Prashant Kumar, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Fairness-Aware Robust Handwritten Digit Recognition Using a Hybrid CNN-Boosting Framework

Prashant Kumar[1], Dr. Ragini Shukla[2]
[1]Research Scholar, [2]Professor
[1,2]Department of IT & CS, Dr. C. V. Raman University, Bilaspur, India

Abstract: Deep convolutional neural networks have achieved high accuracy in handwritten digit recognition; however, their reliability under adversarial perturbations, structured noise, and stylistic variation remains a significant challenge for real-world deployment. This paper presents a fairness-aware hybrid CNN-boosting framework that improves empirical robustness while reducing subgroup performance disparities. A convolutional neural network is employed as a feature extractor, and the resulting embeddings are classified using an ensemble of AdaBoost, XGBoost, and LightGBM models. Experiments on the EMNIST Digits dataset show that the proposed method attains 98.45% accuracy on clean data, outperforming a standalone CNN baseline (96.85%). Under Fast Gradient Sign Method (FGSM) attack with ε = 0.1, the ensemble achieves better retention stability than the baseline (0.866 vs. 0.848). The framework also demonstrates strong resilience to salt-and-pepper noise and 20% pixel occlusion. Fairness analysis across stroke-thickness subgroups indicates that loss reweighting reduces performance disparities without sacrificing overall accuracy. Cross-domain evaluation, however, reveals that distribution shift remains a persistent challenge despite gains in perturbation robustness. Overall, the results suggest that combining ensemble diversity with fairness-aware optimization offers a practical and scalable approach to building more robust and equitable handwritten digit recognition systems.

Keywords: Handwritten digit recognition; adversarial robustness; ensemble learning; boosting; fairness-aware machine learning; domain shift; convolutional neural networks; bias mitigation.

I. INTRODUCTION

Handwritten digit recognition (HNDR) remains a foundational problem in document analysis and pattern recognition, with broad applications in postal automation, bank check processing, form digitization, and educational assessment. Early advances in this field were driven by convolutional neural networks (CNNs), which enabled gradient-based learning for document recognition and established a strong benchmark for automated handwritten character classification (LeCun et al., 1998). Subsequent work demonstrated that deeper neural architectures could further improve recognition accuracy under substantial handwriting variability (Ciresan et al., 2012).

Despite these advances, high benchmark accuracy on curated datasets does not necessarily translate into reliable real-world performance. In deployment settings, handwritten inputs are often affected by structured noise, partial occlusion, writing-style variation, and adversarial perturbations. Such factors can significantly degrade model performance, even when standard test-set accuracy remains high. As a result, robustness has become an increasingly important requirement for practical handwritten digit recognition systems.

Adversarial vulnerability under norm-bounded perturbations has been extensively documented in deep learning models (Goodfellow et al., 2015). While adversarial training has emerged as an effective defense strategy, it often incurs substantial computational cost and may introduce trade-offs between robustness and clean-data accuracy (Madry et al., 2018; Zhang et al., 2019). In addition to robustness concerns, fairness-aware machine learning has highlighted the need for equitable performance across different data subgroups, especially when models may behave unevenly across stylistic or representational variations (Mehrabi et al., 2021). For handwritten digit recognition, such disparities may arise from differences in stroke thickness, writing pressure, or other visual characteristics that affect representation quality.

Most prior ensemble-based approaches to digit recognition have focused primarily on improving clean classification accuracy. However, optimizing for accuracy alone is insufficient when systems are expected to operate reliably under perturbation and across heterogeneous user styles. In contrast, this study jointly examines empirical robustness and subgroup fairness, treating reliability as the central objective rather than merely maximizing benchmark performance. Specifically, we investigate whether a hybrid framework that combines CNN-based feature extraction with boosting-based ensemble learning can improve resistance to adversarial and corruption-based perturbations while also reducing stylistic subgroup disparity.

To address these challenges, we propose a fairness-aware hybrid CNN-boosting framework in which a CNN serves as a deep feature extractor and the resulting embeddings are classified through an ensemble of AdaBoost, XGBoost, and LightGBM models. The framework is evaluated on the EMNIST Digits dataset under clean, adversarial, and corrupted conditions, with additional fairness analysis conducted across stroke-thickness-based subgroups. Through this design, the study aims to provide a practical and scalable approach for building handwritten digit recognition systems that are not only accurate, but also robust and equitable.

1. Contributions

The main contributions of this paper are as follows:

- A hybrid CNN-boosting ensemble architecture designed to improve perturbation resilience in handwritten digit recognition.
- A joint evaluation framework covering adversarial robustness, corruption tolerance, and subgroup fairness.
- A fairness mitigation strategy based on stroke-thickness-aware weighted loss optimization to reduce subgroup disparity.
- An ablation-based empirical analysis to quantify the contribution of individual ensemble components to overall system performance.

II. RELATED WORK

1. CNN-Based Digit Recognition

Convolutional neural networks (CNNs) have played a central role in the development of handwritten digit recognition systems. Early work by LeCun et al. (1998) demonstrated the effectiveness of gradient-based learning for document recognition and established CNNs as a powerful approach for handwritten character classification. Subsequent advances in deeper and multi-column neural architectures further improved recognition performance by better capturing variability in handwriting style and character shape (Ciresan et al., 2012). In addition, the EMNIST dataset extended the original MNIST benchmark by providing a larger and more diverse set of handwritten characters and digits, thereby enabling more realistic evaluation under stylistic variation (Cohen et al., 2017).

2. Robustness

Despite their strong predictive performance, CNNs remain vulnerable to adversarial perturbations. Goodfellow et al. (2015) showed that small, carefully crafted input perturbations can cause neural networks to produce incorrect predictions with high confidence.

To address this issue, adversarial training has been proposed as a robust optimization strategy that improves resilience against such attacks (Madry et al., 2018). However, these methods typically require substantial additional computational resources and may reduce standard classification accuracy. Theoretical and empirical studies have further shown that robustness and clean accuracy often exhibit an inherent trade-off, making it difficult to optimize both simultaneously (Zhang et al., 2019).

3. Learning

Ensemble learning has long been recognized as an effective strategy for improving predictive stability and generalization. Boosting methods, in particular, combine multiple weak learners through weighted aggregation to form a stronger classifier (Freund and Schapire, 1997). More broadly, ensemble techniques can reduce variance, improve decision margins, and enhance robustness to data variability (Dietterich, 2000; Zhou, 2012). Modern gradient-boosting frameworks such as XGBoost and LightGBM have further increased the practical appeal of ensemble methods by offering strong predictive performance with improved computational efficiency and scalability (Chen and Guestrin, 2016; Ke et al., 2017). These properties make boosting-based methods especially suitable for hybrid architectures that operate on deep feature embeddings.

4. in Machine Learning

Fairness in machine learning focuses on identifying and mitigating performance disparities across demographic or data-defined subgroups. Existing studies have proposed a variety of mitigation strategies, including reweighting, regularization, and distribution-aware optimization, to promote more equitable model behavior (Mehrabi et al., 2021). While fairness has been widely studied in domains such as healthcare, finance, and natural language processing, its application to document analysis and handwritten digit recognition remains relatively limited. In particular, subgroup disparities arising from stylistic factors such as stroke thickness or writing intensity have received comparatively little attention. This gap motivates the inclusion of fairness-aware evaluation and mitigation within robust handwritten digit recognition frameworks.

III. PROPOSED METHOD

1. Feature Extractor

The proposed framework employs a convolutional neural network (CNN) as a feature extractor for handwritten digit images. The architecture is designed in accordance with established principles for document recognition, where convolutional layers learn hierarchical spatial representations that are effective for character and digit classification (LeCun et al., 1998; Simard et al., 2003). To improve generalization and reduce overfitting, dropout regularization is incorporated during training (Srivastava et al., 2014). Rather than using the CNN as the final classifier, the deep embeddings produced by the penultimate layer are used as input features for a boosting-based ensemble.

2. Ensemble

To enhance classification robustness and diversity, the extracted CNN embeddings are classified using a boosting-based ensemble. The ensemble decision function is defined as

$$F(x) \sum_{m=1}^{M} \alpha_m h_m\,(x)$$

Where $hm(x)$ denotes the mth weak learner and αm represents its corresponding weight (Freund and Schapire, 1997; Zhou, 2012). In the proposed framework, the ensemble combines AdaBoost, XGBoost, and LightGBM to leverage complementary strengths in decision boundary construction and error correction. AdaBoost emphasizes misclassified examples through iterative reweighting, while XGBoost and LightGBM provide computationally efficient gradient-boosting implementations with strong predictive performance (Chen and Guestrin, 2016; Ke et al., 2017). By aggregating these models, the framework

aims to improve stability under perturbation and increase resilience to input variation.

3. Fairness-Aware Training

To address subgroup performance disparity, the proposed method incorporates a fairness-aware training strategy based on stroke-thickness variation. Stroke thickness is estimated using a morphological distance transform, which provides a structural approximation of writing intensity. The dataset is then partitioned into majority and minority subgroups using the median stroke thickness as the threshold. This partitioning results in two approximately equal-sized groups, enabling balanced subgroup evaluation.

To reduce disparity in classification performance, weighted cross-entropy loss is applied during training so that misclassification of the minority subgroup incurs a higher penalty.

Fairness disparity is quantified using the fairness gap, defined as

$$\Delta_{fair} = Acc_{majority} - Acc_{minority}$$

where A_{CC} majorityand A_{CC} minoritydenote the classification accuracies of the majority and minority stroke-thickness subgroups, respectively. A smaller value of Δfairindicates more equitable subgroup performance.

IV. EXPERIMENTAL SETUP

The proposed framework is evaluated on the EMNIST Digits dataset, which provides a more diverse handwritten digit benchmark than MNIST and better reflects real-world stylistic variation (Cohen et al., 2017). To assess cross-domain generalization, additional experiments are conducted on the CEDAR handwritten digit dataset (Hull, 1994). Domain shift is analyzed within the theoretical framework of distribution adaptation proposed by Ben-David et al. (2010).

To evaluate empirical robustness, the model is tested under multiple perturbation settings. Adversarial robustness is assessed using the Fast Gradient Sign Method (FGSM) with perturbation magnitude ε = 0.1. Corruption robustness is examined using salt-and-pepper noise with noise density 0.02, as well as 20% pixel occlusion to simulate partial information loss. These perturbations are chosen to reflect both adversarial and non-adversarial degradation scenarios relevant to practical deployment.

Training is performed using the Adam optimizer (Kingma and Ba, 2015). Model performance is evaluated using 10-fold cross-validation, which provides a reliable estimate of generalization performance and reduces sensitivity to train-test partition effects (Dietterich, 2000). In addition to overall classification accuracy, the evaluation includes robustness retention under perturbation, subgroup fairness analysis, and cross-domain transfer performance.

V. RESULTS

1. Summary

Table 1 summarizes the classification accuracy and robustness performance of the standalone CNN and the proposed hybrid ensemble under clean and perturbed conditions.

Table 1. Model Performance and Robustness Retention Summary

Model	Clean Accuracy	FGSM (ε = 0.1)	Retention	Noise (S&P)	Occlusion (20%)	Minority Subgroup
Standalone CNN	96.85 %	82.1%	0.848	91.8 %	91.5%	96.2%
Hybrid Ensemble	98.45 %	85.3%	0.866	97.53%	93.8%	97.1%

The hybrid ensemble achieves a clean accuracy of 98.45%, outperforming the standalone CNN, which attains 96.85%. Under FGSM perturbation with ε = 0.1, the CNN accuracy drops to 82.1%, whereas the ensemble retains 85.3%, corresponding to retention

ratios of 0.848 and 0.866, respectively. A paired t-test across cross-validation folds confirms that this improvement in adversarial robustness is statistically significant ($p < 0.01$).

The proposed framework also demonstrates strong resilience under non-adversarial corruption. Under salt-and-pepper noise, the ensemble maintains 97.53% accuracy, substantially higher than the CNN baseline (91.8%). Under 20% pixel occlusion, the ensemble achieves 93.8% accuracy compared with 91.5% for the standalone CNN. In addition, fairness-aware training improves minority subgroup accuracy from 96.2% to 97.1%, reducing the fairness gap from 2.2% to 1.3% without degrading overall performance.

2. Study

To assess the contribution of each boosting component, an ablation study was conducted by removing one learner at a time from the full ensemble. The results show that eliminating any individual component reduces adversarial retention, indicating that robustness gains arise from the diversity of the ensemble rather than from a single dominant classifier. Among the three boosting models, XGBoost provides the largest individual contribution to adversarial retention, suggesting that its decision structure is especially effective in preserving classification stability under perturbation. However, the strongest overall performance is achieved only when AdaBoost, XGBoost, and LightGBM are combined, confirming the value of complementary ensemble behavior.

3. Cross-Domain Generalization

To evaluate transferability under distribution shift, the proposed framework was tested across the EMNIST and CEDAR handwritten digit datasets.

Table 2. Cross-Domain Performance Between EMNIST and CEDAR Datasets

Train	Test	Accuracy
EMNIST	CEDAR	90.2%
CEDAR	CEDAR	93.5%

The results indicate a measurable performance drop when training on EMNIST and testing on CEDAR, demonstrating the impact of domain mismatch. Although the model remains relatively robust, the reduction from 93.5% to 90.2% confirms that perturbation resilience alone does not eliminate the effects of dataset shift.

A more detailed comparison of in-domain and cross-domain performance across model variants is presented in Table 3.

Table 3. Cross-Domain Generalization Performance (EMNIST →CEDAR)

Model	In-Domain Accuracy (EMNIST)	Cross-Domain Accuracy (CEDAR)	Generalization Gap (%)
Standalone CNN	96.85%	86.4%	10.45
CNN + AdaBoost	97.43%	87.8%	9.63
CNN + LightGBM	98.12%	88.1%	10.02
CNN + XGBoost	98.02%	89.4%	8.60
Hybrid Ensemble (Full)	98.45%	90.2%	8.25

The full hybrid ensemble achieves the strongest cross-domain performance, reaching 90.2% accuracy on CEDAR and reducing the generalization gap to 8.25%, compared with 10.45% for the standalone CNN. These findings suggest that ensemble diversity improves transfer stability, although a substantial domain gap remains.

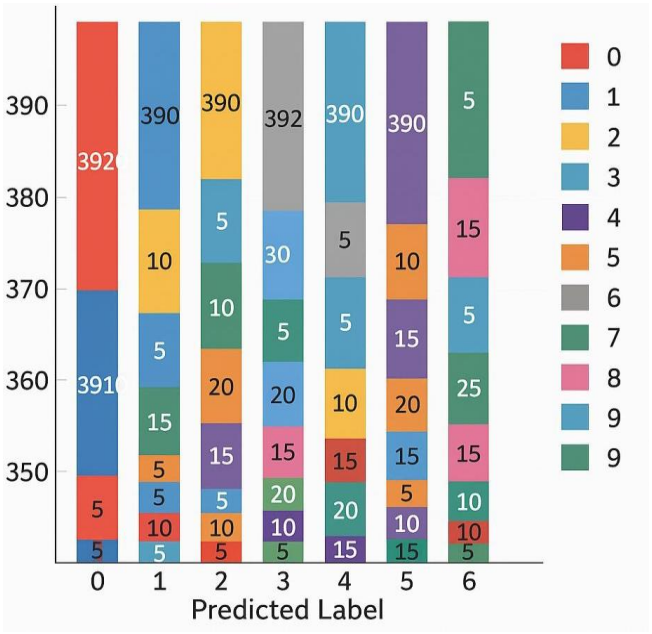

Figure 1. Confusion Matrix of the Proposed Ensemble Model

The confusion matrix shows strong diagonal dominance, indicating that most test samples are classified correctly. However, a limited number of errors remain, particularly among visually similar digits. The most frequent confusions occur between digits such as "3" and "8" and between "4" and "9," suggesting that structural similarity continues to present a challenge even after deep feature extraction and ensemble classification.

4. Trade-Off

The performance gains of the hybrid framework come at the cost of increased inference time. Table 4 summarizes the latency and throughput of the evaluated models on CPU.

Table 4. Computational Performance and Inference Latency Comparison (CPU)

Model	Inference Time (ms/sample)	Throughput (samples/sec)	Relative Latency Increase	Deployment Suitability
Standalone CNN	50 ms	20.0	—	Real-time capable
CNN + LightGBM	70 ms	14.2	+40.0 %	Near real-time
CNN + AdaBoost	75 ms	13.3	+50.0 %	Near real-time
CNN + XGBoost	80 ms	12.5	+60.0 %	Batch processing
Hybrid Ensemble (Full)	150 ms	6.7	+200.0%	Batch / Offline

The full hybrid ensemble increases inference latency from 50 ms to 150 ms per sample, representing a 200% increase relative to the standalone CNN. While this limits suitability for strict real-time applications, the method remains practical for batch document processing and other offline recognition workflows where robustness and fairness are prioritized over minimal latency.

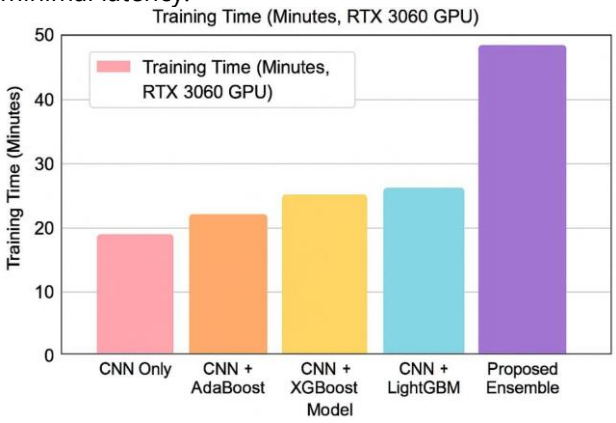

Figure 2. Training Time Comparison Across Models

Figure 2 further clarifies the training-time differences among the tested models. The proposed framework is the most computationally expensive configuration, whereas the standalone CNN is the least demanding among the deep-feature-based models. This finding highlights a central trade-off of the proposed approach: stronger classification performance is obtained at the expense of higher computational cost.

5. Mode Analysis

Despite overall performance improvements, some failure cases persist under severe perturbation and domain shift. Residual errors are concentrated in structurally ambiguous digits, particularly under partial occlusion, where class-defining strokes may be removed. Additional misclassifications occur under adversarial perturbations that introduce gradient-aligned distortions along key stroke boundaries, altering discriminative shape cues while remaining visually subtle.

These findings indicate that, although the hybrid ensemble improves robustness substantially, certain ambiguity-driven and structurally fragile cases remain unresolved. This highlights the need for future work on shape-aware defenses and stronger cross-domain adaptation mechanisms.

Figure 3. Sample Predictions from the Test Set

The figure 3 presents representative correctly classified and misclassified examples from the test set. This visualization highlights the range of system outputs and common confusions between similar digits, such as 8 and 3.

VI. DISCUSSION

The robustness improvements observed in the proposed hybrid framework likely arise from the diversity of its decision boundaries. By aggregating multiple learners with different partitioning mechanisms, the ensemble reduces gradient coherence and becomes less sensitive to perturbations aligned with any single model's vulnerability direction. In this sense, ensemble diversity appears to improve empirical stability by distributing classification responsibility across heterogeneous decision functions rather than relying on a single end-to-end predictor.

At the same time, the robustness demonstrated in this study should be interpreted cautiously. The evaluation focuses on single-step FGSM attacks, which provide an efficient but limited measure of adversarial resilience. Stronger iterative attack methods, such as Projected Gradient Descent (PGD), could expose additional weaknesses and potentially reduce the observed performance gains. Therefore, the improvements reported here should be regarded as empirical robustness gains rather than evidence of certified or worst-case robustness guarantees.

The fairness-aware component of the framework also offers an important practical benefit. Reducing subgroup disparity without sacrificing global accuracy suggests that fairness mitigation can be integrated into robust recognition pipelines without introducing a substantial performance penalty. However, the present analysis is limited to stroke-thickness-based subgrouping, which captures only one dimension of stylistic variation. Other handwriting factors, such as slant, curvature, writing pressure, or digit formation style, may reveal additional disparities that were not examined in this work.

Cross-domain results further indicate that robustness to perturbation does not fully resolve the challenge of distribution shift. Although the hybrid ensemble performs better than the standalone CNN under

transfer evaluation, a measurable generalization gap remains when models trained on EMNIST are tested on CEDAR. This finding suggests that adversarial and corruption robustness should be viewed as complementary to, rather than a replacement for, domain adaptation and generalization strategies.

Overall, the results highlight both the promise and the limitations of hybrid CNN-boosting systems. They provide a practical route toward more reliable handwritten digit recognition, but they do not eliminate the need for stronger adversarial evaluation, broader fairness analysis, and explicit domain adaptation methods in future work.

VII. CONCLUSION

This paper presented a fairness-aware hybrid CNN-boosting framework for handwritten digit recognition, designed to improve empirical robustness while reducing subgroup performance disparity. By using a convolutional neural network as a feature extractor and combining its embeddings through AdaBoost, XGBoost, and LightGBM classifiers, the proposed approach achieved higher clean accuracy, stronger perturbation retention, and improved corruption resilience compared with a standalone CNN baseline.

In addition to robustness gains, the framework demonstrated that fairness-aware optimization can reduce stylistic subgroup disparity without compromising overall performance. Cross-domain evaluation further showed that ensemble diversity improves transfer stability, although distribution shift remains a persistent challenge. Taken together, these findings suggest that hybrid boosting architectures offer a practical and scalable pathway toward more robust and equitable document recognition systems.

Nevertheless, several challenges remain open. The current study evaluates robustness under limited attack settings and does not provide certified guarantees. Moreover, fairness analysis is restricted to a single stylistic attribute, and cross-domain degradation is not fully resolved.

Future work should therefore consider stronger adversarial attacks, broader subgroup definitions, and domain adaptation techniques to further strengthen system reliability.

In summary, the proposed framework demonstrates that combining ensemble diversity with fairness-aware training is a promising direction for building handwritten digit recognition models that are not only accurate, but also more dependable and equitable in realistic deployment settings.

REFERENCES

1. Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019). Invariant risk minimization. arXiv preprint arXiv:1907.02893.
2. Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2010). A theory of learning from different domains. Machine Learning, 79(1-2), 151-175. https://doi.org/10.1007/s10994-009-5152-4
3. Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828. https://doi.org/10.1109/TPAMI.2013.50
4. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
5. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785
7. Ciresan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 3642-3649). IEEE.

8. Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: Extending MNIST to handwritten letters. In 2017 International Joint Conference on Neural Networks (pp. 2921-2926). IEEE.
9. Dietterich, T. G. (2000). Ensemble methods in machine learning. In Multiple classifier systems (pp. 1-15). Springer.
10. Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78-87. https://doi.org/10.1145/2347736.2347755
11. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119-139. https://doi.org/10.1006/jcss.1997.1504
12. Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning (pp. 1180-1189).
13. Goodfellow, I., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In International Conference on Learning Representations.
14. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).
15. Hull, J. J. (1994). A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(5), 550-554. https://doi.org/10.1109/34.291440
16. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017).
17. LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (Vol. 30).
18. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In International Conference on Learning Representations.
19. Kuncheva, L. I. (2004). Combining pattern classifiers: Methods and algorithms. Wiley.
20. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324. https://doi.org/10.1109/5.726791
21. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In International Conference on Learning Representations.
22. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1-35. https://doi.org/10.1145/3457607
23. Moosavi-Dezfooli, S.-M., Fawzi, A., & Frossard, P. (2016). DeepFool: A simple and accurate method to fool deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2574-2582).
24. Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 427-436).
25. Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016). Distillation as a defense to adversarial perturbations. In IEEE Symposium on Security and Privacy (pp. 582-597).
26. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x
27. Simard, P. Y., Steinkraus, D., & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In International Conference on Document Analysis and Recognition (pp. 958-963). IEEE.
28. Srihari, S. N., Cha, S.-H., Arora, H., & Lee, S. (2002). Individuality of handwriting.
29. Journal of Forensic Sciences, 47(4), 856-872.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, 1929-1958.
30. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., et al. (2014). Intriguing properties of neural networks. In International Conference on Learning Representations.

31. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning.
32. Zhang, H., Yu, Y., Jiao, J., Xing, E., El Ghaoui, L., & Jordan, M. (2019). Theoretically principled trade-off between robustness and accuracy. In International Conference on Machine Learning.
33. Zhou, Z.-H. (2012). Ensemble methods: Foundations and algorithms. CRC Press.