

Anomaly Detection in Image Streams with Explainable AI

By Team Bits of Erised



Group Members



Rashmi Perera



Nethmee Sellahewa



Nethmi Wijesinghe

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

Overview

01. Motivation

02. Significance of the Study

03. Novelty of Our System

04. Aim and Objectives

05. Methodology

06. Evaluation and Results

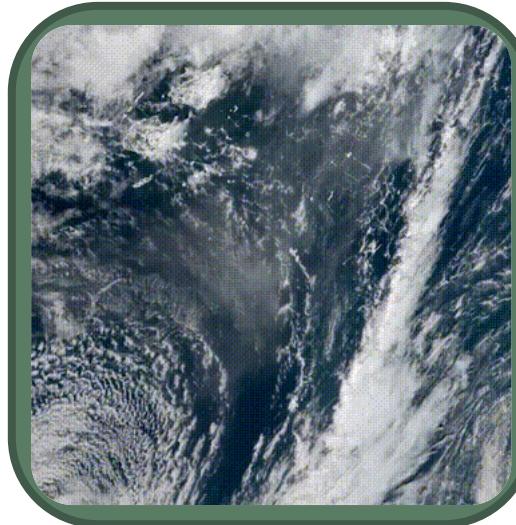
07. Further Work

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

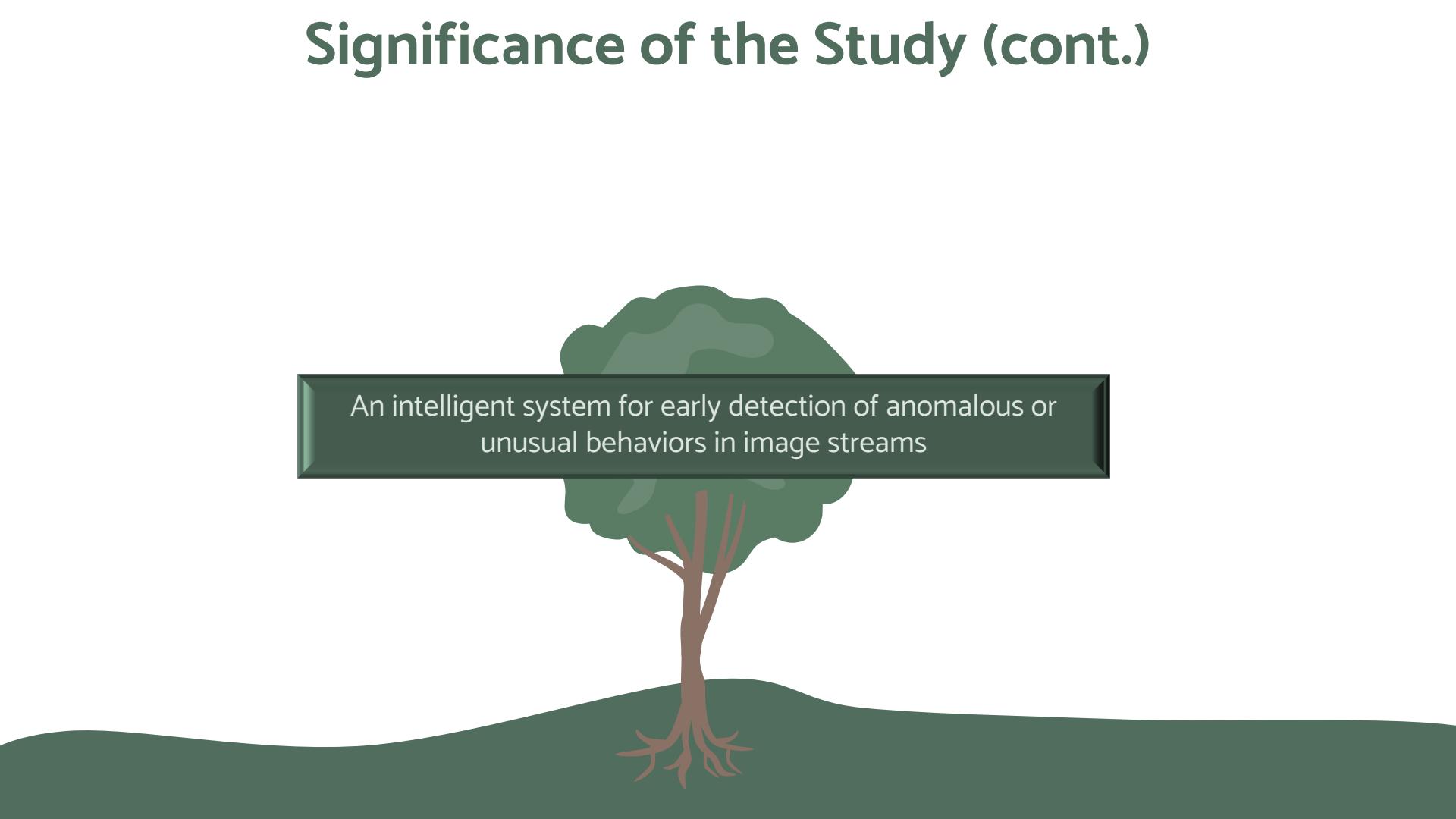
Damage of vegetation communities,
loss of plant and animal habitats
ascribe to **bushfires**.

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
<p>1. Treat anomaly detection problem as a supervised learning problem.</p> <p>Issue: Generalizability is low.</p>	<ul style="list-style-type: none">Proposed a novel anomaly detection framework.In the proposed framework, we treated anomaly detection problem as a one class classification.
<p>2. Most of the existing methods ignore the interdependency between the images (time dependency).</p>	<ul style="list-style-type: none">Integrated computer vision and time series forecasting to capture the interdependency between the images during the model building process.
<p>3. Most of the existing anomaly detection methods use manual thresholds and unrealistic assumptions.</p>	<ul style="list-style-type: none">Novel approach to calculate a data driven anomalous threshold using EVT.

Novelty of Our System (cont.)

Research Gap	Novelty/Main Contribution
<p>4. Existing deep learning-based anomaly detection methods solely focus on the classification task. This leads to lack of explainability.</p> <p>Issue: Reduces the human trust on the system.</p>	<ul style="list-style-type: none">Proposed a novel framework that integrated computer vision, time series forecasting and explainable AI to detect anomalies in image streams.
<p>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.</p> <ul style="list-style-type: none">- Up sampling- Down sampling- Data augmentation	<ul style="list-style-type: none">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



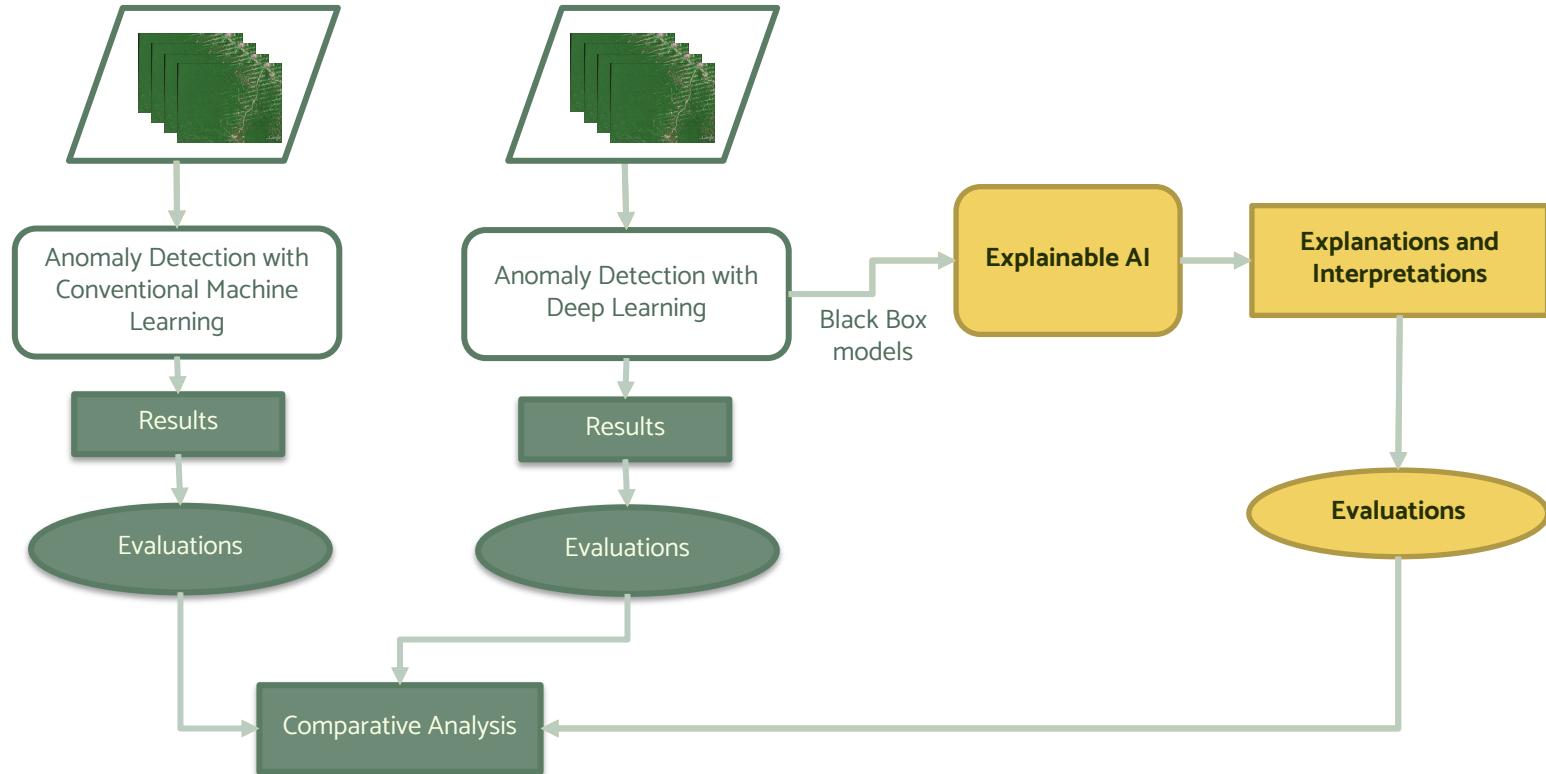
Definition of an Anomaly

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

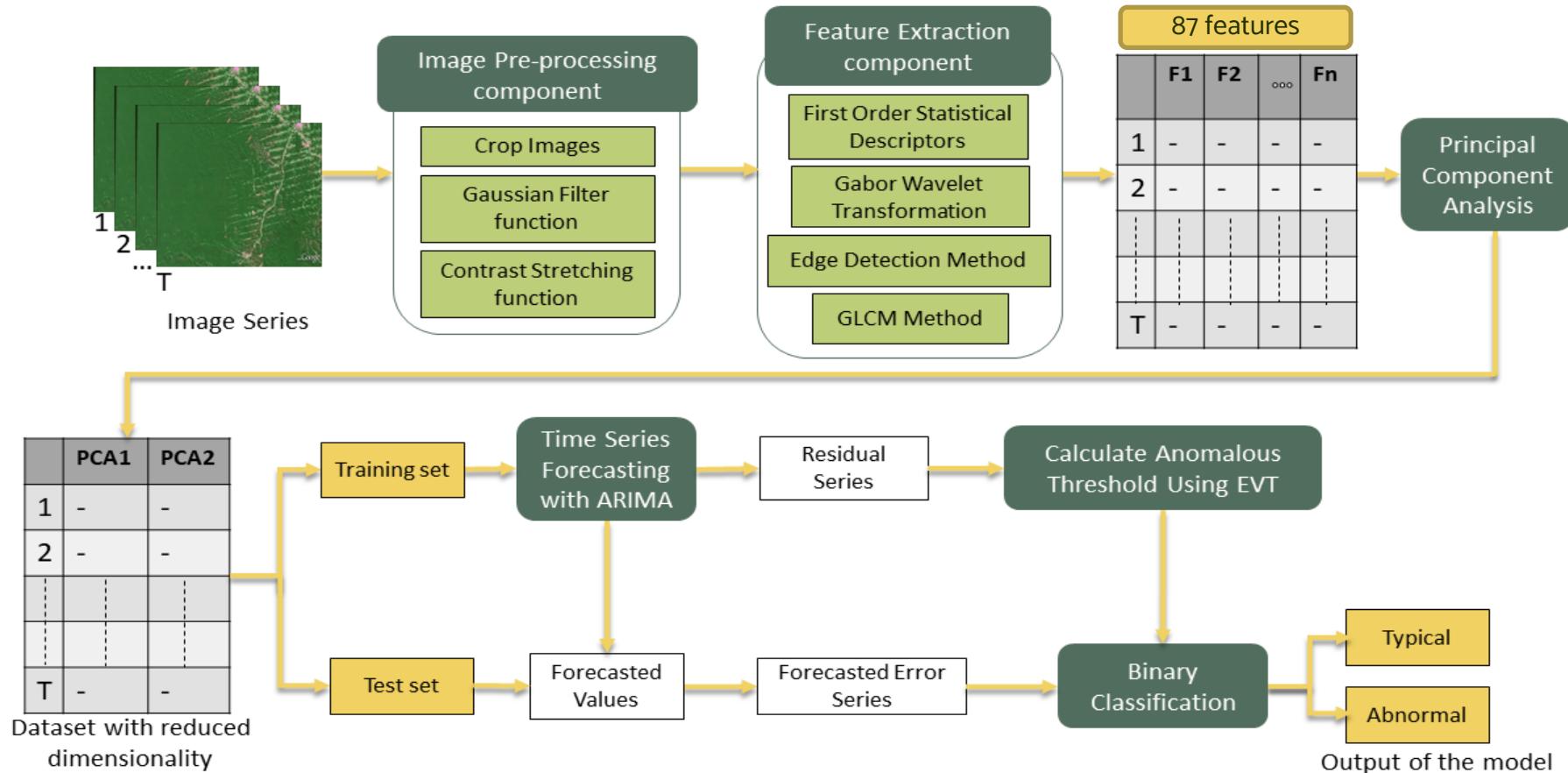
Main Assumptions

- 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.

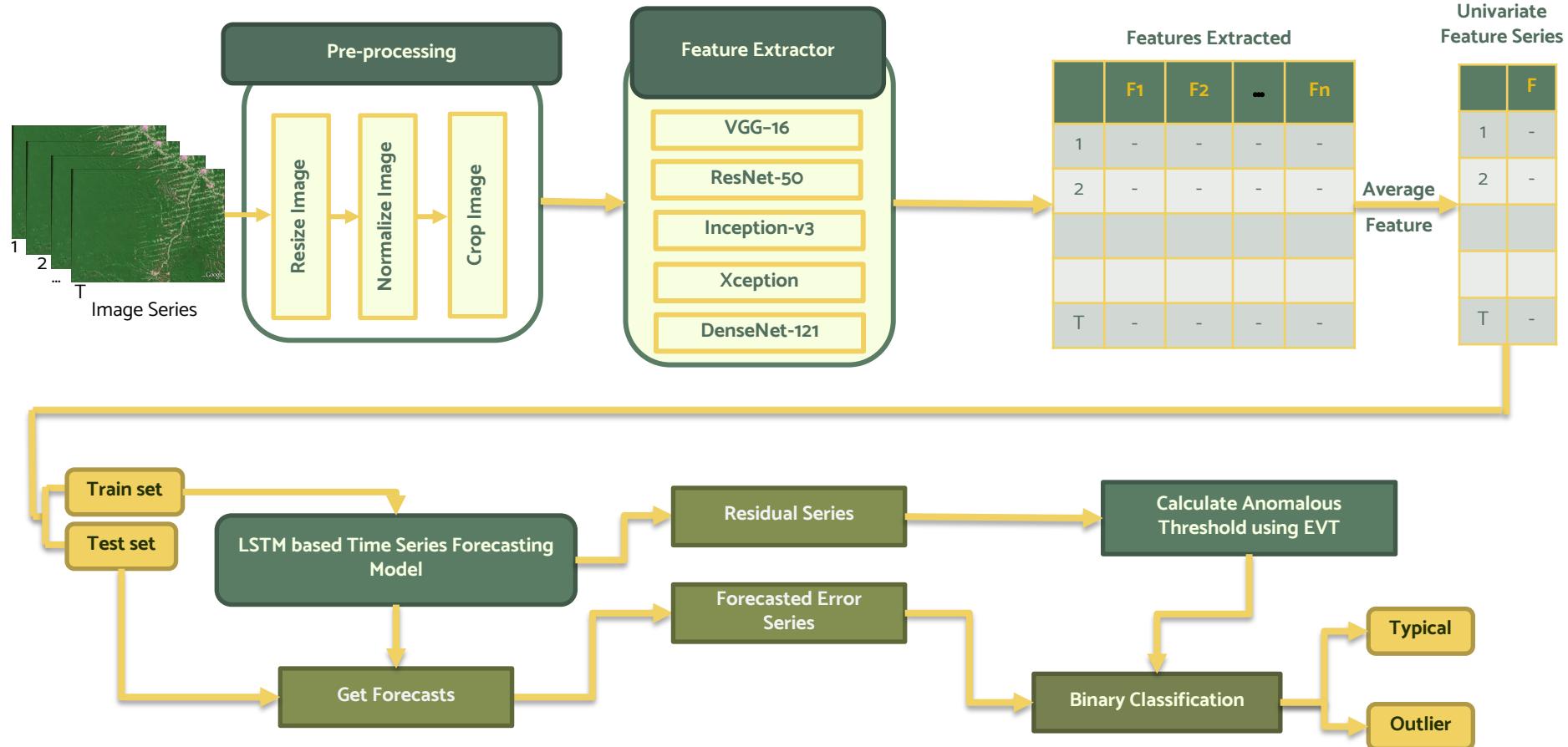
Methodology



Traditional Machine Learning Module



Deep Learning Module



EVT based Anomaly Threshold Calculation

Limitations of EDA based Threshold

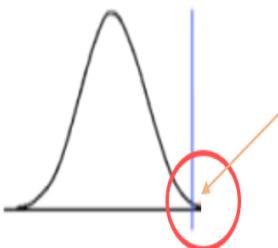
- Assumes the parent distribution conforms to a Gaussian distribution.

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.

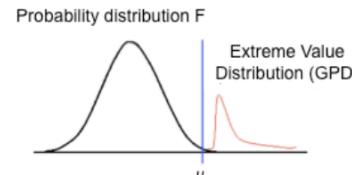
- Frechet
- Weibull
- Gumbell

Theoretical Foundation for EVT

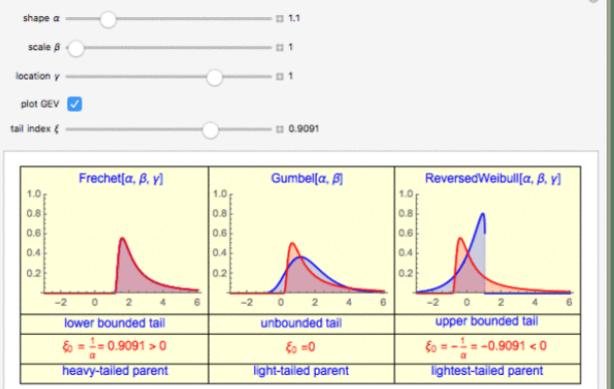


Tail of the distribution where the anomalies are concentrated.

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

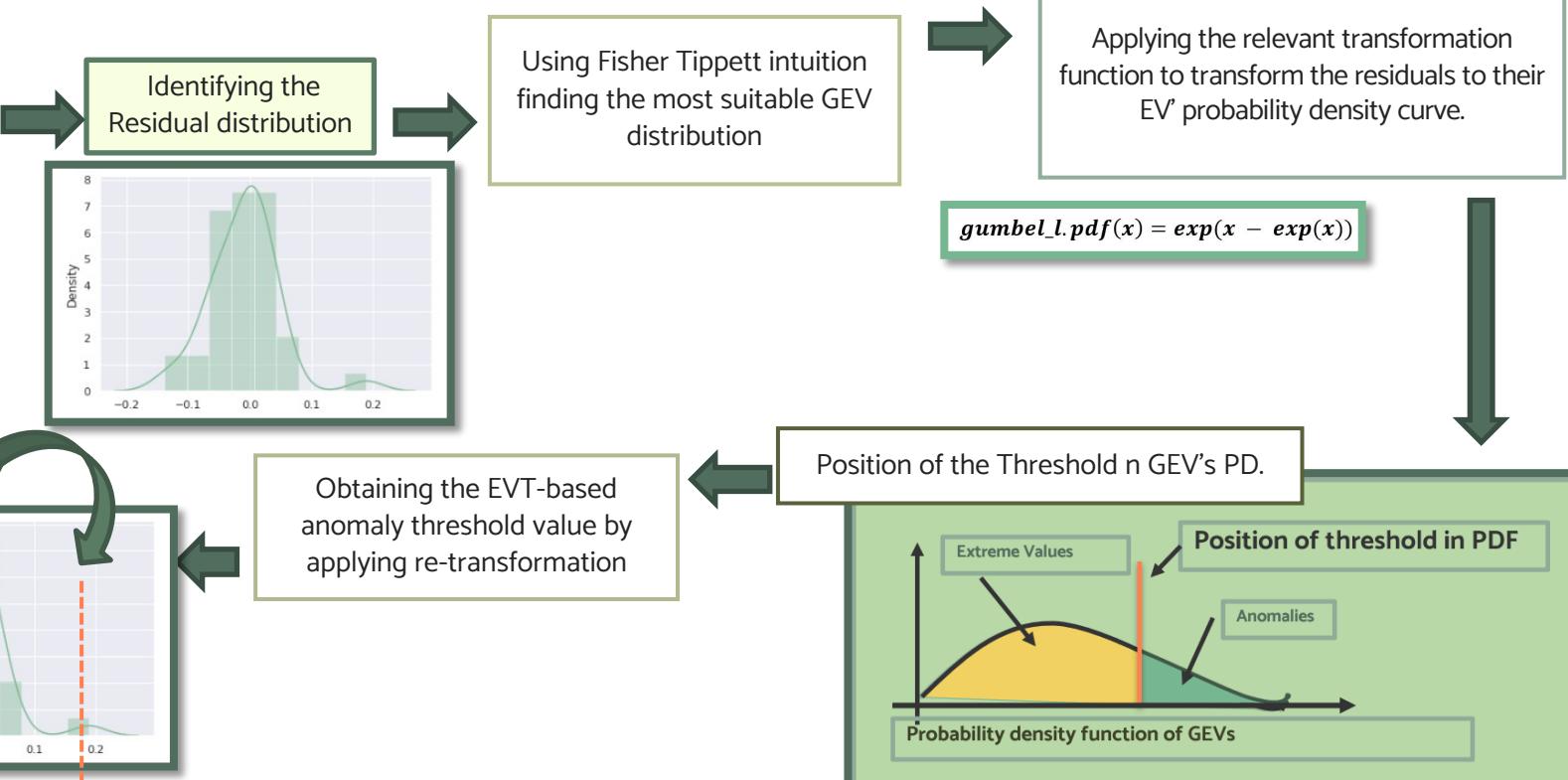


A child curve should be plotted to obtain the accurate anomaly threshold. Value

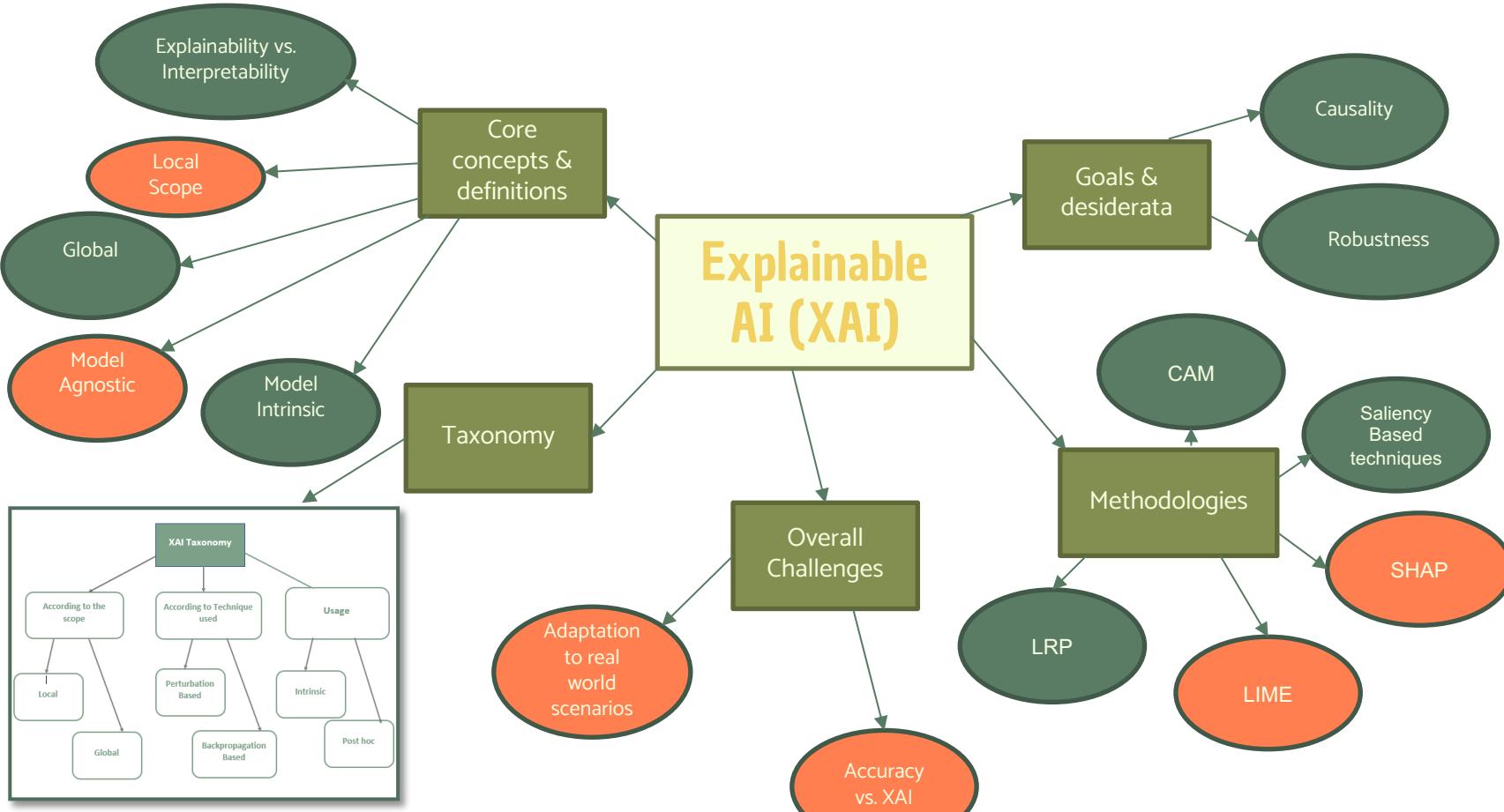


EVT based Anomaly Threshold Calculation

No	Residual Values



Explainable AI Module



Explainable AI Module - (LIME)

Locally Interpretable Model Agnostic Explanations (LIME)

How LIME works with Images?

$$\varepsilon(x) = \text{argmin}$$

Unfaithfulness

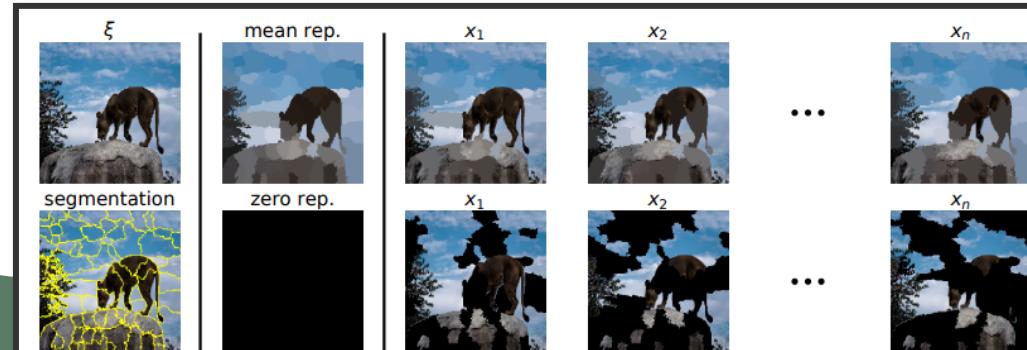
Complexity

Decomposes the target image into homogeneous patches

Create number of images using perturbation by turning on and off the super pixels.

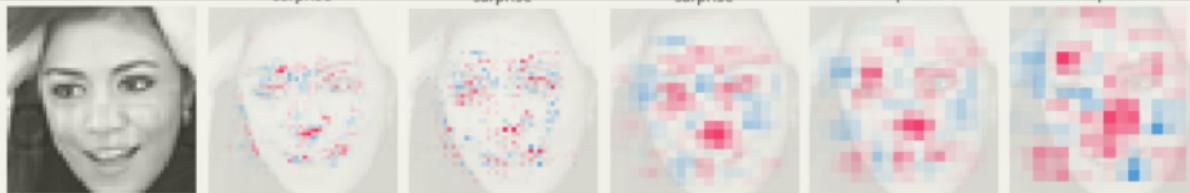
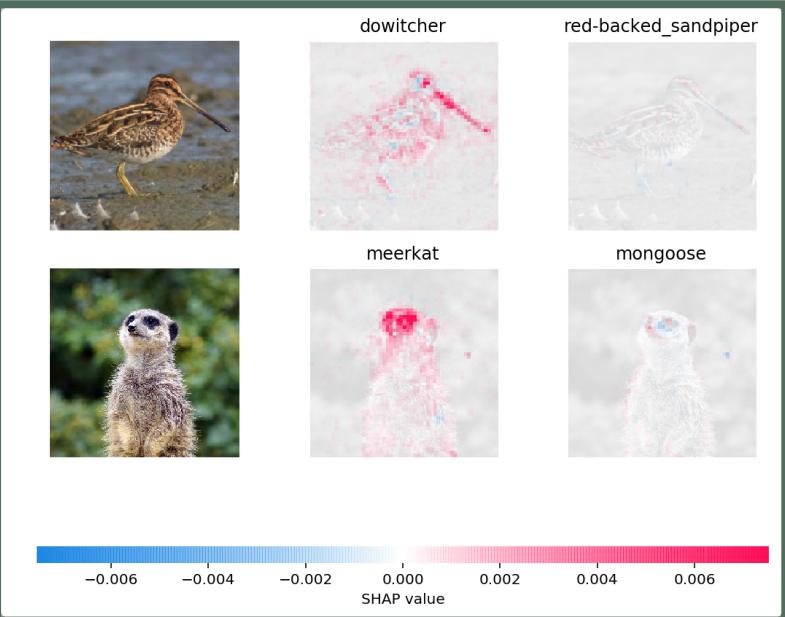
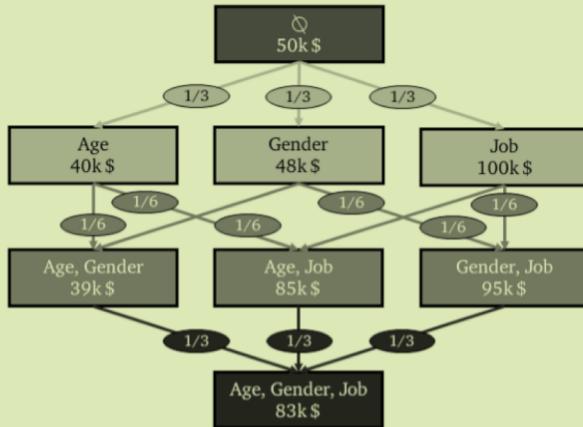
Get predictions of the new image set.

Build the local surrogate model with feature contributions.



Explainable AI Module(SHAP)

	N. of Nodes $\binom{F}{f}$	N. of Edges $f \times \binom{F}{f}$
$f = 0$	1	
$f = 1$		3
$f = 2$	3	
$f = 3$		3
Sum	$2^F=8$	$F \times 2^{F-1}=12$



Base Image

Layer 0

Layer 1

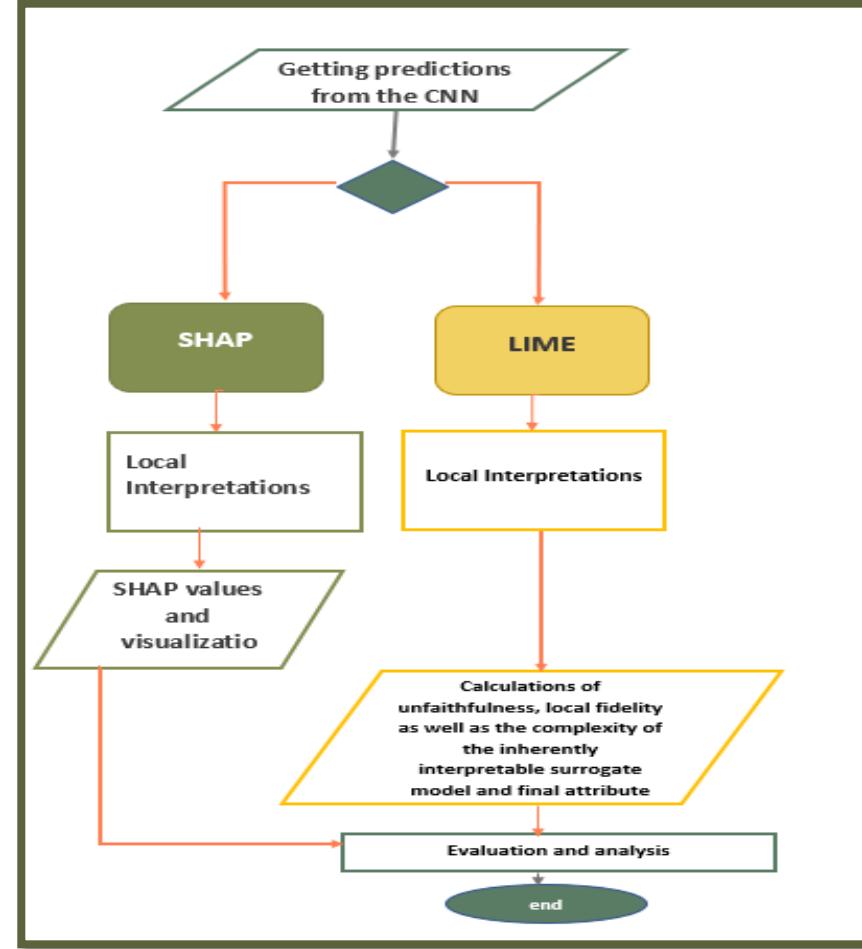
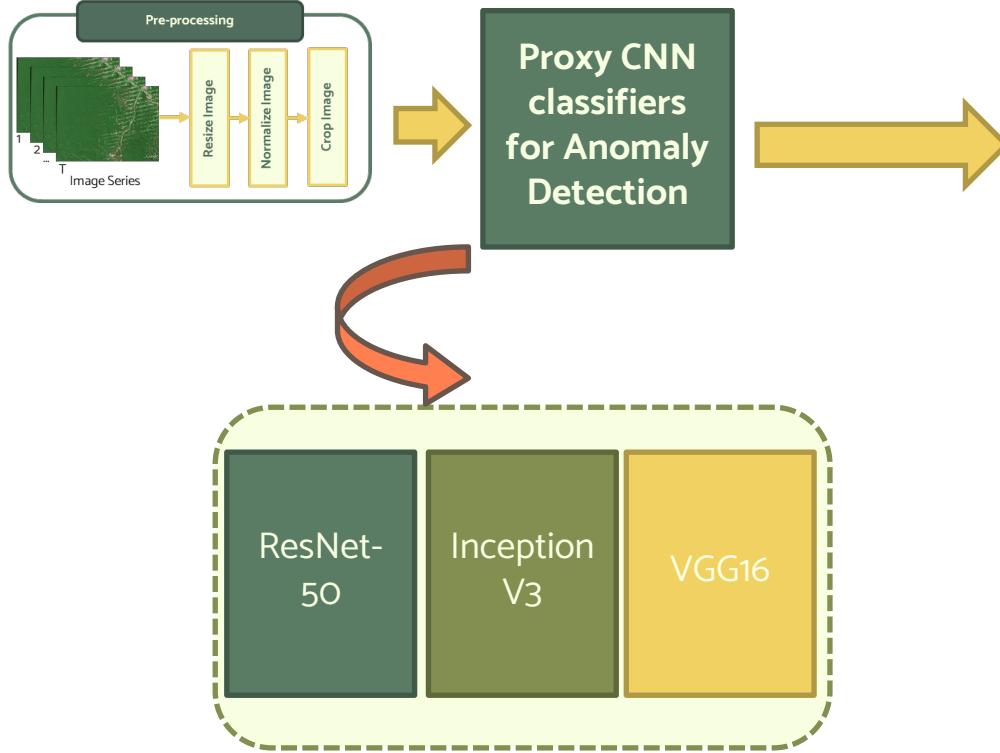
Layer 2

Layer 3

Layer 4

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

Explainable AI Module



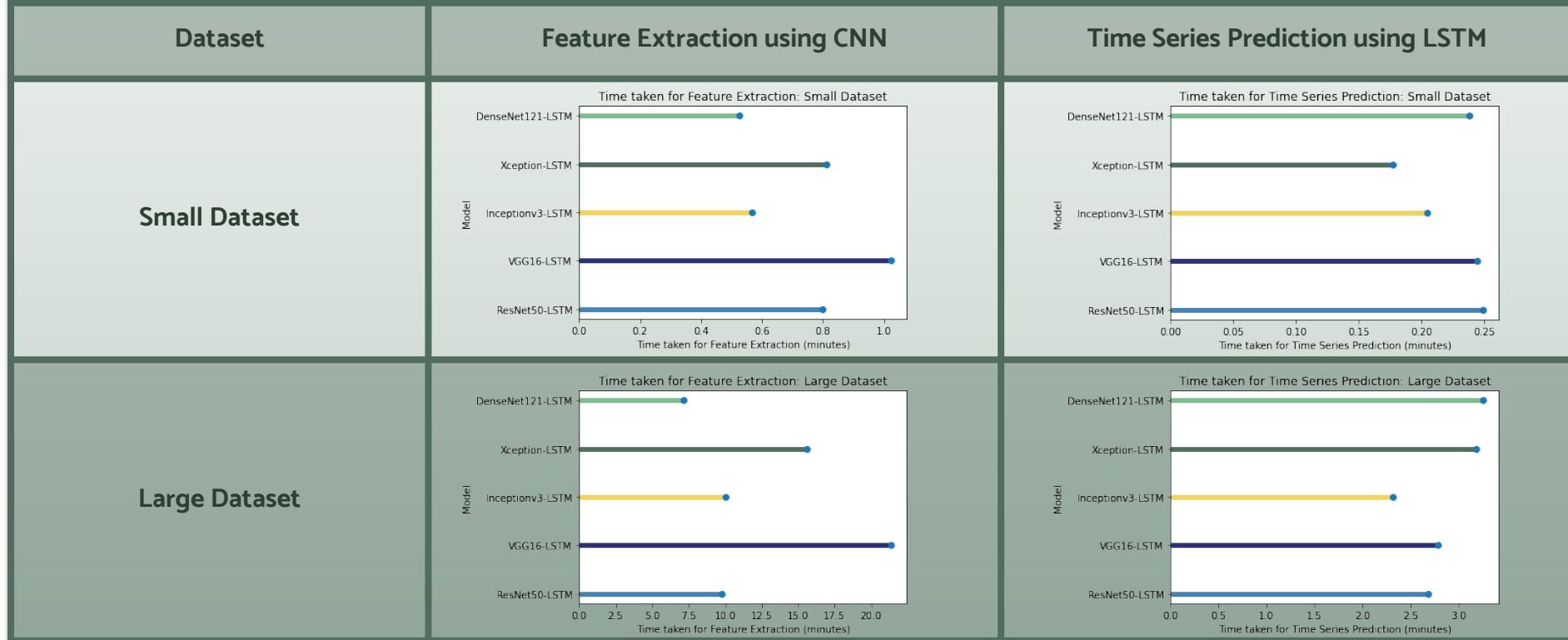
Evaluation



Conventional Machine Learning vs Deep Learning Techniques Comparison												
			Deep Learning Techniques									
	Small Dataset	Large Dataset	Small Dataset				Large Dataset					
	Small Dataset	Large Dataset	InceptionV3	ResNet50	VGG16	Xception	DenseNet121	InceptionV3	ResNet50	VGG16	Xception	DenseNet121
Box Plot Threshold Calculation Technique												
Accuracy	92.98%	80.8%	80%	93.33%	90%	90%	11.67%	95.05%	89.04%	75.67%	70.05%	67,65%
Sensitivity	1.0	1.0	0.77	0.92	0.94	1.0	0.11.	1.0	1.0	1.0	0.96	1.0.
Specificity	0.05	0.016	1.0	1.0	1.0	0.25	0.125	0.85	0.67	0.27	0.18	0.028
Positive Predictive Value	0.92	0.81	1.0	1.0	1.0	0.896	0.46	0.93	0.86	0.73	0.70	0.67
Negative Predictive Value	1.0	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	94.74%	84%	86.7%	92.9%	90%	88.33%	86.67%	99.6%	92.9%	76.2%	70.05%	79.55%
Sensitivity	1.0	1.0	1.0	0.92	0.94	1.0	1.0	1.0	1.0	1.0	0.96	1.0
Specificity	0.63	0.18	0.0	1.0	1.0	0.125	0.0	0.987	0.79	0.29	0.19	0.39
Positive Predictive Value	0.94	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	1.0	1.0	nan	0.67	0.73	1.0	nan	1.0	1.0	1.0	10.69	1.0

Time Comparison

Time Comparison of the CNN based LSTM Model

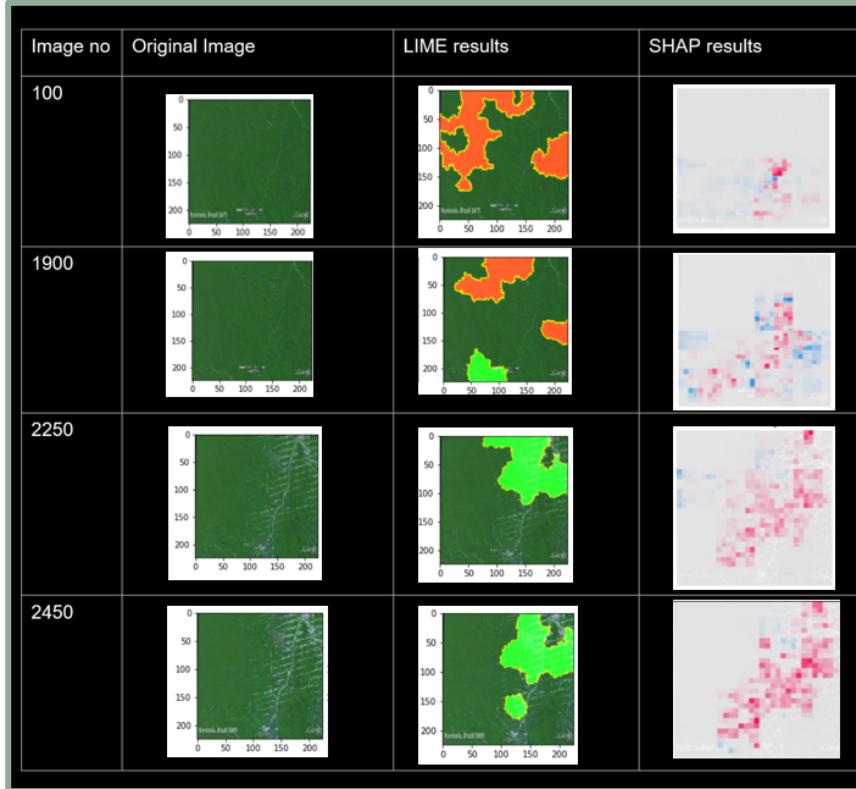


Generalizability

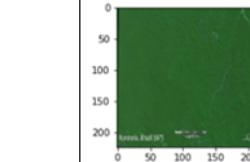
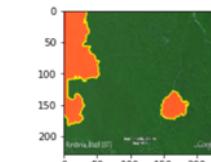
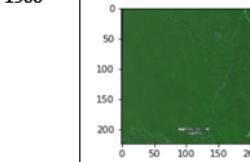
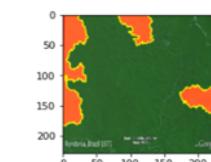
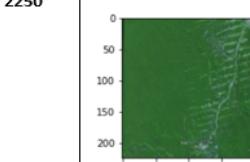
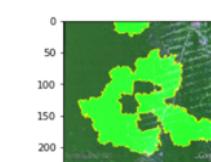
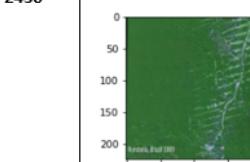
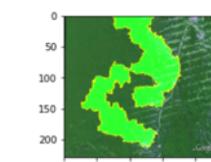
Performance Metrics for Different Datasets

Model	Traditional Machine Learning Model				Deep Learning Model			
	Small Dataset		Large Dataset		Small Dataset		Large Dataset	
Dataset	Deforestation	Volcano Eruption	Deforestation	Volcano Eruption	Deforestation	Volcano Eruption	Deforestation	Volcano Eruption
	96.7%	96.7%	86.3%	91.2%	95%	64.29%	99.6%	94.79%
Accuracy	0.87	0.96	0.47	0.67	0.92	0.47	0.989	0.91
Sensitivity (Sn)	0.98	0.95	1.0	1.0	0.94	1.0	1.0	0.92
Specificity (Sp)	0.9	1.0	0.59	0.74	1.0	0.47	0.987	1.0
Positive Predictive Value (PPV)	0.98	1.0	0.83	0.88	1.0	0.47	0.99	1.0
Negative Predictive Value (NPV)	0.9	0.91	1.0	1.0	0.73	1.0	1.0	0.86

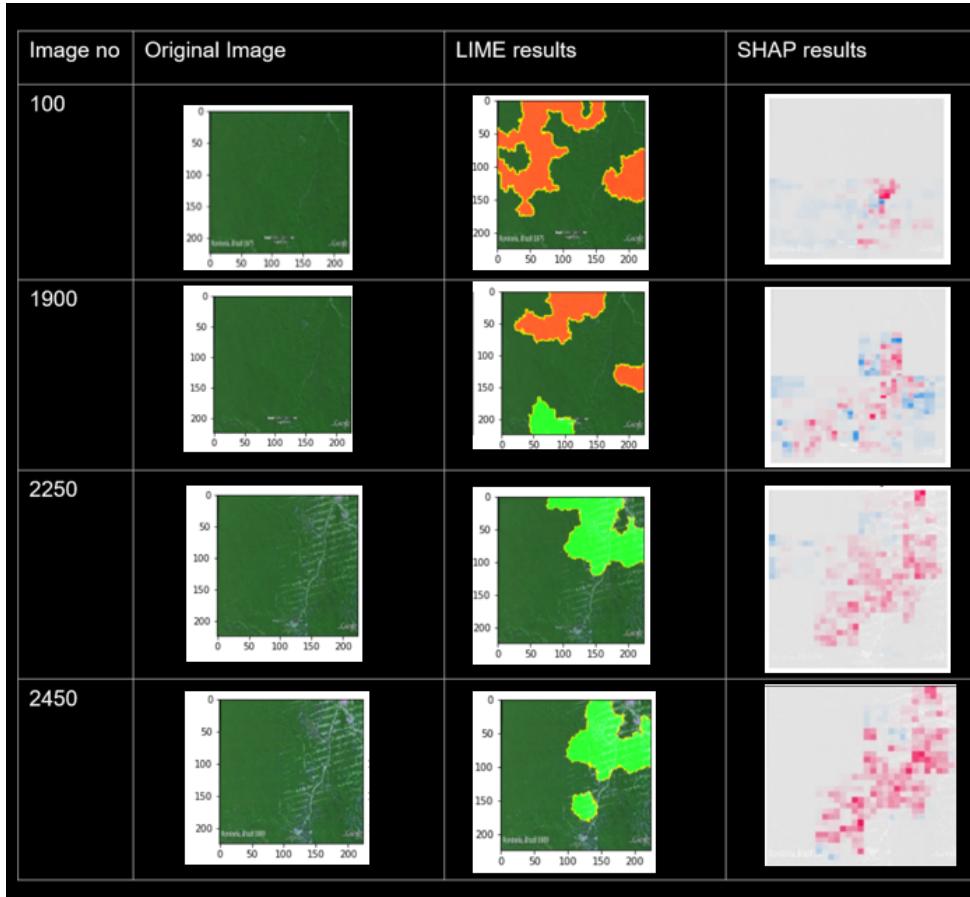
Explainable AI Module - ResNet50



Explainable AI Module - InceptionV3

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

Explainable AI Module - 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, 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: prital.netlify.app

Acknowledgement

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

Q&A

