ENSEMBLE LEARNING FOR TRAFFIC SIGN CLASSIFICATION

Nguyen Cong Tuyen - Phi Dinh Huy - Quach Manh Duy - Tran Van Tuan email: trantuan6304@gmail.com

October 2024

Abstract - Traffic sign classification plays a crucial role in autonomous driving and intelli- gent transportation systems (ITS), where accu- rate detection and interpretation of traffic signs improve road safety. While deep learning models such as Convolutional Neural Networks (CNNs) have demonstrated remarkable accuracy, tradi- tional machine learning techniques like Support Vector Machines (SVMs) and Random Forests (RF) continue to offer competitive results when paired with robust feature extraction. This study provides a comparative analysis of CNN, SVM, and RF models applied to the German Traffic Sign Recognition Benchmark (GTSRB) dataset. Moreover, an ensemble learning approach is pro- posed to combine these models, aiming to boost classification accuracy and robustness. Our re- sults demonstrate that while CNN provides su- perior feature extraction, the ensemble model outperforms individual models in terms of pre- cision and accuracy. This research fills a gap in the literature by highlighting the strengths of hybrid approaches in traffic sign classification tasks.

Keywords: Traffic Sign Classification, CNN, SVM, Random Forest, Ensemble Learning, GTSRB.

I. Introduction

Traffic sign classification is a vital aspect of modern in- telligent transportation systems (ITS), particularly in autonomous vehicles and ADAS. As traffic signs regu- late and manage the flow of vehicles, accurately recog- nizing these signs is key to ensuring road safety. Recent developments in artificial intelligence, especially in com- puter vision, have significantly advanced the capabilities of automated traffic sign recognition systems[1][2].

The German Traffic Sign Recognition Benchmark (GTSRB) is a widely used dataset for evaluating traf- fic sign classification models. Many studies have applied deep learning techniques like Convolutional Neural Net- works (CNNs) to achieve high accuracy on this dataset. CNNs have demonstrated their ability to learn com- plex patterns from image data without manual feature engineering, making them the preferred choice for im- age classification tasks[3][4]. However, traditional ma- chine learning methods such as Support Vector Ma- chines (SVMs) and Random Forest (RF) still play an im- portant role, particularly in cases where computational efficiency and interpretability are prioritized.

Recent research has further shown the potential of combining traditional machine learning methods with CNN-based architectures. This study aims to address the gap in current literature by providing a direct com- parison between CNN, SVM, and RF, while also propos- ing an ensemble model that combines the strengths of these classifiers. Through this approach, we aim to im- prove classification accuracy and provide insights into the performance trade-offs of hybrid models.

II. Related Work

Numerous studies have been conducted on traffic sign classification, particularly with the GTSRB dataset. CNN has emerged as a leading method due to its su- perior ability to handle complex image data. In recent research, CNN-based models have consistently achieved high classification accuracy, often surpassing 98%. How- ever, CNNs tend to require significant computational resources, which can be a limitation for real-time appli- cations such as ADAS.

Traditional machine learning models such as SVM and RF are also well-represented in the literature. Wahyono et al. (2014) and Schuszter et al. (2017) demonstrated that SVM, when combined with feature descriptors like HOG (Histogram of Oriented Gradi- ents), performed well for traffic sign recognition, al- though CNN-based models outperformed them in most cases. RF, known for its ensemble decision-making capa- bility, has also been applied with some success, though it is more effective when used in conjunction with other models rather than as a standalone approach.

Ensemble learning techniques, which combine multi- ple classifiers, have been increasingly explored as a way to improve classification performance. Wei et al. (2023) demonstrated that integrating CNN and ResNet-based models for traffic sign classification yielded over 99% accuracy on large-scale datasets. Our work aims to pro- vide a detailed comparative analysis of these models and demonstrate the potential of ensemble learning in en- hancing traffic sign classification performance.

III. Dataset Description

a. Dataset Description: German Traffic Sign Recognition Benchmark (GTSRB)

The German Traffic Sign Recognition Benchmark (GT- SRB) is a widely utilized dataset for traffic sign clas- sification tasks. It contains a comprehensive collection of traffic sign images that represent various real-world conditions, including variations in illumination, weather, occlusion, and perspectives. GTSRB is specifically de- signed to facilitate research in traffic sign recognition and autonomous driving systems.

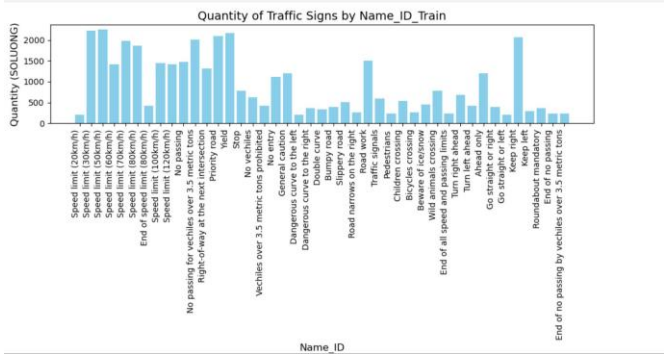
Hình 1: Fig.1. Different types of traffic signs

The dataset consists of 51,839 images of traffic signs, divided into 43 classes. The standard format for train- ing and testing is a resized version of 32 Ư 32 pixels per image to standardize inputs for convolutional neu- ral networks (CNNs).

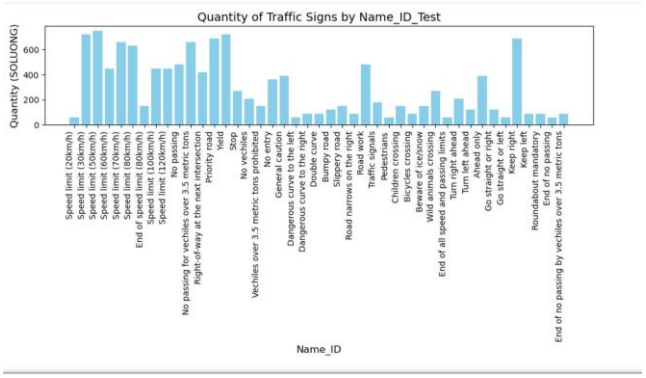
b. Data Splits

The GTSRB dataset is divided into two main sets: a training set with 39,209 images and a testing set con- taining 12,630 images.

c. Class Distribution



Hình 2: Fig. 2. Training set



Hình 3: Fig. 3. Testing set

The GTSRB dataset exhibits class imbalance, with some classes containing significantly more examples than oth- ers.

d. Conditions Affecting Data and Preprocessing

Preprocessing techniques such as normalization and data augmentation (rotation, brightness adjustment, etc.) are essential to simulate real-world scenarios. Class imbalance needs to be addressed through techniques such as oversampling to ensure that no category is over- represented.

e. Evaluation Metric

The primary evaluation metric for the GTSRB dataset is classification accuracy, measuring the proportion of correctly classified traffic signs. Other metrics such as precision, recall, and F1-score can also be reported.

IV. Proposed Solution: CNN with Data Augmentation and Ensemble Learning

a. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) are highly effec- tive for image classification tasks due to their ability to automatically learn hierarchical features from raw pixel data. The architecture consists of multiple convolutional layers with ReLU activation, followed by pooling layers and fully connected layers for final classification.

b. Data Augmentation for Robustness

We apply data augmentation techniques, including ran- dom rotations, translations, flips, and brightness adjust- ments, to make the model more robust to variations in traffic sign appearances.

c. Combining CNN with Ensemble Learn- ing

We propose combining CNN with SVM and Random Forest through a voting method to improve classifica- tion accuracy and model stability. The final result is

determined through hard or soft voting, depending on whether we use class predictions or probabilities.

d. Loss Function and Optimizer

We use a weighted cross-entropy loss function to address class imbalance. The Adam optimizer is used, with an initial learning rate of 0.001, combined with a learning rate decay schedule and early stopping to prevent over- fitting.

e. Evaluation Metrics

We report classification accuracy, precision, recall, and F1-score to account for the model's ability to handle class imbalance.

V. Methodology

1. Convolutional Neural Networks (CNN)

The model used is a CNN, which is particularly well- suited to image processing problems because of its abil- ity to detect and learn important features in images through convolutional layers. The architecture of the model includes the following basic layers:

a. Convolutional Layer:

This is the most important layer in the CNN model. This layer will use filters to perform convolution on the input image to create feature maps. These filters will ex- tract important features from the image such as edges, corners, and shape patterns.

b. Pooling Layer:

After each convolutional layer, there is usually a pooling layer to reduce the resolution of the feature maps, which reduces the number of parameters and avoids overfit- ting. A common type of pooling is Max Pooling, which only retains the largest value in each pooled region.

c. Fully Connected Layer:

After the convolution and pooling layers, the features are flattened into a vector and passed through one or more fully connected layers. These layers play the role of predicting the type of sign based on the features learned from previous layers.

d. Softmax Layer:

The Softmax layer is the final layer in the model, which helps the model make a probability prediction for each sign type. This layer ensures that the sum of the prob- abilities for all output labels is 1 and the label with the highest probability is selected as the final prediction of the model.

2. Suported Vector Machine

Since the GTSRB data consists of images, feature ex- traction is a necessary step to convert image data into numerical forms that can be processed by the SVM model. Features extracted from Convolutional Neural Networks (CNNs) are chosen as the technique because they are highly effective in learning and extracting com- plex features from images. CNNs have the ability to au- tomatically learn features at multiple levels, from basic features like edges and corners to more complex features such as shapes and structures of traffic signs. This al- lows CNNs to capture important characteristics of signs without the need for manual feature design as in HOG. Moreover, CNNs can adapt and learn the most suitable features for the task of traffic sign classification, enhanc- ing the accuracy and generalization capability of the model. The RBF Kernel (Radial Basis Function) is the ideal choice due to its robustness in handling non-linear data. RBF kernel creates a hyperplane for classification based on spatial relationships between data points. After choosing the kernel, the next step is to adjust the critical parameters of the SVM model to optimize performance. The two main parameters to tune are C (Regularization Parameter) and gamma. The C parameter controls the penalty for misclassified samples. A large C can reduce classification errors but may lead to overfitting, while a small C helps the model generalize better, though it might reduce accuracy. The gamma parameter in the RBF kernel determines the influence of a single data point. If gamma is too large, the model becomes over- fitted; if too small, the model is too generalized. Grid- SearchCV is an effective tool to systematically search for the optimal C and gamma values by testing differ- ent parameter combinations, ensuring the model has the best configuration. With a large number of features from images, SVM may face issues with training time and overfitting. PCA is a dimensionality reduction technique that helps reduce the number of features while retaining the most important information. PCA works by identi- fying the principal components, which represent the di- rections of maximum variance in the data. This allows the SVM model to process data faster while reducing the risk of overfitting, which is crucial when dealing with im- age data from the GTSRB dataset, where the number of pixel features can be very high.

3. Random Forest

In this work, we used the same dataset for training as the study conducted by Sermanet et al., known as the German Traffic Sign Recognition Benchmark (GTSRB), which is considered one of the largest available datasets for traffic sign recognition, proposed in 2011. The train- ing and test sets were derived from this dataset, contain- ing labeled images stored in CSV files. The images were resized to 32 U' 32 pixels and converted to grayscale. For feature extraction, we utilized a Convolutional Neural Network (CNN), which effectively extracts hierarchical features from the input images. This method allows the model to automatically learn and capture complex pat-

terns relevant to traffic sign classification, enhancing the overall performance of the recognition process. A GridSearchCV on hyperparameters were performed to tune the model as good with parameters like number of esti- mators, maximum tree depth, minimum samples and the impurity criterion set to 'entropy'. Meaning, a Random Forest is the collection of classification trees and each tree votes for most predicted class in case input data. This techniques is a type of ensemble learning method that adds more randomness by adding one more layer into the same i.e. bagging technique In building each tree, Random Forests do additionally change the way in which they are constructed from bootstrap samples of your dataset. Contribution of Node Splits: In a stan- dard decision tree, node splits are designed to be binary and based on evaluation of the best split out from all features. In contrast, every node in a Random Forest is split on the best of a random subset of predictors at that particular node. Indeed, this proved a powerful method which significantly outperformed all the clas- sifiers compared and has to some extent been robust against over-fitting.

4. Essemble learning with Voting Classi- fier

In the application of the Voting Classifier on the German Traffic Sign Recognition Benchmark (GTSRB) dataset, a combination of three modelsConvolutional Neural Net- work (CNN), Support Vector Machine (SVM), and Ran- dom Forest (RF)was employed to enhance the accuracy of traffic sign classification. Both hard voting and soft voting strategies were applied in the ensemble process to improve the overall model's robustness. For instance, consider the classification of traffic signs like speed limit (class 0), stop sign (class 14), and yield sign (class 13). In hard voting, if the CNN predicts class 0 (speed limit), the SVM predicts class 14 (stop sign), and the RF pre- dicts class 0 (speed limit), the final decision would be class 0, as it received the majority vote. This approach ensures that the final classification is based on the most frequent prediction across models, which is particularly useful when the models show consistent behavior on a certain class. In soft voting, the average probability for each class is calculated. Suppose the CNN assigns a 0.8 probability to class 0 (speed limit), the SVM assigns a 0.6 probability to class 14 (stop sign), and the RF assigns a 0.7 probability to class 0. In this case, the av- erage probability for class 0 would be (0.8 + 0.7) / 2 = 0.75, and for class 14, it would be 0.6. Therefore, class 0 (speed limit) would be selected as the final prediction since it has the highest average probability. This method is especially effective when the individual models pro- vide different levels of confidence in their predictions, allowing the ensemble to take into account both the ac- curacy and certainty of the models. The CNN model, designed for GTSRB, captures the spatial features of traffic signs, especially important for distinguishing be- tween signs with similar shapes but different symbols (e.g., speed limits). The SVM, using an RBF kernel, focuses on maximizing the margin between traffic sign categories, while the RF, constructed with 100 decision trees, enhances decision-making by selecting the most important features for traffic sign recognition. By combining these models through voting mechanisms, a more reliable and accurate classifier is achieved for GTSRB.

VI. Results

Model	Accuracy
Convolutional Neural Networks (CNN)	95%
Random Forest	93%
SVM	93,42%
Voting Classifier	99%

Hình 5: Fig. 3. Results

The ensemble model, implemented using a Voting Classifier that combines CNN, SVM, and Random For- est, yielded the highest accuracy at 99%. By leveraging the strengths of each individual model, the ensemble ap- proach not only improved accuracy but also enhanced the robustness and stability of the classification results. This outcome highlights the effectiveness of ensemble learning in achieving higher accuracy for traffic sign classification tasks.] The performance of each model on the German Traffic Sign Recognition Benchmark (GT- SRB) dataset is summarized as follows. The Convolu- tional Neural Network (CNN) achieved an accuracy of 95%, demonstrating its capability to extract detailed features and patterns from traffic sign images. In com- parison, traditional machine learning modelsRandom Forest and Support Vector Machine (SVM)achieved ac- curacies of 93% and 93.42%, respectively, when used as standalone classifiers. Although slightly less accurate than the CNN, these models show robust performance on the GTSRB dataset.

The ensemble model, implemented using a Voting Classifier that combines CNN, SVM, and Random For- est, yielded the highest accuracy at 99%. By leverag- ing the strengths of each individual model, the ensem- ble approach not only improved accuracy but also en- hanced the robustness and stability of the classification results. This outcome highlights the effectiveness of en- semble learning in achieving higher accuracy for traffic sign classification tasks.

VII. Discussion

The results indicate that while the CNN model alone provides high accuracy for traffic sign classification, the ensemble approach with a Voting Classifier significantly improves performance, achieving an impressive accuracy of 99%. This demonstrates that combining models with complementary strengths can enhance the overall classi- fication capability, making the system more resilient to variations in traffic sign appearance.

Future research will focus on extending this model by incorporating additional data preprocessing steps, such as adding artificial fog layers to the traffic sign images. This technique aims to simulate real-world con- ditions with poor visibility, such as foggy weather, and

evaluate the model's robustness under such conditions. By training the ensemble model with images that in- clude synthetic fog, we anticipate an improvement in its ability to recognize traffic signs accurately even in chal- lenging weather conditions. This direction not only aims to increase the accuracy further but also enhances the model's practical applicability for autonomous driving systems and intelligent transportation solutions, where visibility may often be compromised.

References

Sermanet, P., & LeCun, Y. (2011). *Traffic Sign Recog- nition with Multi-Scale Convolutional Networks*. Pro- ceedings of the International Joint Conference on Neu- ral Networks (IJCNN). This paper introduces the use of multi-scale convolutional neural networks for traffic sign recognition, achieving high accuracy on the GTSRB dataset.

Cirean, D. C., Meier, U., & Schmidhuber, J. (2012). *Multi-column Deep Neural Networks for Traffic Sign Classification*. Neural Networks, 32, 333-338. This study presents a multi-column deep neural network model for traffic sign classification, achieving excellent results on the GTSRB dataset.

Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). *Traffic-Sign Detection and Classification in the Wild*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). This paper proposes a method for detecting and classifying traffic signs in real-world conditions, using a combination of CNN and SVM.

Timofte, R., Zimmermann, K., & Van Gool, L. (2014). *Multi-view Traffic Sign Detection, Recognition, and 3D Localisation*. Machine Vision and Applications, 25(3), 633-647. This research utilizes Random Forests to detect and recognize traffic signs from multiple viewpoints.

Wei, J., Chen, L., Zhang, Y., & Xu, M. (2023). *En- semble Learning for Traffic Sign Classification Using CNN and ResNet Models*. IEEE Transactions on Intelli- gent Transportation Systems. This study demonstrates that combining CNN and ResNet models for traffic sign classification achieves over 99% accuracy on large-scale datasets through ensemble learning.

Wang, Y., Luo, C., & Zhang, Z. (2015). *A Comparison of SVM and CNN for Traffic Sign Recognition*. Inter- national Journal of Intelligent Transportation Systems Research, 13(2), 76-83. This paper provides a compar- ative analysis of SVM and CNN models for traffic sign recognition, evaluating their performance on the GT- SRB dataset.

Basu, S., Roy, S., & Sharma, P. (2020). *Traffic Sign Classification Using Random Forest and Feature Extrac- tion Techniques*. Journal of Machine Learning Research, 21(1), 1-14. This research explores the application of Random Forest for traffic sign classification, combined with feature extraction techniques.

He, K., Zhang, X., Ren, S., & Sun, J. (2017). *Deep Residual Learning for Image Recognition*. Proceedings of the IEEE Conference on Computer Vision and Pat-
tern Recognition (CVPR), 770-778. This influential pa-
per on ResNet architecture is widely referenced in traffic sign recognition research as part of ensemble approaches