INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

The 2024 IEEE Space Computing Conference (IEEE SMC-IT/SCC)

DANIEL LAKEY & TIM SCHLIPPE

ANOMALY DETECTION IN SPACECRAFT TELEMETRY: FORECASTING VS. CLASSIFICATION

Sunnyvale, USA

July 18, 2024

OUTLINE

MOTIVATION : ANOMALY IN SPACECRAFT TELEMETRY

- **Spacecraft telemetry:**
  - process of collecting and transmitting data from a spacecraft back to Earth
  - crucial for mission control to monitor the spacecraft's condition, diagnose potential issues, and ensure the success of the mission

MOTIVATION : ANOMALY IN SPACECRAFT TELEMETRY

- **Spacecraft telemetry**:
  - process of collecting and transmitting data from a spacecraft back to Earth
  - crucial for mission control to monitor the spacecraft's condition, diagnose potential issues, and ensure the success of the mission

- **Telemetry channels:**
  - a telemetry channel consists of one or more telemetry parameters
  - Parameters contain information about the spacecraft system, e.g. temperature, voltage, pressure, or system status

MOTIVATION : ANOMALY IN SPACECRAFT TELEMETRY

- **Anomaly in spacecraft telemetry:**
  - unexpected or irregular data or behavior observed in the telemetry signals
  - E.g. deviations in sensor readings, unusual patterns in data transmission, or unexpected changes in the spacecraft's operational parameters

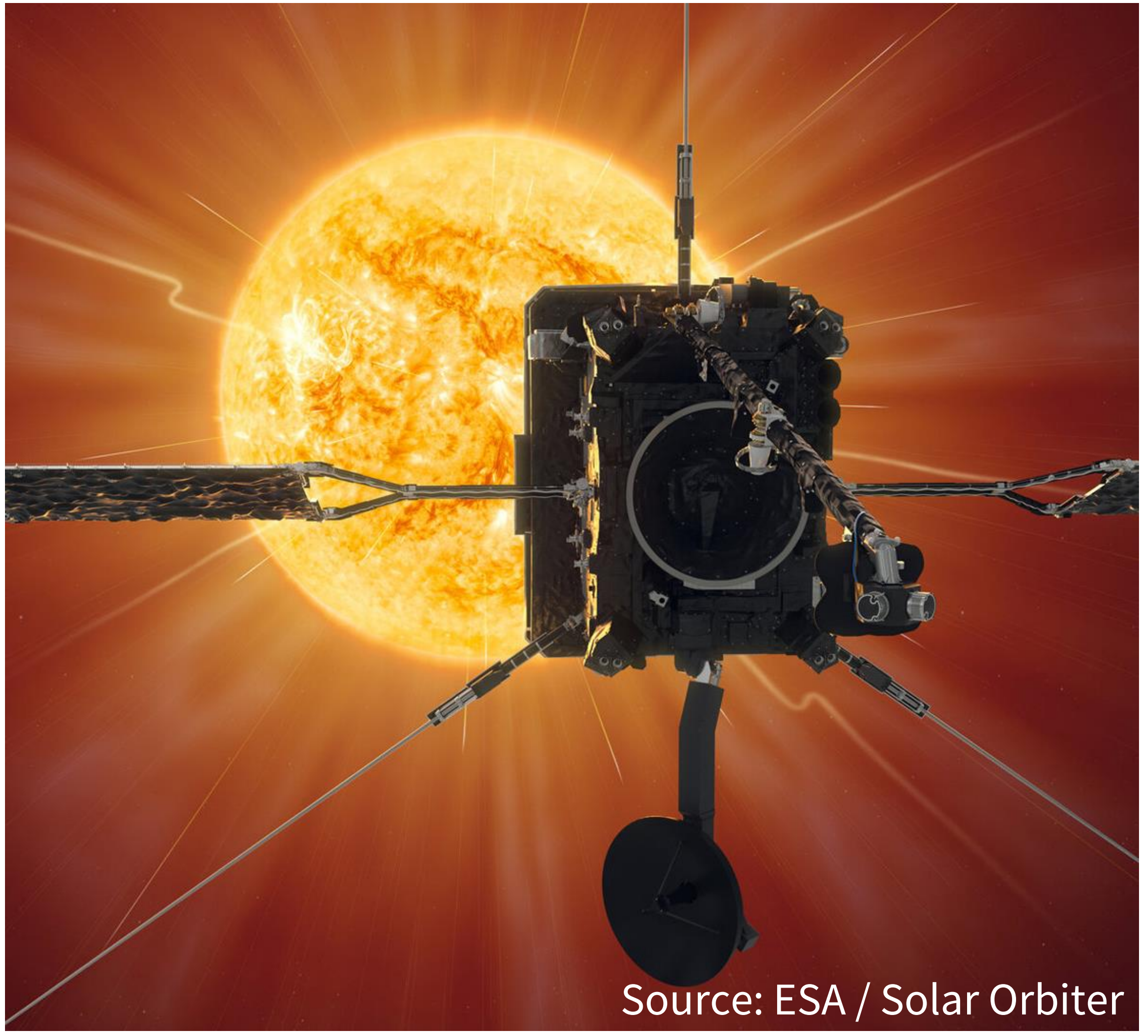
MOTIVATION : ANOMALY IN SPACECRAFT TELEMETRY

- **Anomaly in spacecraft telemetry:**
  - unexpected or irregular data or behavior observed in the telemetry signals
  - E.g. deviations in sensor readings, unusual patterns in data transmission, or unexpected changes in the spacecraft's operational parameters

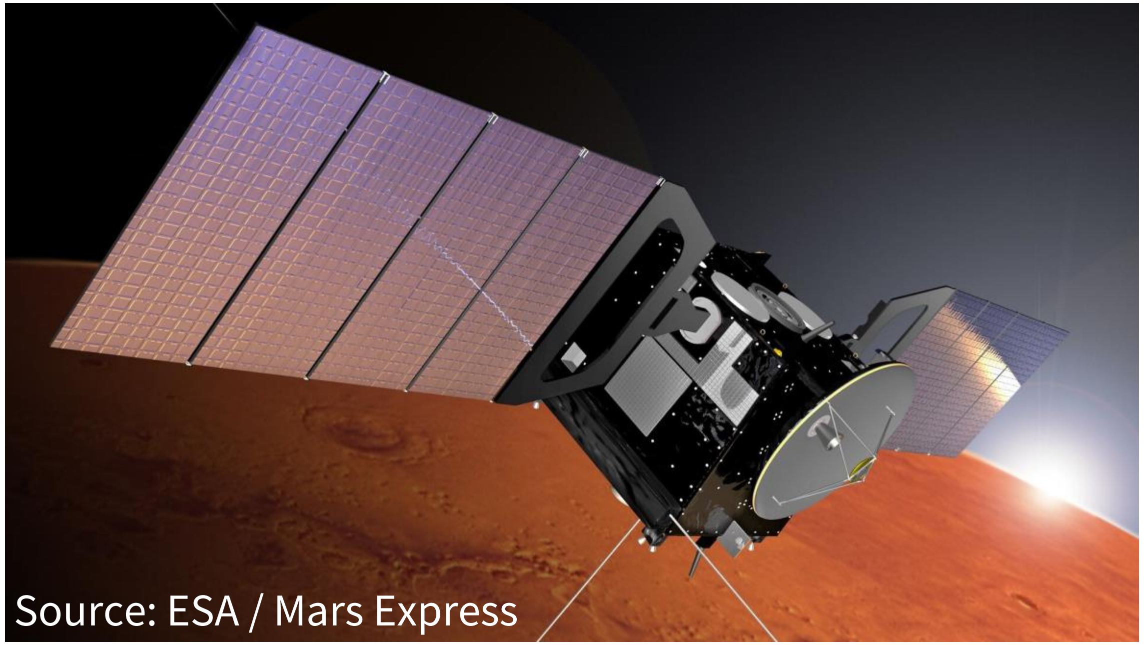
- **Anomaly detection in spacecraft telemetry:**

  ➔ Identifying and addressing these anomalies is crucial to ensure the spacecraft's safety and mission success

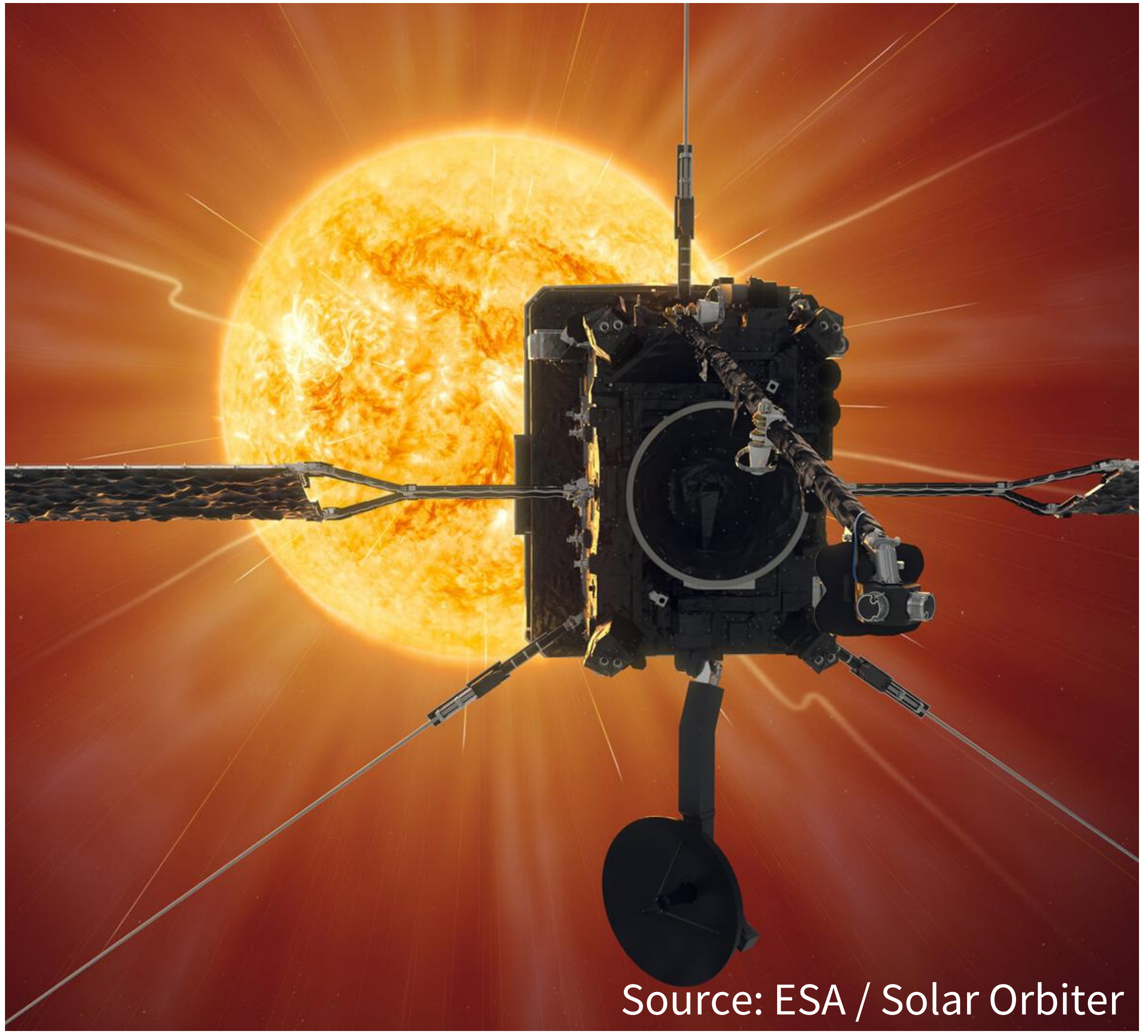
MOTIVATION : SPACECRAFT ARE COMPLICATED



- Spacecraft are complicated
- Monitoring by expert engineers

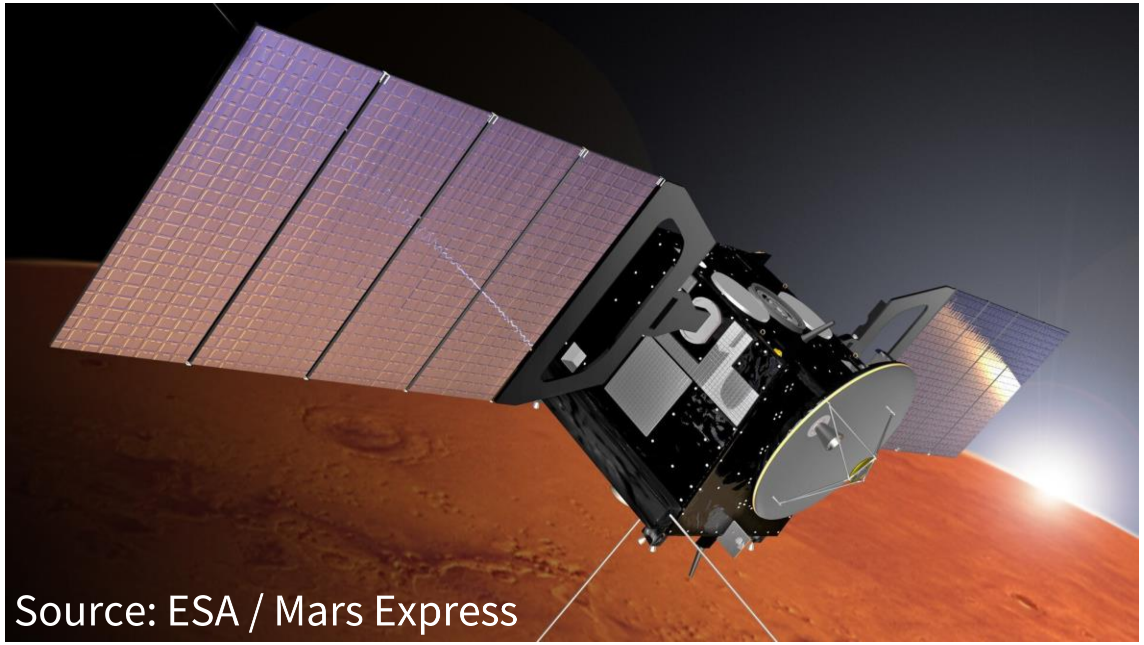


MOTIVATION : SPACECRAFT ARE COMPLICATED

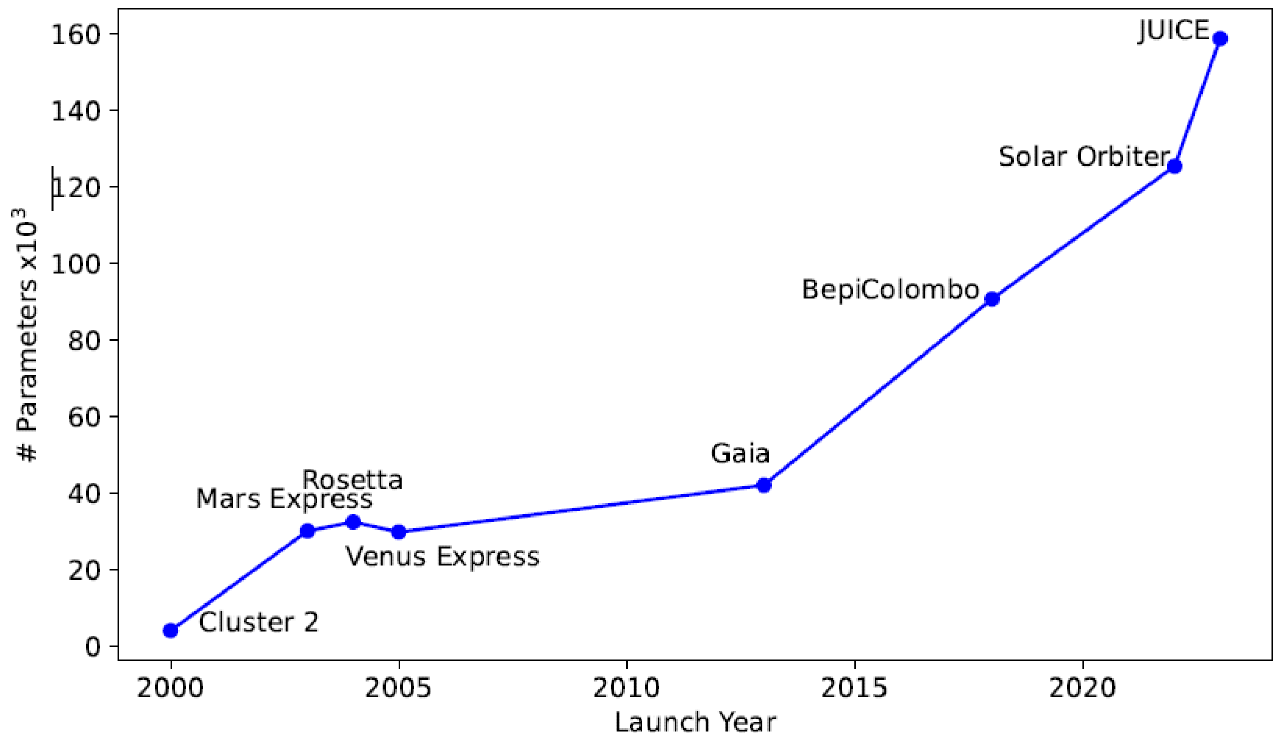


- Spacecraft are complicated
- Monitoring by expert engineers

➔ Pressure to “use AI” in operations

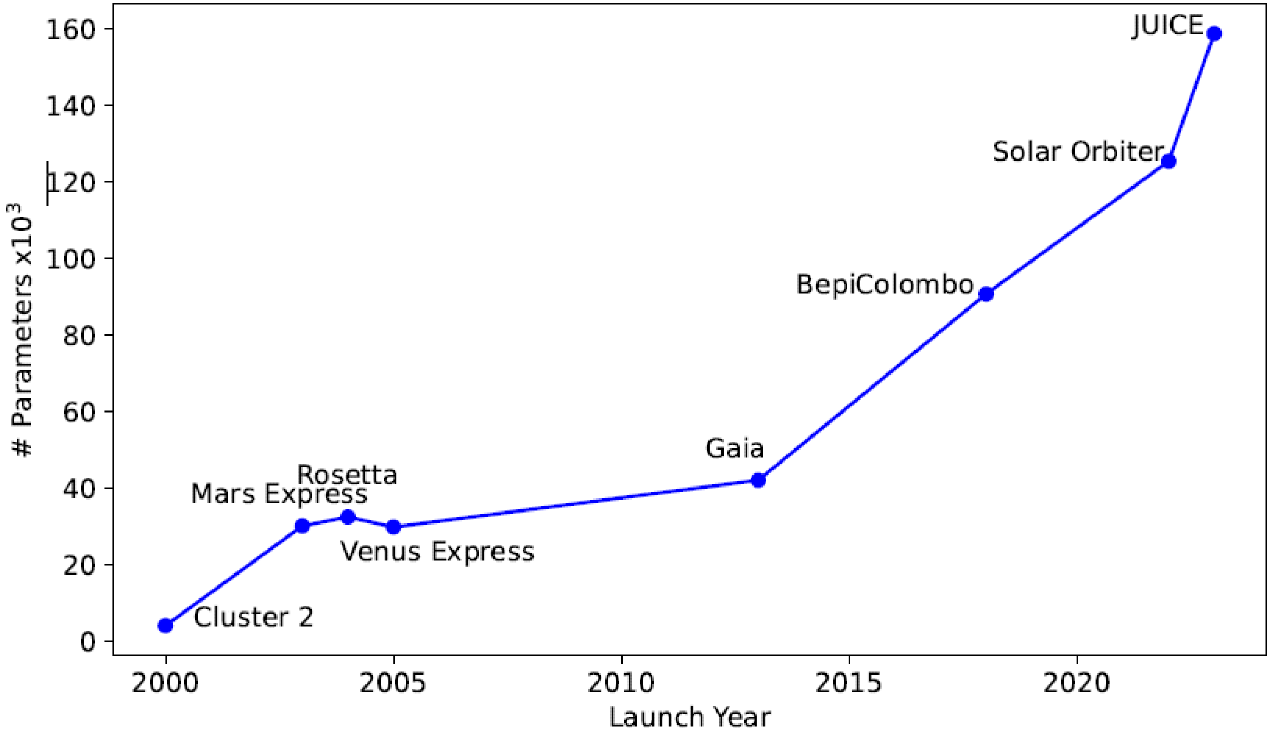


MOTIVATION: IT'S NOT GETTING EASIER



Increasing complexity of spacecraft over time

MOTIVATION: IT'S NOT GETTING EASIER

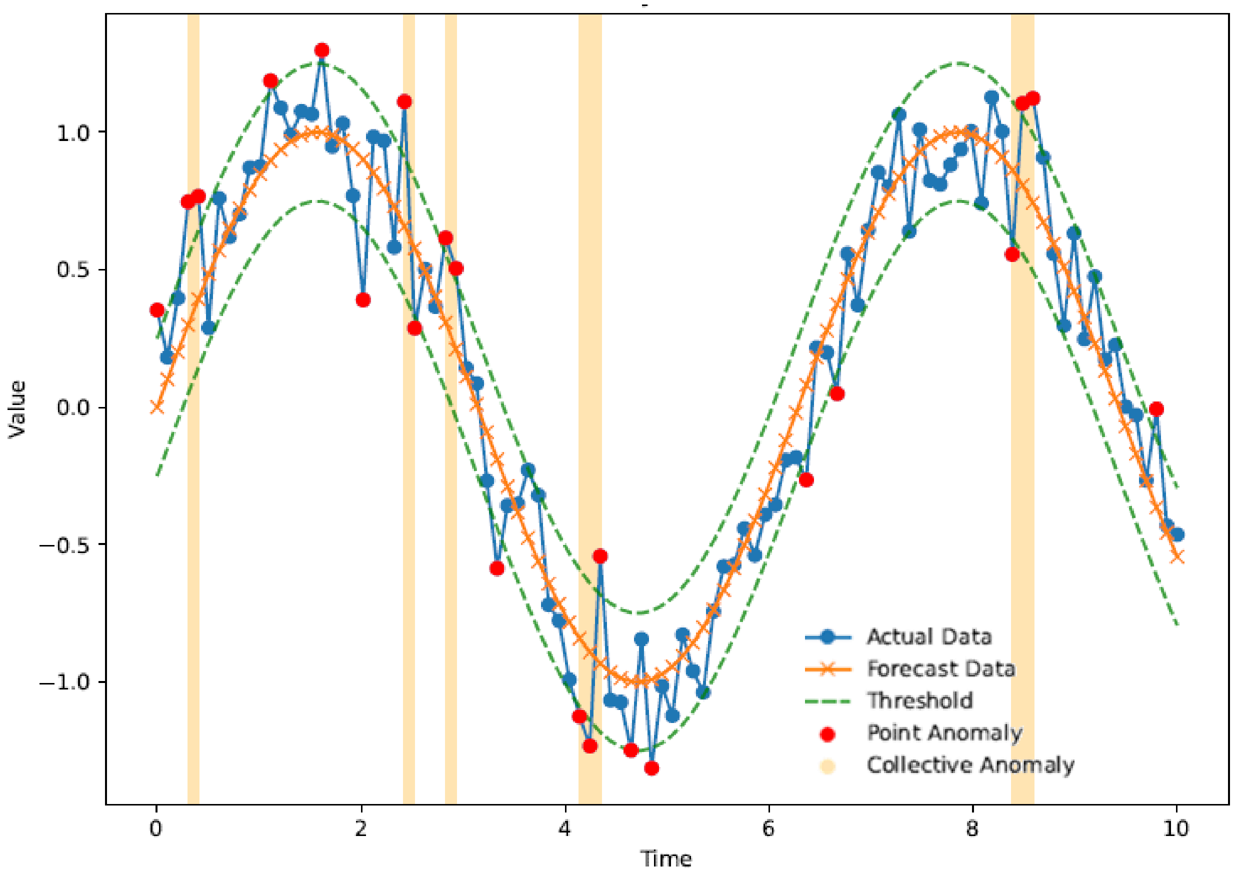


Increasing complexity of spacecraft over time

- **What anomaly detection approaches are out there?**
- **What performance do they have?**
- **Can we do better?**

RELATED WORK: TRADITIONAL METHODS

– ***Forecasting & Threshold***:
1. Predicting future values based on the attitude of past values
2. Identifying anomalies when observed values deviate significantly from predictions

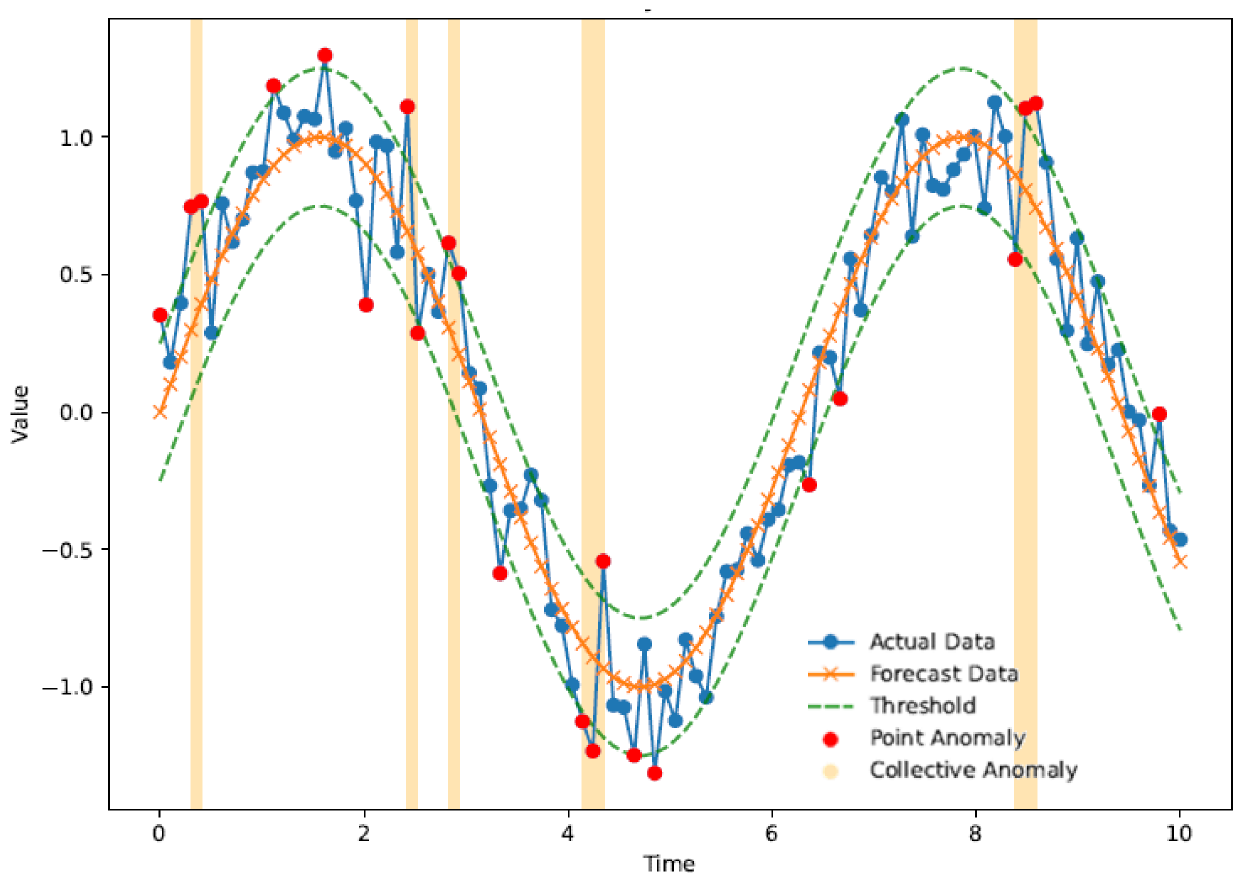


RELATED WORK: TRADITIONAL METHODS

— ***Forecasting & Threshold***:

1. Predicting future values based on the attitude of past values
2. Identifying anomalies when observed values deviate significantly from predictions.

— **Example**: *Forecasting* models like ARIMA & deep learning models like LSTM and CNN.
(Ihler et al., 2006; Hundman et al., 2018; Munir et al., 2019)

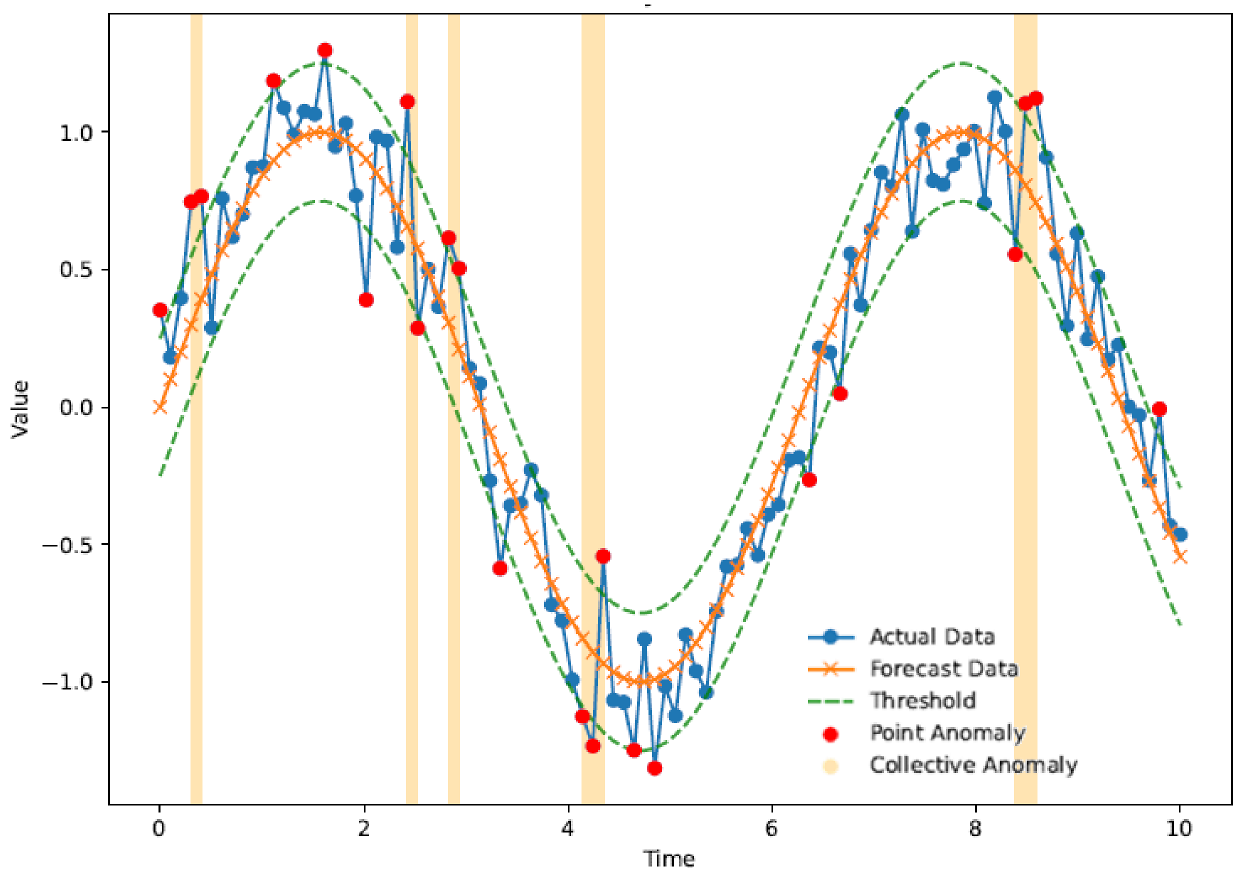


RELATED WORK: TRADITIONAL METHODS

- **Forecasting & Threshold**:
  1. Predicting future values based on the attitude of past values
  2. Identifying anomalies when observed values deviate significantly from predictions.

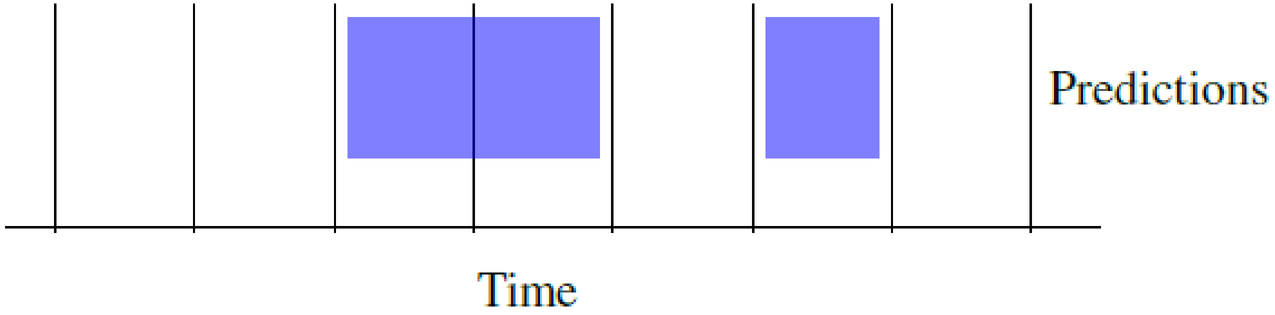
- **Example**: *Forecasting* models like ARIMA & deep learning models like LSTM and CNN.
  (Ihler et al., 2006; Hundman et al., 2018; Munir et al., 2019)

- **Limitations**:
  - ➔ requires manual threshold setting
  - ➔ forecasting inaccuracies can result in false alarms or missed anomalies



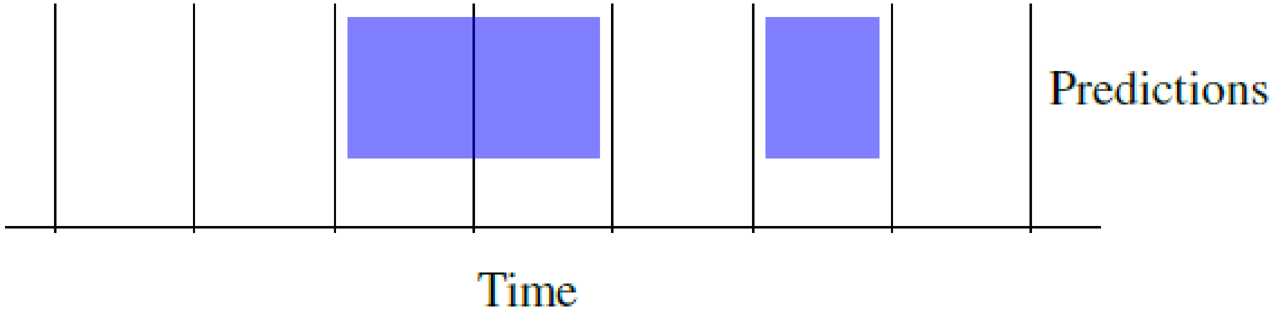
RELATED WORK: INNOVATIVE METHODS

- ***Direct Classification***:
  - Using classifiers to label each time point or window directly as *normal* or *anomalous*.
  - We selected time window-based approach to better capture temporal behavior vs. single time points



RELATED WORK: INNOVATIVE METHODS

- ***Direct Classification***:
  - Using classifiers to label each time point or window directly as *normal* or *anomalous*.
  - We selected time window-based approach to better capture temporal behavior vs. single time points

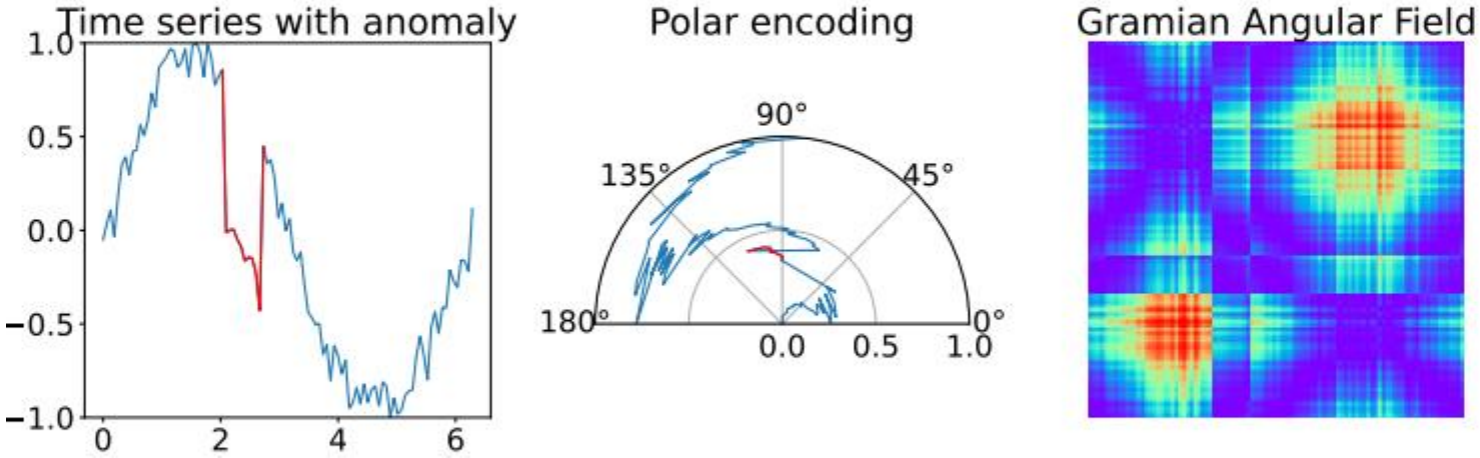


➔ CNNs and LSTMs have shown significant promise in handling time series data with classification
(Wang et al., 2016 ; Karim et al., 2018; Fawaz et al., 2019)

➔ But *direct classification* has not been investigated in the context of spacecraft telemetry anomalies

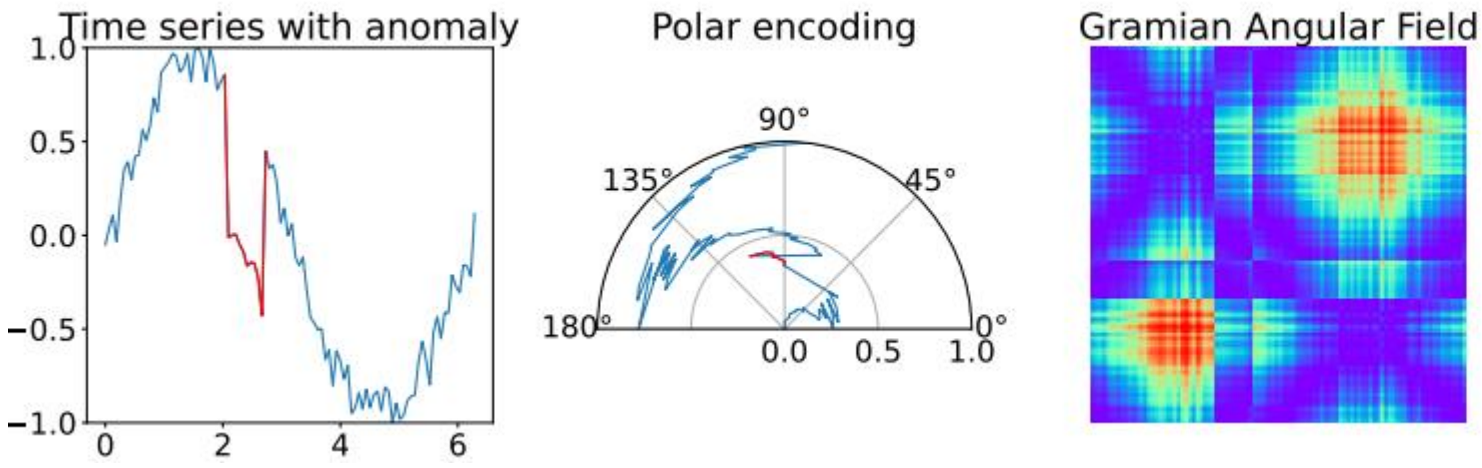
RELATED WORK: INNOVATIVE METHODS

– ***Image Classification***:
1. Transforming time series data into images, e.g. through Gramian Angular Field transforms
2. Applying *image classification* techniques



RELATED WORK: INNOVATIVE METHODS

– ***Image Classification***:
1. Transforming time series data into images, e.g. through Gramian Angular Field transforms
2. Applying *image classification* techniques



➔ Image transformation has shown significant promise in handling time series data with classification (Wang & Oates, 2015; Jiang et al., 2022; Vitry, 2023)

➔ But *image classification* has not been investigated in the context of spacecraft telemetry anomalies

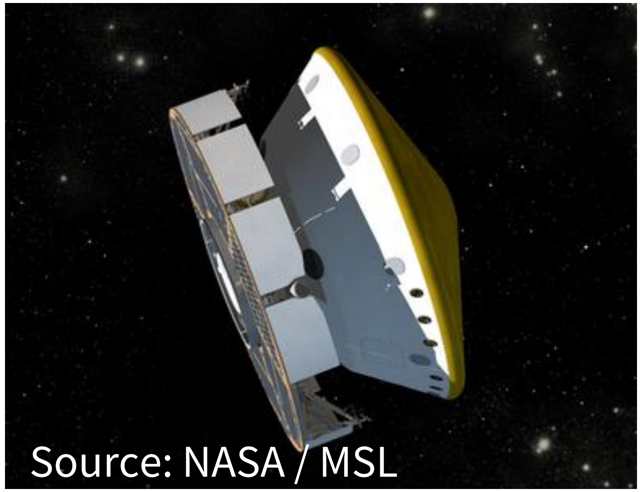
EXPERIMENTAL SETUP: DATA AND PREPROCESSING

- Public data sets of real spacecraft telemetry with well-labeled anomalies are hard to find

- **Dataset**:

  SMAP/MSL dataset (Hundman et al., 2018),
  - includes telemetry data from the Soil Moisture Active Passive (SMAP) spacecraft
  - and the Mars Science Laboratory (MSL) spacecraft


Source: NASA / SMAP


Source: NASA / MSL

EXPERIMENTAL SETUP: DATA AND PREPROCESSING

- Public data sets of real spacecraft telemetry with well-labeled anomalies are hard to find

- **Dataset**:
  SMAP/MSL dataset (Hundman et al., 2018),
  - includes telemetry data from the Soil Moisture Active Passive (SMAP) spacecraft
  - and the Mars Science Laboratory (MSL) spacecraft

- **Challenges**:
  - Data anonymization and imbalance, with very few anomaly cases.
  - 82 multivariate telemetry channels, with around 100 labelled anomalies in total across all channels

EXPERIMENTAL SETUP: DATA AND PREPROCESSING

- Public data sets of real spacecraft telemetry with well-labeled anomalies are hard to find

- **Dataset**:
  SMAP/MSL dataset (Hundman et al., 2018),
  - includes telemetry data from the Soil Moisture Active Passive (SMAP) spacecraft
  - and the Mars Science Laboratory (MSL) spacecraft

- **Challenges**:
  - Data anonymization and imbalance, with very few anomaly cases.
  - 82 multivariate telemetry channels, with around 100 labelled anomalies in total across all channels
  - ➔ 16 telemetry channels with sufficient anomalies (≥ 2) for classification, split each channel into test/train

EXPERIMENTAL SETUP: DATA AND PREPROCESSING

- Public data sets of real spacecraft telemetry with well-labeled anomalies are hard to find

- **Dataset**:
  SMAP/MSL dataset (Hundman et al., 2018),
  - includes telemetry data from the Soil Moisture Active Passive (SMAP) spacecraft
  - and the Mars Science Laboratory (MSL) spacecraft

- **Challenges**:
  - Data anonymization and imbalance, with very few anomaly cases.
  - 82 multivariate telemetry channels, with around 100 labelled anomalies in total across all channels

  ➔ 16 telemetry channels with sufficient anomalies (≥ 2) for classification, split each channel into test/train

- **Preparation**:
  - Data was pre-processed, including windowing for classification tasks

EXPERIMENTAL SETUP: DATA AND PREPROCESSING

- Public data sets of real spacecraft telemetry with well-labeled anomalies are hard to find

- **Dataset**:
  SMAP/MSL dataset (Hundman et al., 2018),
  - includes telemetry data from the Soil Moisture Active Passive (SMAP) spacecraft
  - and the Mars Science Laboratory (MSL) spacecraft

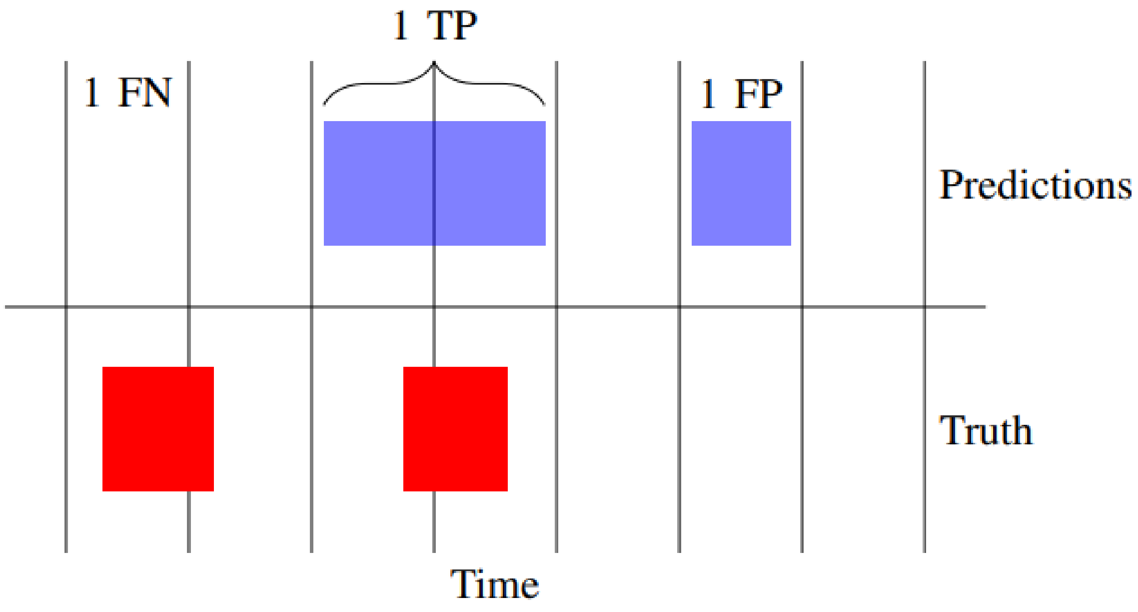
- **Challenges**:
  - Data anonymization and imbalance, with very few anomaly cases.
  - 82 multivariate telemetry channels, with around 100 labelled anomalies in total across all channels

  ➔ 16 telemetry channels with sufficient anomalies (≥ 2) for classification, split each channel into test/train

- **Preparation**:
  - Data was pre-processed, including windowing for classification tasks
  - Ø anomaly length: ~ 240 samples ➔ 64 samples / window

EXPERIMENTAL SETUP: EVALUATION METRICS

- **Key metrics**:
  True Positives (TP), False Positives (FP), False Negatives (FN), F1 Score



- **Per-Anomaly Scoring** (Hundman, 2018):
  Evaluates entire anomaly sequences rather than individual points, providing a more accurate assessment of model performance.

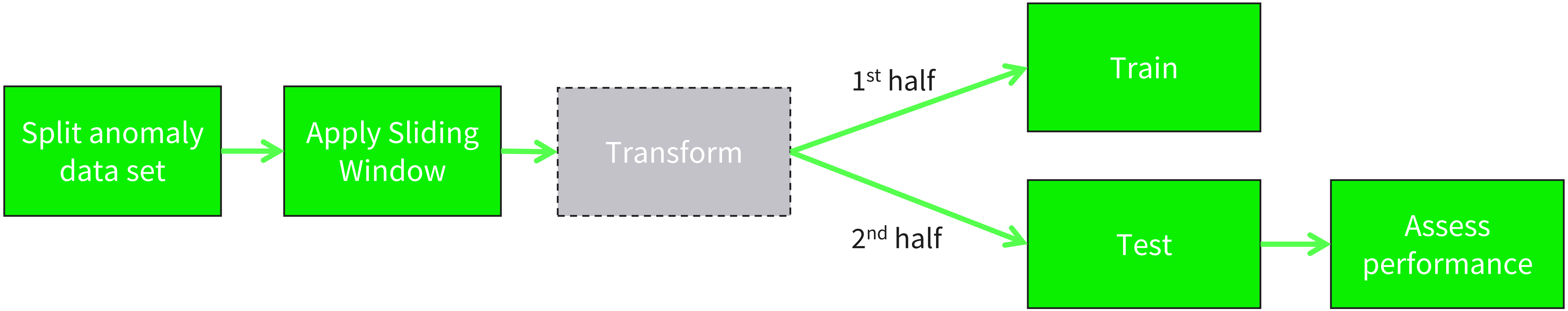
EXPERIMENTAL SETUP: PIPELINE

- **Setup**:
  - Google Colab: T4 GPU,13 deep learning architectures analyzed

EXPERIMENTAL SETUP: PIPELINE

- **Setup**:
  - Google Colab: T4 GPU,13 deep learning architectures analyzed

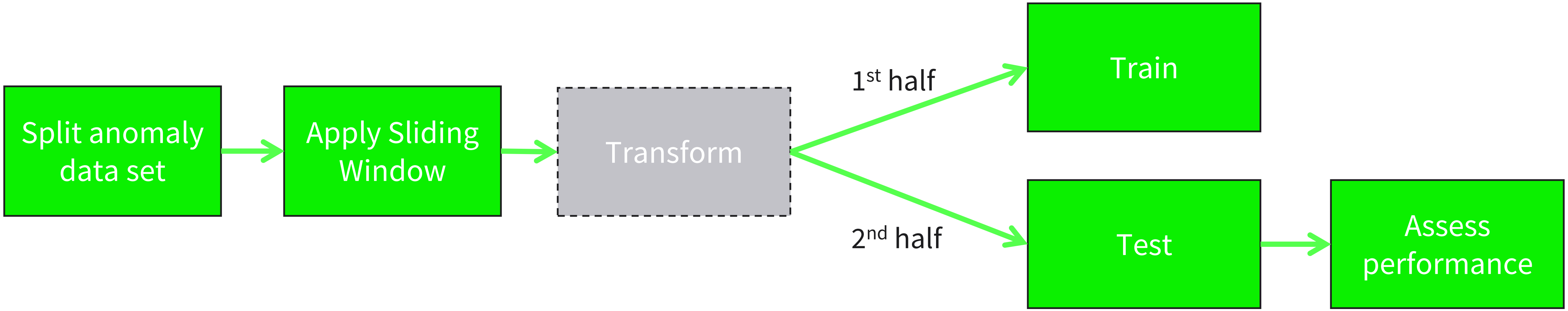
- **Pipeline for *direct / image classification***:
  for data preparation, transformation, model training, and evaluation



EXPERIMENTAL SETUP: PIPELINE

- **Setup**:
  - Google Colab: T4 GPU,13 deep learning architectures analyzed

- **Pipeline for *direct / image classification***:
  for data preparation, transformation, model training, and evaluation



➔ Best performance for *forecasting & threshold / direct classification:* CNN model ***XceptionTimePlus***

➔ Best performance for *image classification*: CNN model ***xResNet34***

RESULTS

- *Forecasting & threshold*
  - low number of **FP** (1)

	Forecasting&threshold				*Direct classification*				**Image classification**			
Spacecraft	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)
SMAP	6	**1**	4	**70.6**	**9**	11	1	60.0	5	52	5	14.9
MSL	0	**0**	6	0.0	6	13	0	48.0	6	16	**0**	**42.9**
Overall	6	**1**	10	52.8	**15**	24	**1**	**54.5**	11	68	5	26.3

Note: Test set contains in total 16 anomalies

RESULTS

- ***Forecasting & threshold***
  - low number of FP (1)
  - good **F1 score** (52.8%), particularly for SMAP (70.6%)

	Forecasting&threshold				*Direct classification*				**Image classification**			
Spacecraft	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)
SMAP	6	**1**	4	70.6	**9**	11	1	60.0	5	52	5	14.9
MSL	0	**0**	6	0.0	6	13	0	48.0	6	16	**0**	**42.9**
Overall	6	**1**	10	52.8	15	24	**1**	**54.5**	11	68	5	26.3

RESULTS

- *Direct classification*
  - high number of **TP** (15), low number of **FN** (1)

	Forecasting&threshold				*Direct classification*				**Image classification**			
Spacecraft	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)
SMAP	6	**1**	4	**70.6**	**9**	11	1	60.0	5	52	5	14.9
MSL	0	**0**	6	0.0	6	13	0	48.0	6	16	**0**	**42.9**
Overall	6	**1**	10	52.8	**15**	24	**1**	**54.5**	11	68	5	26.3

RESULTS

- *Direct classification*
  - high number of **TP** (15), low number of **FN** (1)
  - slightly (but significantly) better **F1 score** (54.5%) than *forecasting & threshold*
  - but higher number of FP (24) than *forecasting & threshold*

	Forecasting&threshold				*Direct classification*				**Image classification**			
Spacecraft	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)
SMAP	6	**1**	4	**70.6**	**9**	11	1	60.0	5	52	5	14.9
MSL	0	**0**	6	0.0	6	13	0	48.0	6	16	**0**	**42.9**
Overall	6	**1**	10	52.8	**15**	24	**1**	**54.5**	11	68	5	26.3

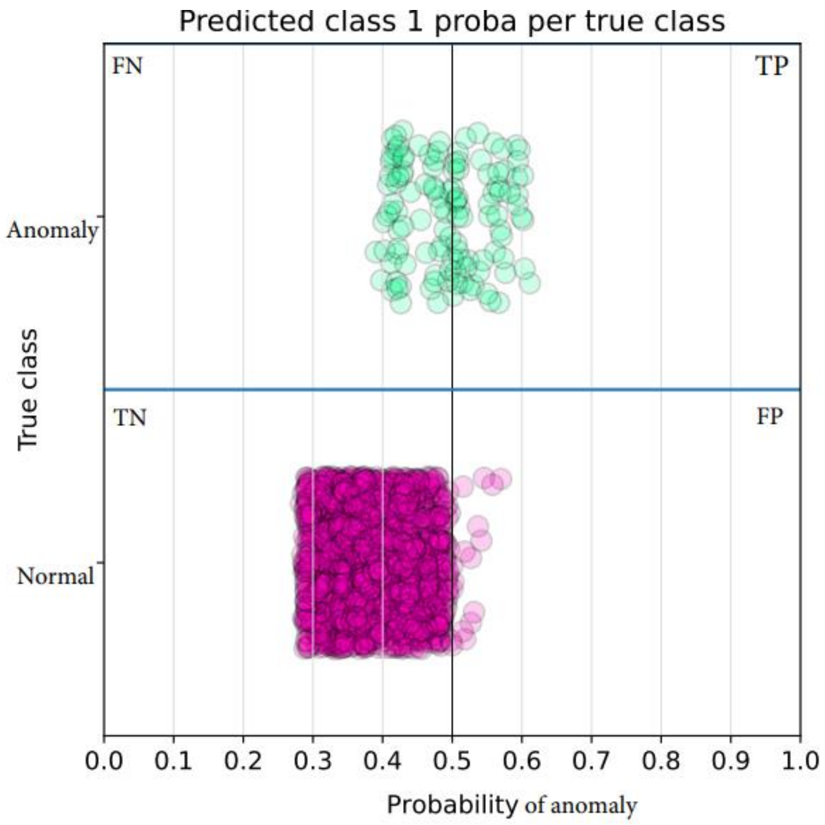
RESULTS

- *Image classification* using Gramian Angular Field transforms
  - less successful
  - Challenges: High computational overhead, high number of FP

	Forecasting&threshold				*Direct classification*				Image classification			
Spacecraft	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)	TP	FP	FN	F1(%)
SMAP	6	**1**	4	**70.6**	**9**	11	1	60.0	5	52	5	14.9
MSL	0	**0**	6	0.0	6	13	0	48.0	6	16	0	42.9
Overall	6	**1**	10	52.8	15	24	**1**	**54.5**	11	68	5	26.3

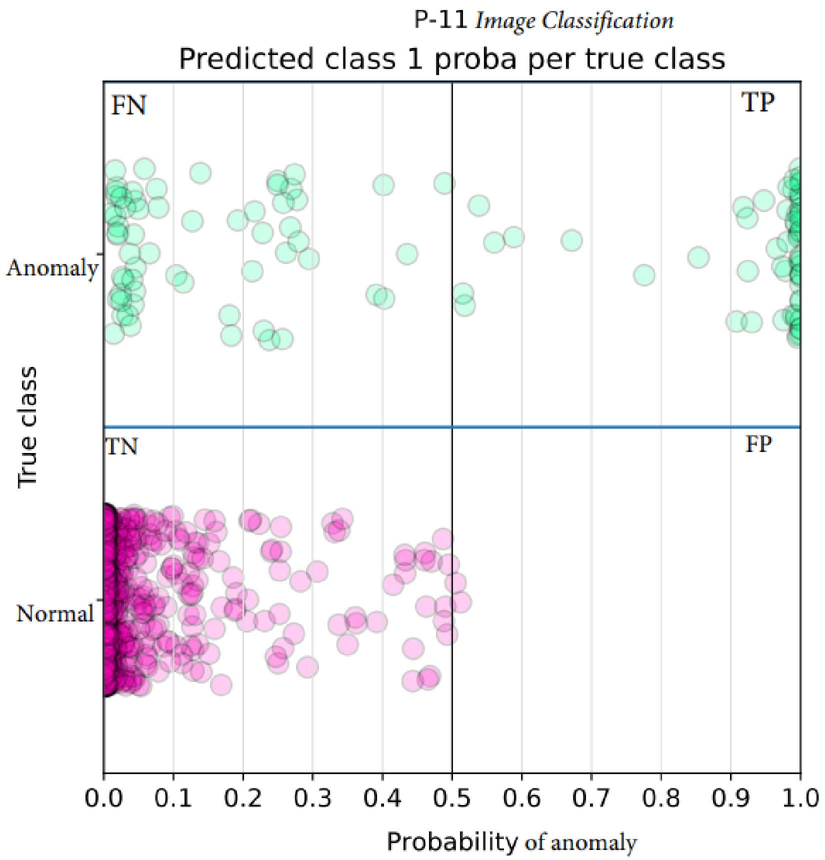
RESULTS: IMAGE CLASSIFICATION

***Direct classification*:** Probability of Anomaly



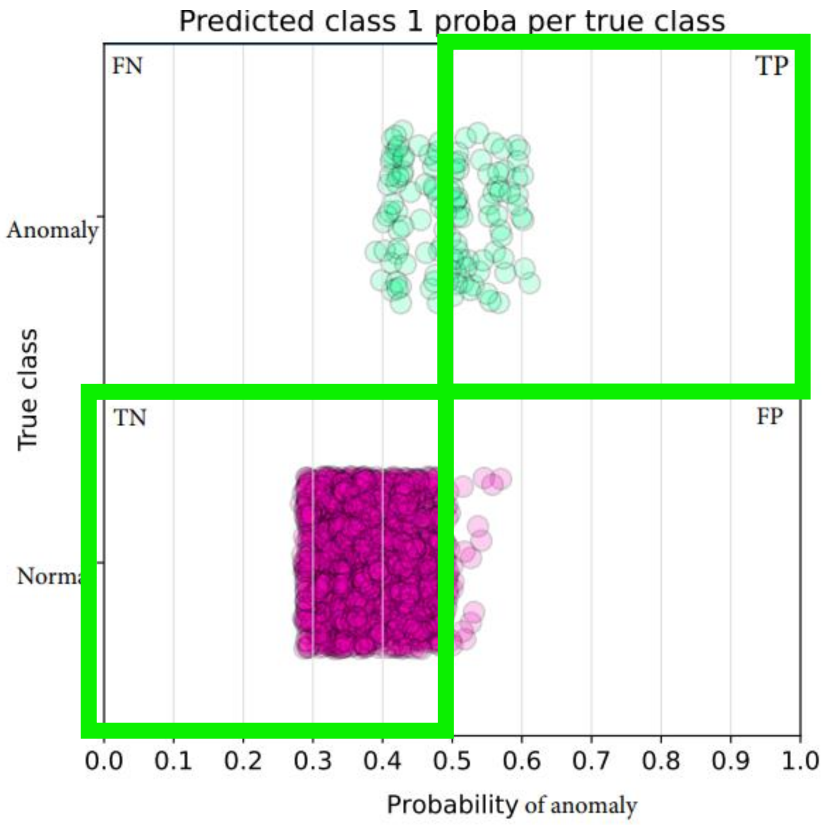
P-11

***Image classification*:** Probability of Anomaly



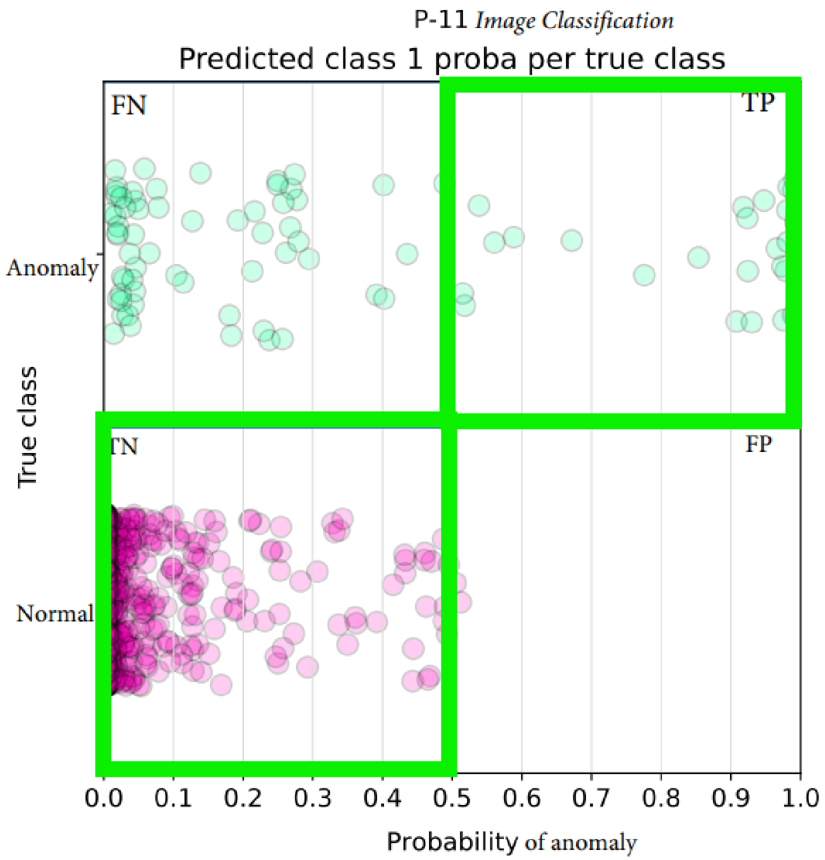
RESULTS: IMAGE CLASSIFICATION

***Direct classification*: Probability of Anomaly**



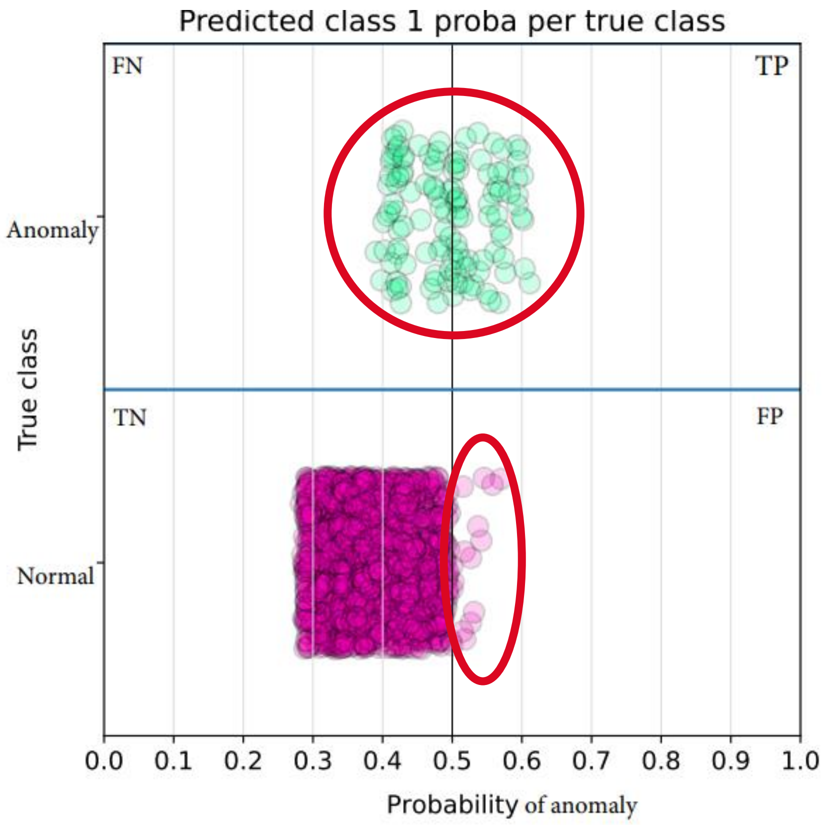
P-11

***Image classification*: Probability of Anomaly**



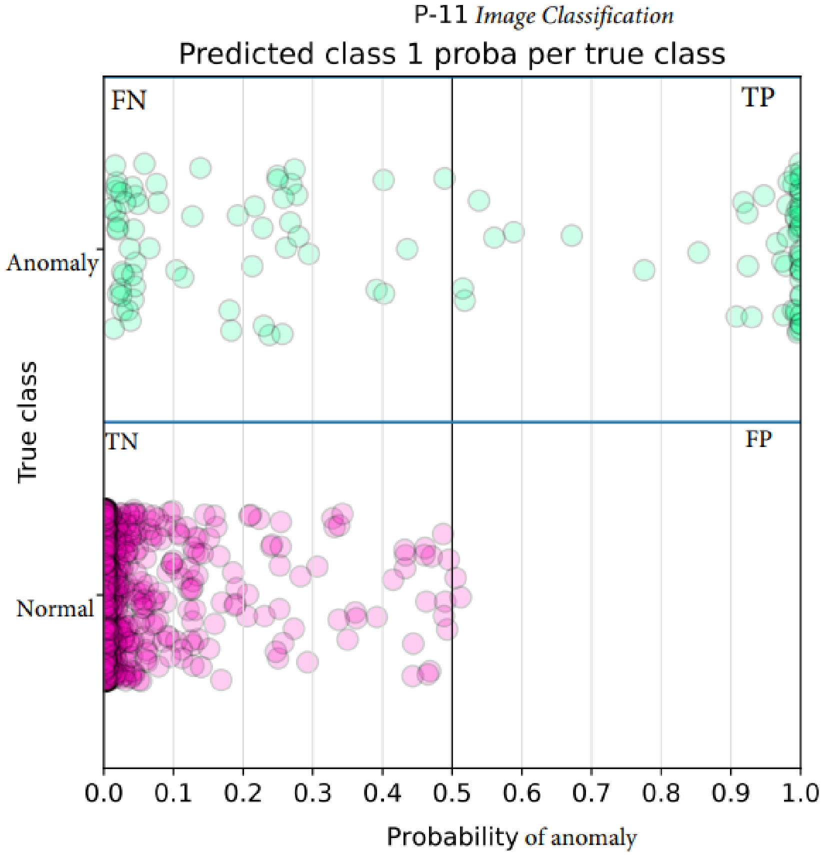
RESULTS: IMAGE CLASSIFICATION

***Direct classification*:** Probability of Anomaly



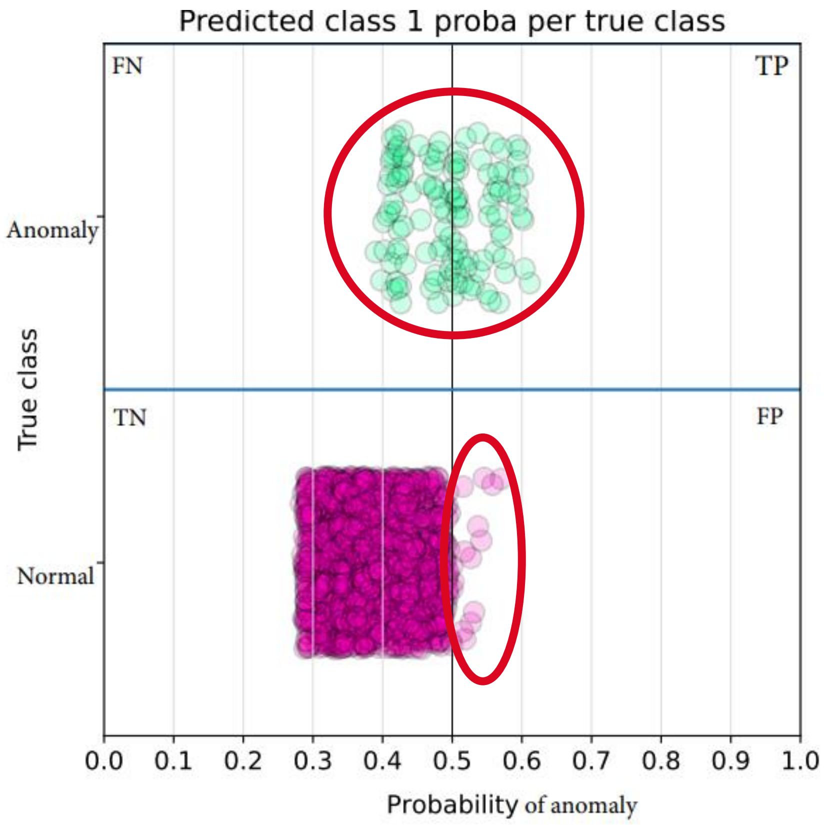
P-11

***Image classification*:** Probability of Anomaly



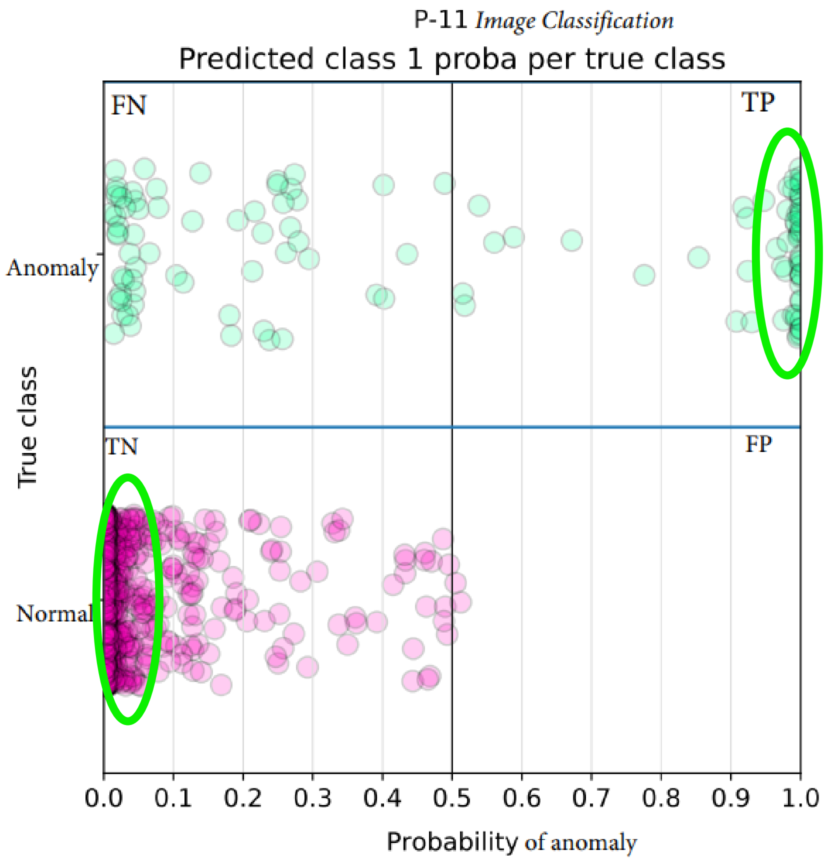
RESULTS: IMAGE CLASSIFICATION

***Direct classification*:** Probability of Anomaly



P-11

***Image classification*:** Probability of Anomaly



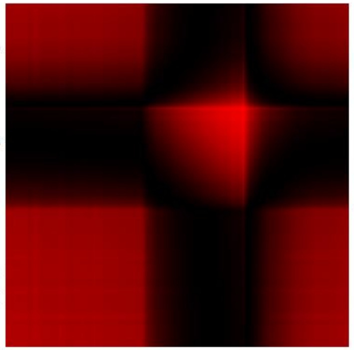
➔ can offer better **class separation / discrimination** than *direct classification* in certain telemetry channels

RESULTS: IMAGE CLASSIFICATION INTERPRETATION

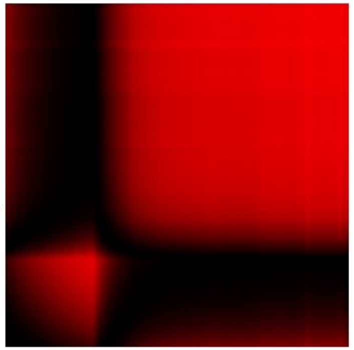
- ***Image classification***
  - Transforming time series data
    - with “high variance” (e.g. T-1) results in not easily identifiable anomaly images
    - with “**low variance**” (e.g. T-3) results in easily identifiable anomaly images

Channel T-1 (“high variance”)

Normal time window

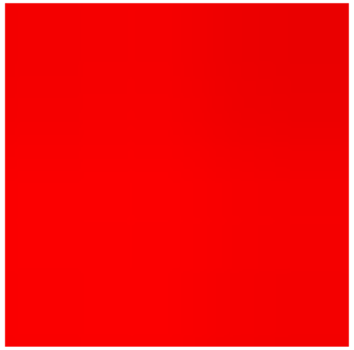


Anomalous time window

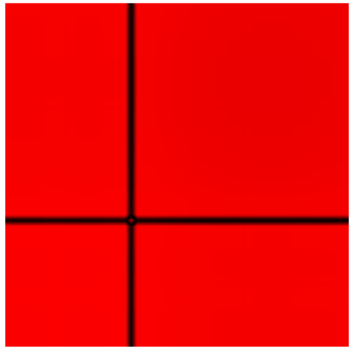


Channel T-3 (“low variance”)

Normal time window



Anomalous time window

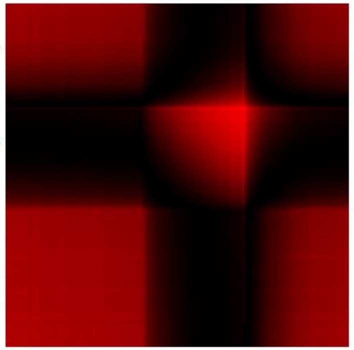


RESULTS: IMAGE CLASSIFICATION INTERPRETATION

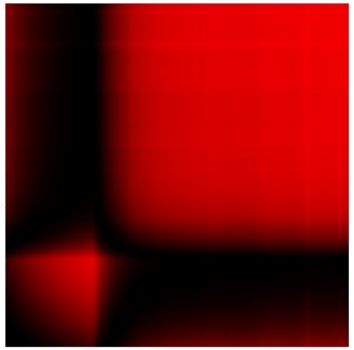
- ***Image classification***
  - Transforming time series data
    - with "**high variance**" (e.g. T-1) results in <u>not</u> easily identifiable anomaly images
    - with "low variance" (e.g. T-3) results in easily identifiable anomaly images

Channel T-1 ("high variance")

Normal time window

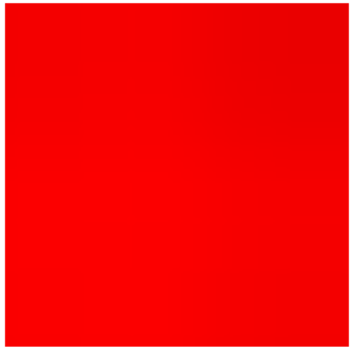


Anomalous time window

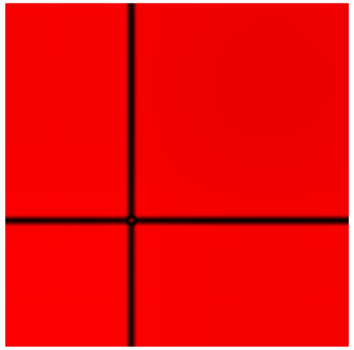


Channel T-3 ("low variance")

Normal time window



Anomalous time window

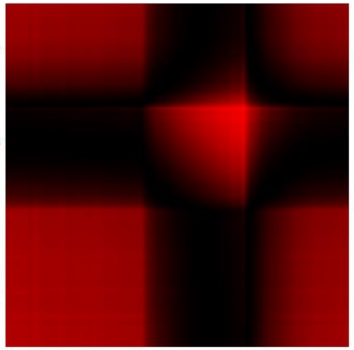


RESULTS: IMAGE CLASSIFICATION INTERPRETATION

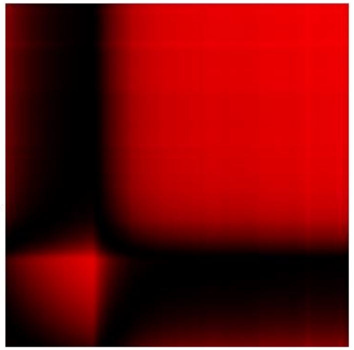
- ***Image classification***
  - Transforming time series data
    - with “high variance” (e.g. T-1) results in <u>not</u> easily identifiable anomaly images
    - with “**low variance**” (e.g. T-3) results in easily identifiable anomaly images

Channel T-1 (“high variance”)

Normal time window

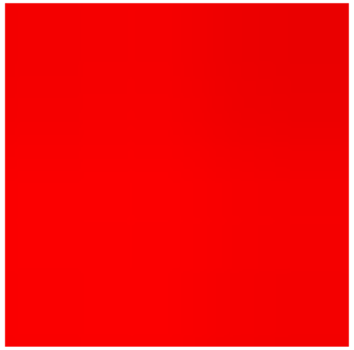


Anomalous time window

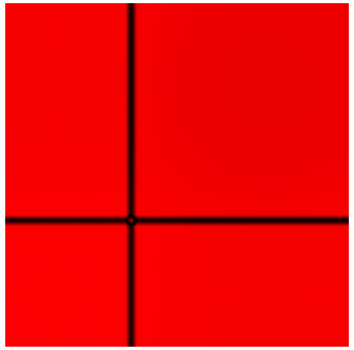


Channel T-3 (“low variance”)

Normal time window



Anomalous time window

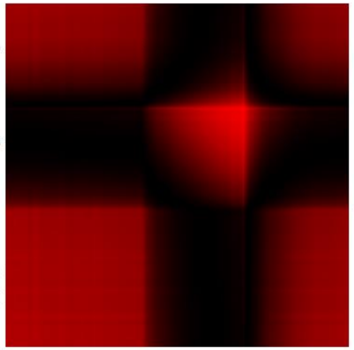


RESULTS: IMAGE CLASSIFICATION INTERPRETATION

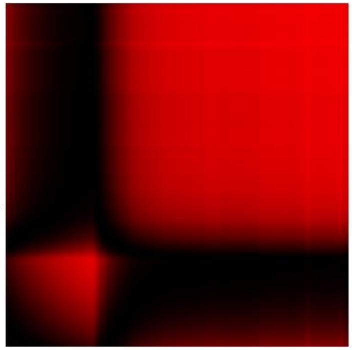
- ***Image classification***
  - Transforming time series data
    - with “high variance” (e.g. T-1) results in not easily identifiable anomaly images
    - with “**low variance**” (e.g. T-3) results in easily identifiable anomaly images

Channel T-1 (“high variance”)

Normal time window

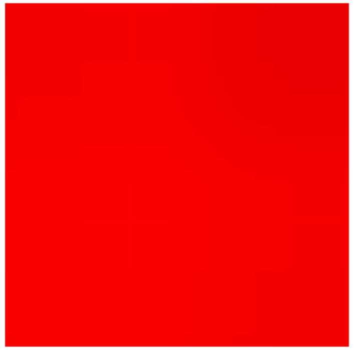


Anomalous time window

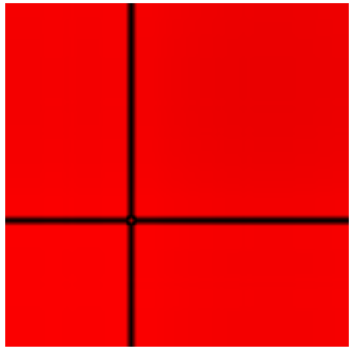


Channel T-3 (“low variance”)

Normal time window



Anomalous time window



➔ *Image classification* may be targeted based on telemetry channel behavior

CONCLUSION

- **Summary:**
  - Comparison of traditional *forecasting & threshold* to *classification* techniques for anomaly detection in spacecraft telemetry

CONCLUSION

- **Summary:**
  - Comparison of traditional *forecasting & threshold* to *classification* techniques for anomaly detection in spacecraft telemetry

- **Key Findings:**
  - *Direct classification* outperforms traditional methods,
    while *image classification* shows potential in specific scenarios

CONCLUSION

- **Summary:**
  - Comparison of traditional *forecasting & threshold* to *classification* techniques for anomaly detection in spacecraft telemetry

- **Key Findings:**
  - *Direct classification* outperforms traditional methods,
    while *image classification* shows potential in specific scenarios

- **Implications:**
  - This research contributes to advancing automated monitoring systems in space missions

CONCLUSION

- **Summary:**
  - Comparison of traditional *forecasting & threshold* to *classification* techniques for anomaly detection in spacecraft telemetry
- **Key Findings:**
  - *Direct classification* outperforms traditional methods,
    while *image classification* shows potential in specific scenarios
- **Implications:**
  - This research contributes to advancing automated monitoring systems in space missions
- **Discussion:**
  - Strengths: *Direct classification* demonstrated better overall performance and potential for operational use
  - Weaknesses: High number of FP in both *direct* and *image classification* needs to be addressed
  - Future Potential: *Image classification* shows promise in specific contexts and warrants further exploration

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle imbalanced data better

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle imbalanced data better
  - Implement oversampling techniques to create more balanced datasets (Liu et al., 2021; Zhao et al, 2022; Baumgartner et al., 2022)

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle imbalanced data better
  - Implement oversampling techniques to create more balanced datasets
    (Liu et al., 2021; Zhao et al, 2022; Baumgartner et al., 2022)

- **Research Directions:**
  - Investigate further image transformation techniques

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle imbalanced data better
  - Implement oversampling techniques to create more balanced datasets (Liu et al., 2021; Zhao et al, 2022; Baumgartner et al., 2022)

- **Research Directions:**
  - Investigate further image transformation techniques
  - Apply different shape-specific models based on the shape of the telemetry channels

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle imbalanced data better
  - Implement oversampling techniques to create more balanced datasets
    (Liu et al., 2021; Zhao et al, 2022; Baumgartner et al., 2022)

- **Research Directions:**
  - Investigate further image transformation techniques

  - Apply different shape-specific models based on the chape of the telemetry channels
    ➔ Tailor the learning to the shape instead of single model for all (Lakey & Schlippe, 2024)

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle in
  - Implement oversampling techniques to create more balance
    (Liu et al., 2021; Zhao et al, 2022; Baumgartner et al., 2022)

- **Research Directions:**
  - Investigate further image transformation techniques
  - Apply different shape-specific models based on the chape of
    ➔ Tailor the learning to the shape instead of single model fo

A Comparison of Deep Learning Architectures for Spacecraft Anomaly Detection

Daniel Lakey
IU International University of Applied Sciences
daniel.lakey@iu-study.org

Tim Schlippe
IU International University of Applied Sciences
tim.schlippe@iu.org

Abstract—Spacecraft operations are highly critical, demanding impeccable reliability and safety. Ensuring the optimal performance of a spacecraft requires the early detection and mitigation of anomalies, which could otherwise result in unit or mission failures. With the advent of deep learning, a surge of interest has been seen in leveraging these sophisticated algorithms for anomaly detection in space operations. Our study aims to compare the efficacy of various deep learning architectures in detecting anomalies in spacecraft data. The deep learning models under investigation include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer-based architectures. Each of these models was trained and validated using a comprehensive dataset sourced from multiple spacecraft missions, encompassing diverse operational scenarios and anomaly types. We also present a novel approach to the rapid assignment of spacecraft telemetry data sets to discrete clusters, based on the statistical characteristics of the signal. This clustering allows us to compare different deep learning architectures to different types of data signal behaviour. Initial results indicate that while CNNs excel in identifying spatial patterns and may be effective for some classes of spacecraft data, LSTMs and RNNs show a marked proficiency in capturing temporal anomalies seen in time-series spacecraft telemetry. The Transformer-based architectures, given their ability to focus on both local and global contexts, have showcased promising results, especially in scenarios where anomalies are subtle and span over longer durations. Additionally, considerations such as computational efficiency, ease of deployment, and real-time processing capabilities were evaluated. While CNNs and LSTMs demonstrated a balance between accuracy and computational demands, Transformer architectures, though highly accurate, require significant computational resources. In conclusion, the choice of deep learning architecture for spacecraft anomaly detection is highly contingent on the nature of the data, the type of anomalies, and operational constraints. This comparative study provides a foundation for space agencies and researchers to make informed decisions in the integration of deep learning techniques for ensuring spacecraft safety and reliability.

TABLE OF CONTENTS

1. INTRODUCTION

The field of space exploration has had significant advancements in recent decades, characterised by the increasing sophistication of spacecraft and the expanding complexity of missions. As mankind expands its presence in outer space, the importance of precise and dependable data from spacecraft systems has become of utmost significance. Time series data, which refers to a sequential arrangement of data points organised in chronological order, holds major significance in the domain of spacecraft telemetry. Spacecraft systems are reflected by telemetry data, which provides information on their state, health, and performance. This data allows for the analysis of both regular and potentially abnormal operations [1].

Anomalies observed in spacecraft telemetry data are unanticipated occurrences that pose potential risks, as they depart significantly from the predicted operational patterns of the system. The quick detection and identification of these abnormalities is of paramount significance in order to avert catastrophic failures, limit risks, and guarantee the durability of space missions. According to [2], the prompt identification and effective detection of these anomalies by operational engineers play a crucial role in enhancing efficiency, minimising expenses, and enhancing safety. As the complexity of spacecraft continues to advance, there is a corresponding growth in the variety of telemetry parameters associated with them. The utilisation of conventional, manual or simple "out-of-limits" techniques are becoming ever more difficult for the purpose of identifying anomalies [3].

In recent years, there has been considerable focus on the advancement of anomaly detection techniques for satellite telemetry data. Numerous advanced algorithms and strategies have been proposed by prominent organisations such as NASA [4], ESA [3], and CNES [5] to tackle this task. Every approach possesses its own set of advantages and disadvantages. There is a clear trend towards deep learning approaches over statistical methods due to their ability to synthesise the complex multivariate temporally-connected data inherent to spacecraft telemetry [6]. The objective of this paper is to investigate and assess different methodologies for anomaly identification in order to determine the most optimal and efficient approach for analysing spacecraft telemetry.

Our work pioneers several notable contributions to the domain of spacecraft anomaly detection, presenting advancements that enhance the understanding of deep learning in this field. Firstly, it unfolds a comprehensive side-by-side comparison of multiple deep learning model architectures, shedding light on their effectiveness in detecting anomalies in spacecraft telemetry. This comparison is distinctively valuable as it incorporates models that, to our knowledge, have not been previously applied to spacecraft anomalies, thereby opening new avenues for exploration and implementation. Secondly, we introduce an innovative unsupervised

FUTURE WORK

- **Improvements:**
  - Explore one-class classification (Fernandez et al., 2018) to handle imbalanced data better
  - Implement oversampling techniques to create more balanced datasets (Liu et al., 2021; Zhao et al, 2022; Baumgartner et al., 2022)
- **Research Directions:**
  - Investigate further image transformation techniques
  - Apply different shape-specific models based on the chape of the telemetry channels
    ➔ Tailor the learning to the shape instead of single model for all (Lakey & Schlippe, 2024)
  - Use a better dataset (more labels, more anomalies, etc.), e.g. ESA Anomaly Dataset (De Canio et al., 2024; Kotowski et al., 2024)

THANK YOU

Tim Schlippe / Daniel Lakey
tim.schlippe@iu.org / daniel.lakey@iu-study.org