Automated CAPTCHA Detection through Machine Learning Models: A Survey

Anjali Kushwaha[1], Prof. Vipin Kasera[2]
Research Scholar[1], Asst. Professor[2]
VITM, Indore, India[1,2]

***Abstract*— As artificial intelligence is improving its purview in matching human level intelligence, it is necessary to identify human and machine learning intelligence. One of the ways is to use a CAPTCHA, or "Completely Automated Public Turing test to tell Computers and Humans Apart," is a security measure that helps distinguish between human users and automated bots. It is widely used across websites to prevent bots from performing certain actions, such as creating accounts or submitting forms, to safeguard against spam, data breaches, and unauthorized access. With advances in machine learning, particularly in computer vision and natural language processing, CAPTCHA systems have been both challenged and enhanced. Machine learning models trained to recognize CAPTCHAs attempt to decode visual and audio puzzles designed specifically to be hard for computers to solve. This paper presents a comprehensive review on machine learning and deep learning algorithms for CATPCHA recognition with the salient features of contemporary work.**

Keywords*— *CAPTCHA, Machine Learning, Deep Learning, Object Detection, Classificaiton Accuracy.

I. INTRODUCTION

CAPTCHAs come in various formats, such as text-based, image-based, and audio-based. Text-based CAPTCHAs often present distorted text images, requiring users to interpret and type the text shown. Image-based CAPTCHAs may ask users to identify images of a specific object from a set, while audio-based CAPTCHAs require interpreting distorted sound files. For machine learning models, recognizing and solving these CAPTCHAs presents numerous challenges. Text-based CAPTCHAs use distortions, rotations, and noise, complicating character recognition. Image-based CAPTCHAs rely on the ability to identify specific objects, demanding a model's understanding of visual patterns and context [1].

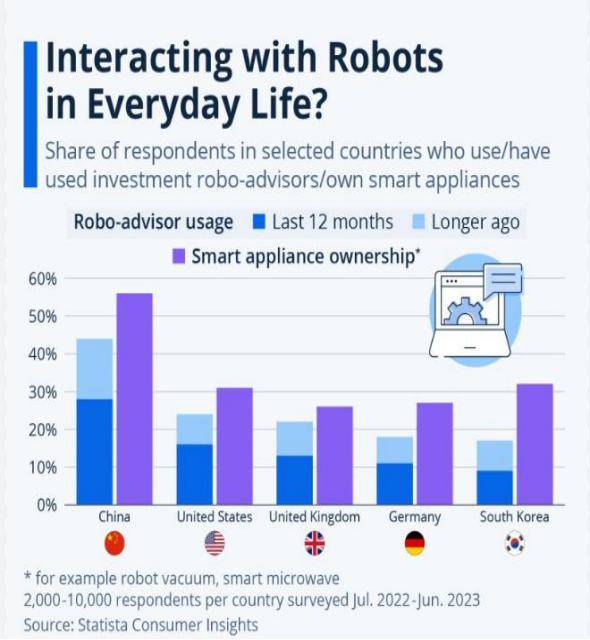

Fig.1 Automated Interactions (Source: Statista)

The application of machine learning to CAPTCHA recognition raises ethical and security concerns. On one hand, CAPTCHA-breaking technology can be exploited by malicious entities, enabling automated bots to bypass security layers, leading to spam, hacking attempts, and data theft. This undermines the purpose of CAPTCHAs and threatens user privacy and system integrity. Conversely, advancing CAPTCHA-breaking techniques helps developers understand the vulnerabilities in CAPTCHA systems, allowing them to create more robust security mechanisms [2]. Researchers in cybersecurity study CAPTCHA-breaking models to refine CAPTCHA designs that are harder for machines yet accessible for human users [3].

II. AOTOMATED OBJECT DETECITON MODEL FOR CAPTCHA RECOGNITION

To bypass CAPTCHAs, researchers have developed machine learning models using techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and more recently, Transformer-based architectures. CNNs are particularly effective for image-based CAPTCHAs, as they can detect shapes and patterns, while RNNs can handle sequential data, which can be beneficial for text-based CAPTCHA recognition. Transformer models, known for their success in natural language processing, are also being adapted for CAPTCHA breaking due to their capacity to handle complex contextual relationships in both text and images. Training these models requires large datasets of CAPTCHAs and computational resources to ensure accuracy in identifying patterns under various distortions [4].

Image Processing:

Prior to computing important parameters or feature of the fundus image, which lays the foundation for the final classification, it is necessary to process the image for the following reasons [5]-[6]:

Illumination Correction: In this part, the inconsistencies in the image illumination are corrected so as to make the image background uniform and homogenous. Illumination inconsistencies occur due to capturing the image from different angles which makes the reflection from the retinal image variable rendering inconsistencies. Inconsistencies in the illumination can be caused due to the position and orientation of the source, the non-homogeneity of wavelengths of the source, the nature of the surface such as smoothness, orientation and material characteristics and finally the characteristics of the sensing device such as resolution, capturing capability and sensitivity [7].

Segmentation: This process of separating the area under interest from the composite image is called segmentation The segmentation is typically a threshold based segmentation since the parts to be separated are not generally regular shapes. The segmentation is generally done adopting the sudden change in pixel characteristics given by the gradient. The gradient based method allows to find the maximum change in the pixel intensities to perform the thresholding so as to separate out the vessels. Further the inpainting can be performed based on the neighbouring pixel information and the stochastic characteristics utilizing the fact that image regions generally comprise of highly redundant values [8].

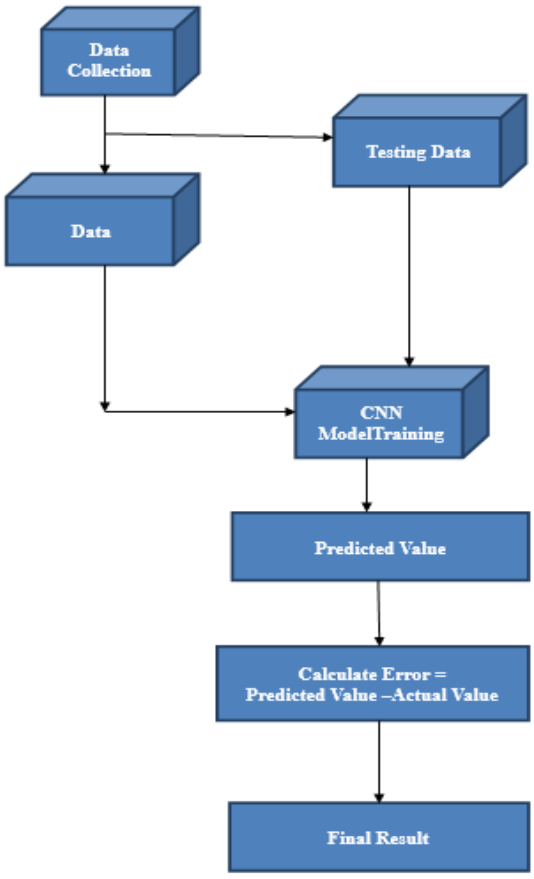

Fig.2 Architecture of Machine Learning Based Models for CAPTCHA Recognition

Figure 2 depicts the generic model for a machine learning or deep learning model for CAPTCHA recognition.

The most common machine learning and Deep Learning models are presented next [9]-[10]:

Convolutional Neural Networks (CNN): Neural network models are very effective for classification problems. If the neural network has multiple hidden layers, then such a neural network is termed as a deep neural network. The training algorithm for such a deep neural network is often termed as deep learning which is a subset of machine learning. Typically, the multiple hidden layers are responsible for computation of different levels of features of the data. Several categories of neural networks such as convolutional neural networks (CNNs),

Recurrent Neural Network (RNNs) etc. have been used as effective classifiers [11].

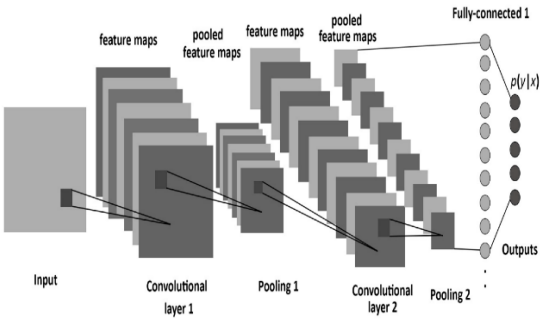

Fig.3 The CNN Model

The CNN is an extremely effective deep learning based classifier which performs pattern recognition in each of its layers based on stochastic computing. The fundamental operation in the CNN hidden layers is the convolution operation mathematically computed using equation 1:

$$x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \qquad (1)$$

Here,
x(t) is the input
h(t) is the system under consideration.
y is the output
*is the convolution operation in continuous domain
For a discrete or digital counterpart of the data sequence, the convolution is computed using equation 2:

$$y(n) = \sum_{-\infty}^{\infty} x(k)h(n-k) \qquad (2)$$

Here
x(n) is the input
h(n) is the system under consideration.
y is the output
*is the convolution operation in discrete domain

Region-Based Convolutional Neural Networks (R-CNNs): Region-based Convolutional Neural Networks (R-CNNs) represent an advancement over standard CNNs, especially in object detection tasks. R-CNNs divide an image into regions and classify each region, which allows for precise identification of individual objects within complex scenes. This model uses selective search to propose possible object regions, and each region is then fed into a CNN for classification. Variants like Fast R-CNN, Faster R-CNN, and Mask R-CNN build upon this approach by making region selection and classification faster and more efficient. In CAPTCHA detection, these models are particularly useful for CAPTCHAs requiring object localization within images, such as selecting all traffic lights or animals [12].

You Only Look Once (YOLO) Models: YOLO, or You Only Look Once, is a real-time object detection model that has gained popularity for its speed and efficiency. Unlike R-CNNs, which process multiple regions, YOLO predicts objects and their bounding boxes in a single pass. This architecture enables faster processing and is suitable for high-throughput CAPTCHA-breaking applications where quick response times are essential. YOLO's capability to detect multiple objects within a single frame makes it highly effective for CAPTCHA images with multiple object instances, such as choosing all images containing specific objects like cars or bicycles.

Single Shot MultiBox Detector (SSD): Single Shot MultiBox Detector (SSD) is another model designed for efficient and real-time object detection, similar to YOLO. SSD combines object classification and bounding box prediction within a single forward pass, providing a balance between detection speed and accuracy. Unlike YOLO, SSD utilizes a multi-scale feature extraction technique, making it more adaptable for detecting small objects or objects with diverse aspect ratios. For CAPTCHAs, this is advantageous as many CAPTCHA images involve detecting smaller objects within cluttered scenes. SSD models are frequently used when object localization must be precise, and processing efficiency is paramount.

Bayesian Networks: Bayesian Networks, or BayesNets, are probabilistic graphical models that represent variables and their conditional dependencies via a directed acyclic graph (DAG). In CAPTCHA detection, Bayesian Networks offer a distinct approach by leveraging probability-based inference rather than traditional deep learning models [13].

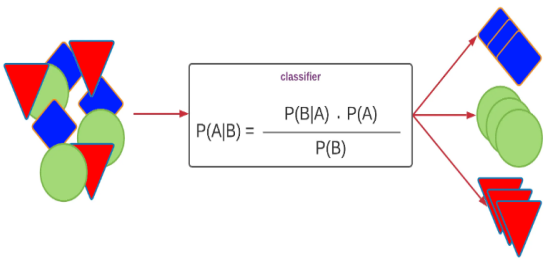

Fig.4 The Bayesian Model

This statistical framework is particularly advantageous in scenarios where there are clear dependencies between different CAPTCHA components, such as sequential or layered character-based CAPTCHAs. By using BayesNets, CAPTCHA detection systems can model uncertainty and make probabilistic decisions, a feature that proves useful in handling the distortions and noise commonly found in CAPTCHA images.

III. EXISTING WORK

This section presents a review on the baseline approaches in the domain.

Hu et al. proposed a method based on Convolutional Neural Network (CNN) model to identify CAPTCHA and avoid the traditional image processing technology such as location and segmentation. The adaptive learning rate is introduced to accelerate the convergence rate of the model, and the problem of over-fitting and local optimal solution has been solved. The multi task joint training model is used to improve the accuracy and generalization ability of model recognition. The experimental results show that the model has a good recognition effect on CAPTCHA with background noise and character adhesion distortion.

Sinha et al. propose a method to crack CAPTCHA using a custom based convolutional neural network (CNN) model called CAP-SECURE. The proposed model aims to distinguish or tell web sites about the weaknesses and vulnerabilities of the CAPTCHAs. The CAP-SECURE model is based on sequential CNN model and it outperforms the existing CNN architecture like VGG-16 and ALEX-net. The model has the potential to solve and explore both numerical and alphanumerical CAPTCHAs. For developing an efficient model, a dataset of 200000 CAPTCHAs has been generated to train our model. In this exposition, we study CNN based deep neural network model to meet the current challenges, and provide solutions to deal with the issues regarding CAPTCHAs. The network cracking accuracy is shown to be 94.67 percent for alpha-numerical test dataset. Compared to traditional deep learning methods, the proposed custom based model has a better recognition rate and robustness.

Kimbrough et al. proposed deep learning techniques to recognize text-based CAPTCHAs. In particular, the work focuses on generating training datasets using different CAPTCHA schemes, along with a pre-processing technique allowing for character-based recognition. We have encapsulated the CRABI (CAPTCHA Recognition with Attached Binary Images) framework to give an image multiple labels for improvement in feature extraction. Using real-world datasets, performance evaluations are conducted to validate the efficacy of our proposed approach on several neural network architectures (e.g., custom CNN architecture, VGG16, ResNet50, and MobileNet). The experimental results confirm that over 90% accuracy can be achieved on most models.

Kumar et al. presented a review of multiple schemes especially based on the English language, are successfully broken with a success rate that ranges from 2 to 100%. The techniques that are used to break these schemes include shape context matching, distortion estimation, Log Gabor 2D filter, horizontal and vertical projection (for a segment the letters) are used. For recognition CNN, KNN, DNN and MCDNN are used. Almost 15 images-based CAPTCHAs are discussed that are designed with usability and security range 90–100 and 17–100%, respectively. Out of these 5 schemes are successfully broken with a success rate ranging between 7 and 100%. The K-NN and SVM are mostly used algorithms to recognize the images. Audio based CAPTCHAs (5 designs) are discussed with usability and security range from 68.5 to 100 and 100%, respectively. The broken rate of these audio schemes is also 45–75%. These schemes are broken with SVM and K-NN algorithms. The paper also discusses 4 popular video-based designs that provide usability and security that ranges from 75 to 100 and 98 to 100, respectively. These schemes are also compromised with broken rate 16–10% using SIFT, NN and simple OCR techniques. The paper can be a benchmark to precede any specific research to dive into any one of these types.

Mathew et al. proposed a DOCR-CAPTCHA model has presented which is deep learning approach based on an optical character recognition (OCR) to address the concerns of lower efficiency and inadequate performance of existing CAPTCHA detection algorithms. First, the DOCR-CAPTCHA model preprocesses the images to enhance the quality by following gray scale conversion,

cropping and resizing the image. Second, it extracts the character to create a dictionary and mapping each character with labeling. Next, it performs classification using OCR technique and train the model. It also performed the validation on the same data. The simulation has done on the CAPTCHA images dataset. The recognition rate of this simulated model results achieved a high accuracy rate of 99.98 percent and minimized error rate of 0.0051 for the CAPTCHA train dataset. It has also compared with the existing YOLO technique and found that it has outperformed than YOLO.

IV. PERFORMANCE METRICS AND CHALLENGES:

The performance metrics of the classifiers are generally computed based on the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values which are used to compute the following parameters [14]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

$$Precisiosn = \frac{TP}{TP+FP} \quad (6)$$

$$F - Measure = \frac{2.Precision.Recall}{Precision+Recall} \quad (7)$$

The aim of any designed approach is to attain high values of accuracy of classification along with other associated parameters

As machine learning models evolve, CAPTCHA systems will likely need to adapt to remain effective. Future CAPTCHAs may integrate advanced elements such as behavioral biometrics, like mouse movements or typing patterns, which are harder for bots to mimic. Alternatively, more interactive CAPTCHAs, involving logical puzzles or games, could present further challenges to automated bots. With the rapid progression in AI, CAPTCHAs are expected to leverage AI themselves, making real-time adjustments based on the latest bot-detection insights. The continuous race between CAPTCHA technology and machine learning capabilities highlights the dynamic nature of cybersecurity [15].

CONCLUSION: CAPTCHA recognition using machine learning is a double-edged sword, providing insights into CAPTCHA vulnerabilities while challenging the security system's integrity. While machine learning techniques can decode complex CAPTCHAs, they also push for the development of stronger, more adaptive CAPTCHA systems. In the future, CAPTCHAs will need to balance usability for humans with increased complexity for machines, ensuring they remain effective against evolving AI-driven attacks. This interplay between CAPTCHA and machine learning reflects the broader, ongoing battle in cybersecurity, where adversarial innovations shape stronger defense strategies.

References

1. Yu Hu , Li Chen , Jun Cheng, “A CAPTCHA recognition technology based on deep learning”, IEEE 2023.
2. S. Sinha and M. I. Surve, "CAPTCHA Recognition And Analysis Using Custom Based CNN Model - Capsecure," 2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), Hyderabad, India, 2023, pp. 244-250.
3. T. Kimbrough, P. Tian, W. Liao, E. Blasch and W. Yu, "Deep CAPTCHA Recognition Using Encapsulated Preprocessing and Heterogeneous Datasets," IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), New York, NY, USA, 2022, pp. 1-6.
4. M Kumar, MK Jindal, M Kumar, “A systematic survey on CAPTCHA recognition: types, creation and breaking techniques”, Archives of Computational Methods in Engineering, Springer 2022, vol.107, pp.1107-1136.
5. A. Mathew, A. Kulkarni, A. Antony, S. Bharadwaj and S. Bhalerao, "DOCR-CAPTCHA: OCR Classifier based Deep Learning Technique for CAPTCHA Recognition," 2021 19th OITS International Conference on Information Technology (OCIT), Bhubaneswar, India, 2021, pp. 347-352.
6. Gaihuan An, Wanjun Yu, “CAPTCHA Recognition Algorithm Based on the Relative

Shape Context and Point Pattern Matching", IEEE 2021

7. Ye Wang ,Yuanjiang Huang ,Wu Zheng ,Zhi Zhou ,Debin Liu ; Mi Lu, "Combining convolutional neural network and self-adaptive algorithm to defeat synthetic multi-digit text-based CAPTCHA", IEEE 2020
8. Suphannee Sivakorn, Iasonas Polakis and Angelos D. Keromytis, " I Am Robot: (Deep) Learning to Break Semantic Image CAPTCHAs", IEEE 2019
9. Honey Mehta, Sanjay Singla, Aarti Mhajan , "Optical Character Recognition (OCR) System for Roman Script & English Language using Artificial Neural Network (ANN) Classifier", IEEE 2018
10. Zhuoyao Zhong, Lianwen Jin, Zecheng Xie , "High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps", IEEE 2017
11. Fabian Stark, Caner Hazırba¸s, Rudolph Triebel, and Daniel Cremers, "CAPTCHA Recognition with Active Deep Learning", Research Gate 2016.
12. R. Hussain, H. Gao, R. A. Shaikh and S. P. Soomro, "Recognition based segmentation of connected characters in text based CAPTCHAs," 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), Beijing, China, 2016, pp. 673-676.
13. S. Tingre and D. Mukhopadhyay, "An approach for segmentation of characters in CAPTCHA," International Conference on Computing, Communication & Automation, Greater Noida, India, 2015, pp. 1054-1059.
14. Q. Li, "A computer vision attack on the ARTiFACIAL CAPTCHA", Multimedia Tools and Applications, Springer 2015, vol. 75, pp. 4583–4597.
15. V Novák, P Hurtík, H Habiballa, M Štepnička, "Recognition of damaged letters based on mathematical fuzzy logic analysis", Journal of Applied Logic, Elsevier, 2015, vol.13, no.2, pp. 94-104.