

Second AI and Data Science Workshop for Earth and Space Science, Jet Propulsion Laboratory

Multi-Class Anomaly Detection in Flight Data Using Semi-Supervised Explainable Deep Learning Model

Milad Memarzadeh and Bryan Matthews
Data Sciences Group, NASA Ames Research Center

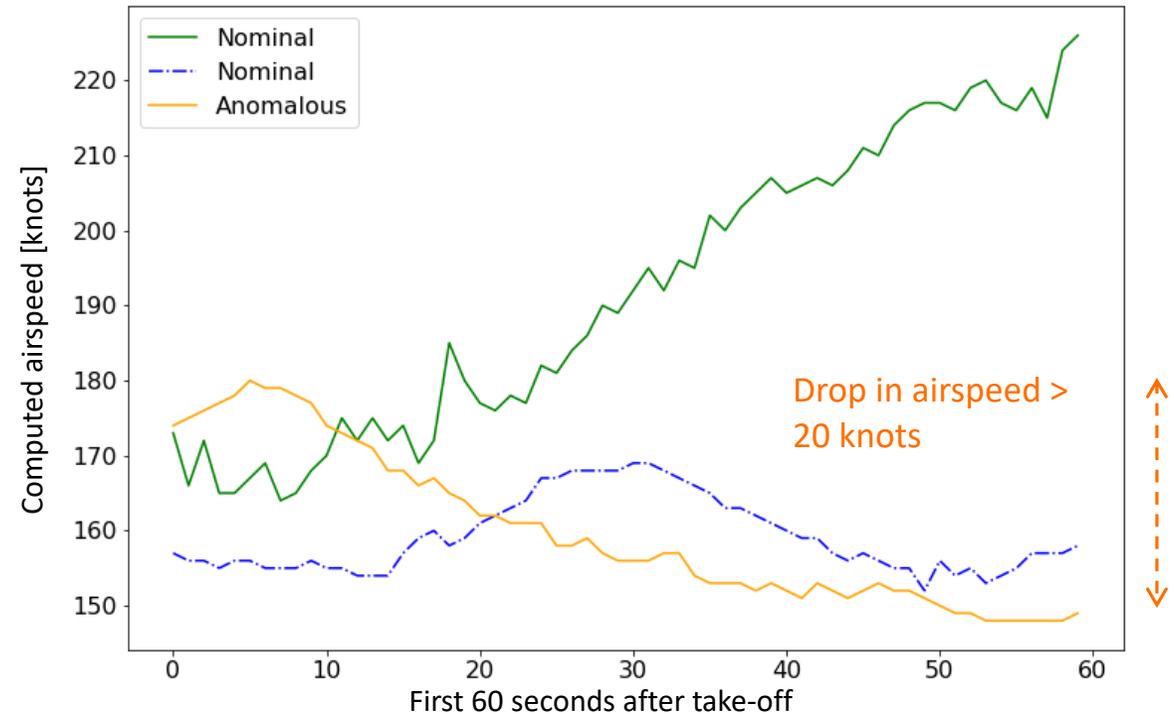
Other contributors: Ilya Avrekh and Thomas Templin



Aviation anomaly detection literature

Exceedance detection:

Comparing against the **pre-defined thresholds**, which are identified by subject-matter experts.



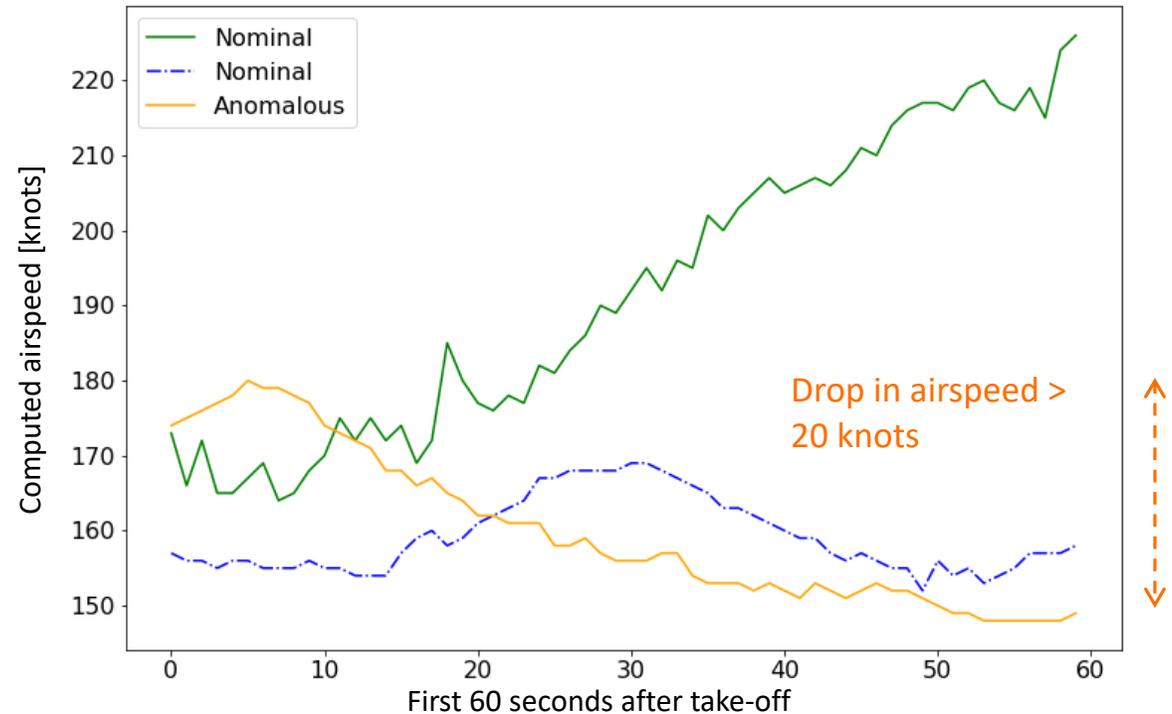
Aviation anomaly detection literature

Exceedance detection:

Comparing against the **pre-defined thresholds**, which are identified by subject-matter experts.

Cons:

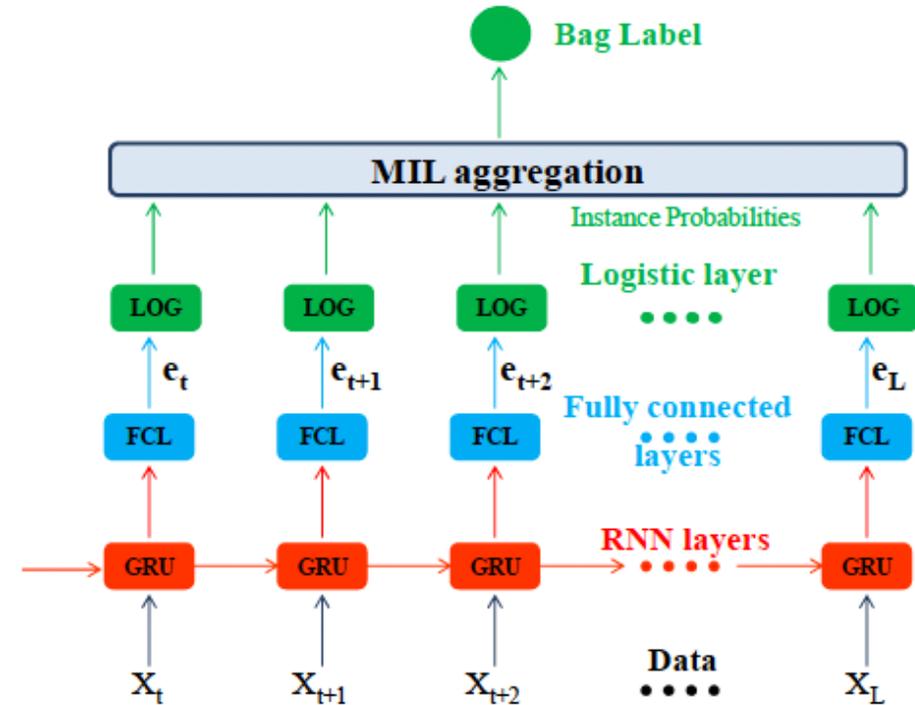
- complete reliance on domain knowledge.
- requires extensive reviews of entire data.
- can only identify known anomalies.



Aviation anomaly detection literature

Supervised learning:

Taking advantage of recent developments in deep learning and recurrent neural networks to tackle the reliance on the domain knowledge.



Aviation anomaly detection literature

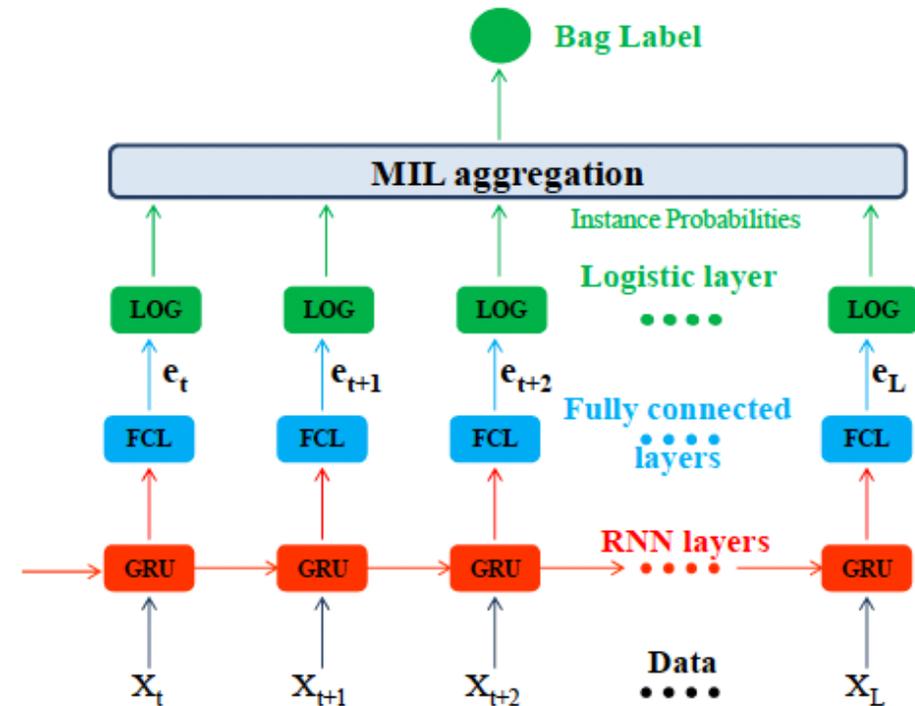
Supervised learning:

Taking advantage of recent developments in deep learning and recurrent neural networks to tackle the reliance on the domain knowledge.

Cons:

- can only identify **known anomalies**.
- **creating labels** for data requires huge effort from subject-matter experts and is largely expensive and impractical.

Hence, **unsupervised** learning or **semi-supervised** learning are the only feasible choices.



Aviation anomaly detection literature

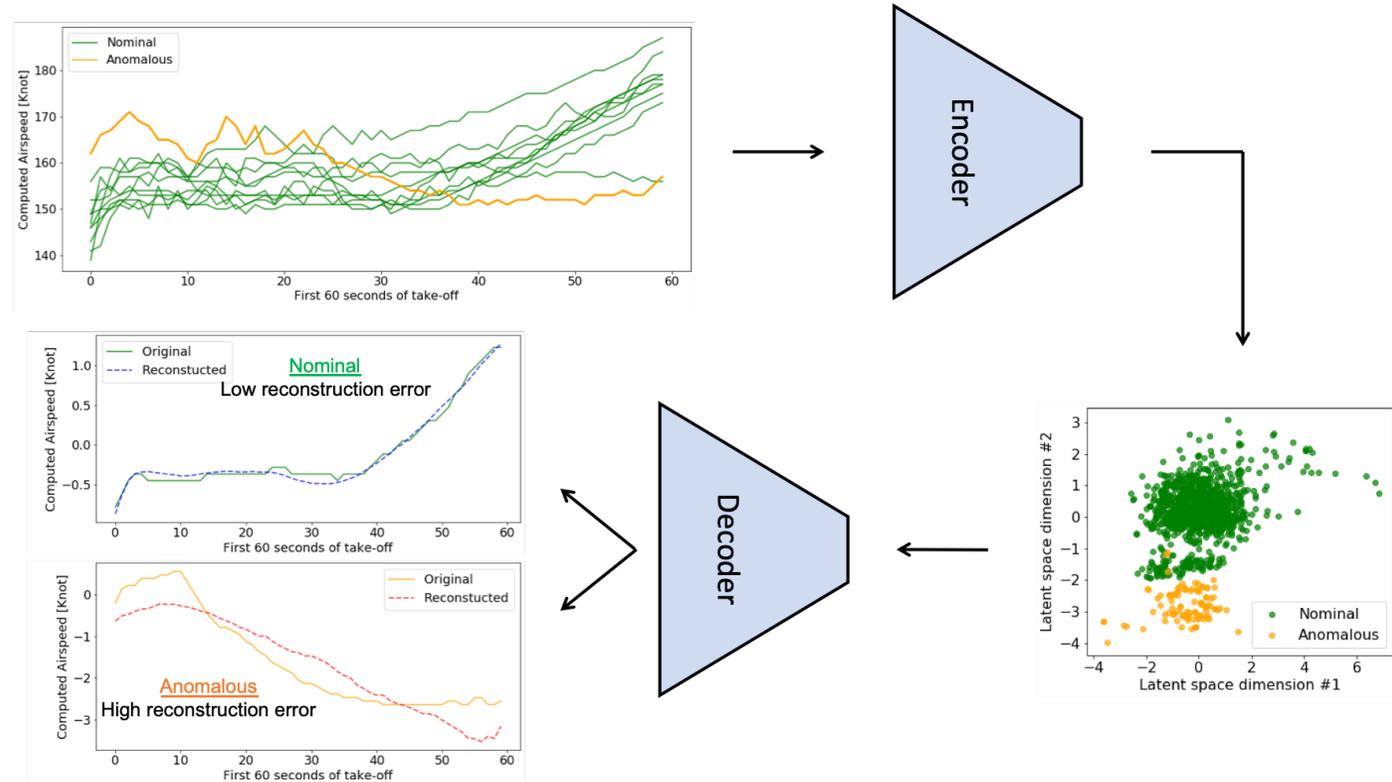
Unsupervised learning:

Using deep auto-encoders to identify anomalies without the need for labels.

reconstruction quality

$$\mathcal{J}_{\text{CVAE}} = \mathbb{E}_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] - \beta \text{KL}(q_{\phi}(z|x) || p_{\theta}(z))$$

distance of posterior and prior



Acronyms:

CVAE: Convolutional Variational Auto-Encoder



Aviation anomaly detection literature

Unsupervised learning:

Using deep auto-encoders to identify anomalies without the need for labels.

reconstruction quality

$$\mathcal{J}_{\text{CVAE}} = \mathbb{E}_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] - \beta \text{KL}(q_{\phi}(z|x) || p_{\theta}(z))$$

distance of posterior and prior

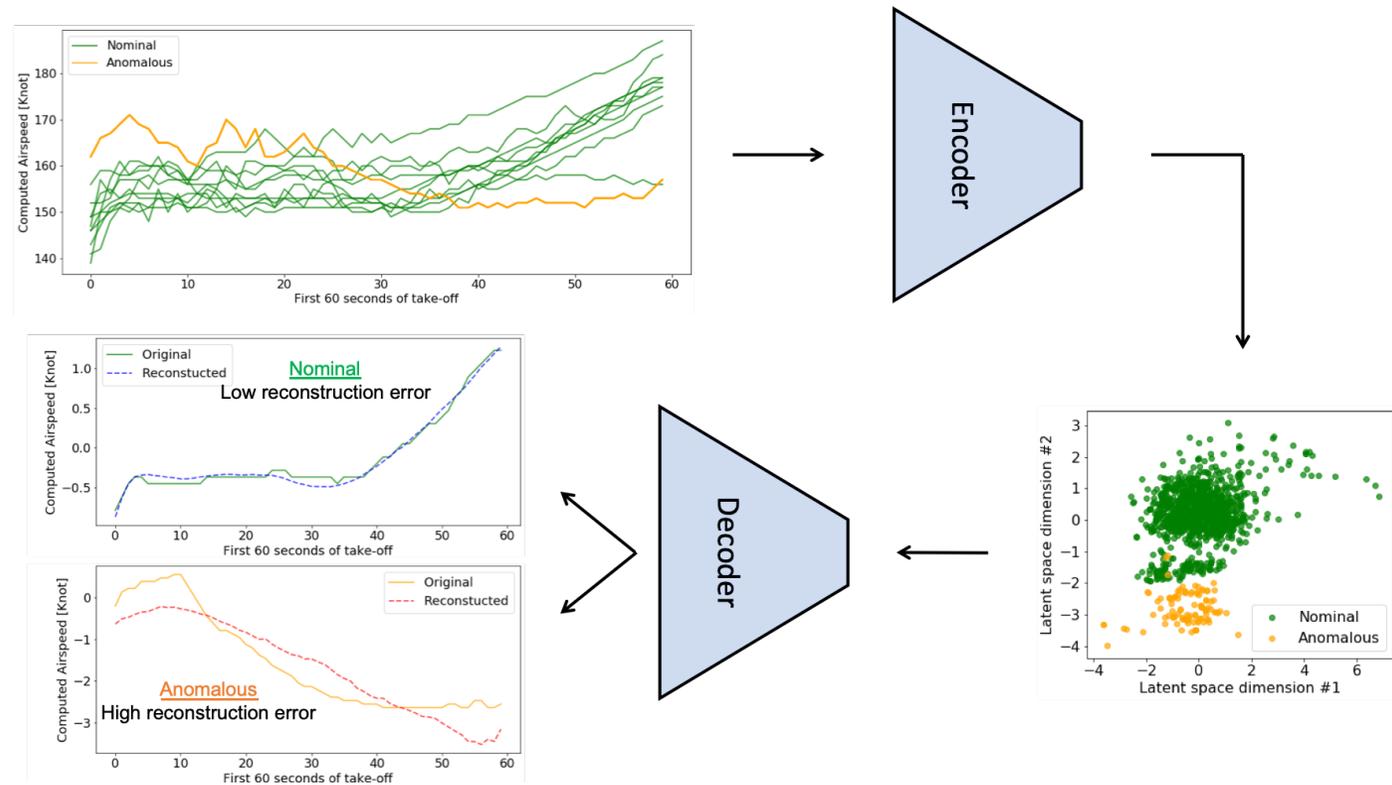
Identifying anomalies:

$$\zeta_i = \|x_i - \hat{x}_i\|_2^2, i \in \{1, \dots, N\}$$

$$\text{thr} = \mathbb{E}[\zeta] + \alpha\sigma(\zeta)$$

Acronyms:

CVAE: Convolutional Variational Auto-Encoder



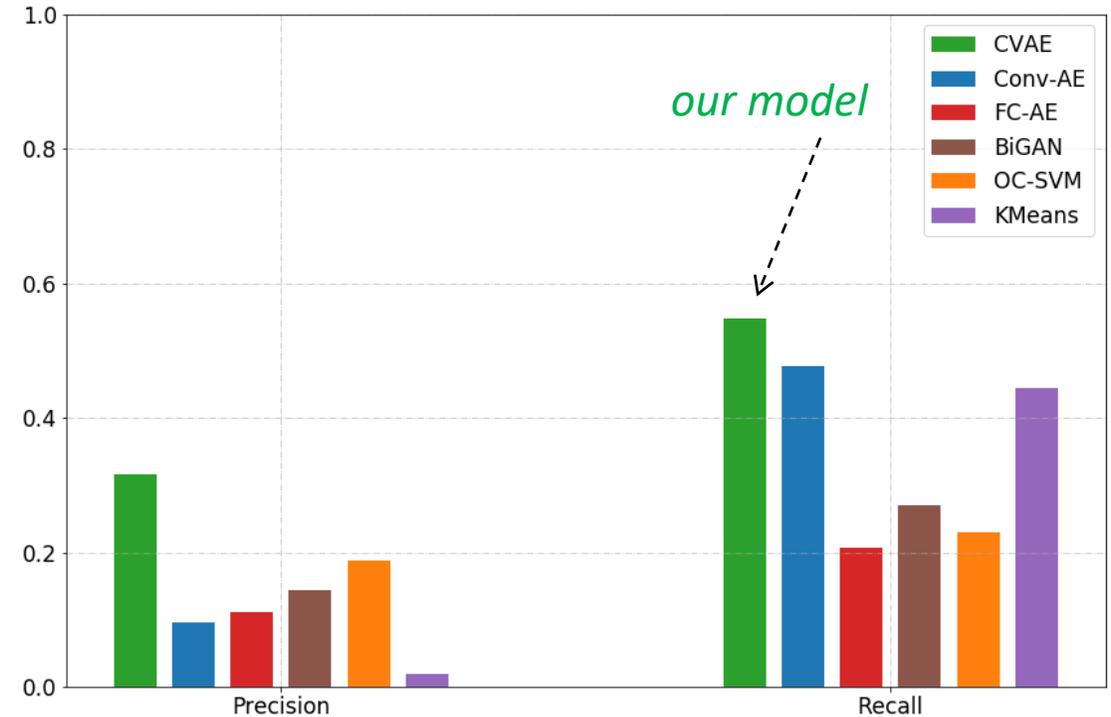
Aviation anomaly detection literature

Unsupervised learning:

Using deep auto-encoders to identify anomalies without the need for labels.

Cons:

- **Low precision**, which means high number of false positives and low reliability.
- It is not easy to extend to **multi-class** anomaly detection.



How to improve the reliability of unsupervised learning

Training CVAE (our model) only on nominal data improved the performance significantly:

- 36.8pp higher precision
- 27.3pp higher recall

Takeaway: how to take advantage of **minimally labelled data** that are available?



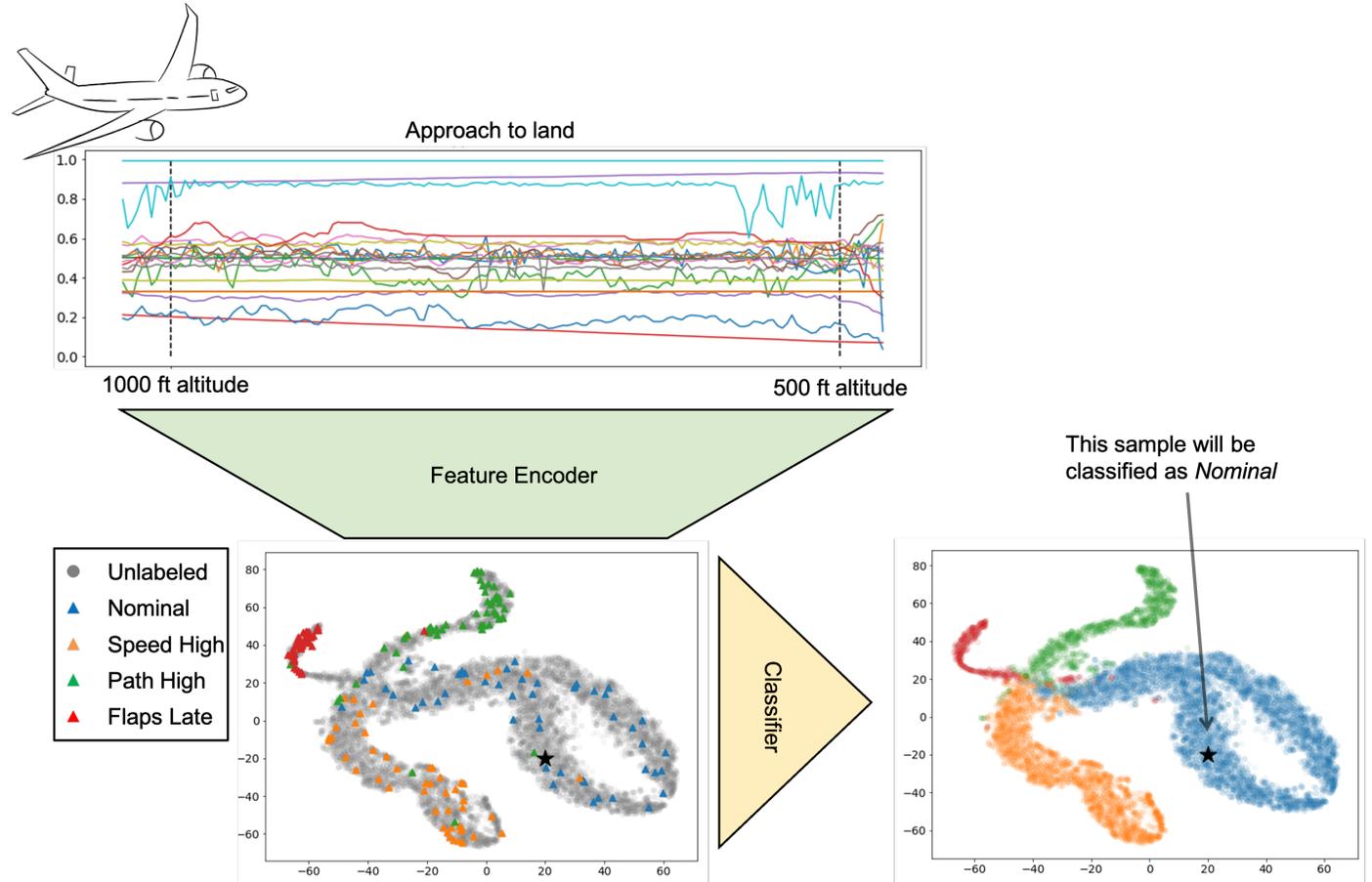
Semi-supervised learning: does it actually work?

Research objective: how to take advantage of **all available data**.

The data, X , is divided into:

- Labelled set, (X_L, y_L)
- Unlabelled set, X_U
- $|X_U| \gg |X_L|$

Unsupervised learning ignores y_L ,
while **supervised** learning ignores X_U .



Semi-supervised learning: does it actually work?

Research objective: how to take advantage of **all available data**.

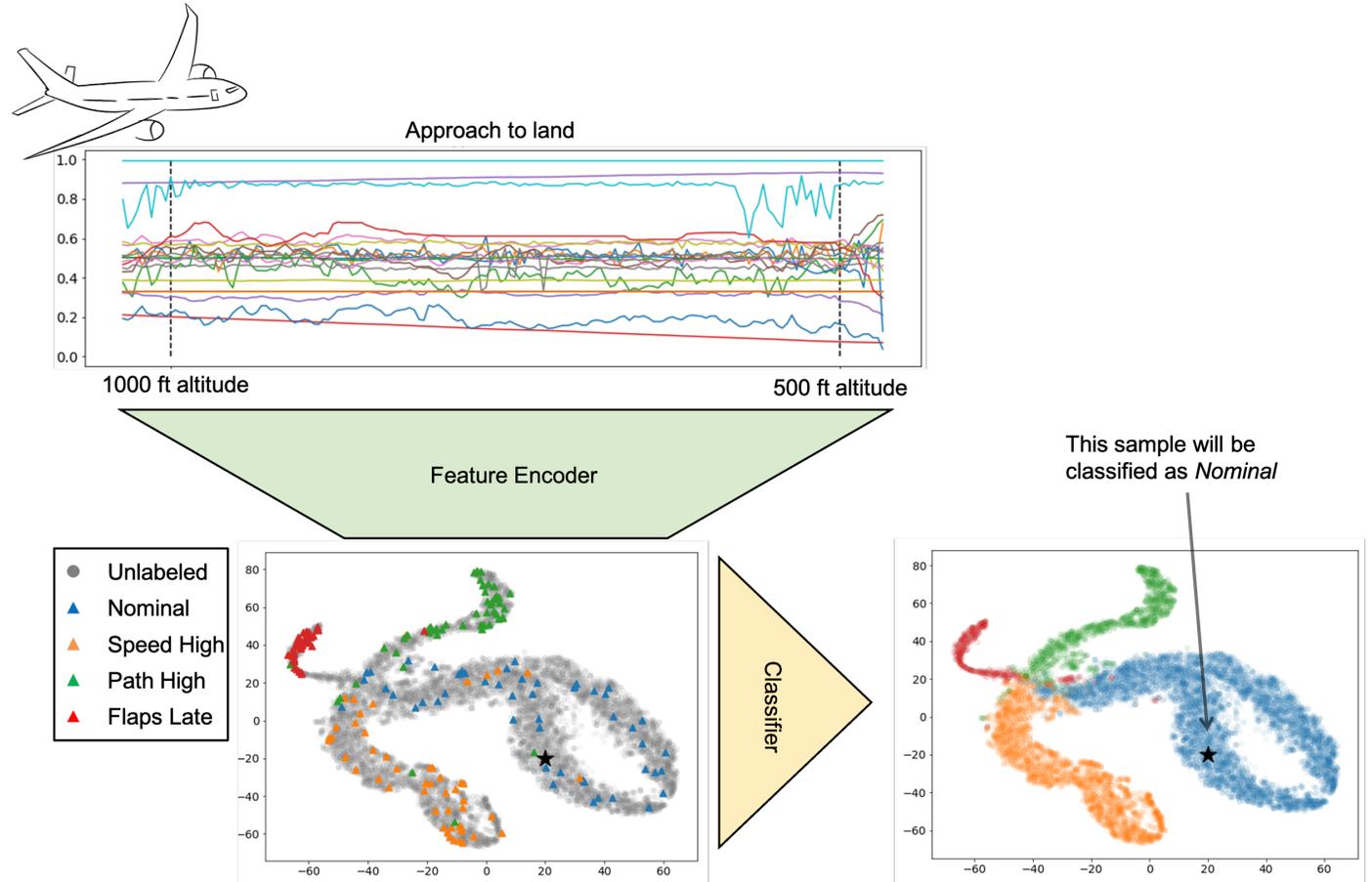
The data, X , is divided into:

- Labelled set, (X_L, y_L)
- Unlabelled set, X_U
- $|X_U| \gg |X_L|$

Unsupervised learning ignores y_L ,
while **supervised** learning ignores X_U .

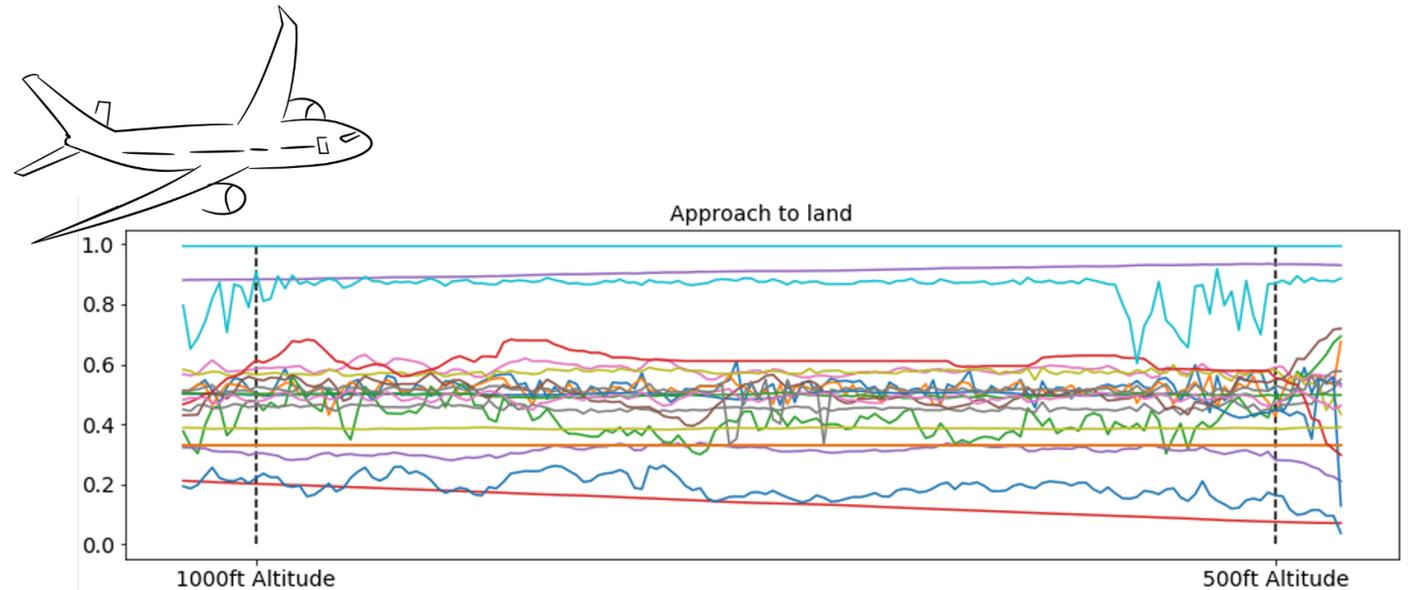
We implement two models:

- (1) **M1+M2**: encoding of the data is unsupervised.
- (2) **CCLP**: propagates label from X_L to X_U using graph theory and enforces compact clustering of data of same class in the feature space.



Anomaly detection case study based on real flight data

Each data instance is 160-s recording of 19 variables during **approach of a commercial aircraft to landing**. Attributes cover a variety of systems, including the state and orientation of the aircraft, positions and inputs of the control surfaces, engine parameters, and auto pilot modes and corresponding states.



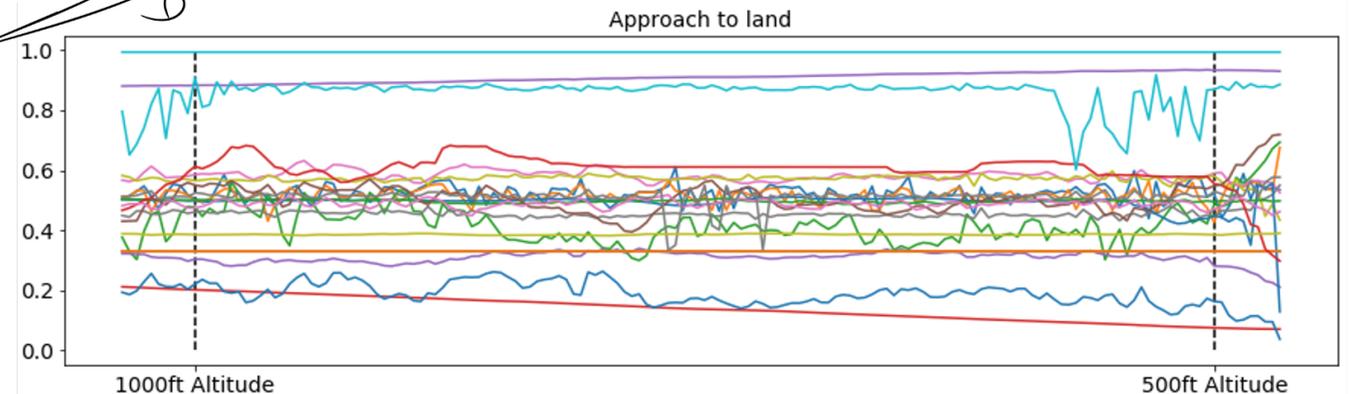
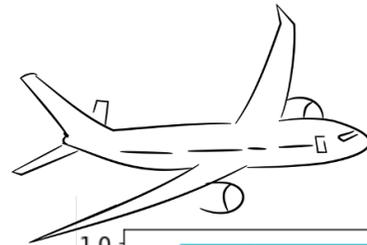
Anomaly detection case study based on real flight data

Each data instance is 160-s recording of 19 variables during **approach of a commercial aircraft to landing**. Attributes cover a variety of systems, including the state and orientation of the aircraft, positions and inputs of the control surfaces, engine parameters, and auto pilot modes and corresponding states.

Training data consists of 18,313 samples falling into four classes:

1. **Nominal** (66.7%)
2. **Speed High** (22.9%)
3. **Path High** (7.2%)
4. **Flaps Late** (3.2%)

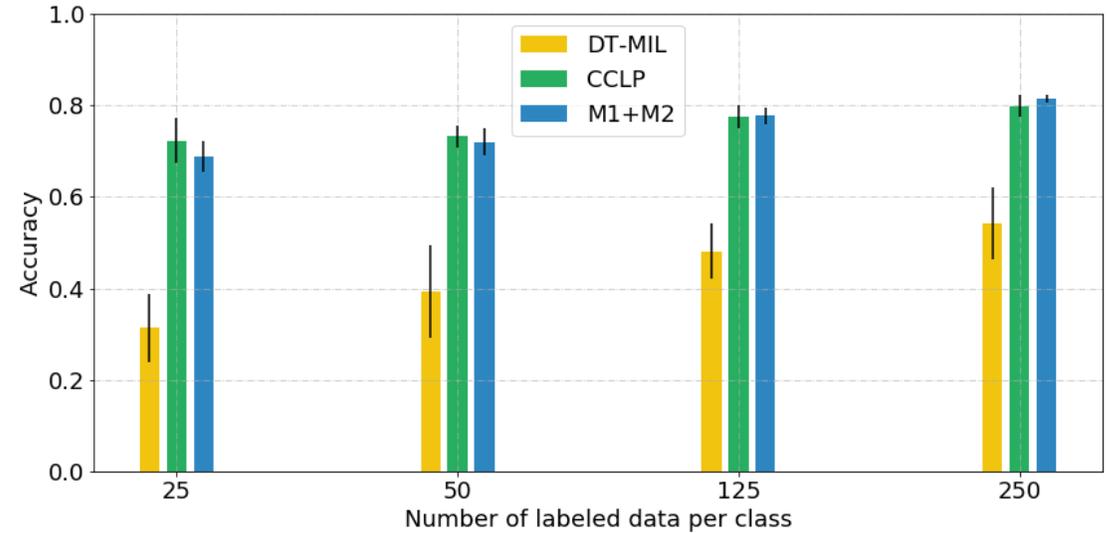
Separate **test data** of 6105 samples is used for evaluating the models.



Anomaly detection performance

Both **semi-supervised** models outperform the state-of-the-art **supervised** model:

- **CCLP** achieve 72.2% accuracy **with only 100 labeled data (0.55% of total)**, while **DT-MIL** reaches only 31.4%.

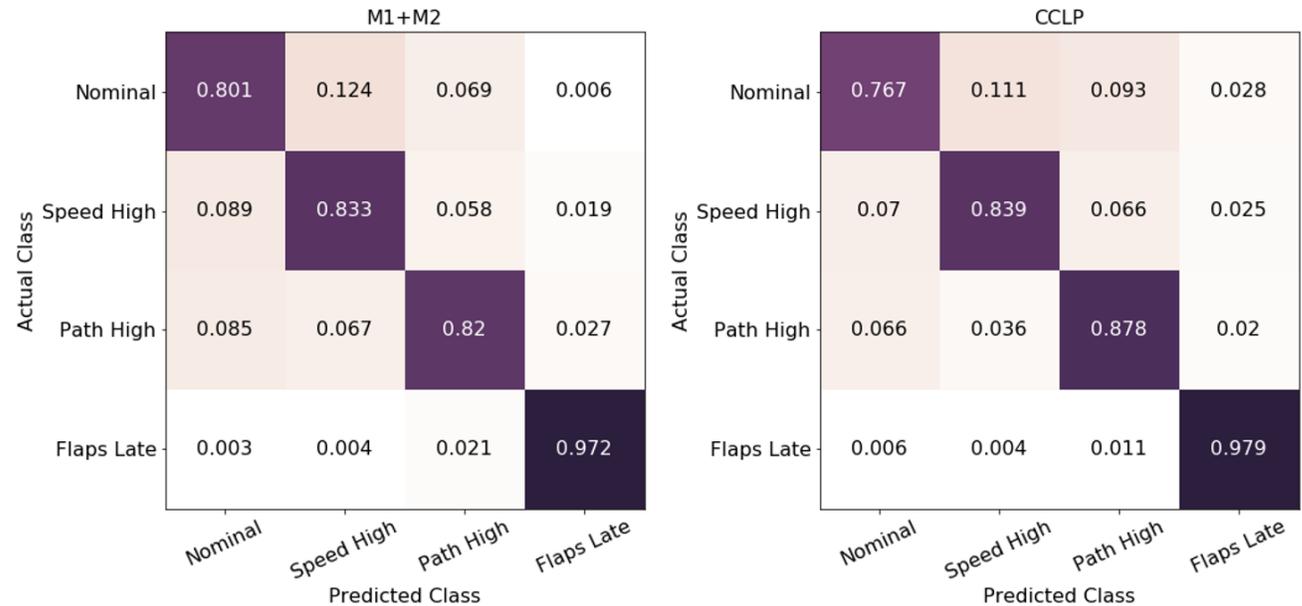
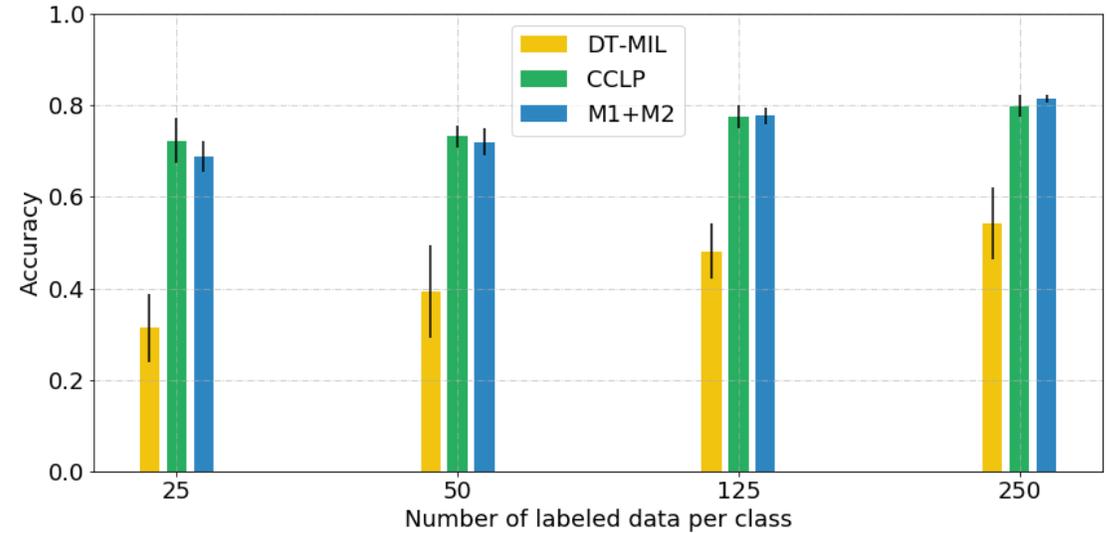


Anomaly detection performance

Both **semi-supervised** models outperform the state-of-the-art **supervised** model:

- **CCLP** achieve 72.2% accuracy **with only 100 labeled data (0.55% of total)**, while **DT-MIL** reaches only 31.4%.

The confusion matrix comparison between **M1+M2** and **CCLP** models show that **CCLP** performs better on anomaly classes.



Feature importance with random permutation

Using the trained model, we can identify **the most important features** for each class of anomaly.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

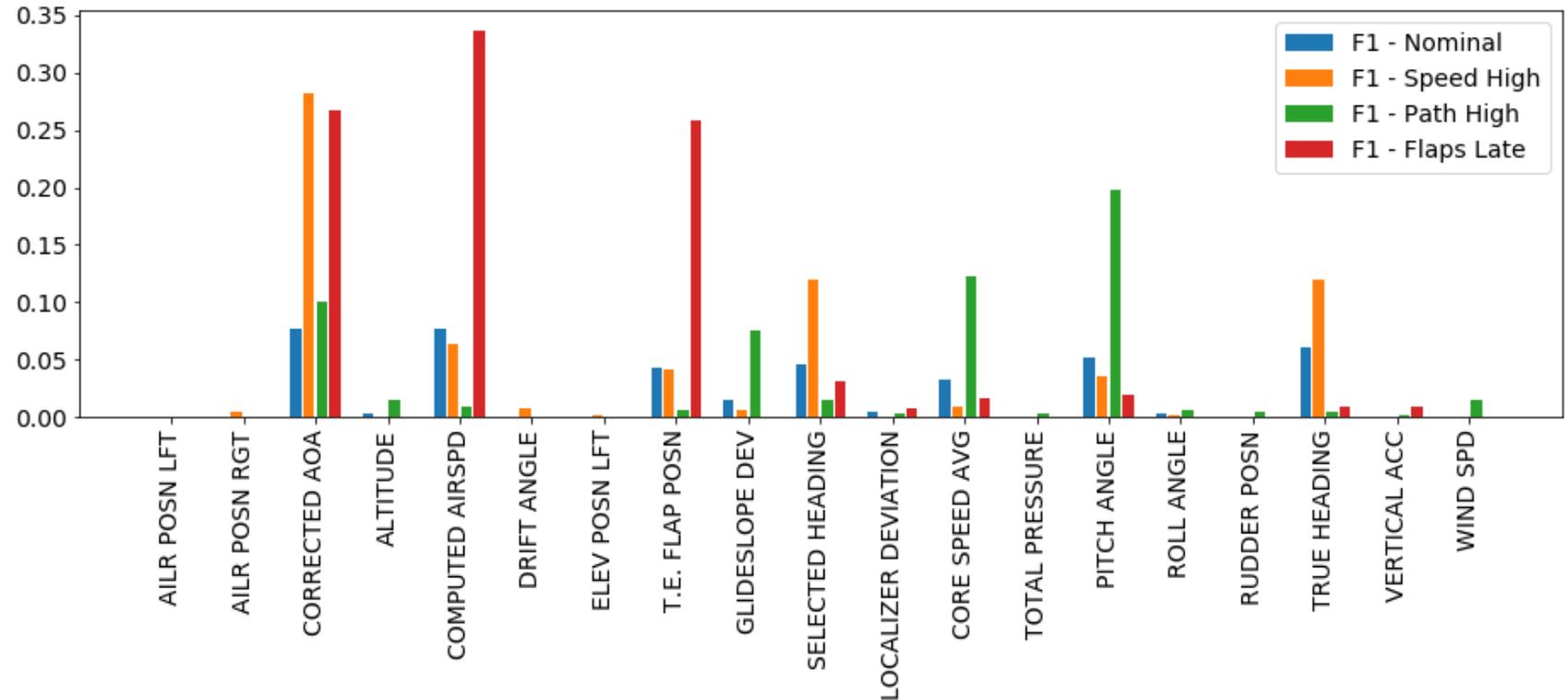
Acronyms:

TP: True Positive

FP: False Positive

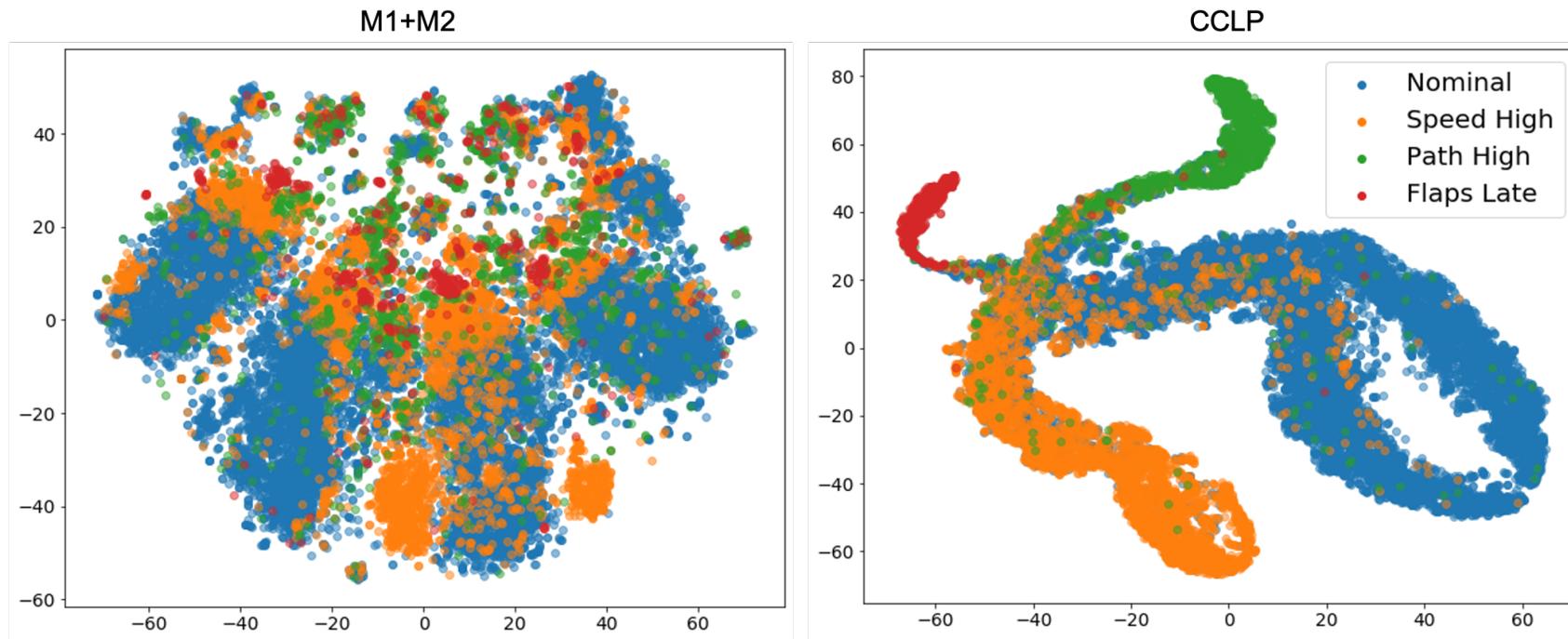
TN: True Negative

FN: False Negative



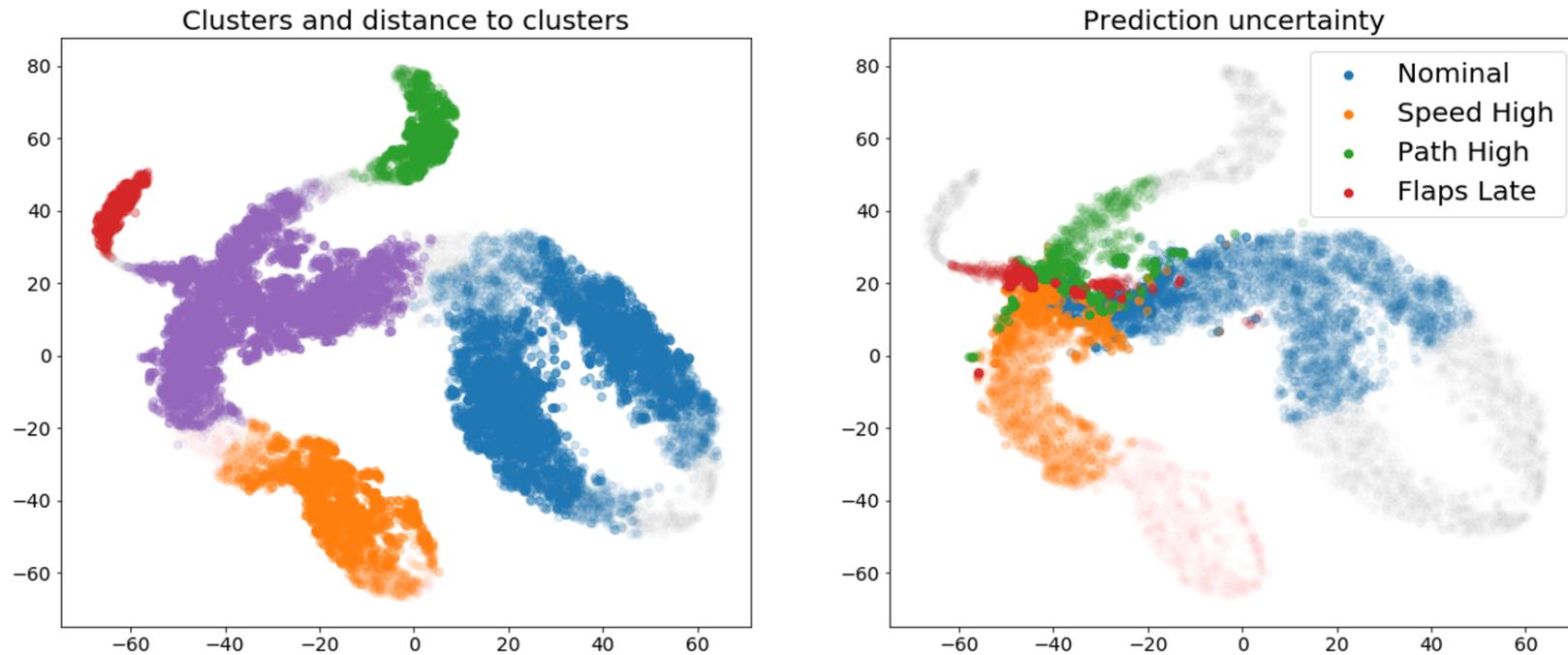
Latent space configuration: the superiority of the CCLP approach

Both figure show **2D visualization of the 256D latent space** of M1+M2 and CCLP models using t-Distributed Stochastic Neighbor Embedding (t-SNE), color-coded based on the actual class of the data.



Latent space configuration: thinking about the next steps

We evaluate the relationship between clusters shaped in the latent space and the prediction uncertainty of the classifier. These results suggest a **novel active learning strategy** for selecting the **most informative data** to be labeled in future efforts.



Our newest endeavor: best of both worlds

Key idea: Do not sacrifice the compact clustering for the reconstruction quality.

$$\begin{aligned}
 \mathcal{J}_{\text{DET}} = & \underbrace{-w_s \mathbb{E}_{(x_l, y_l)} \left[\mathcal{H} \left(y_l, c_\psi \left(y \mid q_\phi(z \mid x_l) \right) \right) \right]}_{\text{classification loss}} - \underbrace{w_c \frac{1}{S} \sum_{s=1}^S \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N -T_{ij} \log H_{ij}^{(s)}}_{\text{compact clustering loss}} - \\
 & \underbrace{w_r \mathbb{E}_{x \in X_L \cup X_U} \left[\mathcal{L}_{\text{rec}} \left(x, p_\theta \left(x \mid q_\phi(z \mid x) \right) \right) \right]}_{\text{reconstruction loss}}
 \end{aligned}$$

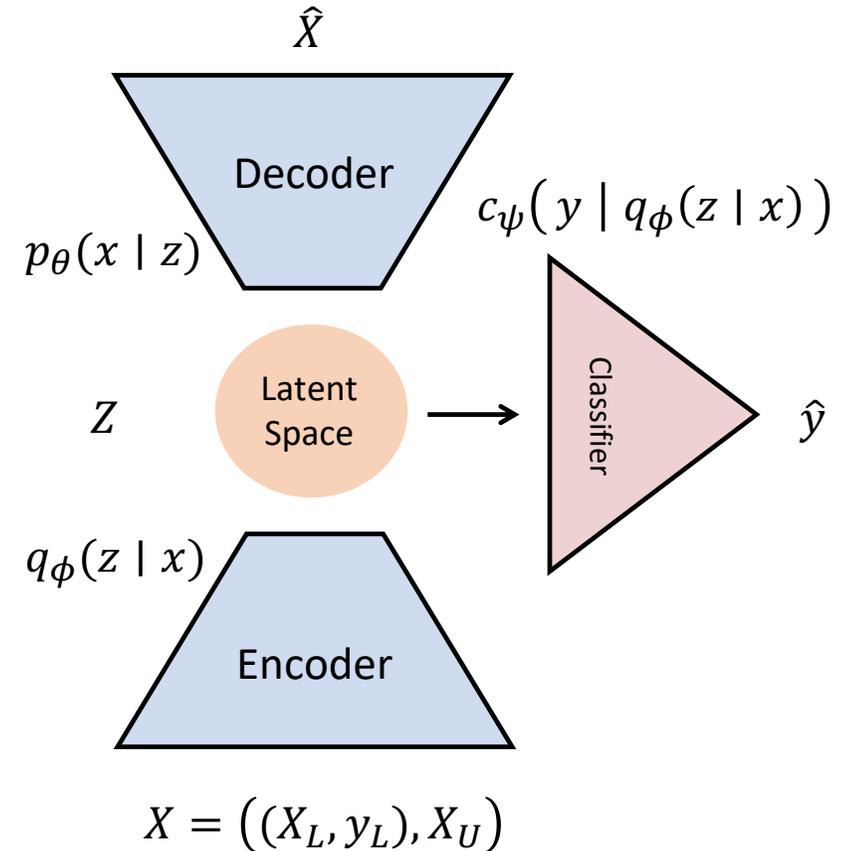
Nomenclature:

$N = N_L + N_U$: Total number of data.

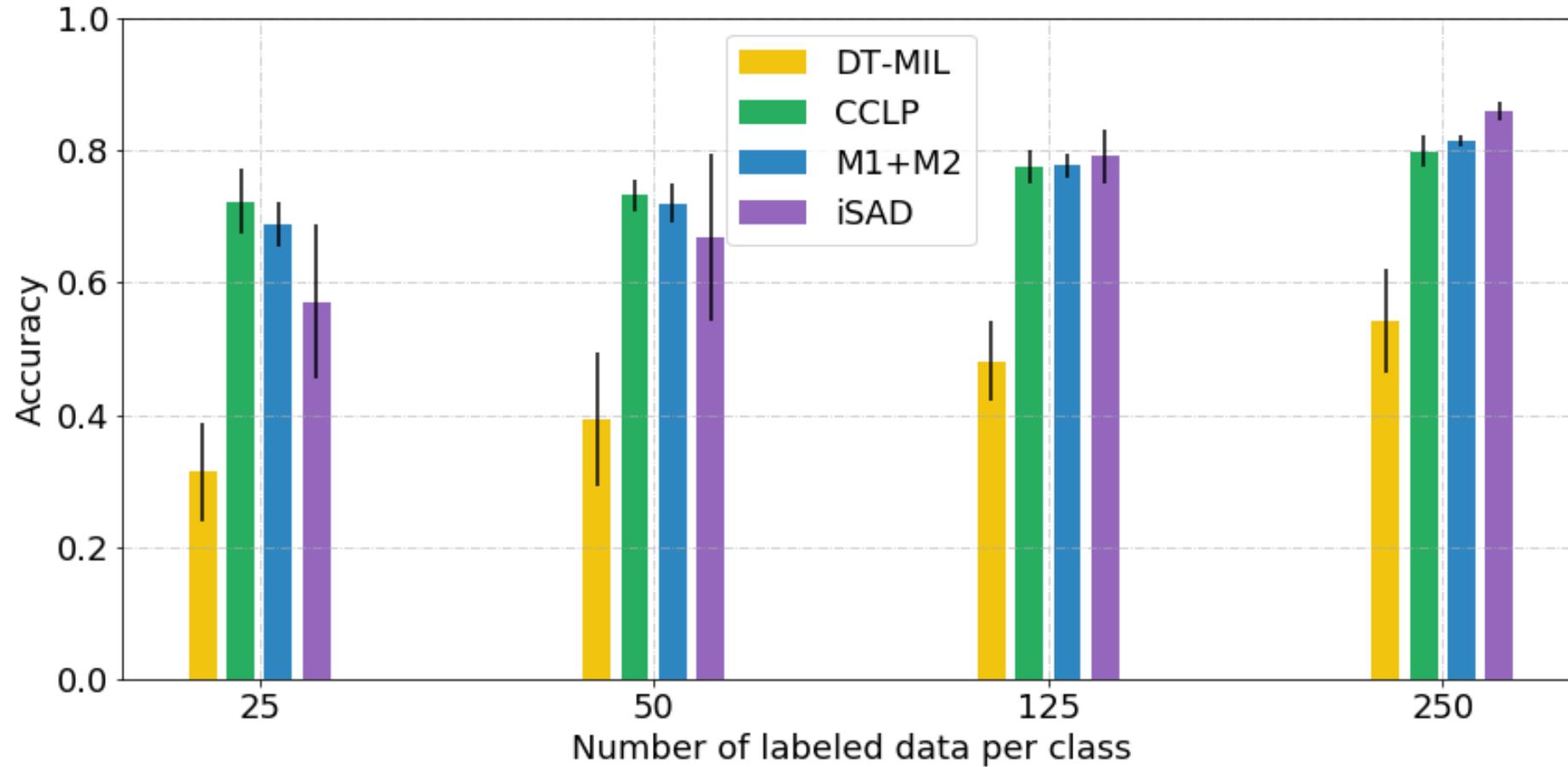
T, H : Optimal and estimated transition functions, respectively.

S : Step of the Markov chain on the graph.

w_s, w_c, w_r : Hyper-parameters tuning the weights of classification, clustering, and reconstruction losses, respectively.



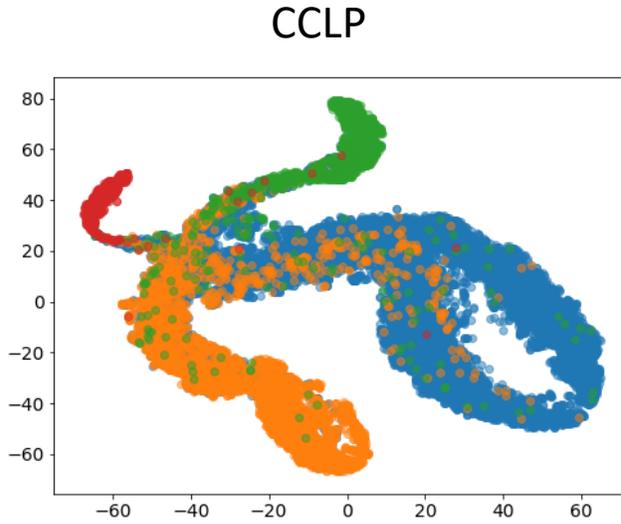
Improvement in accuracy of anomaly detection



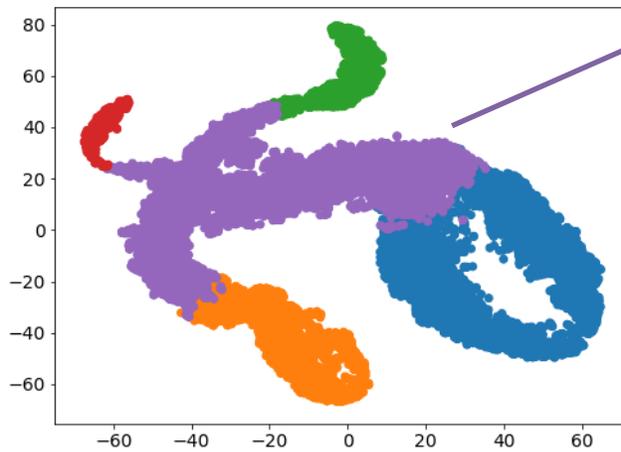
Improvement in the latent space configuration: training set



Color-coded by the actual class

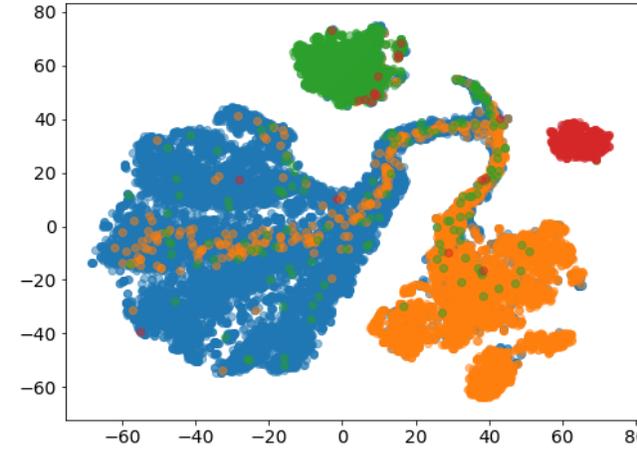


Color-coded by the formed clusters

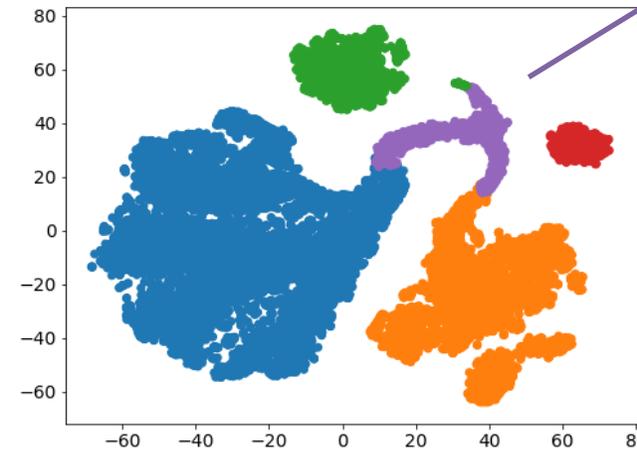


40% of data lied in the uncertain cluster

iSAD



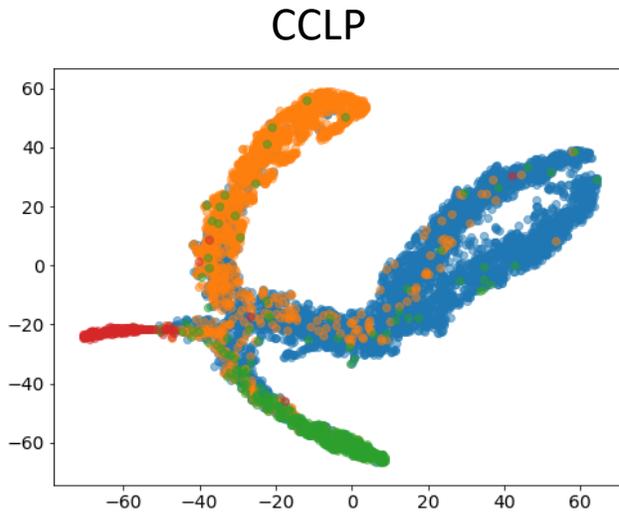
7.5% of data lied in the uncertain cluster



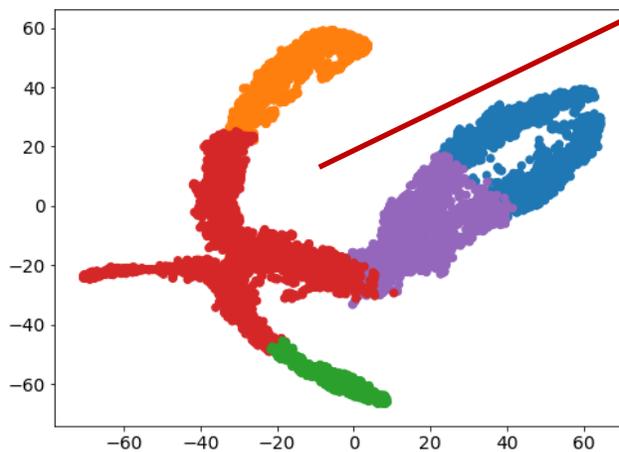
Improvement in the latent space configuration: testing set

- Nominal
- Path High
- Path High
- Flaps Late
- Uncertain

Color-coded by the actual class

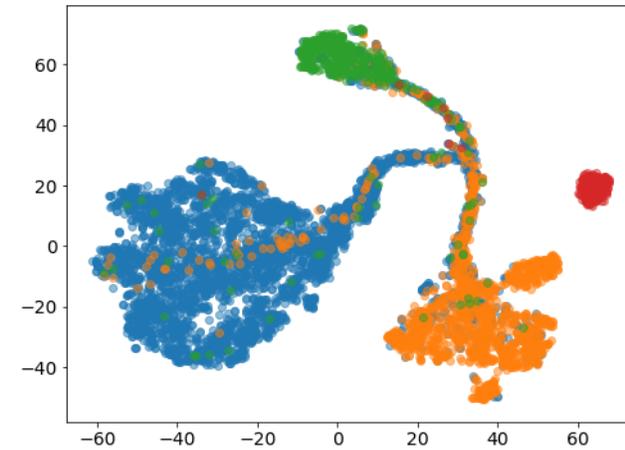


Color-coded by the formed clusters

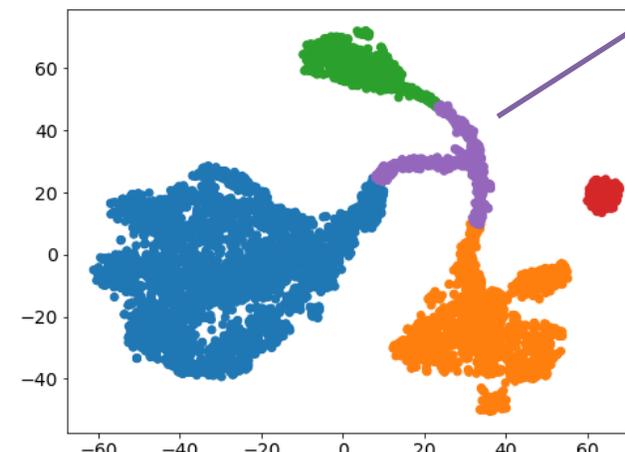


Central cluster is mixed with the minority class cluster

iSAD



8.6% of data lied in the uncertain cluster



Concluding remarks and next steps

We showed that **semi-supervised learning works**, and it is superior to supervised learning when only a limited number of labeled data is available:

- CCLP achieves 72.2% accuracy **with only 0.55% of data labeled**, while the performance of DT-MIL is 31.4%.

Concluding remarks and next steps

We showed that semi-supervised learning works, and it is superior to supervised learning when only a limited number of labeled data is available:

- CCLP achieves 72.2% accuracy with only 0.55% of data labeled, while the performance of DT-MIL is 31.4%.

Our newest model iSAD reaches **86% accuracy with 5% of data labeled**, and outperforms M1+M2 (81%), CCLP (80%) and DT-MIL (54%).

Concluding remarks and next steps

We showed that semi-supervised learning works, and it is superior to supervised learning when only a limited number of labeled data is available:

- CCLP achieves 72.2% accuracy with only 0.55% of data labeled, while the performance of DT-MIL is 31.4%.

Our newest model iSAD reaches 86% accuracy with 5% of data labeled, and outperforms M1+M2 (81%), CCLP (80%) and DT-MIL (54%).

The combination of **enforcement of compact clustering in the latent space** via graph theory and **improving the reconstruction quality** further enhanced the interpretability and explainability of the model:

- Latent space configuration opens avenues for **deploying an active learning strategy** to identify the **most informative data** for future labeling by subject matter experts.

Concluding remarks and next steps

We showed that semi-supervised learning works, and it is superior to supervised learning when only a limited number of labeled data is available:

- CCLP achieves 72.2% accuracy with only 0.55% of data labeled, while the performance of DT-MIL is 31.4%.

Our newest model iSAD reaches 86% accuracy with 5% of data labeled, and outperforms M1+M2 (81%), CCLP (80%) and DT-MIL (54%).

The combination of enforcement of compact clustering in the latent space via graph theory and improving the reconstruction quality further enhanced the interpretability and explainability of the model:

- Latent space configuration opens avenues for deploying an active learning strategy to identify the most informative data for future labeling by subject matter experts.

The reconstruction capability of the new model allows us to evaluate the **robustness to perturbations in the input space** and improve it accordingly.

Acknowledgement and references

We acknowledge funding of this research from the NASA System-Wide Safety Project under contracts 80ARC020D0010 and NNA16BD14C.

References:

Memarzadeh, M., Matthews, B., and Templin, T., “Multi-Class Anomaly Detection in Flight Data Using Semi-Supervised Explainable Deep Learning Model”, AIAA Scitech 2021 Forum, 2021, <https://doi.org/10.2514/6.2021-0774>.

Memarzadeh, M., Matthews, B., and Avrekh, I., “Unsupervised Anomaly Detection in Flight Data Using Convolutional Variational Auto-Encoder,” Aerospace, Vol. 7, 2020, p. 115.

Janakiraman, V. M., “Explaining Aviation Safety Incidents Using Deep Temporal Multiple Instance Learning,” KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 406–415.

Kingma, D., Rezende, D., Mohamed, S., and Welling, M., “Semi-supervised learning with deep generative models,” Advances in Neural Information Processing Systems (NeurIPS), 2014. URL <https://arxiv.org/abs/1406.5298>.

Kamnitsas, K., Castro, D., Le-Folgoc, L., Walker, I., Tanno, R., Rueckert, D., Glocker, B., Criminisi, A., and Nori, A., “Semi-supervised learning via compact latent space clustering,” Proceedings of the 35th International Conference on Machine Learning (ICML), 2018, pp. 2459–2468. URL <https://arxiv.org/abs/1406.5298>.