Anomaly Detection in Image Streams with Explainable AI

We define an anomaly as an unlikely occurrence that deviates from a typical behavior [1]. An anomaly could be a defect in a production line, sudden stock market fluctuations or natural disasters such as deforestation, volcanic eruptions, or floods [2] [3]. The assistance of an intelligent system to identify such disturbances would be very beneficial to initiate methods to prevent such situations in the early stages. This study forwards an AI based anomaly detection system and its testing stages primarily focused on the detection of deforestation, where when deforestation occurs, it shows an anomalous scenario which deviates from the typical sights of lush green forests.

How AI can help prevent deforestation?

Note that the topic of this research contains the phrase "image streams". An image stream is a continuous and real-time sequence of images or frames captured over time, typically by a camera, or an imaging system [4]. The temporal aspect of image streams plays a pivotal role in time series analysis for anomaly detection and this study has integrated computer vision and time series forecasting to capture the interdependency (or the time dependency) between the images during the model building process. Using time-series data and models for analysis in this study is important since it helps us understand, predict, and make decisions based on how phenomena change over time. If we consider the example of deforestation, by using image streams of a forest cover that has been captured over time, we can train a model for the typical behavior of lush greenery. Using this historical information about the forest cover, our model will be able to detect any anomalies such as deforestation or forest fires that occur in future timestamps.

The detection of anomalies has been done using Artificial Intelligence extensively in many domains and researchers have experimented with technologies such as Machine Learning, Deep Learning, ensemble methods and others. This research developed two anomaly detection frameworks with the use of Machine Learning and Deep Learning in order to conduct a comparison between the technologies to determine the more optimized approach for anomaly detection. Moreover, Explainable AI plays a major role in providing reasoning to the outcomes of the black-box models of Deep Learning. Machine learning involves the development of algorithms and models that use data to recognize patterns and enable computer systems to learn and make predictions or decisions based on the data. Deep learning, being a subset of Machine Learning, has the ability to learn through large amounts of data. Refer the diagram below for an overview of the anomaly detection system developed through this research.

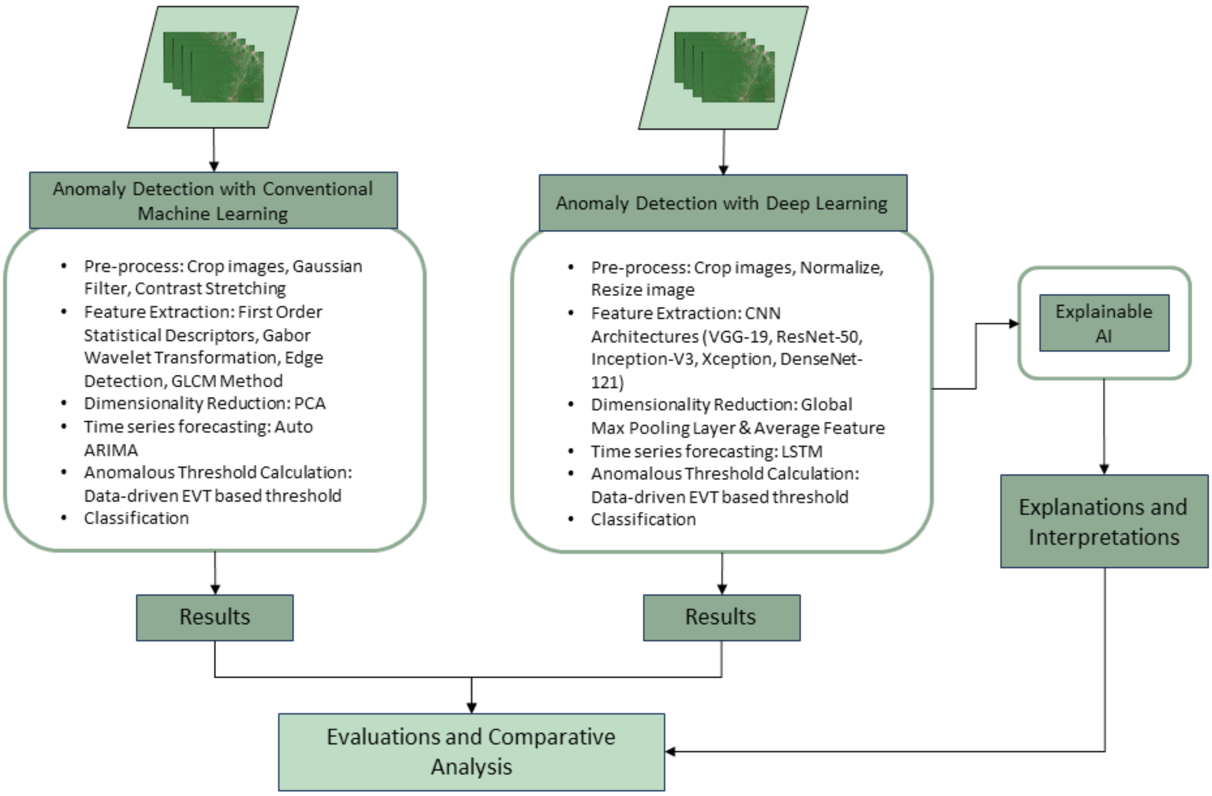

Figure 1: Overview of the Methodology

How we use Machine Learning and Deep Learning in the application of anomaly detection is quite clear when we breakdown the process. The image streams collected through years are the data we use, and the system tries to identify unique characteristics in the data that indicate the typical forest covers where deforestation has not occurred. These characteristics are used in analyzing and for forecasting the future behavior and thus predicting if an anomaly such as deforestation has occurred. This process in detail can be viewed as follows.

Pre-processing: The initial step that involves tasks like cleaning, scaling, and transforming data. It's crucial because it enhances data quality, normalizes, and prepares it for accurate modeling, ultimately improving the effectiveness and reliability of analytical or predictive tasks [5].

Feature Extraction: The process of selecting and transforming elements that constitute to meaningful patterns, textures, or structures from the image data to create a feature-set which will be an informative representation for further analysis or other tasks. The Machine Learning module extracts edge features, Gabor Wavelet features and texture-based features using First Order Statistical Descriptors and the GLCM method [6]. Meanwhile, the Deep Learning method used pre-trained Convolutional Neural Network architectures (VGG-19, ResNet-50, Inception-V3, Xception, DenseNet-121) [7] [8].

Dimensionality Reduction: The feature set that has been extracted from the Feature Extraction component results in a large number of features for each image in the image series dataset. Such a high number of features can lead to several problems, including the Curse of Dimensionality problem. Therefore, we narrow down the large, extracted feature set of time-series data into a more manageable dataset that retains the important information from the original dataset. The Machine Learning module used Principal Component Analysis (PCA) and the Deep Learning module added a Global Max Pooling layer to the CNNs to get a single feature vector for each image and the obtained the average of the feature vector for each image to obtain a univariate dataset.

Time-series Forecasting: A statistical technique used to predict future values or trends in a sequence of data points ordered chronologically (time-series data). The univariate dataset obtained after dimensionality reduction is split into training and test datasets and the training set is used to train the time series forecasting model to find the fitted values of the future timestamps. The Machine Learning model uses the Auto-ARIMA model, and the Deep Learning model used an LSTM for this purpose. Once we get the forecasted values, we also calculate the residuals and the forecasted error series from the test dataset. The residuals are the difference between the actual train y and the fitted values and forecast error series is the difference between the actual test y and the forecasted value.

Threshold Calculation: The residuals from the previous component are used to calculate a data driven threshold which can be used for classification. The method used for this purpose is an Extreme Value Theory (EVT) based anomalous threshold calculation since many existing threshold calculation methods are based on unrealistic assumptions. As anomalies have a low probability of happening, extreme values can be considered to represent anomalies. Therefore, in this technique a child distribution for generalized extreme values is plotted and using that distribution, threshold calculation is conducted [9].

Binary Classification: Any data point of the forecasted series which is above the upper anomalous threshold or below the lower anomalous threshold is defined as an anomalous behavior and anything in between is defined as a typical behavior.

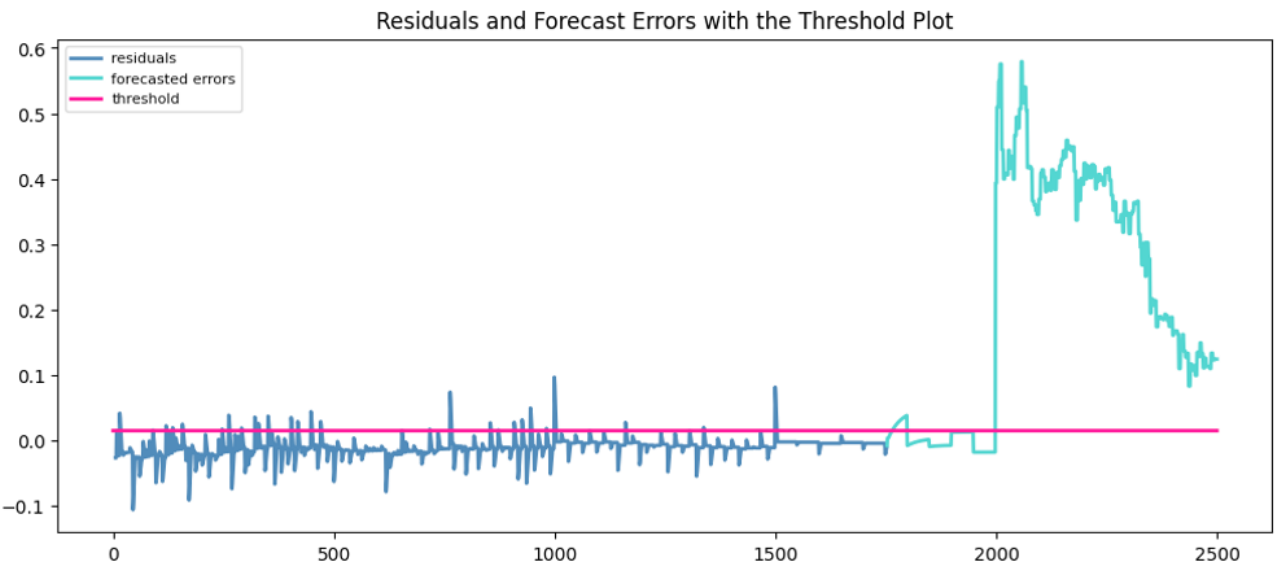

Figure 2: Plot of the Residuals and Forecast Errors with the Threshold

How Explainable AI can help build trust

The algorithms used in the Machine Learning module are interpretable, however, the algorithms used is Deep Learning such as CNNs are not interpretable and hence are called black-box models. Their complex architectures make it challenging to interpret how they arrive at specific predictions. This is problematic as users may find it difficult to understand, trust, and validate, thereby compromising transparency and accountability. As a solution to this problem, we use Explainable AI to help us understand and interpret predictions made by the models. Two model agnostic models, LIME and SHAP have been used for this purpose. SHAP gives a relative degree about how positively or negatively each pixel's contribution was and LIME gives a depiction of which pixels/patches contributed positively or negatively to the outcome that was categorized as an anomaly [10]. The results of the XAI module can be seen in Figure 3.

Key Takeaways from this research

The novelty of the system can be briefly listed as follows,

- A novel anomaly detection framework was proposed by treating anomaly detection as a one class classification, hence increasing generalizability of the framework.
- The interdependency between the images is captured by using time series forecasting.
- Calculating a data-driven EVT based anomalous threshold is a novel approach in the field.
- Explainable AI helps increase human trust on the system by interpreting black-box models.
- The class imbalance problem is countered through this research by treating the anomaly detection problem as a one-class classification problem and modeling for the typical behavior.

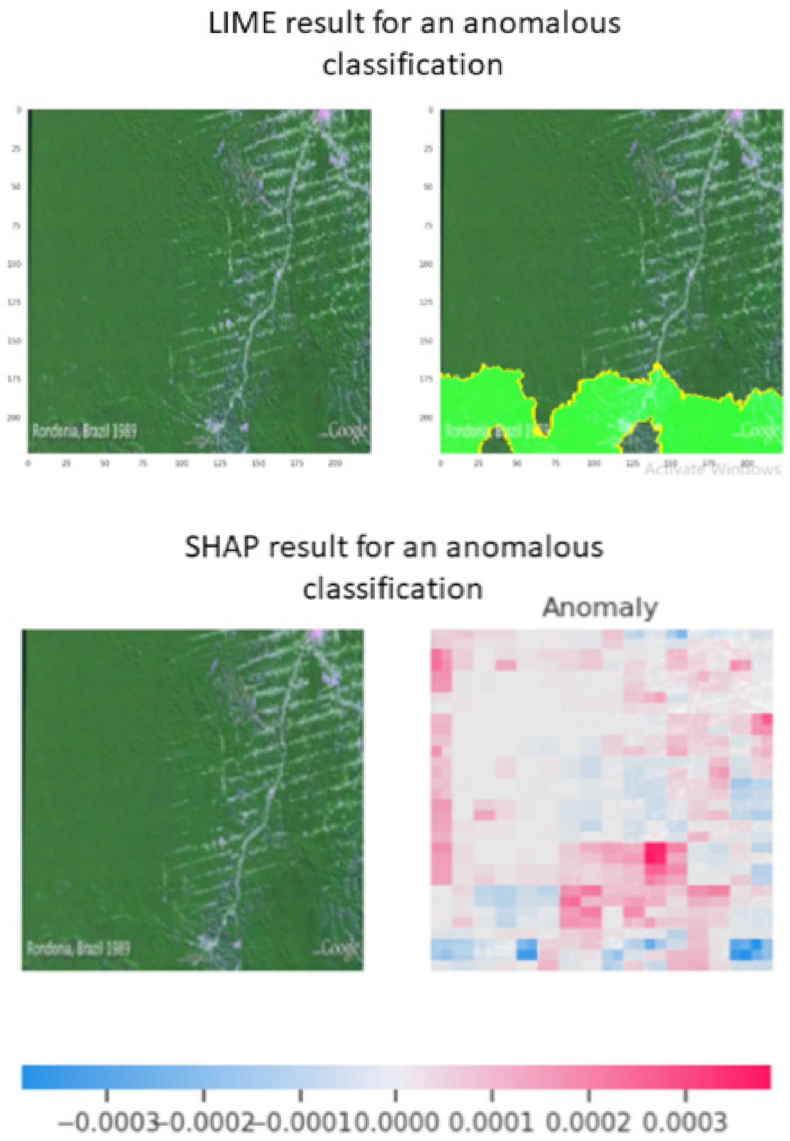

Figure 3: XAI Results of LIME and SHAP for an anomalous classification

References

[1] C. Raghavendra and C. Sanjay, "Deep learning for anomaly detection: A survey," arXiv preprint arXiv:1901.03407, 2019.

[2] A. Leman and F. Christos, "Anomaly, event, and fraud detection in large network datasets," in Proceedings of the sixth ACM international conference on Web search and data mining, 2013, pp. 773--774.

[3] D. R. Sowmya, and D. P. Shenoy, K. R. Venugopal, "Remote sensing satellite image processing techniques for image classification: a comprehensive survey," International Journal of Computer Applications, vol. 161, no. 11, pp. 24--37, 2017.

[4] S. Arturo, R. Benoit, W. Joseph, F. Ji-Lu, L. Peijun, C. Mei Mei, B. A. Stephanie, "Estimation of the distribution of Tabebuia guayacan (Bignoniaceae) using high-resolution remote sensing imagery," Sensors, vol. 11, no. 4, pp. 3831--3851, 2011.

[5] B. Catarina, C. M. Emre and M. S. Jorge "A survey of feature extraction in dermoscopy image analysis of skin cancer," IEEE journal of biomedical and health informatics, vol. 23, no. 3, pp. 1096--1109, 2018.

[6] N. Nooshin and K. Miroslav, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features," Computers \& Electrical Engineering, vol. 45, pp. 286--301, 2015.

[7] H. Kaiming, Z. Xiangyu, R. Shaoqing and S. Jian, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770--778.

[8] H. Gao, L. Zhuang, V. Der Maaten, W. Q. Kilian, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700--4708.

[9] Xiang, Gang and Lin, Ruishi, "Robust Anomaly Detection for Multivariate Data of Spacecraft Through Recurrent Neural Networks and Extreme Value Theory," IEEE Access, vol. 9, pp. 167447--167457, 2021.

[10] Dikshit, Abhirup and Pradhan, Biswajeet, "Interpretable and explainable AI (XAI) model for spatial drought prediction," Science of the Total Environment, vol. 801, 2021.

Article by

Nethmi Wijesinghe, Rashmi Perera, Nethmee Sellahewa, Priyanga Talagala

Department of Computational Mathematics, Faculty of Information Technology, University of Moratuwa, Sri Lanka