Anomaly Detection in Image Streams with Explainable AI

By Team Bits of Erised

Meet Our Team

		
Rashmi Perera	Nethmee Sellahewa	Nethmi Wijesinghe

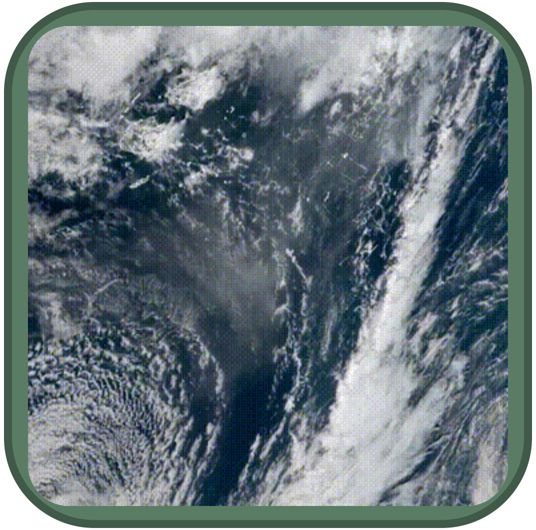
Supervised By: Dr. (Ms.) Priyanga D. Talagala

Motivation

- Benefits of extracting information and detecting anomalies from streaming data is prevalent with the availability of high-quality image streams

Forest Coverage of Rondônia, Brazil

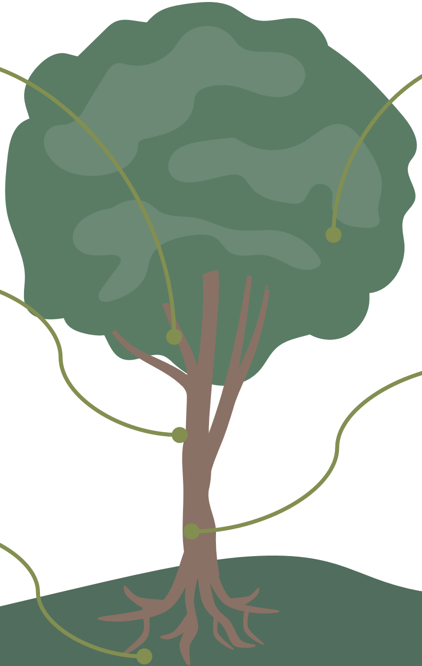
1975 - 2001



Undersea Volcano Eruption near Tonga

Sri Lankan Flood Situation

Significance of the Study



01

Deforestation in Sri Lanka – average 1.14% annual rate

02

High density populations and inappropriate land usage due to **improper urban planning** have an impact on climate change.

03

Habitat degradation and diseases such as ‘Sena caterpillar’ problem.

04

Damage of vegetation communities, loss of plant and animal habitats ascribe to **bushfires**.

05

Forest disturbances caused by both natural and man-made activities.

Significance of the Study (cont.)

An intelligent system for early detection of anomalous or unusual behaviors in image streams

Novelty of Our System

Research Gap	Novelty/Main Contribution
1. **Treat anomaly detection problem as a supervised learning problem.** **Issue:** Generalizability is low.	• Proposed a **novel anomaly detection framework**. • In the proposed framework, we treated anomaly detection problem as a **one class classification**.
2. **Most of the existing methods ignore the interdependency between the images (time dependency.**	• Integrated computer vision and **time series forecasting** to capture the interdependency between the images during the model building process.
3. **Most of the existing anomaly detection methods use manual thresholds and unrealistic assumptions.**	• Novel approach to calculate a **data driven anomalous threshold using EVT**.

Novelty of Our System (cont.)

Research Gap	Novelty/Main Contribution
4. **Existing deep learning-based anomaly detection methods solely focus on the classification task. This leads to lack of explainability.** **Issue:** Reduces the human trust on the system.	• Proposed a novel framework that integrated computer vision, time series forecasting and **explainable AI** to detect anomalies in image streams.
5. **Existing methods suffers from class imbalance and sometimes in order to address this some researchers has used different approaches. These methods have their own limitations.** - Up sampling - Down sampling - Data augmentation	• We treat the anomaly detection problem as a one class classification problem and **model for the typical behavior**.

Aim & Objectives

The aim of this research is to develop a novel framework that detects and interprets anomalies in images streams using computer vision, time series forecasting, extreme value theory, and explainable AI.

Objectives:

- Defining a pool of features that encapsulate the signal information content in image streams using computer vision.
- Propose a novel framework to detect anomalies in image streams in different application domains using time series forecasting and extreme value theory based anomalous threshold.
- Implementing a suitable explainable model for anomaly detection in image streams.

Methodology

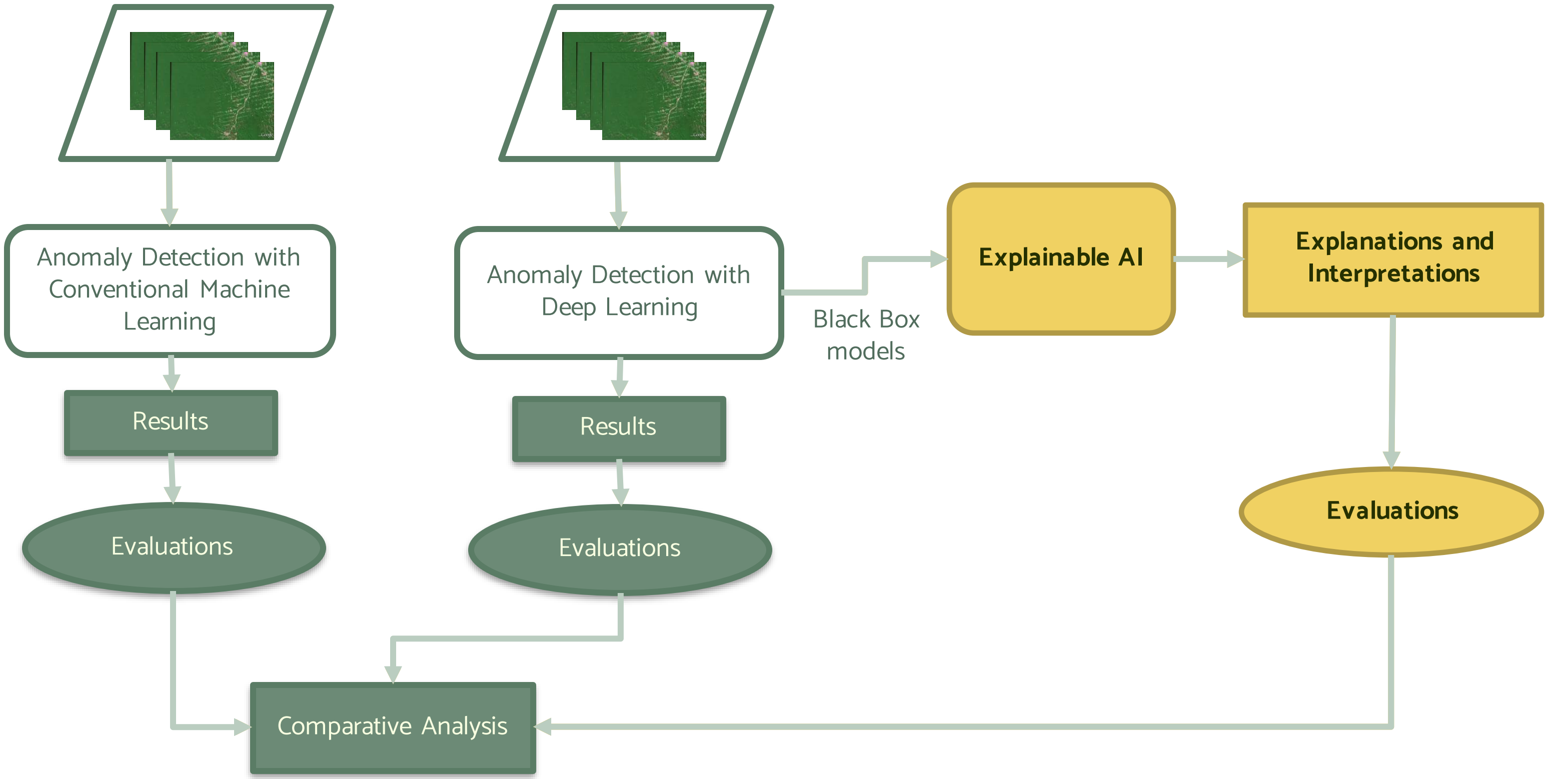
Definition of an Anomaly

In our context, we define anomaly as an observation that is very unlikely given the forecast distribution.

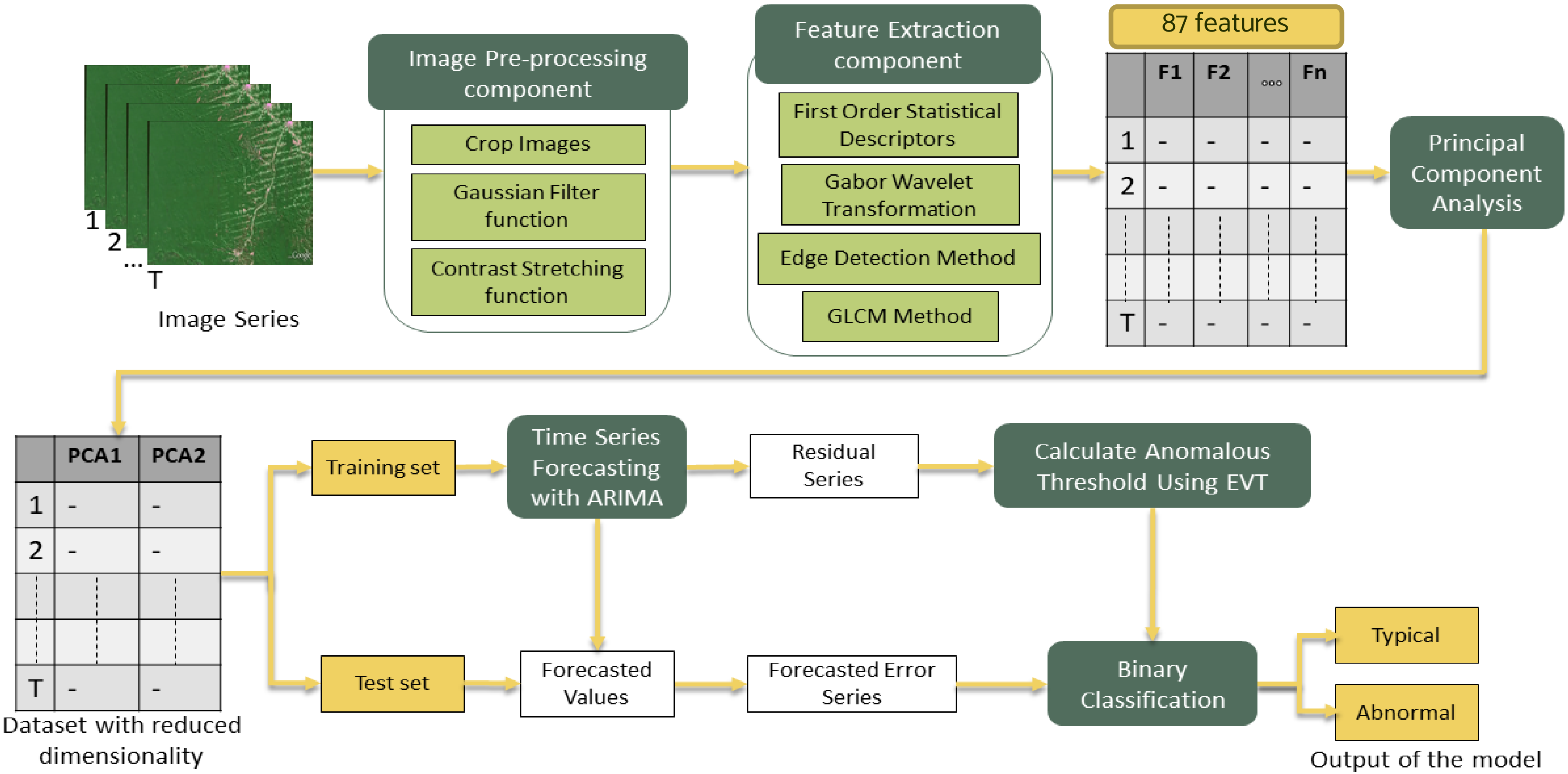
Main Assumptions

- Anomalies show a significant deviation from the typical behavior of a given system.
- A representative dataset of the system's typical behavior is available to define a model for the typical behavior of the image streams generated by a given system.

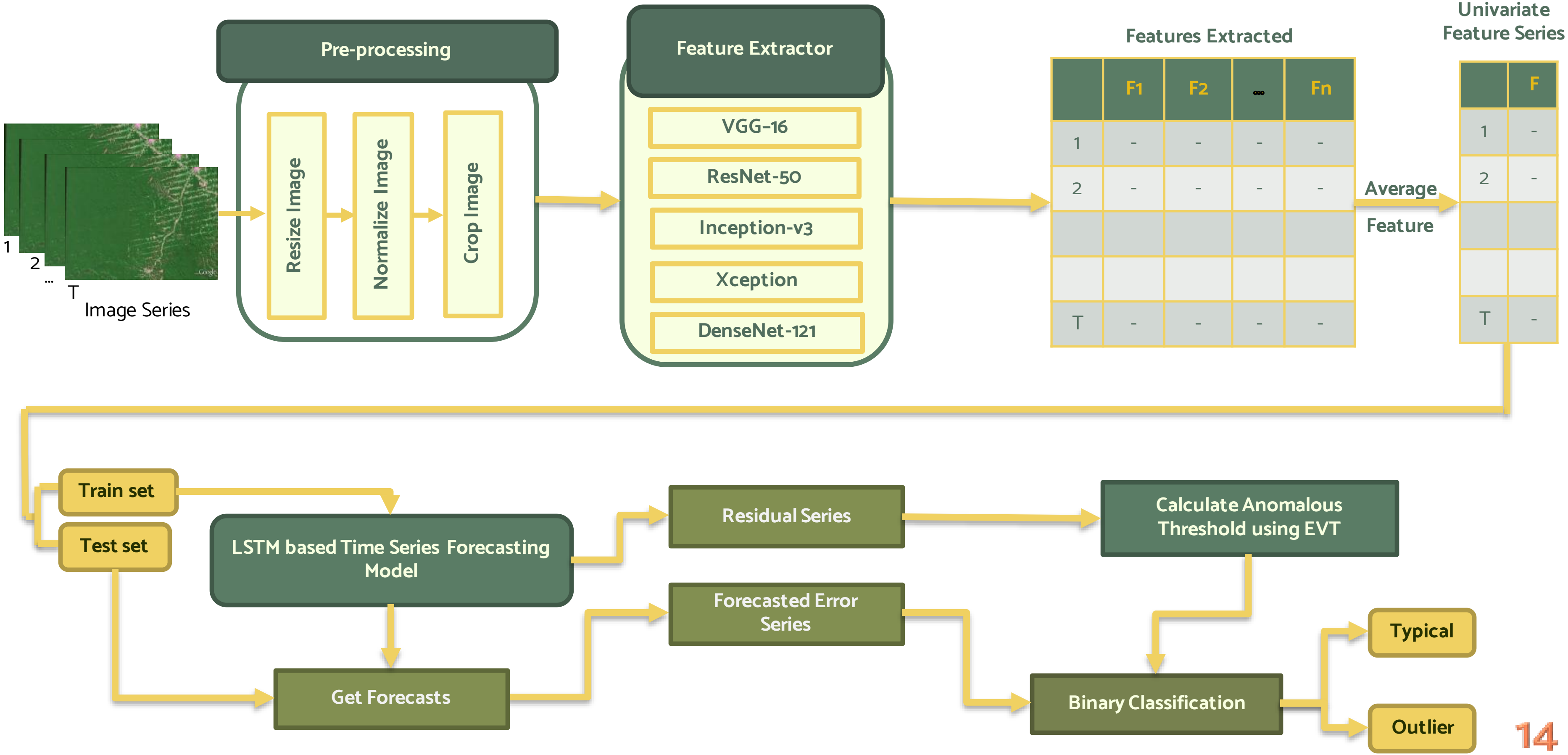
Methodology



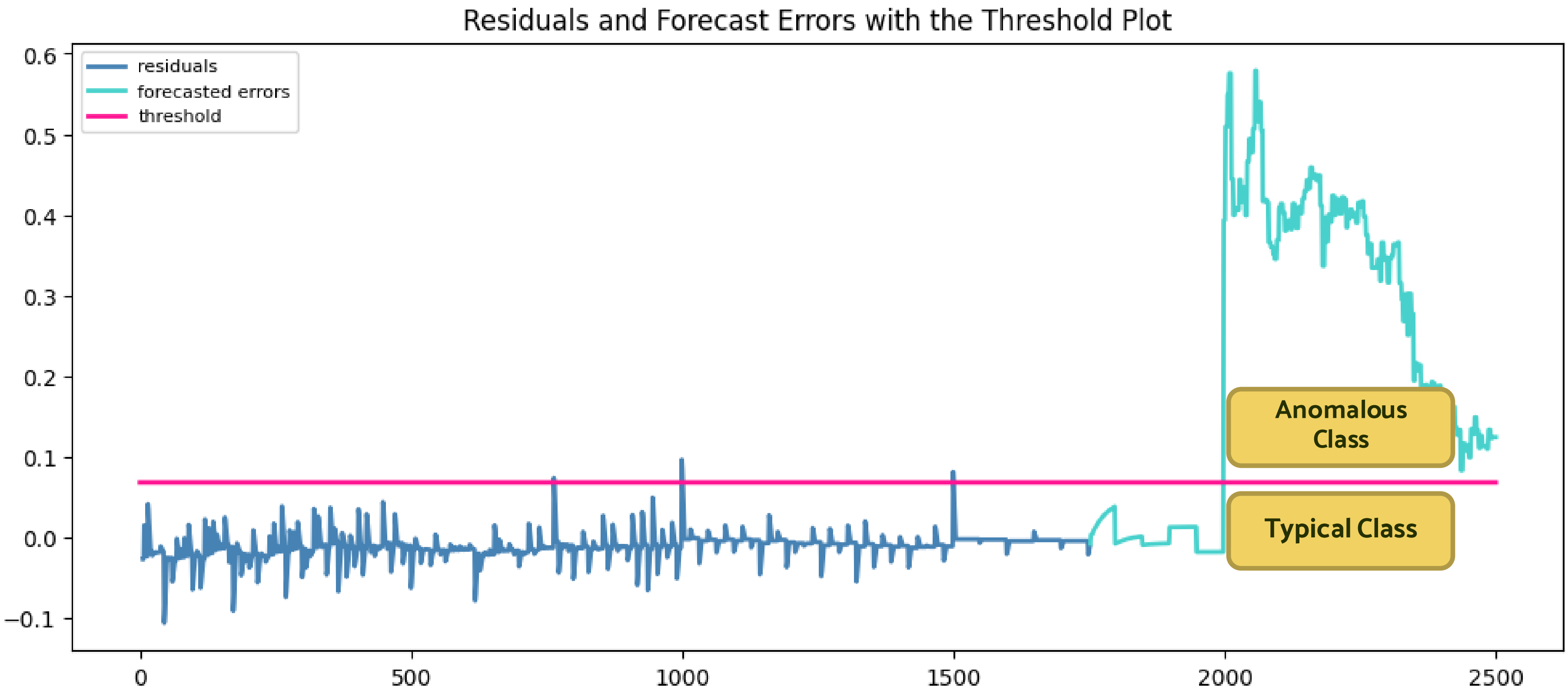
Traditional Machine Learning Module



Deep Learning Module



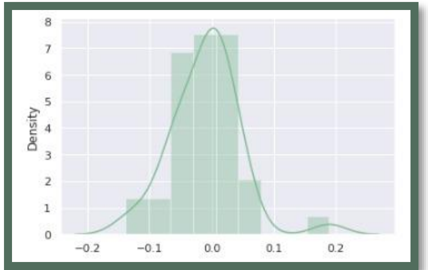
Binary Classification using EVT based Threshold



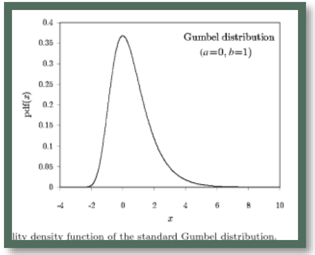
EVT based Anomaly Threshold Calculation

No	Residual Values

Identifying the Residual distribution



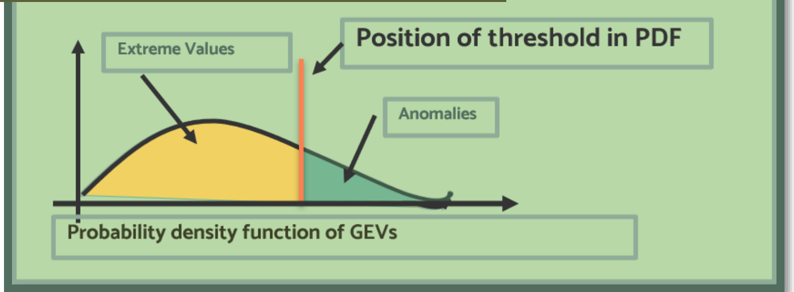
Using Fisher Tippett intuition finding the most suitable GEV distribution



Applying the relevant transformation function to transform the residuals to their EV' probability density curve.

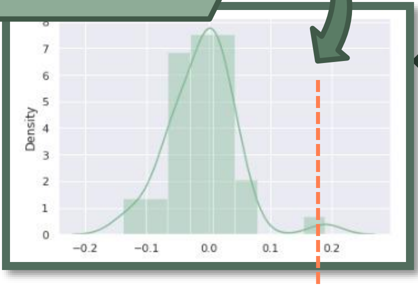
$$gumbel\_l.pdf(x) = exp(x - exp(x))$$

Position of the Threshold n GEV's PD.

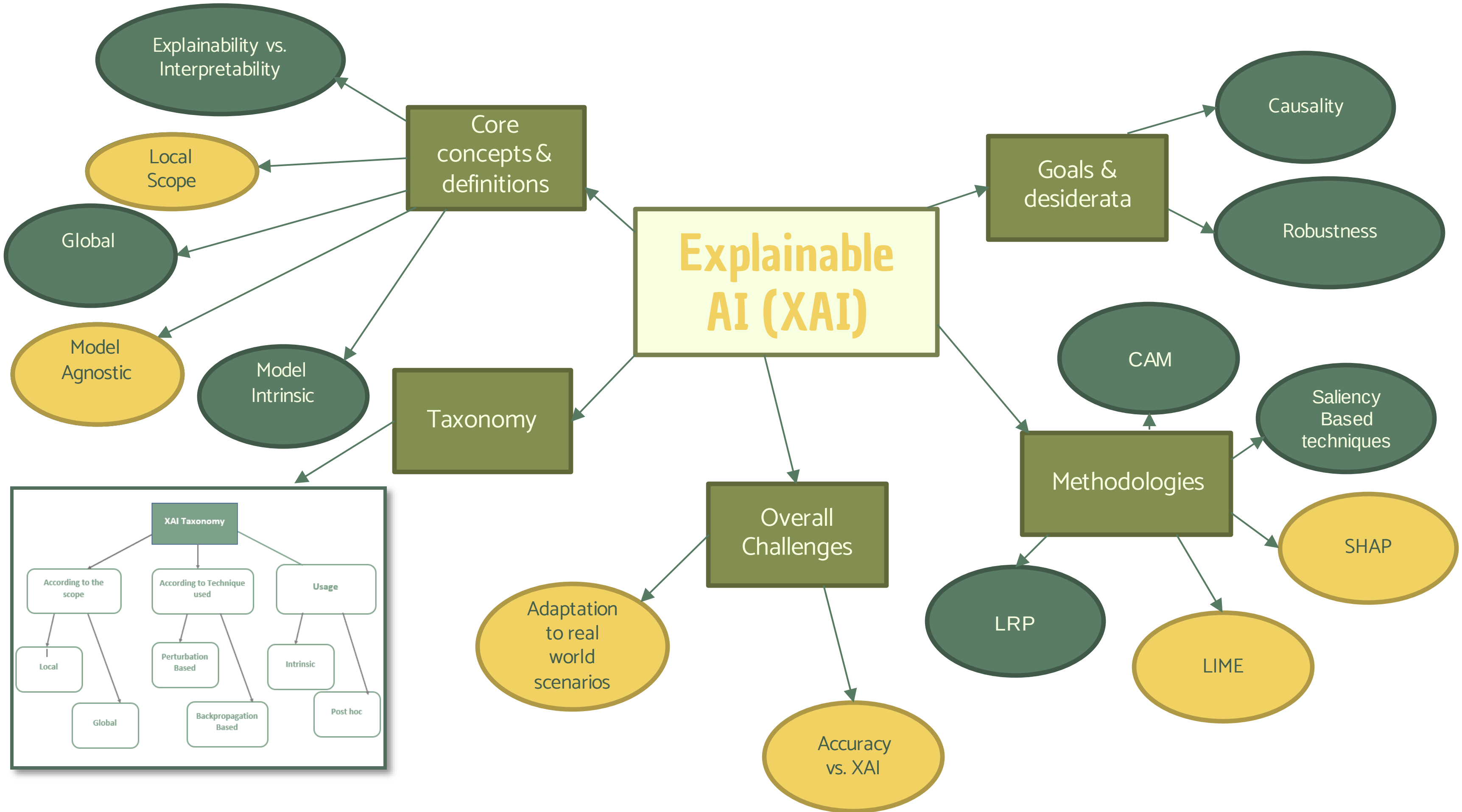


Obtaining the EVT-based anomaly threshold value by applying re-transformation

Threshold value



Explainable AI Module



Explainable AI Module - (LIME)

Shap

Universal set of features (F1,F2,F3)

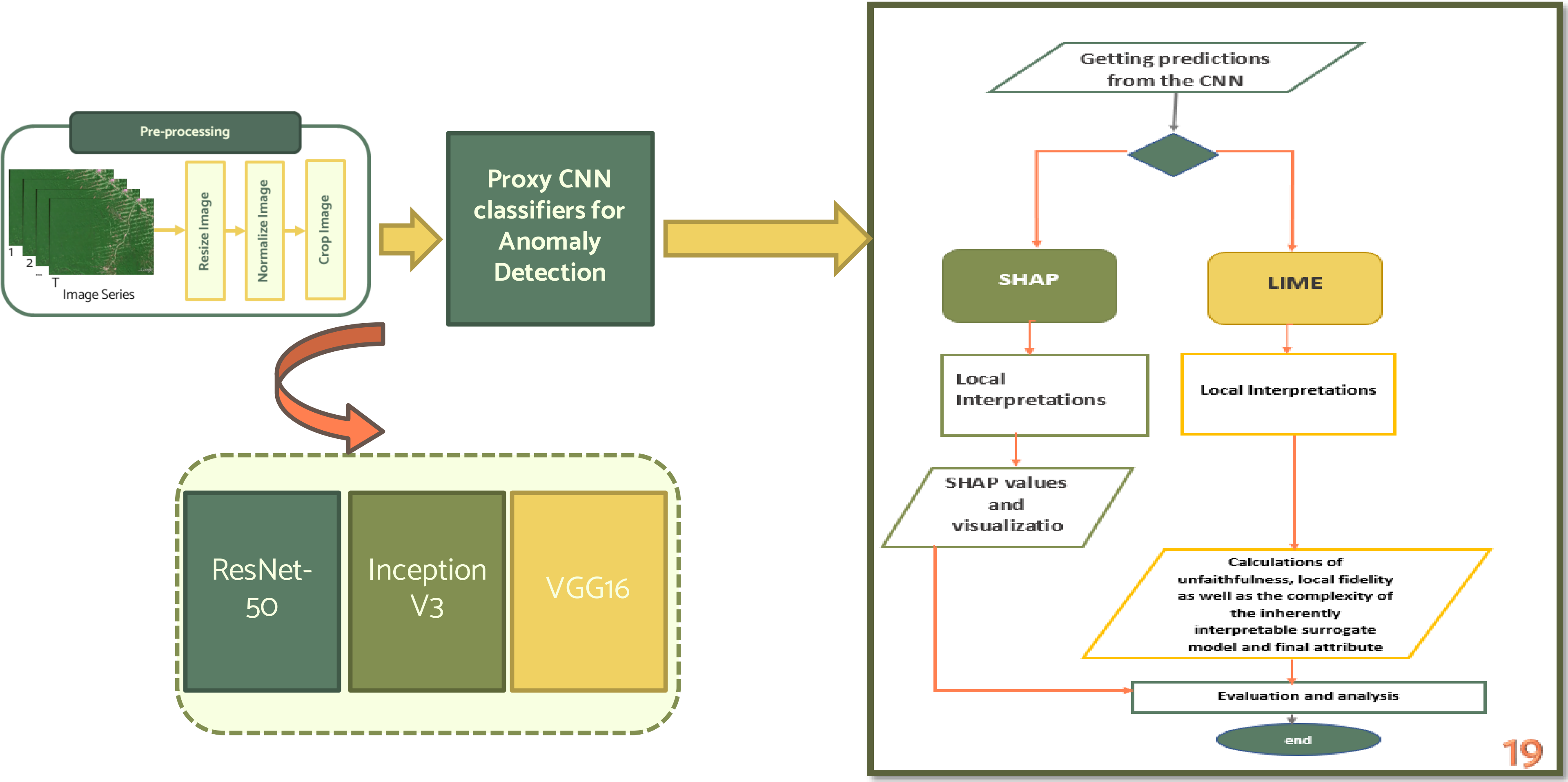
Feature combinations	Predictions from the black box model for Class A
{0}	0
{F1}	0.15
{F2}	0.65
{F3}	0.8
{F1, F2}	0.68
{F1, F3}	0.83
{F3, F2}	0.94
{F1, F2, F3}	0.96

- Calculates the marginal contributions of each feature towards class A
- Total Feature contribution using SHAP values
- Feature Map visualization with the degree of feature importance.

(LIME)

- Decomposes the target image into homogeneous patches
- Create number of images using perturbation by turning on and off the super pixels.
- Get predictions of the new image set.
- Build the local surrogate model with feature contributions.

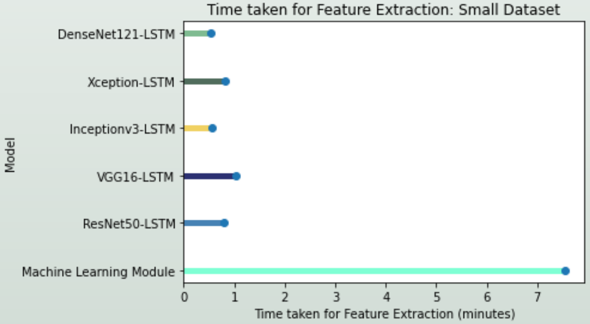
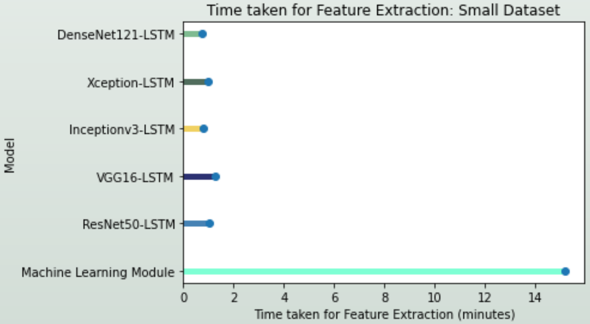
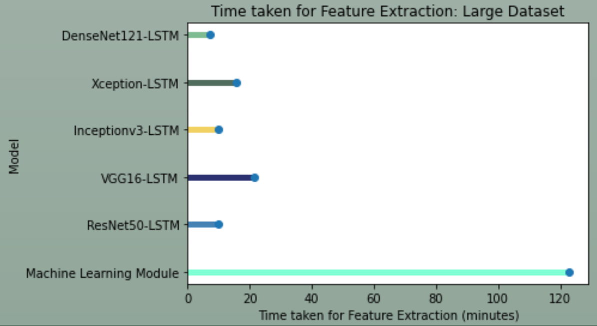
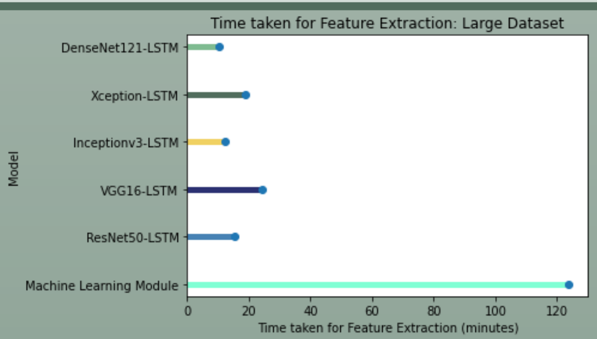
Explainable AI Module



Evaluation

	Conventional Machine Learning		Deep Learning Techniques										
	Small Dataset	Large Dataset	Small Dataset					Large Dataset				
			InceptionV3	ResNet50	VGG16	Xception	DenseNet121	InceptionV3	ResNet50	VGG16	Xception	DenseNet121
Box Plot Threshold Calculation Technique												
Accuracy	98.3%	73.5%	80%	93.33%	90%	90%	11..67%	95.05%	89.04%	75.67%	70.05%	67,65%
Sensitivity	0.98	1.0	0.77	0.92	0.94	1.0	0.11.	1.0	1.0	1.0	0.96	1.0.
Specificity	1.0	0.204	1.0	1.0	1.0	0.25	0.125	0.85	0.67	0.27	0.18	0.028
Positive Predictive Value	1.0	0.72	1.0	1.0	1.0	0.896	0.46	0.93	0.86	0.73	0.70	0.67
Negative Predictive Value	0.91	1.0	0.4	0.67	0.73	1.0	0.02	1.0	1.0	1.0	0.69	1.0
EVT Threshold Calculation Technique (with 95% confidence)												
Accuracy	98.3%	86.27%	86.7%	92.9%	90%	88.33%	86.67%	99.6%	92.9%	76.2%	70.05%	79.55%
Sensitivity	0.98	1.0	1.0	0.92	0.94	1.0	1.0	1.0	1.0	1.0	0.96	1.0
Specificity	1.0	0.59	0.0	1.0	1.0	0.125	0.0	0.987	0.79	0.29	0.19	0.39
Positive Predictive Value	1.0	0.83	0.87	1.0	1.0	0.88	0.87	0.99	0.90	0.74	0.70	0.77
Negative Predictive Value	0.91	1.0	nan	0.67	0.73	1.0	nan	1.0	1.0	1.0	10.69	1.0

Time Comparison

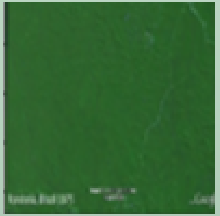
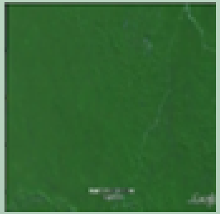
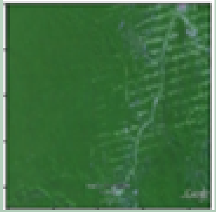
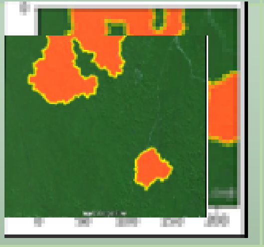
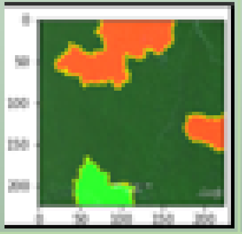
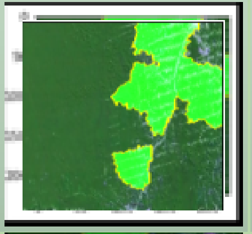
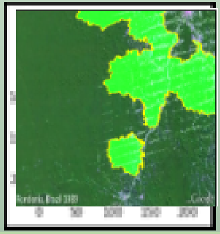
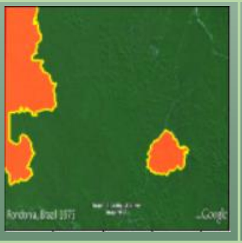
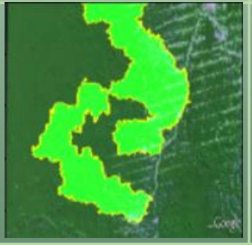
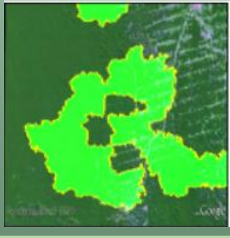
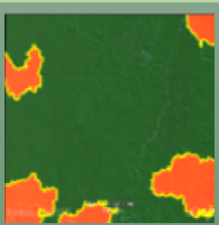
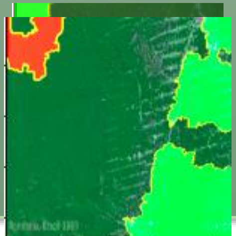
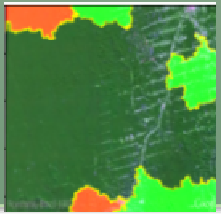
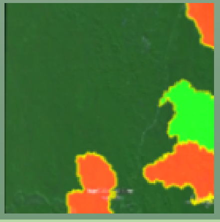
<table>
<tr><th colspan="3">Time Comparison of the CNN based LSTM Model</th></tr>
<tr><th>Dataset</th><th>Time Taken for Feature Extraction</th><th>Total Time Taken by the Model</th></tr>
<tr><td>Small Dataset</td><td>
</td><td>
</td></tr>
<tr><td>Large Dataset</td><td>
</td><td>
</td></tr>
</table>

Generalizability

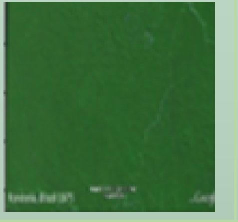
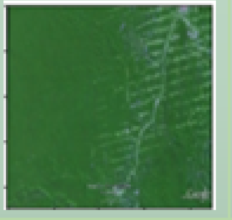
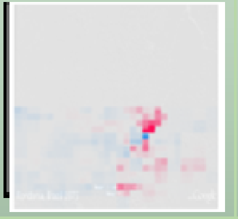
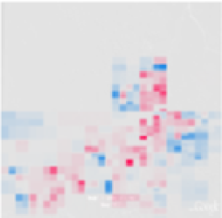
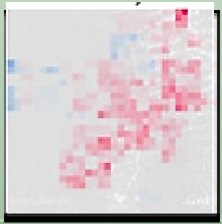
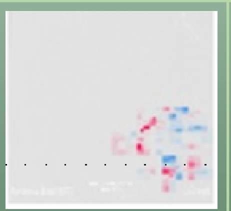
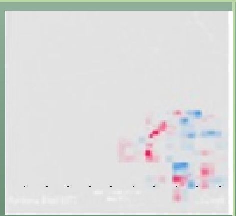
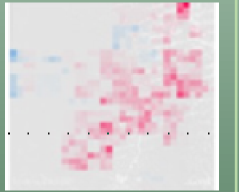
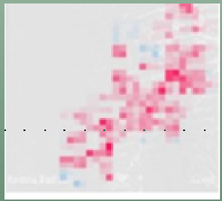
Performance Metrics for Different Datasets

Model	Traditional Machine Learning Model				Deep Learning Model			
Dataset	Small Dataset		Large Dataset		Small Dataset		Large Dataset	
	Deforestation	Volcano Eruption	Deforestation	Volcano Eruption	Deforestation	Volcano Eruption	Deforestation	Volcano Eruption
Accuracy	**98.3%**	**96.7%**	86.3%	91.2%	95%	64..29%	**99.6%**	**94.79%**
Sensitivity (Sn)	**0.98**	**1.0**	1.0	1.0	0.94	1.0	**1.0**	**0.92**
Specificity (Sp)	**1.0**	**0.9**	0.59	0.74	1.0	0.47	**0.987**	**1.0**
Positive Predictive Value (PPV)	**1.0**	**0.95**	0.83	0.88	1.0	0.47	**0.99**	**1.0**
Negative Predictive Value (NPV)	**0.91**	**1.0**	1.0	1.0	0.73	1.0	**1.0**	**0.86**

Explainable AI Module - LIME

	100 	1900 	2250 	2450 
ResNet50				
InceptionV3				
				

Explainable AI Module - SHAP

	100		1900		2250		2450	
ResNet50								
InceptionV3								
VGG16								

Further Work

- **Machine Learning Module**: Can sport a more optimized feature extraction and selection procedure to increase the performance.

- **Deep Learning Module**: Can be optimized to accommodate smaller datasets, feature selection and hyperparameter tuning.

- **Explainable AI Module**:
    - Explainable AI techniques used in this study are both perturbation-based techniques and as a further study backpropagation based XAI technique such as LRP can be applied in comparison with the existing methods.
    - A proper evaluation method for XAI techniques can be applied for the XAI methods used.

THANK YOU!

Slides available at:
https://prital.netlify.app/talks/OCTAVE2022/OCTAVE2022.pdf

ACKNOWLEDGEMENT

RETINA research and innovation project funded by UNESCO and the International Development Research Centre, Ottawa.

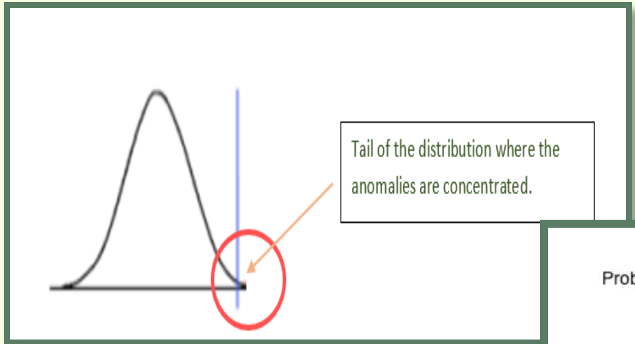
Q&A

EVT based Anomaly Threshold Calculation

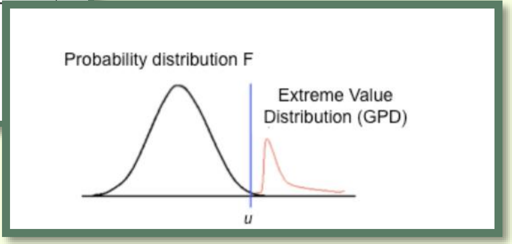
Limitations of EDA based Threshold

Assumes the parent distribution conforms to a Gaussian distribution.

Theoretical Foundation for EVT



Anomalies resides on the tails of the original distributions thus, its is difficult to correctly identify the anomaly threshold from the original curve.

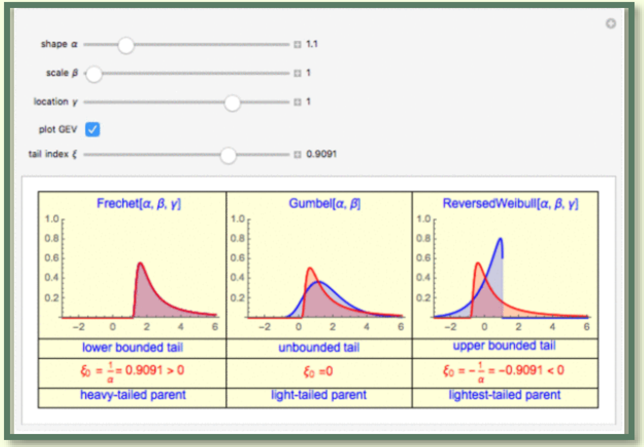


A child curve should be plotted to obtain the accurate anomaly threshold value.

Fisher Tippet Theorem

According to the Fisher Tippett Theorem any GEV distribution of any natural distribution that satisfies a weak condition (Described in Implementation) should take one of the following shape.

- Frechet
- Weibull
- Gumbell



Explainable AI Module - (LIME)

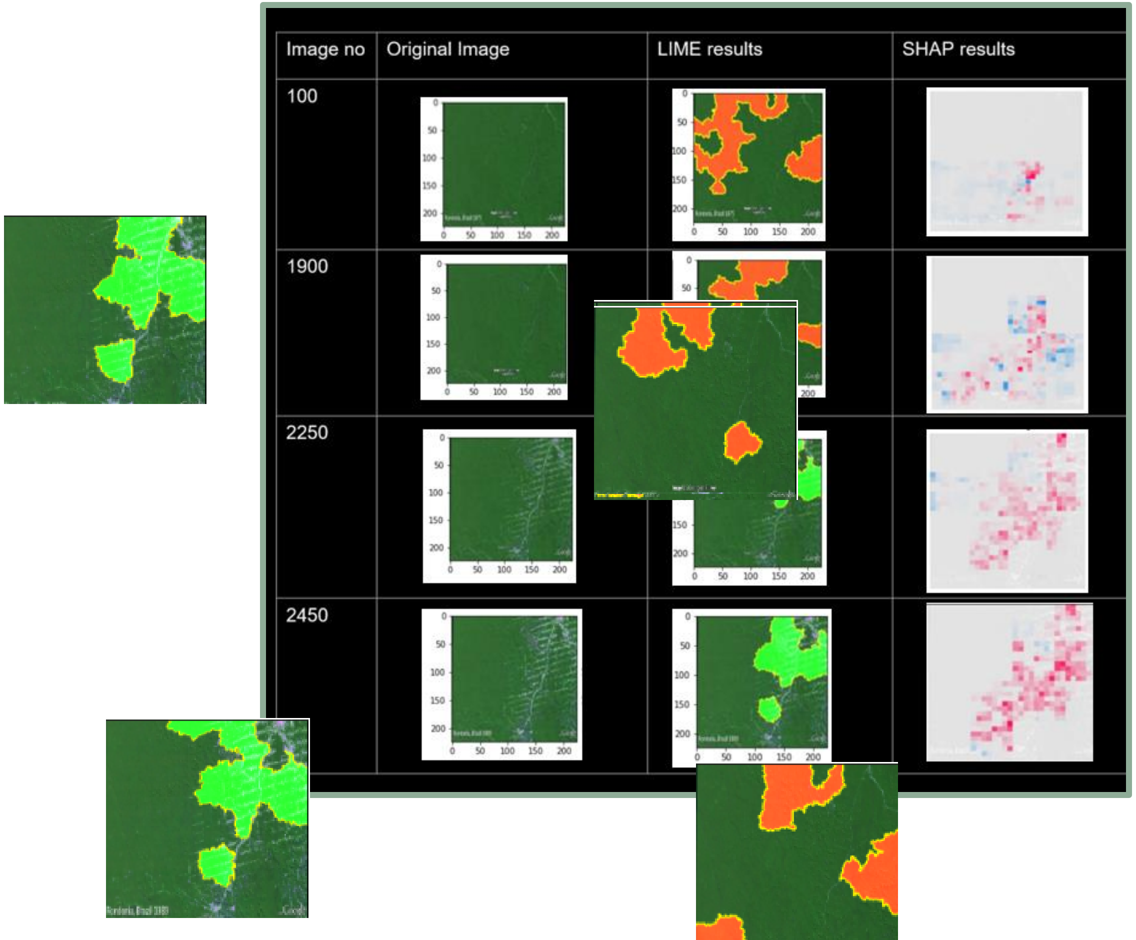
Locally Interpretable Model Agnostic Explanations (LIME)

$$\varepsilon(x) = argmin \; \boxed{\text{Unfaithfulness}} + \boxed{\text{Complexity}}$$

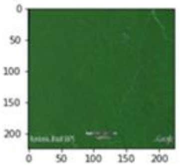
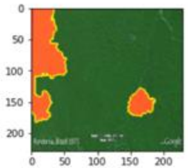
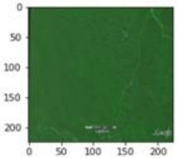
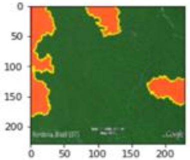
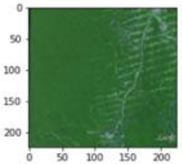
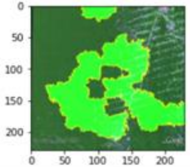
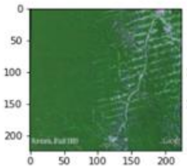
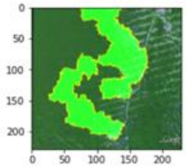
How LIME works with Images?

Decomposes the target image into homogeneous patches
Create number of images using perturbation by turning on and off the super pixels.
Get predictions of the new image set.
Build the local surrogate model with feature contributions.

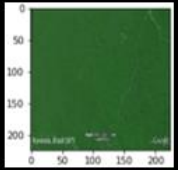
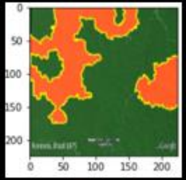
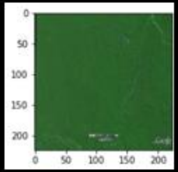
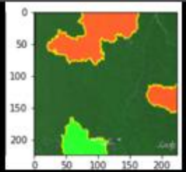
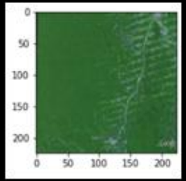
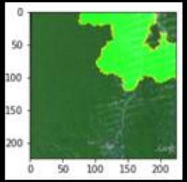
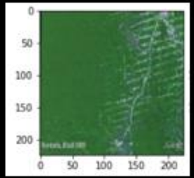
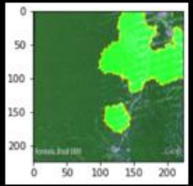
Explainable AI Module - ResNet50



Explainable AI Module – InceptionV3

Image No:	Original Image	LIME Result	SHAP Result
100			
1900			
2250			
2450			

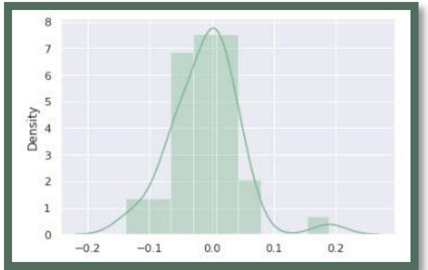
Explainable AI Module – VGG16

Image no	Original Image	LIME results	SHAP results
100			
1900			
2250			
2450			

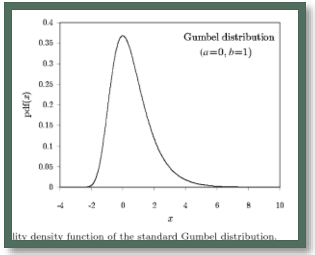
EVT based Anomaly Threshold Calculation

No	Residual Values

Identifying the Residual distribution



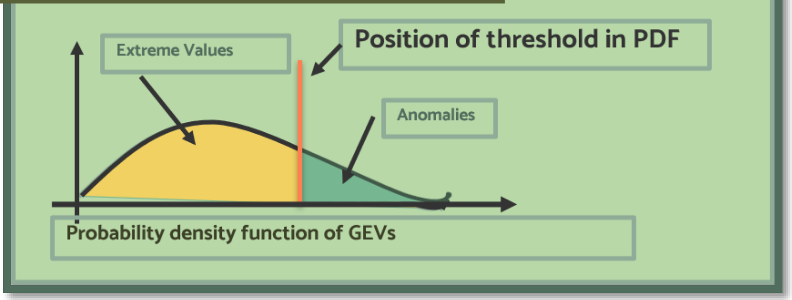
Using Fisher Tippett intuition finding the most suitable GEV distribution



Applying the relevant transformation function to transform the residuals to their EV' probability density curve.

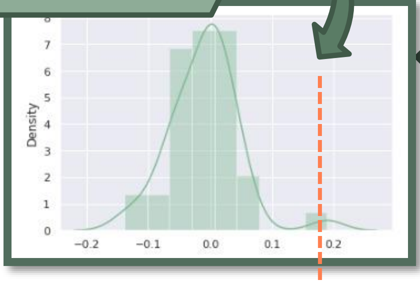
$$gumbel\_l.pdf(x) = exp(x - exp(x))$$

Position of the Threshold n GEV's PD.

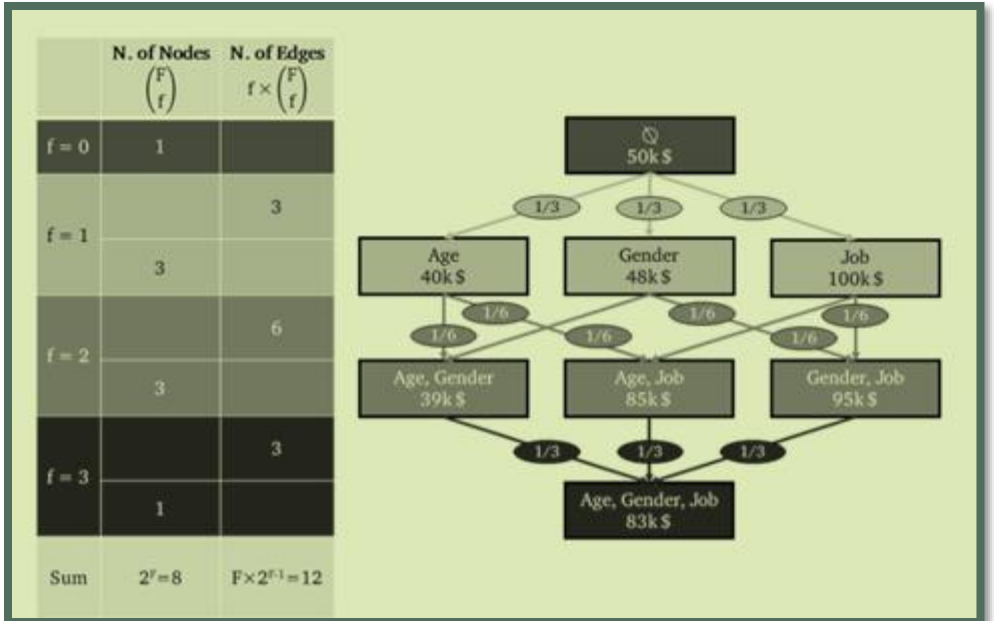


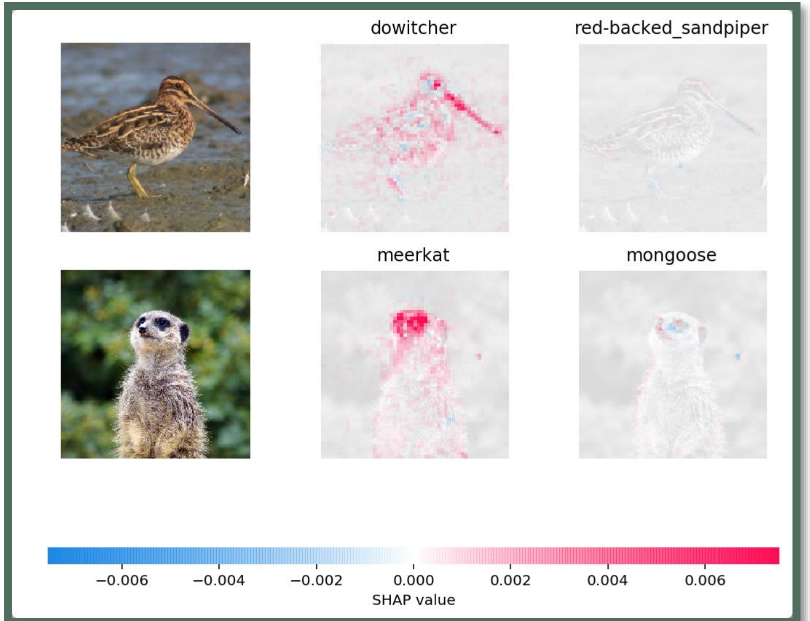
Obtaining the EVT-based anomaly threshold value by applying re-transformation

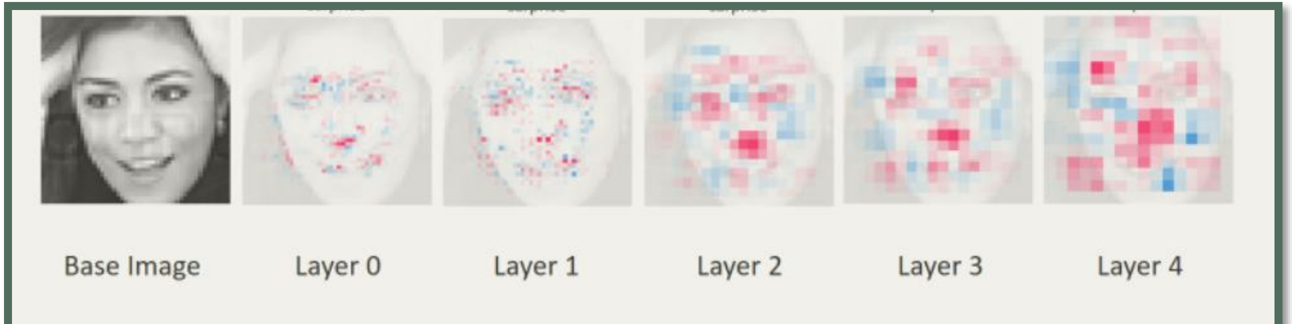
Threshold value



Explainable AI Module - (SHAP)







For Images super pixels and pixels can be used as features for which the contribution towards the model's prediction is calculated.

Definition of an Anomaly

In our context, we define anomaly as an observation that is very unlikely given the forecast distribution.