

INVESTIGATION ON THE HYPERSPECTRAL AUGMENTATION TO CONSTRUCTION MATERIALS CLASSIFICATION

Zheng Liu ^{a,*}, Patrick Hunhold ^b, Ziran He ^a, Galina Polte ^a, Elske Linß ^b, Janice Kielbassa ^d,
Maik Rosenberger ^a, Gunther Notni ^{a,c}

^aTechnology university of Ilmenau, Germany

^bMaterialforschungs- und -prüfanstalt Weimar, Germany

^cFraunhofer Institute for Applied Optics and Precision Engineering, Germany

^dInstitut für Angewandte Bauforschung gGmbH, Germany

*Corresponding author: z.liu@tu-ilmenau.de

Abstract The recycling of construction waste faces challenges in material identification due to class imbalance in hyperspectral datasets. To address this, we propose integrating a data augmentation module into the classification workflow for construction materials using short-wavelength infrared (SWIR) reflectance spectra. Experiments were conducted with Random Forest (RF) and 1D-CNN classifiers across multi-class and binary classification tasks, where the latter targeted classes commonly confused with minority categories. Various augmentation methods were tested, with the self-attention-based WGAN (SA-WGAN) showing the most notable improvement. It increased the recall of the minority class by up to 60 and 48 percentage points in the multi-class and binary classification tasks, respectively, while maintaining stable performance on the majority classes.

Keywords: hyperspectral augmentation, construction waste recycling, construction materials classification, machine learning, deep learning

1. INTRODUCTION

In the construction industry, the efficient recycling and utilization of construction waste represent significant challenges and key areas of research. The high-quality recycling of waste into usable recycled aggregates is crucial for promoting sustainable economic development. To achieve this, accurate reclassification and identification of construction materials are essential steps in the recycling process. Spatial resolved reflectance hyperspectral data, covering various spectral ranges, capture the unique properties of different construction materials. Therefore, the measured reflectance spectra of the construction materials are widely used for the reclassification task. To analyse the characteristics of these measured spectra, both traditional machine learning and deep learning methods have been widely employed as tools in this field. However, the high cost of acquiring reflectance hyperspectral data and the difficulty in obtaining sufficient samples for certain types of construction materials often lead to class imbalance, where some categories of construction materials are severely underrepresented. This imbalance significantly reduces the performance of both traditional and deep learning-based classification methods, which in turn negatively

impacts downstream processes, resulting in economic and labour inefficiencies.

Therefore, this research proposes the integration of a data augmentation module into the construction materials classification workflow. This module can effectively expand and enrich the dataset by leveraging the limited hyperspectral information available from minority classes, thereby addressing the data imbalance problem and improving classification outcomes.

2. METHODS AND PROCEDURES

Methods for augmenting reflectance spectra can generally be categorized into traditional spectral processing techniques [1][2][3] and generative model-based approaches [4][5][6][7]. Traditional techniques aim to expand the dataset by introducing slight distributional variations to the original spectra while preserving the essential spectral characteristics. In contrast, generative model-based approaches leverage deep learning, particularly generative adversarial networks (GANs), to perform adversarial training that fully exploits the distributional properties of existing data. These models generate synthetic spectra that closely resemble the real data distribution, thereby enriching the dataset and effectively increasing its diversity. In this study, four data augmentation techniques were investigated: uniform noise addition, uniform noise addition with local perturbation, Wasserstein GAN (WGAN) [5], and self-attention-based WGAN (SA-WGAN) [7].

To evaluate the applicability of data augmentation in addressing data imbalance in construction materials classification based on reflectance spectra, this study constructed a practical short-wavelength infrared (SWIR) reflectance hyperspectral dataset comprising eight categories of construction materials. Among these categories, concrete had significantly fewer samples than the other seven and was therefore selected as the minority class for augmentation and further experimental analysis.

Based on this dataset, multi-class and binary classification experiments were conducted to examine the impact of the aforementioned data augmentation techniques separately on the performance of Random Forest (RF) and convolutional neural network (1D-CNN).

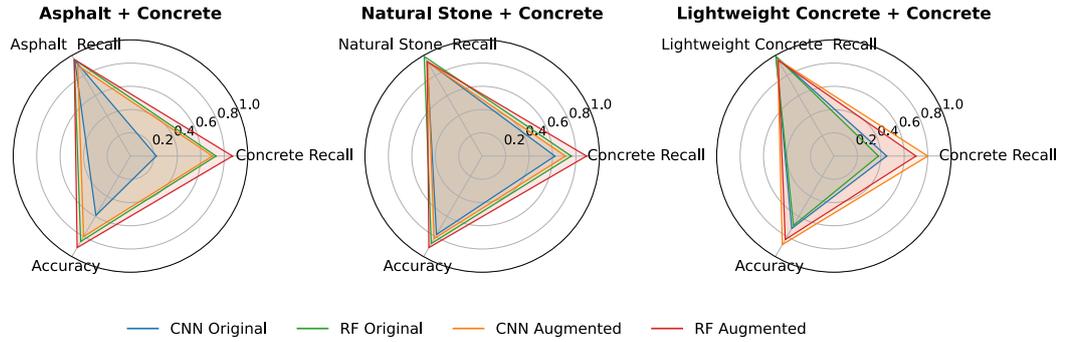


Figure 1 Performance comparison between two original and SA-WGAN-augmented classifiers in three binary scenarios.

Specifically, in the multi-class classification scenario, all eight categories were used, while in the binary scenarios, three binary classification tasks were designed by pairing the concrete with the three categories it was most frequently confused with in the original multi-class classification setting. In both scenarios, the dataset was first divided into training and test sets. To ensure the representativeness of the test set, for each construction material category, the test set was composed of reflectance spectra collected from 30 randomly selected samples. The reflectance spectra from the remaining samples were used to form the original training set. Subsequently, four data augmentation techniques were applied to the concrete in the training set, resulting in balanced training sets. The classifiers were then trained on both the original and the augmented training sets and evaluated on the same test set. The performance of the classifiers on the test set reflects the effectiveness of each data augmentation technique in addressing the issue of class imbalance.

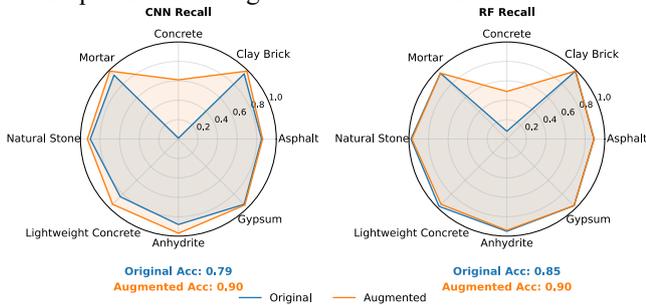


Figure 2. Performance comparison between two original and SA-WGAN-augmented classifiers in the multi-class scenario

3. RESULTS AND DISCUSSION

The results of the above experimental design demonstrated that, after augmenting the minority class using data augmentation techniques, the performance of both classifiers improved across all binary and multi-class scenarios, particularly in cases involving highly confusable categories. Among the evaluated methods, the SA-WGAN yielded the most notable improvements. In the binary classification tasks, as illustrated in Figure 1, the recall for the minority class increased by up to 48 percentage points, while the recall for the majority classes remained largely unchanged. In the eight-class classification scenario, as shown in Figure 2, the recall for the minority class increased by up to 60 percentage

points, with the majority classes showing similarly stable performance.

4. CONCLUSIONS

These results indicate that incorporating SA-WGAN as a data augmentation module into the classification pipeline for construction materials classification can effectively address the problem of poor classification performance caused by data imbalance, significantly enhancing the recognition of underrepresented categories.

ACKNOWLEDGMENTS

This work was funded by the State of Thuringia and the Thüringer Aufbaubank (TAB) by a research grant within "HypLIPS" grant number 2023 FGR 0088.

REFERENCES

- [1] Bjerrum, Esben Jannik, Mads Glahder, and Thomas Skov, **Data augmentation of spectral data for convolutional neural network (CNN) based deep chemometrics**. arXiv preprint arXiv:1710.01927 (2017)
- [2] Chakravartula, Swathi Sirisha Nallan, et al. **Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee**. Food Control 135 (2022): 108816.
- [3] Momeny, Mohammad, et al. **Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks**. Ecological Informatics 71 (2022): 101829.
- [4] Xin, Xiaowei, et al. **Combination of near-infrared spectroscopy with Wasserstein generative adversarial networks for rapidly detecting raw material quality for formula products**. Optics Express 32.4 (2024): 5529-5549.
- [5] McHardy, Rose G., et al. **Augmentation of FTIR spectral datasets using Wasserstein generative adversarial networks for cancer liquid biopsies**. Analyst 148.16 (2023): 3860-3869.
- [6] Jiawei, N. I. U., et al. **Conditional self-attention generative adversarial network with differential evolution algorithm for imbalanced data classification**. Chinese Journal of Aeronautics 36.3 (2023): 303-315.
- [7] Zhang, Qingjie, Xiaoying Wang, and Chunhui Li. **SA-WGAN Based Optimization Method for Network Traffic Feature Camouflage**. IEEE Access (2024).