Evaluation of Combined Machine Learning Algorithms in Digit Image Recognition

Lê Minh Giang
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
dtc1954802010088@ictu.edu.vn

***Abstract*—This thesis evaluates the performance of several machine learning algorithms, including Random Forest, AdaBoost, and XGBoost, in the context of digit image recognition. The author describes the preprocessing steps, model training, and evaluation process, highlighting the importance of data preprocessing and the combination of multiple algorithms to improve recognition accuracy. The preprocessing steps involve data normalization and splitting the dataset into training, validation, and testing sets. The author applies and fine-tunes various machine learning algorithms and uses a voting classifier to combine their results. The combined model achieved high performance on both the training and testing datasets, indicating good generalization capabilities. However, the exact performance metrics are not provided. The author discusses the advantages and disadvantages of each algorithm, noting that the combined model improved recognition accuracy. However, the interpretability of the model results is limited due to the complexity of the algorithms. Future work could focus on improving the interpretability of the models.**

***Index Terms*—Machine Learning, Random Forest, AdaBoost, XGBoost, Digit Recognition, Voting Classifier**

I. Introduction

The thesis provides an overview of supervised machine learning, discussing its applications, types, and future prospects. It focuses on the evaluation of combined machine learning algorithms for digit image recognition, emphasizing the importance of data preprocessing and the integration of multiple algorithms to enhance recognition accuracy. The author reviews the basic concepts, applications, and future directions of machine learning, providing a foundation for the subsequent technical sections. The introduction also highlights the significance of combining different algorithms to improve the overall performance of the digit image recognition system [1].

II. Methods

The author describes the preprocessing steps, including data normalization and splitting the dataset into training, validation, and testing sets. Various machine learning algorithms, such as Random Forest, AdaBoost, and XGBoost, are applied and fine-tuned. A voting classifier is used to combine the results of these algorithms. The preprocessing steps involve normalizing the pixel values of the images to ensure that they lie within the range of 0 to 1. The dataset is then split into training, validation, and testing sets to ensure that the model is evaluated on unseen data. The author applies and fine-tunes the algorithms to optimize their performance on the training set. The voting classifier combines the predictions of the individual algorithms to produce a final output. This approach allows for the integration of multiple algorithms, potentially improving the overall recognition accuracy [1].

Supervisor: Trần Thị Xuân, Thai Nguyen University of Information and Communication Technology (ttxuan@ictu.edu.vn).

III. Results

The thesis documents preprocessing and an experimental comparison of Random Forest, AdaBoost, XGBoost, and a voting classifier for digit-image recognition. Its narrative describes the combined classifier as performing well on the data used in the experiment. This summary does not assign an accuracy value because the extractable narrative does not state a verified aggregate metric for each compared model. The result supports a local comparison of classifiers, while broader generalization would require the exact dataset split and reproducible evaluation figures [1].

IV. Discussion

The author discusses the advantages and disadvantages of each algorithm, noting that the combined model improved recognition accuracy. However, the interpretability of the model results is limited due to the complexity of the algorithms. The author suggests that future work could focus on improving the interpretability of the models. The Random Forest algorithm is noted for its ability to handle high-dimensional data and provide feature importance, while AdaBoost and XGBoost are powerful for handling complex data and providing high accuracy. However, these algorithms can be computationally intensive and may suffer from overfitting. The voting classifier combines the strengths of these algorithms, potentially improving recognition accuracy. However, the complexity of the algorithms makes it challenging to interpret the results, which could be a limitation for practical applications [1].

V. Conclusion

The study applies several machine-learning classifiers and a voting combination to digit-image recognition. It documents the preprocessing, model-training, and comparison workflow and describes the combined approach as promising. The available narrative does not establish a quantified improvement

over every constituent classifier, so that stronger conclusion is avoided here. Further work should report exact held-out metrics and the experimental split to support a reproducible comparison [1].

References

[1] Lê Minh Giang, ”Evaluation of Combined Machine Learning Algorithms in Digit Image Recognition,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] - Lee, J., Lee, J., & Kim, J. (2020). A Study on the Impact of Service Quality on

[3] Customer Satisfaction and Loyalty in Theme Parks. Journal of Hospitality & Tourism Research, 44(1), 131-150

[4] - Smith, A., & Johnson, B. (2019). Understanding Customer Experience: A Review

[5] and Research Agenda. Journal of Service Research, 22(1), 5-22

[6] - Dogan, O., Ayçin, E., & Bulut, Z. A. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry. International

[7] Journal of Contemporary Economics and Administrative Sciences, 8(1), 1-19

[8] - Chen, Y., Wang, Y., & Nevo, S. (2019). Effects of Website Quality on Customer Online Purchase Intention: The Mediating Roles of Trust and Satisfaction. Information & Management, 56(1), 103-117

[9] - Viksne, K., Salkovska, J., Gaitniece, E., & Puke, I. (2016). Comparative analysis

[10] of customer behaviour models. Trong Proceedings of the 2016 International conference on Economic Science for Rural Development (tr. 231-231)