13th International Conference on Applied Informatics
Eszterházy Károly Catholic University
Eger, Hungary, February 19–21, 2026

Robust CAPTCHA Recognition with a CNN–RNN–CTC Ensemble on Real-World Data*

Ádám Kovács[ab], **Tibor Tajti**[ab]

[a]Eszterházy Károly Catholic University, Institute of Mathematics and Informatics
kovacs2.adam@uni-eszterhazy.hu
tajti.tibor@uni-eszterhazy.hu

[b]University of Debrecen, Doctoral School of Informatics

Abstract

Text-based CAPTCHA schemes are widely used to differentiate human users from automated bots, but advances in deep learning have rendered many of these schemes increasingly vulnerable [1, 7]. This paper presents a robust end-to-end CAPTCHA recognition system built upon a CNN–RNN architecture with a Connectionist Temporal Classification (CTC) decoder [3], enhanced by an ensemble voting strategy. The system is evaluated on the real-world ComprasNet CAPTCHA dataset [2], comprising 66,000 images across five visually distinct CAPTCHA styles. Our approach does not require prior segmentation or CAPTCHA-type classification, offering a generalizable solution for distorted sequence recognition tasks [8].

The core model architecture integrates convolutional layers for spatial feature extraction, bidirectional LSTM layers for modeling temporal dependencies, and a CTC output layer for flexible alignment-free decoding. Thirty independently initialized CNN–RNN–CTC models were trained. During evaluation, ensemble predictions were obtained by randomly selecting 21 models per run to ensure statistical robustness and to evaluate ensemble stability, and aggregating their outputs using per-character, position-wise majority voting. This ensemble method effectively

*This research was supported by the QoS-HPC-IoT Laboratory and the TKP2021-NKTA project at the University of Debrecen, Hungary. Project no. TKP2021-NKTA-34 has been implemented with support from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NKTA funding scheme.

mitigates common OCR confusions such as ‘F’ vs. ‘5’, ‘C’ vs. ‘2’, and ‘H’ vs. ‘7’, which often occur due to font and distortion overlap in CAPTCHA design [5].

Experiments were conducted using an 80-10-10 training-validation-test split of the dataset. A single model achieved an average sequence-level (word-level) accuracy of 98.95%, while the 21-model ensemble consistently reached 99.70%, reducing the overall error rate by over 70%. Character-level accuracy remained high across all string positions, with only slight performance drops at edge positions due to geometric distortion. Analysis of ensemble disagreement further revealed that ambiguous characters and later string positions were more prone to recognition uncertainty, though majority voting consistently resolved such cases [4].

Compared to prior CAPTCHA solving systems (Deep-CAPTCHA and segmentation-free CNN–RNN models), our method achieves state-of-the-art performance on a more complex and diverse real-world dataset. Table 1 summarizes related CAPTCHA solvers and their reported accuracy across benchmarks.

Table 1. Benchmark comparison of CAPTCHA solvers across architectures and datasets. Results correspond to word-level accuracy.

Reference	**Architecture**	**Dataset**	**Accuracy**
Noury & Rezaei (2020)	CNN classifier	Synthetic (500k)	98.9%
Kovács & Tajti (2023)	CNN ensemble	Public CAPTCHA (1k)	92.4%
Khatavkar et al. (2024)	CNN–RNN + CTC	Public CAPTCHA (1k)	95.0%
Divya et al. (2025)	BiConvLSTM + CTC	Public CAPTCHA (1k)	97.0%
Current Paper	**CNN–RNN–CTC ensemble**	**ComprasNet (66k)**	**99.7%**

The study also highlights the importance of ensemble diversity: despite all models sharing identical architectures and training data, random initialization induced sufficient variation to boost collective accuracy.

Ethically, the work is framed as a security analysis rather than a tool for circumvention. Our goal is to demonstrate vulnerabilities in common CAPTCHA systems and advocate for more secure alternatives. We recommend that CAPTCHA designers adopt multimodal, adaptive, or behavior-based verification schemes that are less susceptible to machine learning attacks [6]. As AI capabilities continue to improve, static-image-based CAPTCHA will require substantial redesign to remain effective in real-world deployments [4, 6].

References

[1] L. von Ahn, M. Blum, N. J. Hopper, J. Langford: *CAPTCHA: Using Hard AI Problems for Security*, in: Advances in Cryptology — EUROCRYPT, Springer, 2003, pp. 294–311, doi: 10.1007/3-540-39200-9_18.

[2] J. Carvalho: *66k Captchas (ComprasNet) Dataset*, https://www.kaggle.com/datasets/jasoncarvalho/comprasnet-captchas, Kaggle dataset, accessed: 2025-01, 2021.

[3] A. Graves, S. Fernandez, F. Gomez, J. Schmidhuber: *Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks*, in: Proceedings of the 23rd International Conference on Machine Learning (ICML), ACM, 2006, pp. 369–376, doi: 10.1145/1143844.1143891.

[4] D. C. Hoang, B. Ousat, A. Kharraz, C. V. Nguyen: *EnSolver: Uncertainty-Aware Ensemble CAPTCHA Solvers with Theoretical Guarantees*, arXiv preprint arXiv:2307.15180 (2023).

[5] V. Khatavkar, M. Velankar, S. Petkar: *Segmentation-Free CTC Loss Based OCR Model for Text CAPTCHA Classification*, arXiv preprint arXiv:2402.05417 (2024).

[6] H. D. Nguyen, K. Subramani, B. Acharya, R. Perdisci, P. Vadrevu: *C-FRAME: Characterizing and Measuring In-the-Wild CAPTCHA Attacks*, in: Proceedings of the IEEE Symposium on Security and Privacy (SP), IEEE, 2024, pp. 277–295, doi: 10.1109/SP54263.2024.00200.

[7] Z. Noury, M. Rezaei: *Deep-CAPTCHA: A Deep Learning Based CAPTCHA Solver for Vulnerability Assessment*, arXiv preprint arXiv:2006.08296 (2020).

[8] B. Shi, X. Bai, C. Yao: *An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence 39.11 (2017), pp. 2298–2304, doi: 10.1109/TPAMI.2016.2646371.