

Using ensemble learning to analyze hurricane images to identify damage

K.P.S.D. Kumara Pathirana , D. Kulasiri and S. Samarasinghe

*Centre for Advanced Computational Solutions, Lincoln University, Christchurch, New Zealand
Email: samantha.kumarapathirannahalage@lincolnuni.ac.nz*

Abstract: The classification of hurricane images looking for damages includes analysing satellite or radar images of the areas affected by hurricanes and categorizing them into two classes (damaged/not damaged) based on their characteristics. This can be done using deep learning techniques where a deep neural network is trained on a large dataset of satellite images labelled as damaged or not damaged to recognize patterns and make predictions on new images. Convolutional Neural Networks (CNNs) are a type of neural network commonly used for image classification tasks. The classification process includes image preprocessing, feature extraction, and classification. In image preprocessing, the images are usually resized or filtered to remove noise and enhance quality. Feature extraction involves identifying relevant features in the images, such as building structures, that can be used to distinguish the damages of hurricanes.

The primary objective of this research is to find whether ensemble learning can be used to improve the accuracy of the classification of hurricane images. Ensemble learning is a machine learning technique that combines multiple models to improve the accuracy and robustness of predictions by taking advantage of the complementary strengths of each model.

We used the hurricane image dataset from <https://ieee-dataport.org/open-access/detecting-damaged-buildings-post-hurricane-satellite-imagery-based-customized> (Cao and Choe 2018). The training dataset has 7000 images (50% - damaged), and the testing dataset has 1200 images (50% - damaged). The experiments are performed with CNN and TPOT (Tree-based Pipeline Optimization Tool) (Olson, et al. 2016), an automated machine learning tool. It is designed to automate the entire machine-learning pipeline, including data pre-processing, feature selection, and hyperparameter tuning, to find the best machine-learning model for a given problem. TPOT cannot be used to analyse colour images since it does not take 3-dimensional datasets. Hence, we combined the CNN with TPOT, where the 3-dimensional dataset is sent into a model developed using CNN and features were extracted. The features were then given to the TPOT classifier as the input. The best pipeline selected by the TPOT classifier uses Recursive Feature Elimination for feature selection, two ExtraTreesClassifiers with 100 trees, and StackingEstimator to stack the models together into a single ensemble model. In ExtraTreesClassifier, extra-trees build an ensemble of decision trees by fitting each tree to a random subset of the training data. The overall predictive accuracy is improved by averaging the output of each tree.

The accuracy of the model built with 5 convolutional 2D layers, 3 max-pooling layers, 5 batch normalization layers, 5 activation layers, and 1 flatten layer was 0.89. The new approach replaced the output layer with the ensemble model, and the accuracy could be improved to 0.95. The CNN model took 8,910 seconds for training and the ensembled model took 36,408 seconds. Both approaches took 10 seconds to predict 1200 images. Despite of the increase in training time, time for prediction is not affected by the new development.

Hurricane damage identification is an important task for analyzing potential damages to buildings and tracking the path of hurricanes, and for studying the characteristics and behaviour of these powerful storms.

REFERENCES

- Cao, Q. D., and Y. Choe. 2018. "Detecting Damaged Buildings on Post-Hurricane Satellite Imagery Based on Customized Convolutional Neural Networks." IEEE Dataport. doi: <https://dx.doi.org/10.21227/sdad-1e56>.
- Olson, R. S., R. J. Urbanowicz, P. W. Andrews, N. A. Lavender, L. R. Kidd, and J. H. Moore. 2016. "Automating Biomedical Data Science Through Tree-Based Pipeline Optimization." Applications of Evolutionary Computation: 19th European Conference, EvoApplications, Proceedings, Part I (Springer International Publishing) 123-137. doi:10.1007/978-3-319-31204-0_9.

Keywords: *Hurricanes, image classification, ensemble learning, convolutional neural networks*