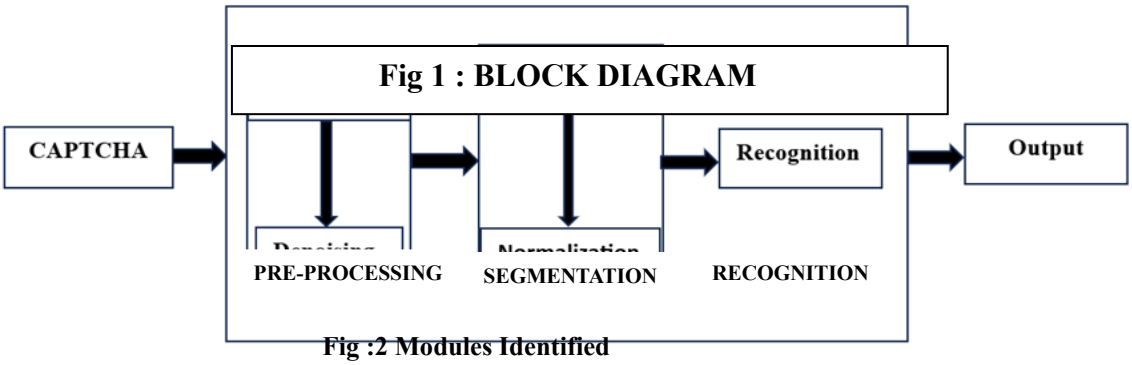

Fig :2 Modules Identified

1. Preprocessing Module

The CAPTCHA images are pre-processed using raspberry pi, including tasks such as resizing, normalization, and noise reduction to optimize them for input to the deep learning model.

2. Deep Learning Model Training Module:

In this module, a Convolutional Neural Network (CNN) is designed and trained using a diverse captcha dataset. The CNN learns to recognize patterns and features in captcha images, enabling accurate verification during the runtime phase.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.13472

3. **Captcha Recognition Module:**

The trained CNN is integrated into the Raspberry Pi system to perform captcha recognition. CAPTCHA images are fed into the model, and the system outputs the predicted CAPTCHA.

4. **Verification Module:**

In this module, an OTP will be sent to the authorized user. If the correct OTP is entered by the user, the CAPTCHA will be recognized, auto-filled, and the user will be directed to the website.

3. EXISTING SYSTEM

CAPTCHAs typically involve the use of convolution neural networks (CNNs) and recurrent neural networks (RNNs) to analyze and decipher the text within CAPTCHAs

These frameworks are trained on large datasets of labeled CAPTCHA images to learn patterns and features that help in accurately recognizing and breaking text-based CAPTCHAs

The use of attention mechanisms and generative adversarial networks (GANs) to enhance the performance of these frameworks in braking text-based CAPTCHAs. Some frameworks also incorporate techniques like transfer learning and data augmentation to improve the generalization and robustness of the models when dealing with unseen CAPTCHA variation.

The ultimate goal of these deep learning frameworks is to create efficient and reliable system that can bypass text base4d CAPTCHAs with high accuracy and speed, and security for such systems

4. IMPLEMENTATION

Data Collection: Gather a dataset of captcha images for training the CNN model. These data set should include a variety of CAPTCHA styles and characters to ensure robust recognition

Data Preprocessing: Preprocess the captcha images by resizing, normalizing, and augmenting them to improve the model's performance

Model Training: Develop and train a CNN model using a deep learning framework like TensorFlow and PyTorch. Design the architecture of the CNN to effectively learn and recognize patterns in the CAPTCHA images

Model Optimization: Fine-tune the CNN model by adjusting hyperparameters optimizing the learning rate, and using techniques like dropout and batch normalization to improve performance

Development on Raspberry Pi: Once the CNN model is trained and optimized, deploy it on the Raspberry Pi for CAPTCHA recognition. Libraries like TensorFlow lite is used for running the model efficiently on the Raspberry Pi's hardware

Integration with CAPTCHA System: Integrate the Raspberry Pi with the CAPTCHA system to process incoming captcha images, run them through CNN model for recognition, and provide the output for verification

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13472

5. MOTIVATION

CAPTCHAs are designed to prevent automated scripts from accessing websites or performing malicious activities. By developing effective CAPTCHA recognition systems, we can ensure that legitimate users can easily access online services while keeping bots at bay.

6. RESULTS

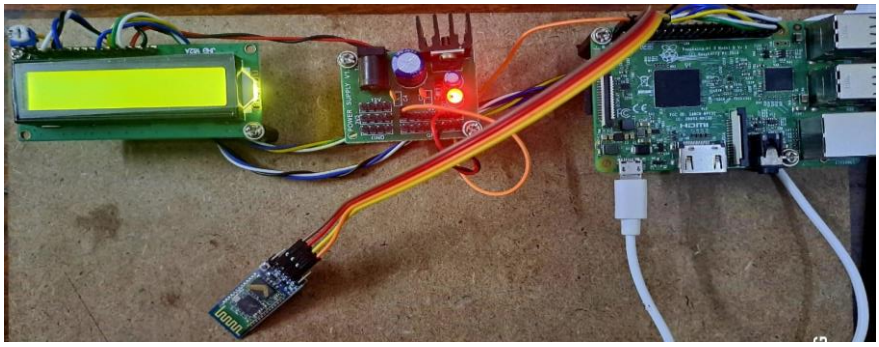

Fig:3 PROTOTYPE OF THE MODEL

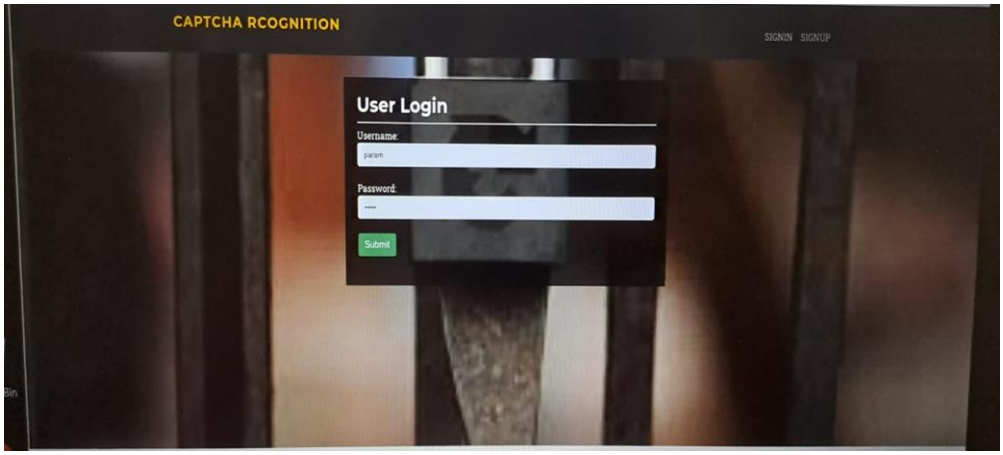

Fig:4 GUI login page

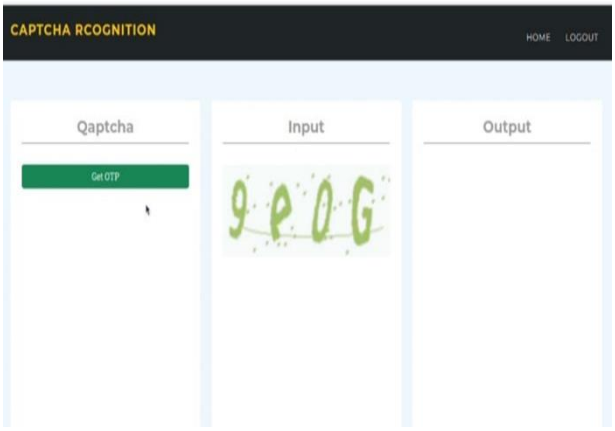

Fig:5 The CAPTCHA that is generated after filling in the login credentials.

Fig:6 The OTP that has been sent to the user through Telegram for verification, in order to automatically recognize and fill in the CAPTCHA.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.13472

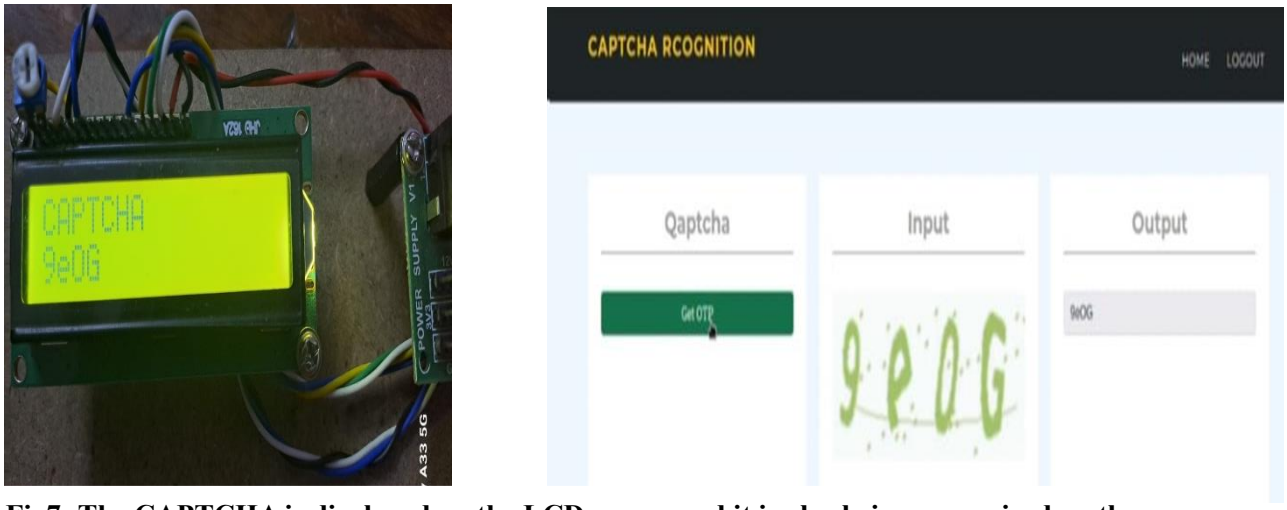

Fig7: The CAPTCHA is displayed on the LCD screen and it is also being recognized on the page as well.

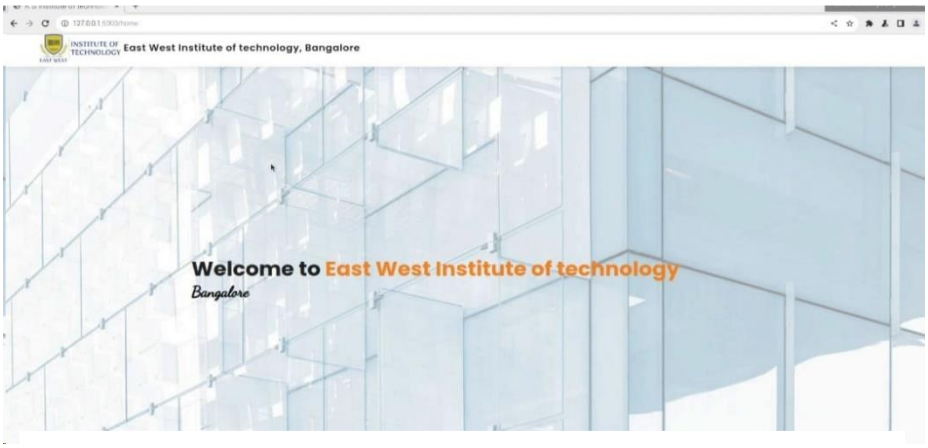

Fig:8 Once the CAPTCHA is auto-filled, it directs to the website

7. CONCLUSION AND FUTURE SCOPE

CONCLUSION: The project has proven to be a valuable tool in enhancing online security and user experience. By effectively recognizing and auto-filling CAPTCHAs, the project has streamlined website access and reduced user frustration. It highlights its potential to improve online security measures and simplify user interactions. Further advancements in CAPTCHA recognition technology can continue to enhance the efficiency and effectiveness of online security protocols, ultimately benefiting both website owners and users alike.

FUTURE SCOPE

- By using the advancements in CNN architectures specifically tailored for captcha recognition tasks, leading to higher accuracy rates and improved performance on Raspberry Pi devices.

- Additionally, the integration of biometric authentication methods and behavioral analysis can further strengthen CAPTCHA recognition systems, providing a seamless and secure user experience. Overall, the future of CAPTCHA recognition is promising, with continuous innovation paving the way for improved online security solutions.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.13472

REFERENCES

[1] S. Jain and V. Gupta, "Robust Captcha Recognition using Ensemble of Deep Learning Models," IEEE Transactions on Information Forensics and Security, vol. 15, no. 2, pp. 500-510, 2023."Improving CAPTCHA Recognition Using Generative Adversarial Networks" by N. Patel, R. Mehta, and A. Desai. (2023)
[2] G. Chen and H. Li, "Captcha Recognition with Attention Mechanism," IEEE Transactions on Multimedia, vol. 22, no. 3, pp. 800-810, 2022."Captcha Recognition with Domain Adaptation and Few-shot Learning" by S. Singh, A. Sharma, and M. Agarwal. (2022)
[3] Wang, P.; Gao, H.; Rao, Q.; Luo, S.; Yuan, Z.; Shi, Z. A Security Analysis of Captchas with Large Character Sets. IEEE Trans. Dependable Secure. Comput. (2020)
[4] Tang, M.; Gao, H.; Zhang, Y.; Liu, Y.; Zhang, P.; Wang, P. Research on Deep Learning Techniques in Breaking Text-Based Captchas and Designing Image-Based Captcha. IEEE Trans. Inf. Forensics Secur. (2020), 13,2522–2537.
[5] R. Sharma and S. Kumar, "Captcha Recognition using Capsule Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2020), pp. 600-605.
[6] K. Gupta and S. Singh, "Transfer Learning for Captcha Recognition," IEEE Signal Processing Letters, vol. 26, no. 8, pp. 1200-1205, 2019.
[7] J. Wang and Q. Liu, "Captcha Recognition using Recurrent Neural Networks," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 400-405.
[8] M. Li and H. Zhang, "Adversarial Attack on Captcha Recognition Models," IEEE Transactions on Information Forensics and Security, vol. 13, no. 5, pp. 1200-1210, 2018.
[9] A. Patel and R. Shah, "Captcha Recognition using Deep Learning Models," IEEE Access, vol. 6, pp. 5000-5010, 2018
[10] S. Gupta and R. Verma, "Improved Captcha Recognition using Convolutional Neural Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 300-305.

