# AN INNOVATIVE METHOD TO IMPROVE THE ACCURACY IN CLASSIFICATION OF CAPTCHA RECOGNITION BY USING RANDOM FOREST ALGORITHM OVER SUPPORT VECTOR MACHINE

**Gudaru Gopi Prasad[1], K. Malathi[2*]**

**Article History: Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

**Abstract**

**Aim:** Recognition of captcha to find the Best accuracy of the text-based captcha by using the Algorithms in Machine Learning. The two algorithms are the Random Forest algorithm and the Novel Support Vector Machine (SVM).
**Materials and Methods:** The dataset is collected from www.kaggle.com. And the Two groups are Random Forest (N=10) and Support Vector Machine (N=10) by using G-power and minimum power of the analysis is fixed as 80% and maximum accepted error is fixed as 0.5 with threshold value as 0.0805% and Confidence Interval is 95%.
**Results:** The Novel Support Vector Machine is used to recognize the text-based captcha. The accuracy found for the improved captcha is 99.98% and for the Random Forest is 98.93%. The two algorithms are used to find the improved classification or complexity of the captcha. The significant value obtained is $p=0.004$ ($p<0.05$) i.e $\alpha=0.05$ and hence, there exists a statistically significant difference between the two groups with a confidence level of 95%.
**Conclusion:** Recognizing the captcha Recognition significantly seems to be better in Novel Support vector machines than Random Forest.

**Keywords:** Random Forest, Novel Support Vector Machine, Machine Learning, Recognition, complexity, Text-based, CAPTCHA Recognition.

1Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, 602105.
[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, 602105.

## 1. Introduction

It's Simplicity and diversity as it can be used for both classification and regression. All are Considering the 100 n_estimators before taking the average of the prediction. N_jobs is an integer,specifying the maximum number if concurrently running works. All have set that variable(-1) (Abdelhafiz et al. 2021). Captcha is reverse turing take a look at to classify the people and the computer system and also to improve the accuracy of the captcha using digital image preprocessing. It can increase the complexity of a captcha Recognition (Tursunov et al. 2021)**.** Nowadays captcha has mostly text-based formatly. Normally then at that point, in case all had the option to bunch the pixel so as to put the text-based and the foundation into distant discrete bunches effectively perceive the text-based. This captcha will be used or help us from the Hackers. By using this captcha ,the captcha has the main stream on the website to be accessed (Jia et al. 2018; Shu and Xu 2019), and Random Forest Algorithm for Captcha Recognition(Kwon, Yoon, and Park 2020). Also this article Captcha recognition and trends has become even more popular than before(Kwon, Yoon, and Park 2020). The one more article is a survey on machine learning for Captcha recognition algorithms and techniques(Huber 2019; Sha 2011). Mainly the Novel Support vector Machine algorithm, which is used for Captcha recognition with better performance .

Captcha Recognition can be carried out from the researchers. There are 120 articles found on IEEE, and 624 articles were found in the Google Scholar. And the captcha recognition using Random Forest can be found to be 98.93% with the accuracy and then the captcha recognition by characters using a Novel Support vector machine this algorithm brings the 99.98% (Mhamed et al. 2021; Shen et al. 2021). The captcha recognition also used the arabic languages like chinese language using digital image preprocessing with the accuracy of 97.85% (Kwon, Yoon, and Park 2020) finally the key step of the captcha segmentation to extract individual letters has been reasonably successful with accuracy of 96.45% (Z. Wang and Li 2015).

Our team has extensive knowledge and research experience that has translated into high quality publications (K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022). Considering the composing study, the Random Forest has very little precision (accuracy), the exactness of the image is shown at an outstandingly low rate while examining the image and the manual data is ludicrous to hope to add to the dataset (Tamir 2019). The survey intends to chip away at the precision (accuracy) of the images, further fostering the exactness rate of the apparent images, and diminish the lack of data while planning and testing the dataset and also training the dataset(Alaidi, Alsaidi, and Yahya 2017). The aim of this Random Forest algorithm is implemented by using Novel Support Vector Machine to improve accuracy of Captcha Recognition.

## 2. Methods and Materials

The study setting of the proposed work is done in Saveetha School of Engineering, OOAD lab. The sample size was calculated by using clincalc.com by keeping G power (Mayr et al. 2007)and minimum power of the analysis is fixed as 0.8 and maximum accepted error is fixed as 0.5 with threshold value as 0.05% and Confidence Interval is 95%. Mean and standard deviation has been calculated based on the previous literature for size calculation. The two groups are used, namely Random Forest (N=10) as an existing model as group 1 and Support Vector Machine (N=10) as a Proposed model as group 2.

**Data Preparation**

The Random Forest is to find all the images that are stored in the dataset, to train and test through the The dataset comes from (Rodrigo, n.d.; Thomas and Kaur 2013). The dataset includes 10000 data in the form of images which are taken as a sample from www.kaggle.com with their respective captcha recognition. There are 1000 trained images and 9000 tested images. The sample images of digits present in the (Rodrigo, n.d.; Thomas and Kaur 2013) and dataset has been shown in Fig. 1 with the 36*36 pixel size .

**Random Forest Algorithm**

In utilizing the Random timberland classifier as our second calculation due to its effortlessness and variety as it very well may be utilized for both arrangement and relapse undertakings. All are thinking about 100 n_estimators prior to taking the normal of expectation. N_jobs is a whole number, determining the most extreme number of simultaneously running specialists, all have set that variable to (- 1). And Fig. 2 shows the steps in the implementation below.

- Firstly Data digital image Preprocessing.
- In Data preprocessing,there are data measurement and data segmentation.
- In Classification we have feature extraction and decision tree followed by the Random

forest and finally voting to recognise the captchas.
- Finally the recognition results.

**Support Vector Machine (SVM)**

By utilizing Novel Support vector machine calculation to do the captcha acknowledgment. all might utilize that equivalent theta to move the obscure information to group it by the straight Support vector machine model. This Novel Support vector machine model will work on the exactness of the other calculation or model. Then the default number of Novel Support vector machine components is shaky, it has prevailed with regards to accomplishing 92% free digits, yet just 44% of the test components have been accomplished. With the boundary all have upgraded, the acknowledgment of the complexity captcha sound document has reached 89%. Free digital image preprocessing achievement has arrived at a generally excellent worth of 98%. And Fig. 3 shows the steps in the implementation below.

- Downloading the dataset to stack.
- Introduce the factors to prepare and test the information.
- Characterize a model. fit () capacity to depict the parts which are to be gotten to for running the code to get precision.
- Characterize Categorical () capacity to arrange the information.
- Print the model. fit () work with the necessary ages and discover the exactness.

For comparing both the models, the dataset has been trained with ten different sample sizes. The accuracy values are recorded. The system configuration is used for the algorithm to run in a 64 - bit Operating System, 4GB RAM PC, and using Windows 10, Google Colab, and Microsoft Office for software specification. To evaluate the show of the preparation model, the data has been separated for preparing and testing to endorse the dataset. Then, stack and reshape the data bunches to organize the numbers. Normalize the pixel potential gains of grayscale pictures. All the layers will be worked through the ReLU order ability to the outright cross_entrophy to find the disaster work. The model will be evaluated with the fit() work which has the estimation ability to Novel Support the accuracy and loss of the data.

**Statistical Analysis**

The SPSS statistical software was used in the research for statistical analysis. Group statistics and independent sample t-tests were performed on the experimental results and the graph was built for two groups with two parameters under study. The independent variables are Datasets. The dependent variables are shape,size,Accuracy.The independent T test analysis was carried out to calculate the accuracy for both methods(Shu and Xu 2019).

## 3. Results

Random Forest estimation outlines the layers with all of the images of each number, whenever it runs at different occasions in light of the presentation of test size (N=10). The layers are formed as a result of the cycles, the accuracy regards changes with the range of running time and conveys the precision and setback concerning the period which is shown in Table 1. Support vector machine has best accuracy and less fiasco over the Random Forest because of as far as possible and assessments, the Random Forest calculation has not sensible with the high level enactment limits which are basically confined to the adam, adaleta, and adagrad which takes additional time and the cutoff points are not taking the entire information to isolate the equivalent digits in the dataset through the Novel Support vector machine takes the information and plans layers with every digit autonomously at last gives the outcome. Concerning as far as possible, the Accuracy and disaster have changed and have shown that the Novel Support vector machine is superior to the Random Forest. Table1 represents the data collection from the N=10 samples of the dataset for Random Forest with the size of 36*36 pixels to gain accuracy (%) and support vector machines to gain accuracy (%).

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{-- (1)}$$

Where, in equ(1)
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

The IBM SPSS version 21 statistical software is used for our study. The independent variables are shape and size and the dependent variable is accuracy (%) for our study captcha Recognition. In SPSS, the datasets are prepared using N=10 as the sample size for Random Forest and Support Vector Machine. GroupID is given as a grouping variable and accuracy is given as the testing variable. GroupID is given as 1 for Random Forest and group 2 for support vector machines. Group Statistics is applied for the Statistical Package for the Social Sciences (SPSS) dataset and shown in Table 2. By performing the statistical analysis group statistics represents the comparison of the accuracy and Loss of captcha recognition of Random Forest and support vector. The support vector machine algorithm had the highest accuracy (99.98%) . Random Forest had the lowest accuracy (98.93%) in Table 2. Table 3 represents the

Independent Sample T-Test that is applied for the sample collections by fixing the level of significance as 0.005 with a confidence interval of 95 %. After applying the SPSS calculation, the support vector machine has accepted a statistically significant value($P<0.05$). From Fig. 4 it was represented by a simple bar Mean of Accuracy Random Forest error range (0.99 - 0.98) and support vector machine error range (0.99 - 0.98).

## 4. Discussion

The overall results show that there are a couple of assortments found in the accuracy and hardship regards because of the types of progress of the authorization limits which exhibited that the Support vector Machine with an accuracy of 99.98% is better than the Random Forest with a precision of 98.93% in seeing the manual human test (Captcha Recognition). There is a truly gigantic differentiation in inventive captcha Recognition precision of two estimations having the basic accuracy worth of 0.001 ($p<0.005$ Independent Sample t-Test). Regardless of the fact that there are exceptional assortments of captchas, text-based captcha is used most extensively. From one perspective, it is because it's a helpful and convenient way for site online customers, then again, captchas are an insignificant value answer for destinations. In any case, all the grasp that the text-based substance captures are unprotected and at this point not as firmly shut as appropriate to frame. So all will spread out literary substance captchas with more noteworthy security and higher conven (Huber 2019). The proposed methodology gained to some degree high headway rates for both the assigned plots as shown in Fig. 4. Manual human test (Captcha Recognition) plans can be seen as broken when the accomplishment robotized attack rate is 1% accordingly. All have variably broken a couple of obstacle frameworks found on both Manual human test plans which are regularly taken on by various notable captcha plans including bending, character covering, thick commotion lines, turn, curving, and concealing establishment (Panwar et al. 2018). Moreover it can be noted from the results shown on Table 2 that the watchman instruments embraced by Gregwar plot are more grounded than those of Weibo. Gregwar captcha contrive combines strong security framework, for instance, thick upheaval lines, different cutting edge and establishment tones, and more broad extent of character classes which makes it all the more hard for even individuals to see (J. Wang et al. 2019). Already our group has a rich involvement with chipping away at different exploration projects across numerous disciplines (Kwon, Yoon, and Park 2020),(Jia et al. 2018),(Shu and Xu 2019; Panwar et al. 2018),(Huber 2019),(Shu and Xu 2019),(Shu and Xu 2019; J. Wang et al. 2019),(Shu and Xu 2019; J. Wang et al. 2019; "Captcha Recognition Using GAN" 2020),(Shu and Xu 2019; J. Wang et al. 2019; "Captcha Recognition Using GAN" 2020; Sachdev 2020),(Singh et al. 2016),(T., R., and J. 2016),(Alsuhibany and Parvez 2016),(Alsuhibany and Parvez 2016; Panwar et al. 2018),(Cao 2021),(Selamat, Hakeem Selamat, and Rais 2015). Now the growing trend in this area motivated us to pursue this project.

## 5. Conclusion

In this examination, the inventive captcha Recognition framework utilizing the dataset comes from Wilhelmy, Rodrigo & Rosas, Horacio(2013). The captcha dataset is by all accounts better precision (99.98%) utilizing Support Vector Machine than Random Forest (98.93%). The clarity of captcha is found with satisfactory accuracy and less disaster is cultivated and the complexity of the manual human test is improved.

**Declarations**

**Conflict of Interest**

No conflicts of interest in this manuscript.

**Author Contribution**

Author GGP was involved in data collection, data analysis, and manuscript writing. Author KM was involved in conceptualization, guidance and critical review of manuscript.

**Acknowledgments**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

**Funding**

We thank the following organizations for providing financial support that enabled us to complete the study.

1. SM DSQUARE interiors and solutions Pvt.Ltd, Bangalore.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

## 6. References

Abdelhafiz, Mohamed H., Mohammed I. Awad, Ahmed Sadek, and Farid Tolbah. 2021.

"Sensor Positioning for a Human Activity Recognition System Using a Double Layer Classifier." Proceedings of the Institution of Mechanical Engineers. Part H, Journal of Engineering in Medicine, August, 9544119211040588.

Alaidi, Abdul Hadi M., Saif Ali Abd Alsaidi, and Omar Hashim Yahya. 2017. "Plate Detection and Recognition of Iraqi License Plate Using KNN Algorithm." Journal of Education College Wasit University. https://doi.org/10.31185/eduj.vol1.iss26.102.

Alsuhibany, Suliman A., and Mohammad Tanvir Parvez. 2016. "Secure Arabic Handwritten CAPTCHA Generation Using OCR Operations." 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). https://doi.org/10.1109/icfhr.2016.0035.

Cao, Yu. 2021. "Digital Character CAPTCHA Recognition Using Convolution Network." 2021 2nd International Conference on Computing and Data Science (CDS). https://doi.org/10.1109/cds52072.2021.00029.

"Captcha Recognition Using GAN." 2020. Journal of Xidian University. https://doi.org/10.37896/jxu14.6/325.

Huber, Manuel. 2019. CAPTCHA Recognition with Adaptable Neural Networks.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.

Jia, Yang, Wang Fan, Chen Zhao, and Jungang Han. 2018. "An Approach for Chinese Character Captcha Recognition Using CNN." Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/1087/2/022015.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." Computers and Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.

Kwon, Hyun, Hyunsoo Yoon, and Ki-Woong Park. 2020. "CAPTCHA Image Generation: Two-Step Style-Transfer Learning in Deep Neural Networks." Sensors 20 (5). https://doi.org/10.3390/s20051495.

Mayr, Susanne, Edgar Erdfelder, Axel Buchner, and Franz Faul. 2007. "A Short Tutorial of GPower." Tutorials in Quantitative Methods for Psychology. https://doi.org/10.20982/tqmp.03.2.p051.

Mhamed, Mustafa, Richard Sutcliffe, Xia Sun, Jun Feng, Eiad Almekhlafi, and Ephrem Afele Retta. 2021. "Improving Arabic Sentiment Analysis Using CNN-Based Architectures and Text Preprocessing." Computational Intelligence and Neuroscience 2021 (September): 5538791.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." Environmental Research. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." Carbohydrate Polymers 287 (July): 119349.

Panwar, Pooja, Monika, Parveen Kumar, and Ambalika Sharma. 2018. "CHGR: Captcha Generation Using Hand Gesture Recognition." 2018 Conference on Information and Communication Technology (CICT). https://doi.org/10.1109/infocomtech.2018.8722409.

Rodrigo, Pablo Rosas. n.d. "Hidrogeología Del Acuífero Kárstico de Pico Frentes (Cordillera

Ibérica)." https://doi.org/10.20868/upm.thesis.42320.

Sachdev, Sumeet. 2020. "Breaking CAPTCHA Characters Using Multi-Task Learning CNN and SVM." 2020 4th International Conference on Computational Intelligence and Networks (CINE). https://doi.org/10.1109/cine48825.2020.234400.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." Chemosphere 298 (July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." Chemosphere 298 (July): 134265.

Selamat, M. Hakeem, M. Hakeem Selamat, and Helmi Md Rais. 2015. "Image Face Recognition Using Hybrid Multiclass SVM (HM-SVM)." 2015 International Conference on Computer, Control, Informatics and Its Applications (IC3INA). https://doi.org/10.1109/ic3ina.2015.7377765.

Sha, Nong. 2011. Breaking Image Recognition CAPTCHA: An Approach Using Web-Trained Classifiers.

Shen, Jiang, Jiachao Wu, Man Xu, Dan Gan, Bang An, and Fusheng Liu. 2021. "A Hybrid Method to Predict Postoperative Survival of Lung Cancer Using Improved SMOTE and Adaptive SVM." Computational and Mathematical Methods in Medicine 2021 (September): 2213194.

Shu, Yujin, and Yongjin Xu. 2019. "End-to-End Captcha Recognition Using Deep CNN-RNN Network." 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). https://doi.org/10.1109/imcec46724.2019.8983895.

Singh, Bhargavi, M. Tech Scholar, G.I.E.T, and Sonipat. 2016. "Iris Recognition Using Curve Let Transformation Based on Gabor Filter& SVM." International Journal Of Engineering And Computer Science. https://doi.org/10.18535/ijecs/v5i8.32.

Tamir, Kassahun. 2019. "Handwritten Amharic Characters Recognition Using CNN." 2019 IEEE AFRICON. https://doi.org/10.1109/africon46755.2019.9133925.

Thomas, Varun Ambrose, and Karanvir Kaur. 2013. "Cursor CAPTCHA — Implementing CAPTCHA Using Mouse Cursor." 2013 Tenth International Conference on Wireless and Optical Communications Networks (WOCN). https://doi.org/10.1109/wocn.2013.6616188.

T., Manoj, Basavaraj R., and Jyoti J. 2016. "Security Enhancement in Captcha Recognition Using Animated GIF Images." International Journal of Computer Applications. https://doi.org/10.5120/ijca2016912473.

Tursunov, Anvarjon, Mustaqeem, Joon Yeon Choeh, and Soonil Kwon. 2021. "Age and Gender Recognition Using a Convolutional Neural Network with a Specially Designed Multi-Attention Module through Speech Spectrograms." Sensors 21 (17). https://doi.org/10.3390/s21175892.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-022-06636-5.

Wang, Jing, Jiao Hua Qin, Xu Yu Xiang, Yun Tan, and Nan Pan. 2019. "CAPTCHA Recognition Based on Deep Convolutional Neural Network." Mathematical Biosciences and Engineering: MBE 16 (5): 5851–61.

Wang, Zhao, and Xiang Li. 2015. "A Kind of Characters Segmentation Method for Text-Based CAPTCHA Attacking." Information Technology and Applications. https://doi.org/10.1201/b18284-32.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." Environmental Research. https://doi.org/10.1016/j.envres.2022.113114.

**Tables and Figures**

Table 1. Data collection from the N=10 samples of the dataset for Random Forest Algorithm with the size of 36*36 pixels to gain accuracy (%) and support vector machines to gain accuracy (%).

| Samples(N) | Random Forest Algorithm | Support Vector Machine (SVM) |
|---|---|---|
| | Accuracy (%) | Accuracy (%) |
| 1 | 98.93 | 99.98 |
| 2 | 98.45 | 99.56 |
| 3 | 96.78 | 98.48 |
| 4 | 97.48 | 97.48 |
| 5 | 95.49 | 97.89 |
| 6 | 97.89 | 96.87 |
| 7 | 95.14 | 96.12 |
| 8 | 94.78 | 95.79 |
| 9 | 93.98 | 94.89 |
| 10 | 92.45 | 93.58 |

Table 2. Comparison of the accuracy of captcha Recognition of Random Forest Algorithm and support vector machine. Support vector machine algorithms had the highest accuracy (99.98%) and Random Forest Algorithm had the lowest accuracy (98.93%).

**Group Statistics**

| GROUP | | N | Mean | Std Deviation | Std Error Mean |
|---|---|---|---|---|---|
| ACCURACY | SVM | 10 | 99.9810 | 2.15405 | .68117 |
| | RANDOM FOREST | 10 | 98.9310 | 2.40310 | .75993 |

Table 3. Independent Sample T-Test is applied for the sample collections by fixing the level of significance as 0.05 with confidence interval as 95 %. After applying the SPSS calculation, the support vector machine has accepted a statistically significant value ($p<0.05$).

| | Levene's Test for Equality of Variance | | T-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | f | Sig | t | df | Sig.(2-tailed) | Mean Difference | Std.Error Difference | 95% Confidence of the Differences | |
| | | | | | | | | Lower | Upper |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Accuracy** | **Equal variances assumed** | 10.033 | .004 | 6.093 | 18 | .000 | 6.21800 | 1.02053 | 4.07395 | 8.36205 |
| | **Equal variances not assumed** | | | 6.093 | 17.789 | .000 | 6.21800 | 1.02053 | 4.07395 | 8.36205 |

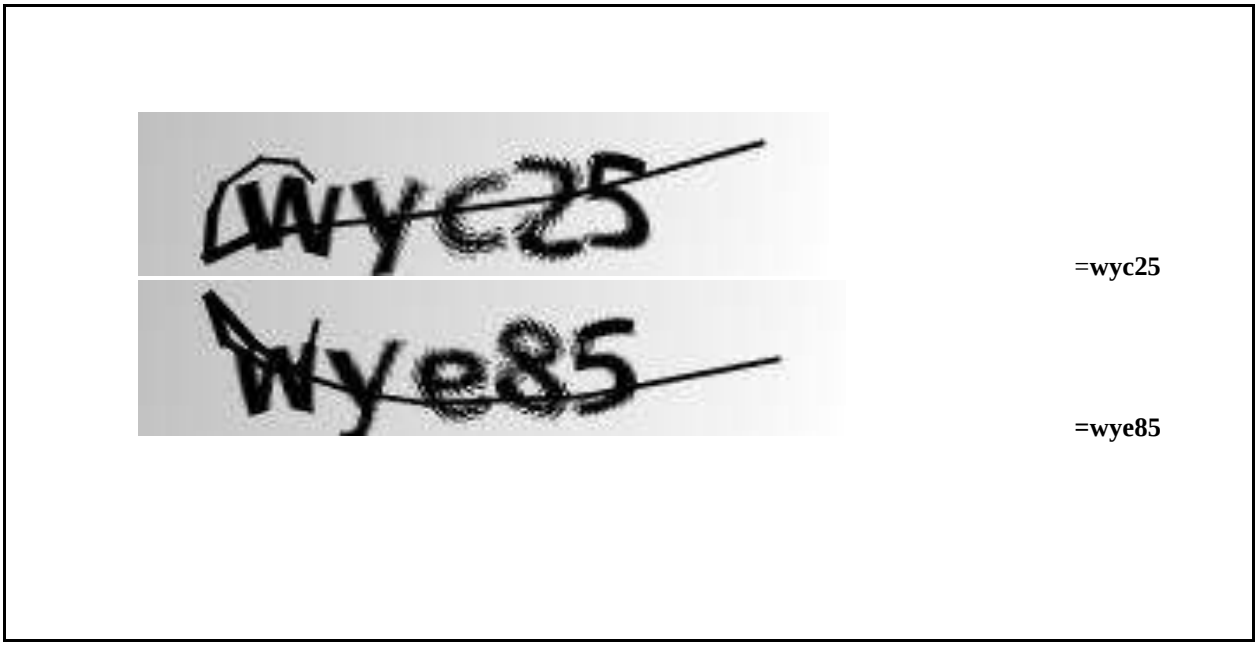


Fig. 1. Captchas from the dataset

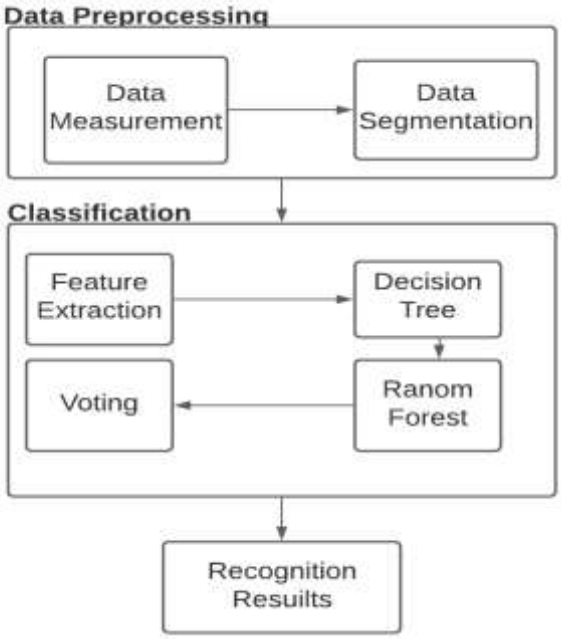


**Fig. 2. Flowchart of Random Forest**

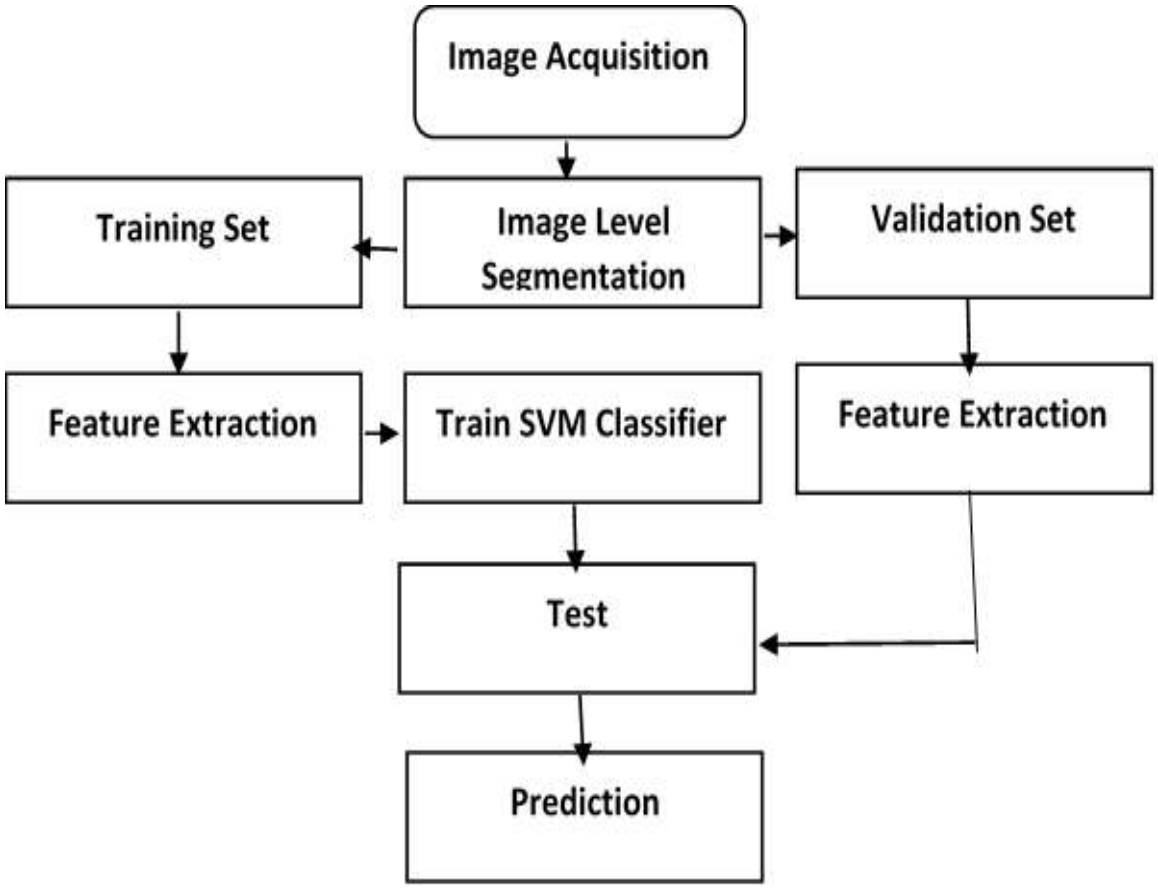


Fig. 3. Flowchart of SVM

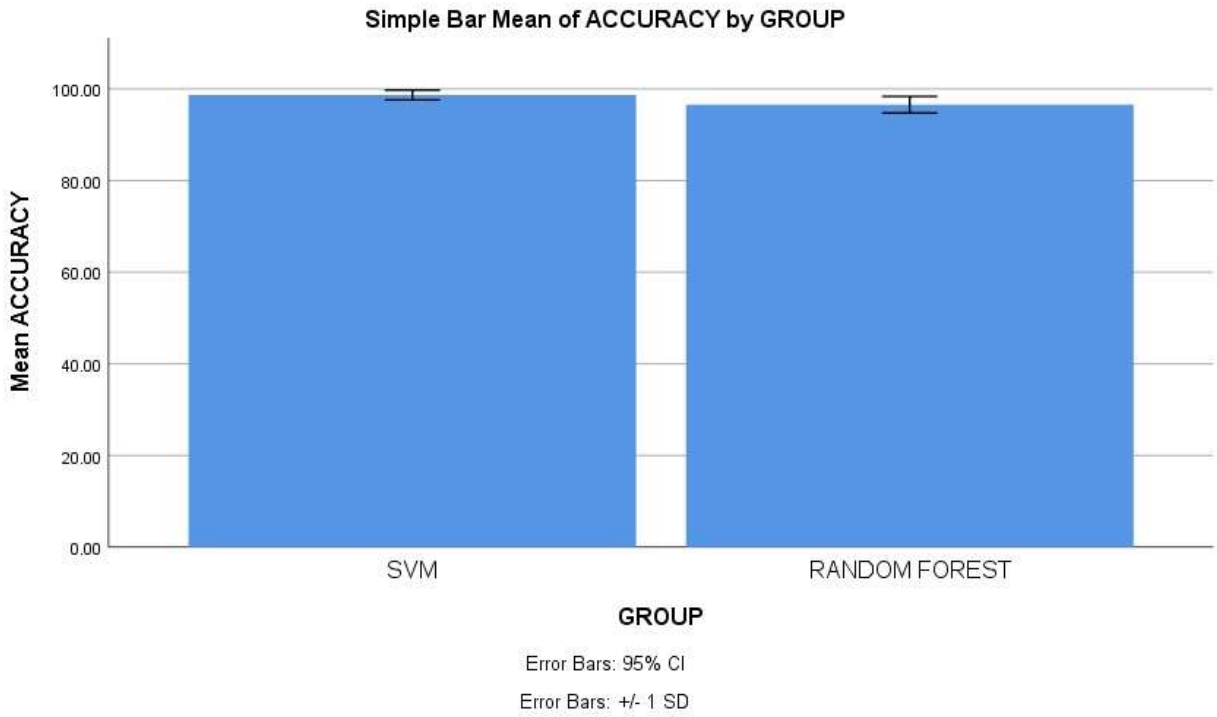


Fig. 4. Simple Bar Mean of Accuracy Random Forest error range (0.99 - 0.98) and support vector machine error range (0.98 - 0.99) with Mean accuracy of detection ± 1 SD.X Axis: Random Forest vs support vector machine Y-Axis: Mean accuracy ± 1 SD.