INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE

Anomaly detection on flight simulator

Supervisors: Joël POTIN - SII Sud-Ouest
Florian MONTELS - SII Sud-Ouest
Students: Phuc Luan Nguyen
Thibault Lauilhé
Department: Mathematical Modeling and AI
Academic year: 2024 - 2025

January 28, 2025

Introduction

Develop an AI model able to detect anomalies in a realistic flight simulator.

Outline

I. Exploring Literature to choose the model

II. Dataset preprocessing

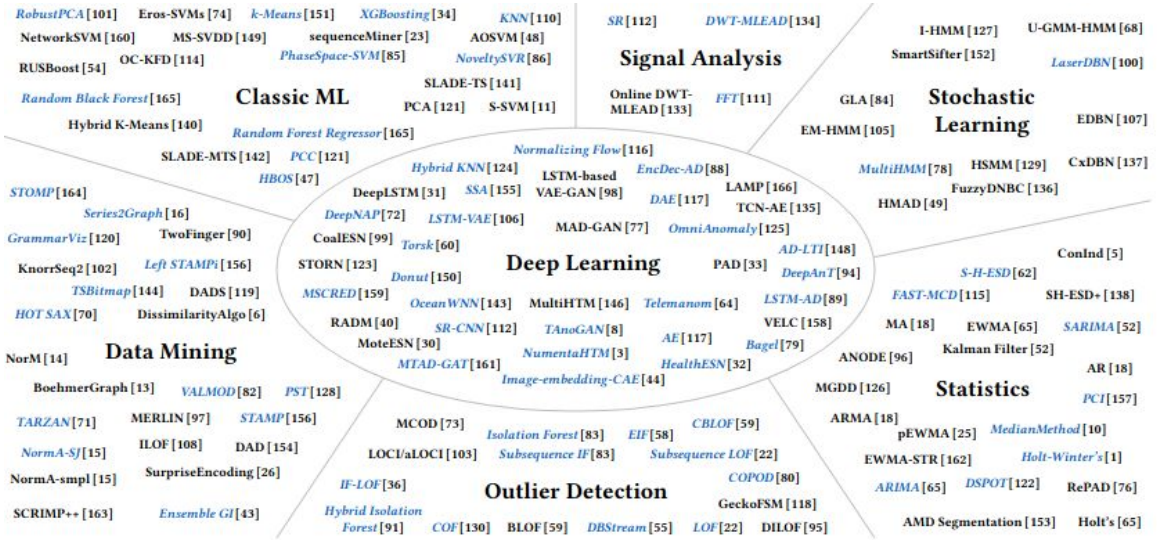
III. Configuration and training of the models

IV. Results and comparison the models

V. Conclusion and perspectives

I. Exploring Literature to choose the best model

Model families for anomaly detection



Deep Learning

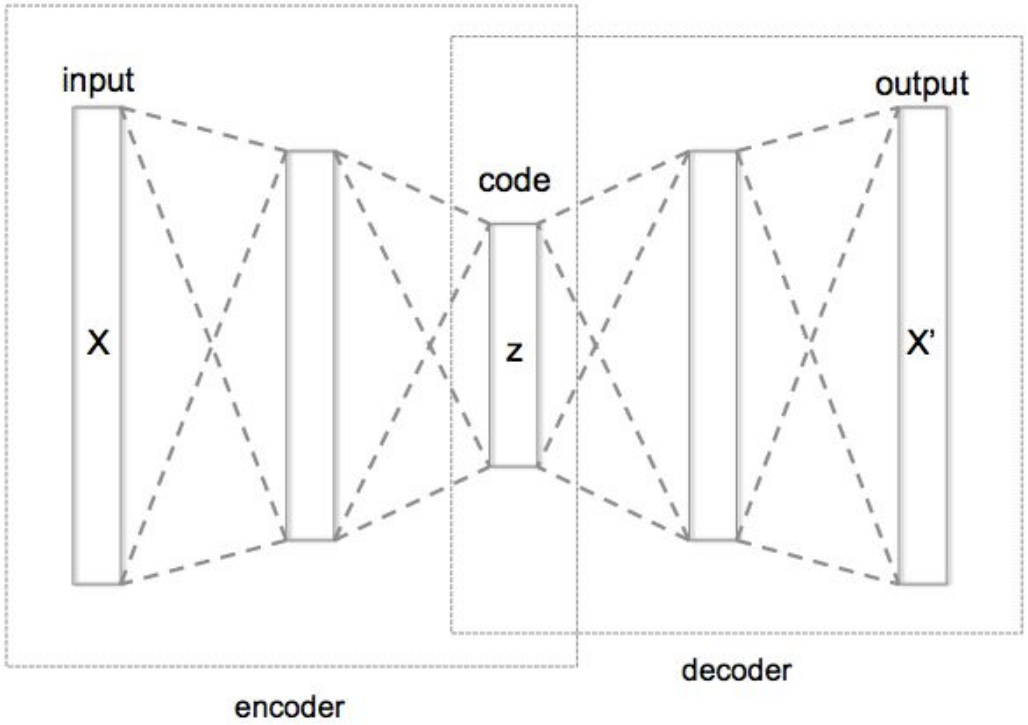
Machine Learning

Outlier detection

Data mining

Signal Analysis

What is an Auto Encoder ?



LSTM

- about long sequences
- captures dependencies between parameters

UNet-LSTM

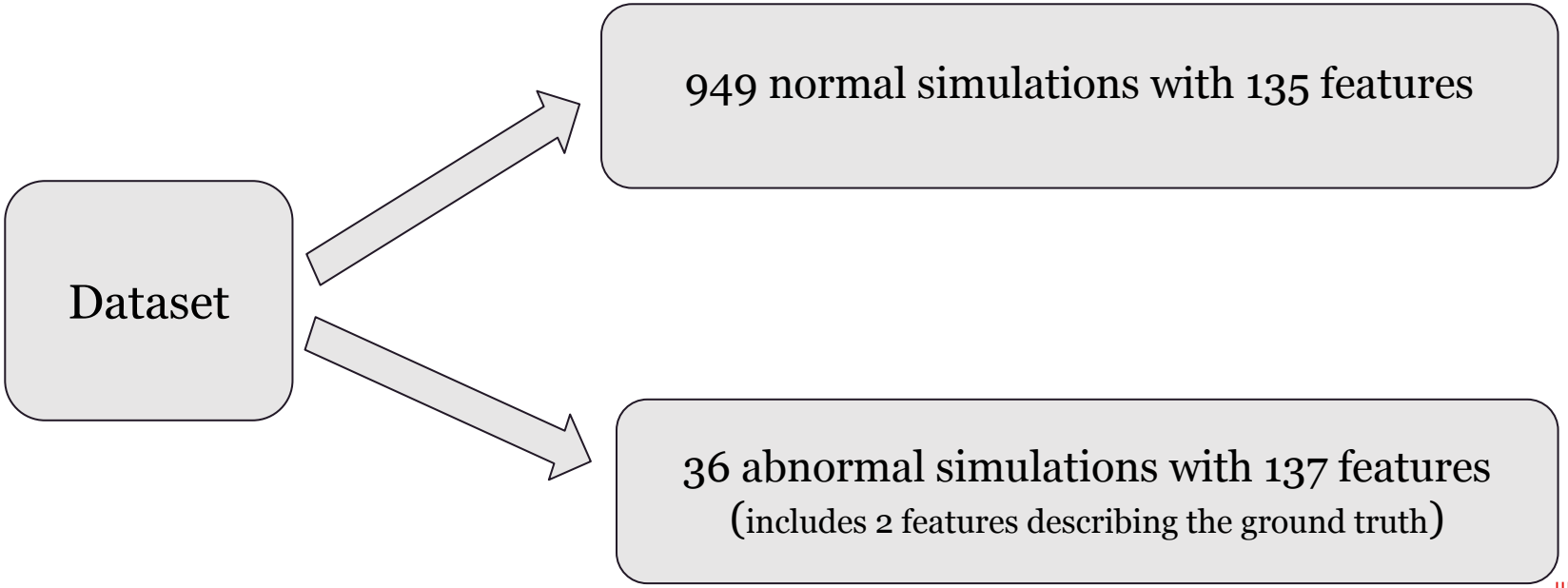
- combined with UNet
- skip connections to guide reconstruction

TRANSFORMERS

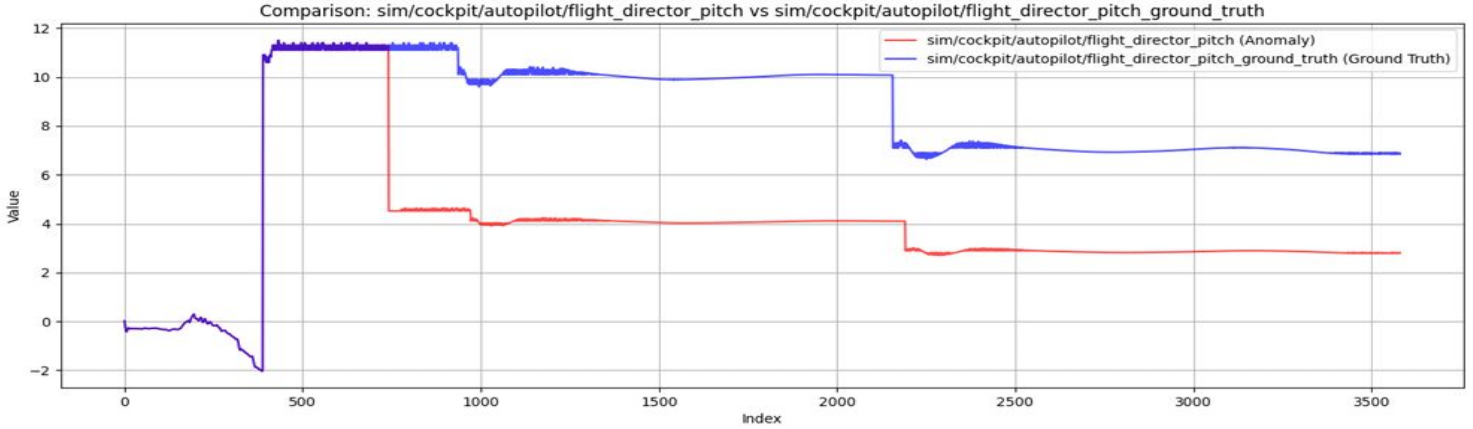
- about long sequences
- fast training

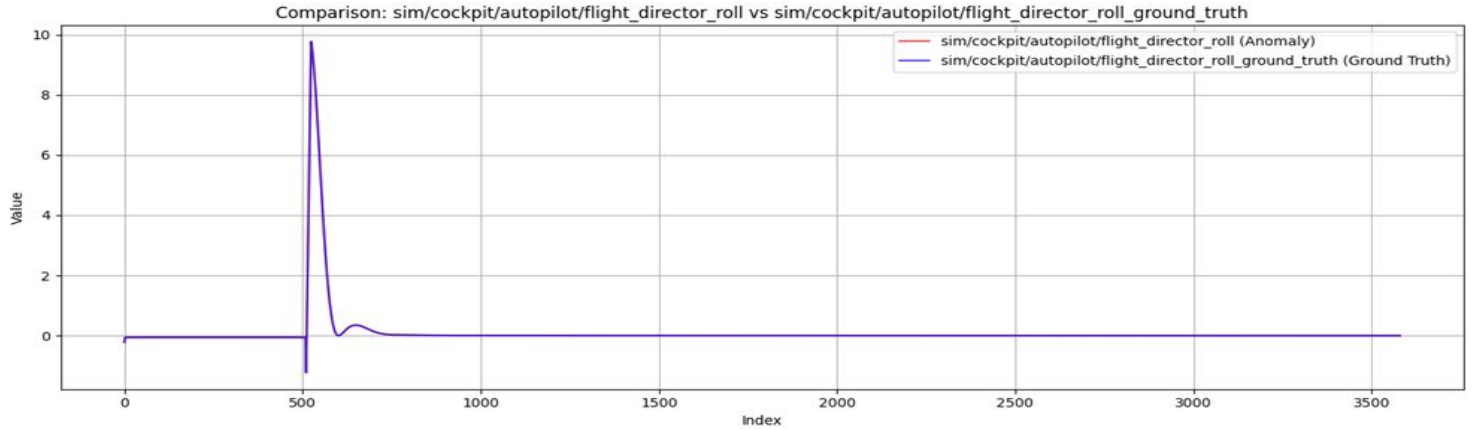
II. Dataset preprocessing

Dataset


Dataset
949 normal simulations with 135 features
36 abnormal simulations with 137 features
(includes 2 features describing the ground truth)

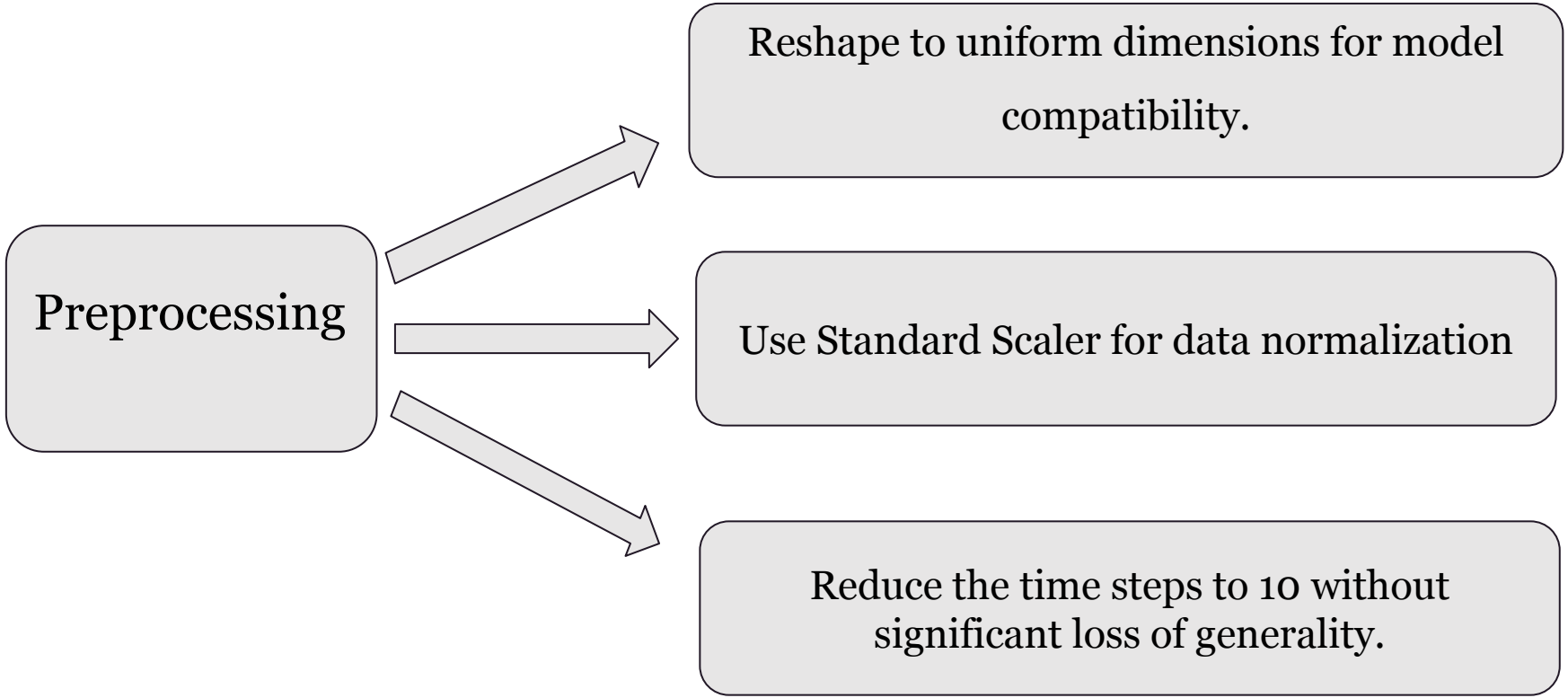
Dataset



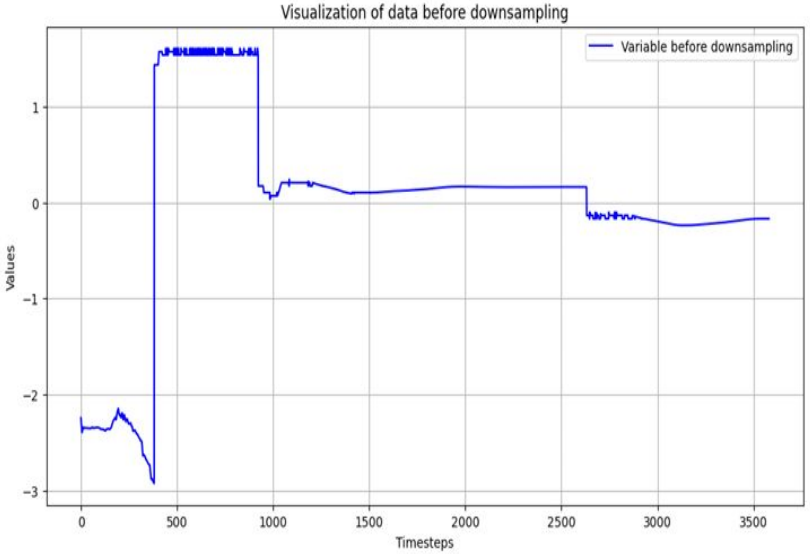


The difference between ground truth variables and anomalous variables

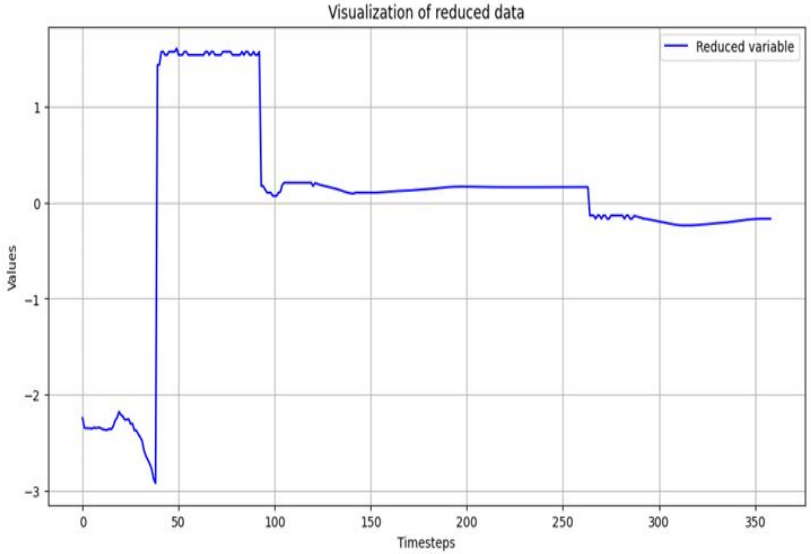
Dataset preprocessing



Dataset preprocessing

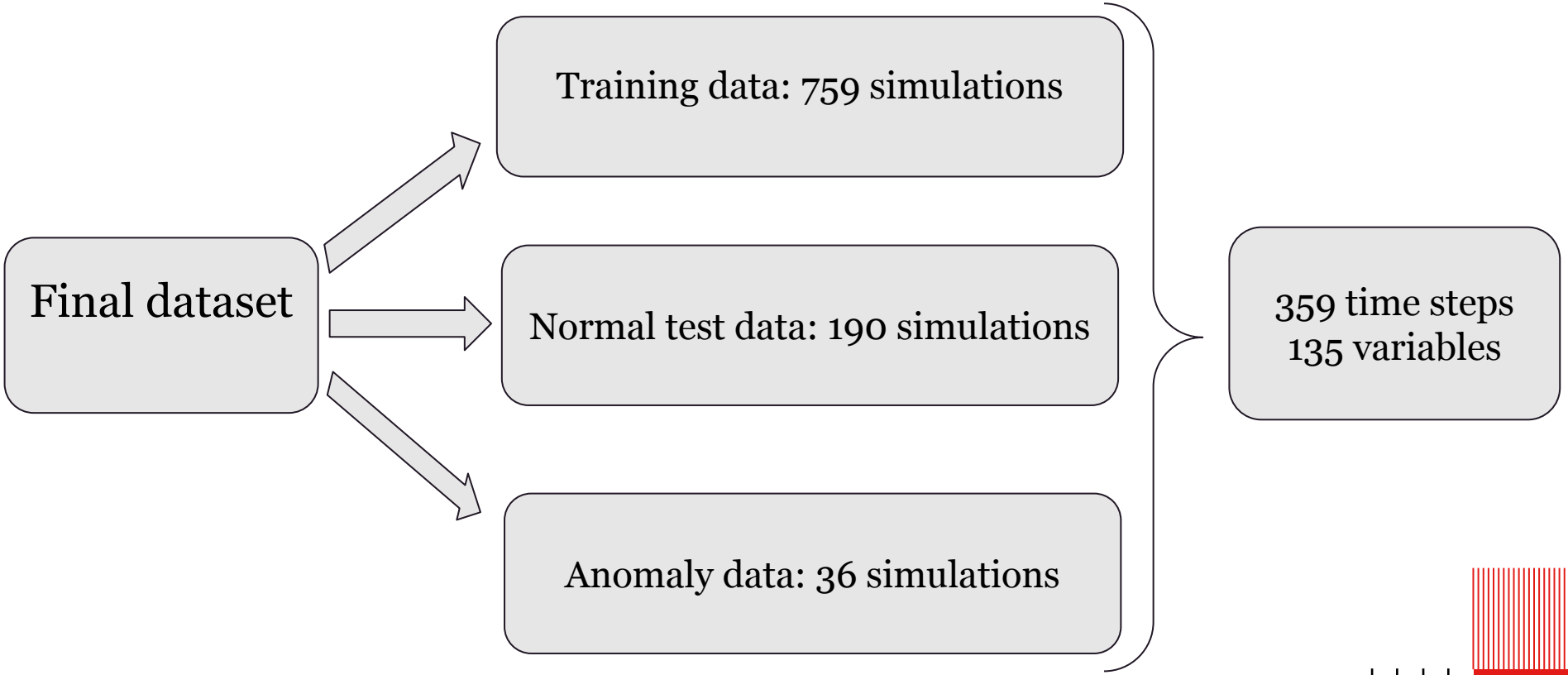


(a) Before downsampling



(b) After downsampling

Dataset preprocessing



III. Configuration and training of the models

LSTM Autoencoder

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 359, 135)]	0
lstm (LSTM)	(None, 359, 512)	1327104
lstm_1 (LSTM)	(None, 256)	787456
dense (Dense)	(None, 128)	32896
repeat_vector (RepeatVector)	(None, 359, 128)	0
lstm_2 (LSTM)	(None, 359, 256)	394240
lstm_3 (LSTM)	(None, 359, 512)	1574912
time_distributed (TimeDistributed)	(None, 359, 135)	69255

Total params: 4,185,863
Trainable params: 4,185,863
Non-trainable params: 0

⟹ Input layer

⟹ Encoder

⟹ Bottleneck (latent space)

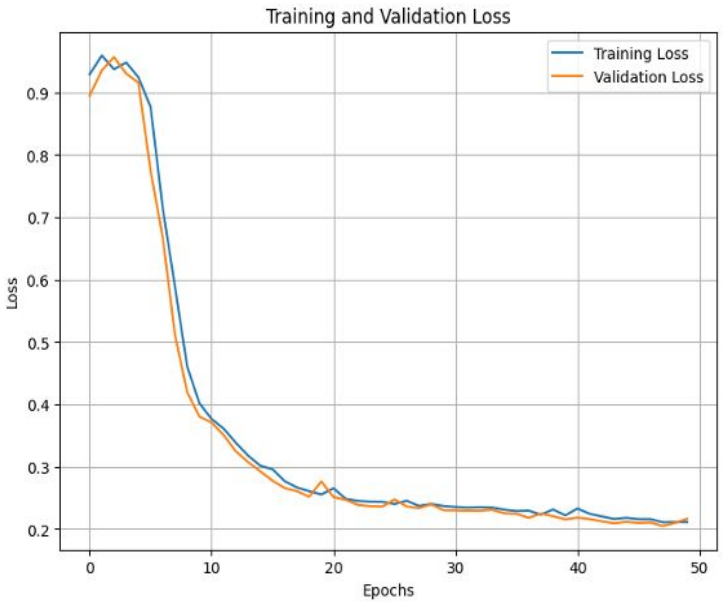
⟹ Decoder

⟹ Output layer

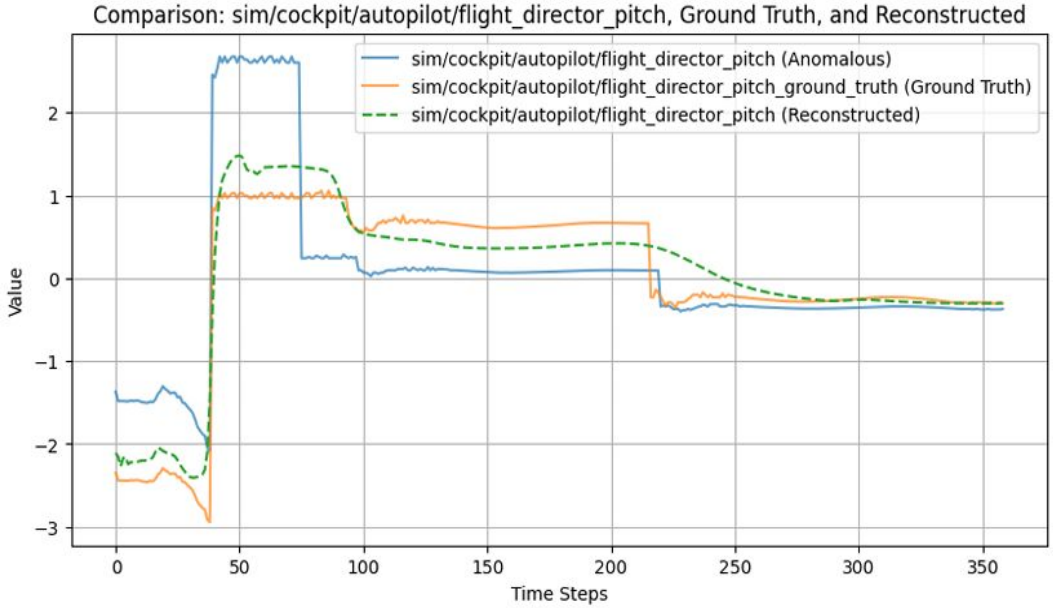
- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam with a learning rate of 0.001

Autoencoder model

LSTM Autoencoder

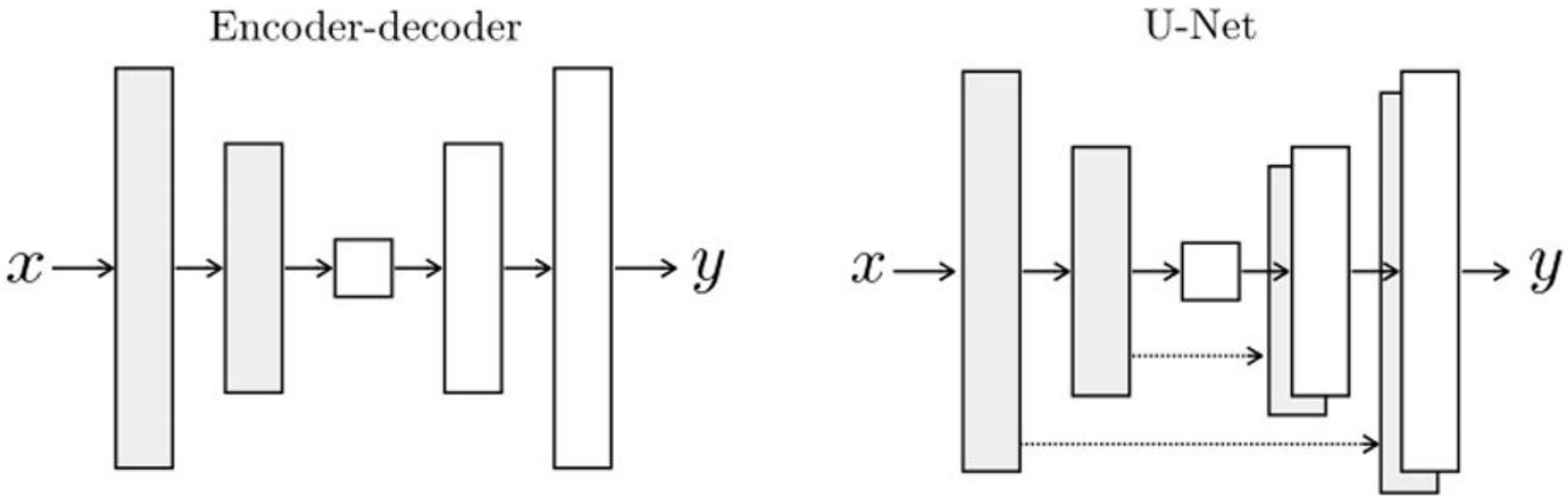


(a) Training and validation loss



(b) Compare the reconstructed data with other datasets.

LSTM UNet Autoencoder



Autoencoder model and Unet Autoencoder model

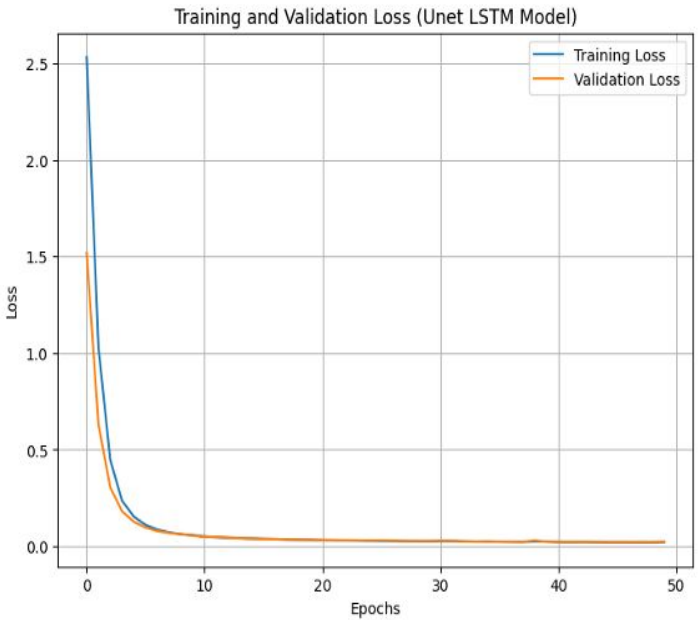
LSTM UNet Autoencoder

Model: "model_1"

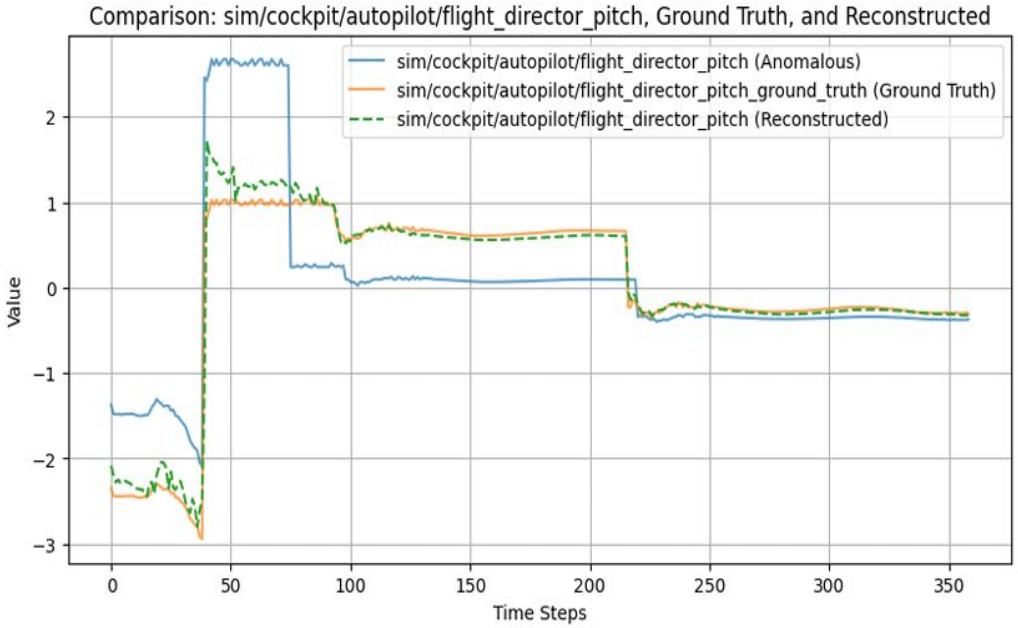
Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 359, 135)]	0	[]
lstm_4 (LSTM)	(None, 359, 512)	1327104	['input_2[0][0]']
lstm_5 (LSTM)	(None, 359, 256)	787456	['lstm_4[0][0]']
lstm_6 (LSTM)	(None, 128)	197120	['lstm_5[0][0]']
dense_2 (Dense)	(None, 64)	8256	['lstm_6[0][0]']
repeat_vector_1 (RepeatVector)	(None, 359, 64)	0	['dense_2[0][0]']
lstm_7 (LSTM)	(None, 359, 128)	98816	['repeat_vector_1[0][0]']
concatenate (Concatenate)	(None, 359, 384)	0	['lstm_7[0][0]', 'lstm_5[0][0]']
lstm_8 (LSTM)	(None, 359, 256)	656384	['concatenate[0][0]']
concatenate_1 (Concatenate)	(None, 359, 768)	0	['lstm_8[0][0]', 'lstm_4[0][0]']
lstm_9 (LSTM)	(None, 359, 512)	2623488	['concatenate_1[0][0]']
time_distributed_1 (TimeDistributed)	(None, 359, 135)	69255	['lstm_9[0][0]']

Unet Autoencoder model

LSTM UNet Autoencoder



(a) Training and validation loss



(b) Compare the reconstructed data with other datasets.

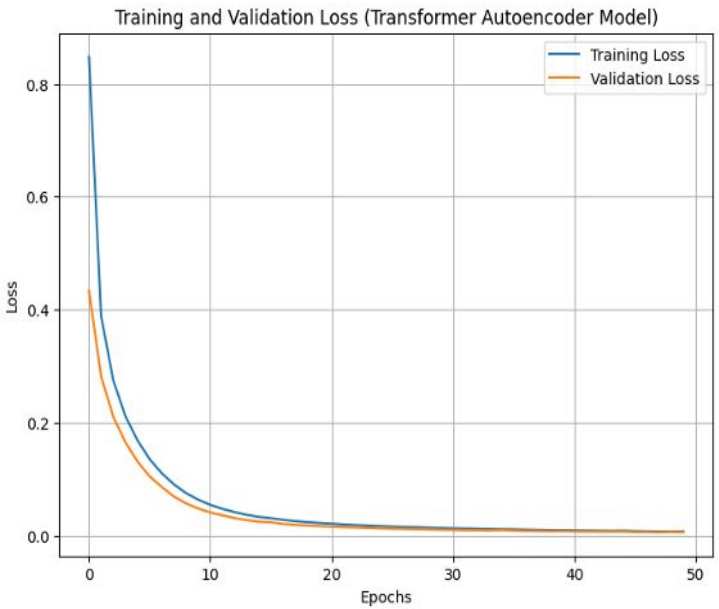
Transformer Autoencoder

Layer (type)	Output Shape	Param #	Connected to
input_4 (InputLayer)	[(None, 359, 135)]	0	[]
multi_head_attention_2 (MultiHeadAttention)	(None, 359, 135)	293355	['input_4[0][0]', 'input_4[0][0]']
dropout_4 (Dropout)	(None, 359, 135)	0	['multi_head_attention_2[0][0]']
add_4 (Add)	(None, 359, 135)	0	['input_4[0][0]', 'dropout_4[0][0]']
layer_normalization_4 (LayerNormalization)	(None, 359, 135)	270	['add_4[0][0]']
dense_11 (Dense)	(None, 359, 128)	17408	['layer_normalization_4[0][0]']
dense_12 (Dense)	(None, 359, 135)	17415	['dense_11[0][0]']
dropout_5 (Dropout)	(None, 359, 135)	0	['dense_12[0][0]']
add_5 (Add)	(None, 359, 135)	0	['layer_normalization_4[0][0]', 'dropout_5[0][0]']
layer_normalization_5 (LayerNormalization)	(None, 359, 135)	270	['add_5[0][0]']
dense_13 (Dense)	(None, 359, 64)	8704	['layer_normalization_5[0][0]']
dense_14 (Dense)	(None, 359, 135)	8775	['dense_13[0][0]']
multi_head_attention_3 (MultiHeadAttention)	(None, 359, 135)	293355	['dense_14[0][0]', 'layer_normalization_5[0][0]']
dropout_6 (Dropout)	(None, 359, 135)	0	['multi_head_attention_3[0][0]']
add_6 (Add)	(None, 359, 135)	0	['dense_14[0][0]', 'dropout_6[0][0]']
layer_normalization_6 (LayerNormalization)	(None, 359, 135)	270	['add_6[0][0]']
dense_15 (Dense)	(None, 359, 128)	17408	['layer_normalization_6[0][0]']
dense_16 (Dense)	(None, 359, 135)	17415	['dense_15[0][0]']
dropout_7 (Dropout)	(None, 359, 135)	0	['dense_16[0][0]']
add_7 (Add)	(None, 359, 135)	0	['layer_normalization_6[0][0]', 'dropout_7[0][0]']
layer_normalization_7 (LayerNormalization)	(None, 359, 135)	270	['add_7[0][0]']
dense_17 (Dense)	(None, 359, 135)	18360	['layer_normalization_7[0][0]']

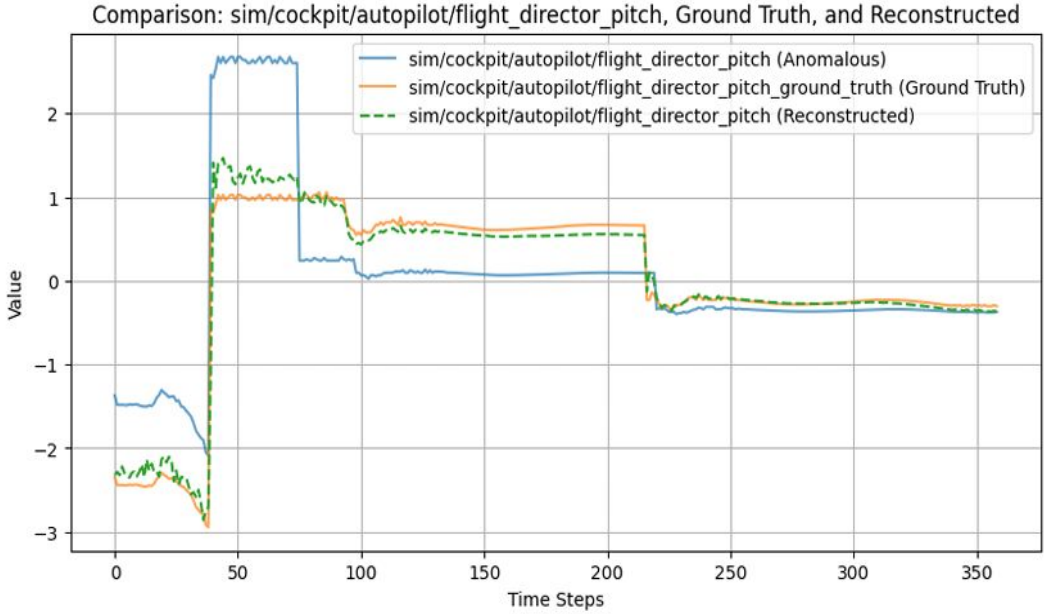
Total params: 693,275
Trainable params: 693,275
Non-trainable params: 0

Transformer Autoencoder model

Transformer Autoencoder



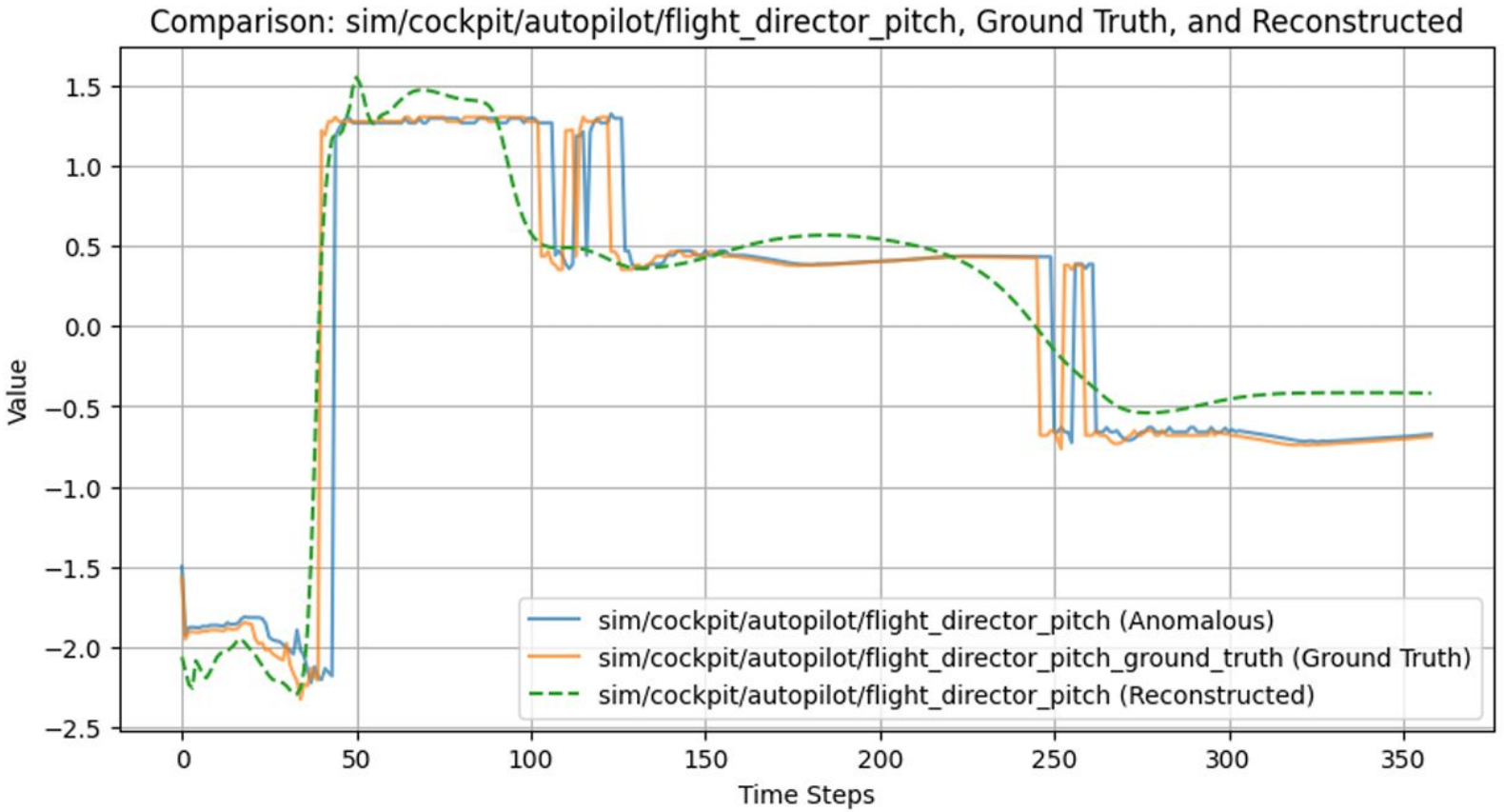
(a) Training and validation loss



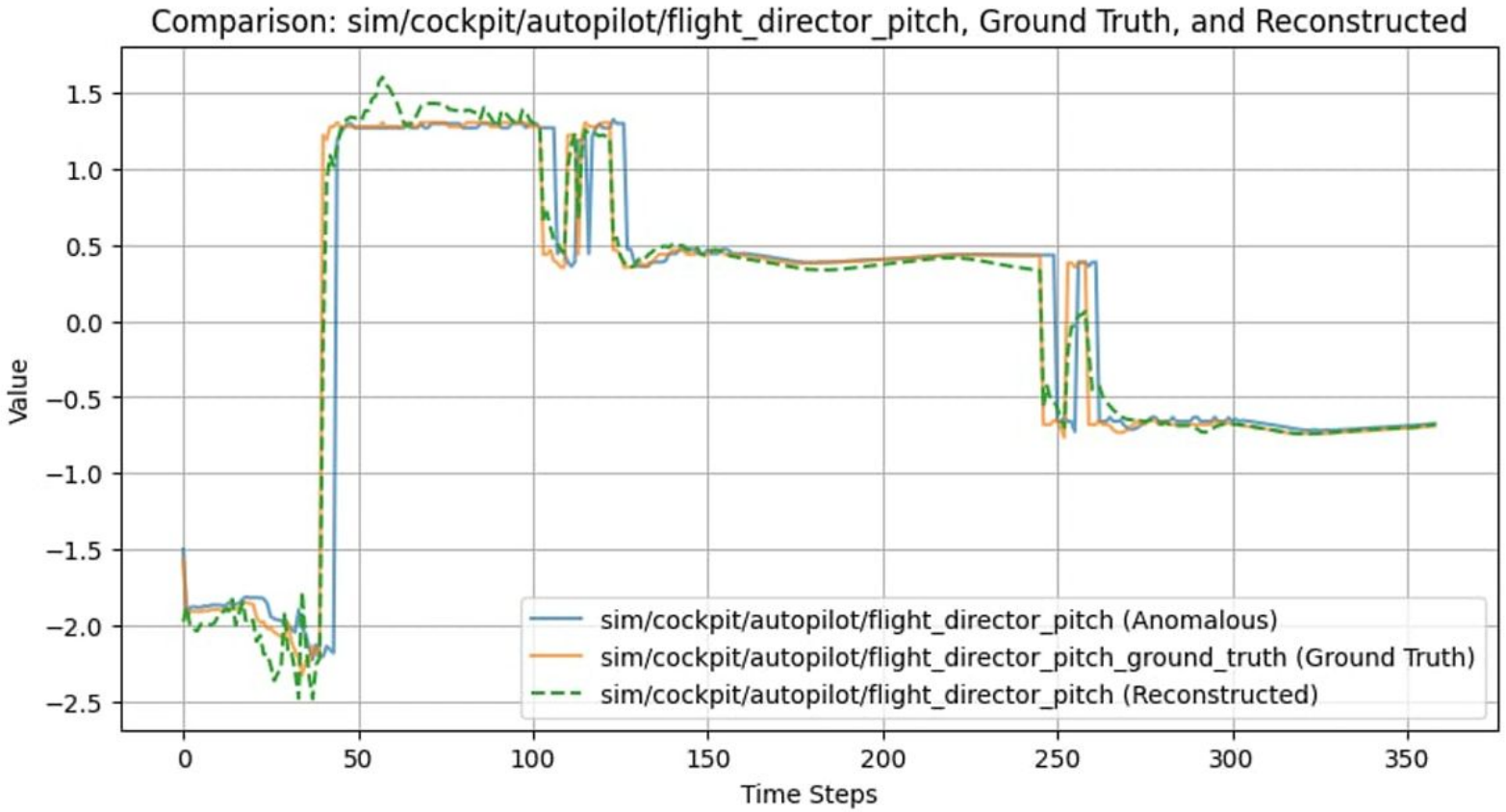
(b) Compare the reconstructed data with other datasets.

IV. Results and comparison the models

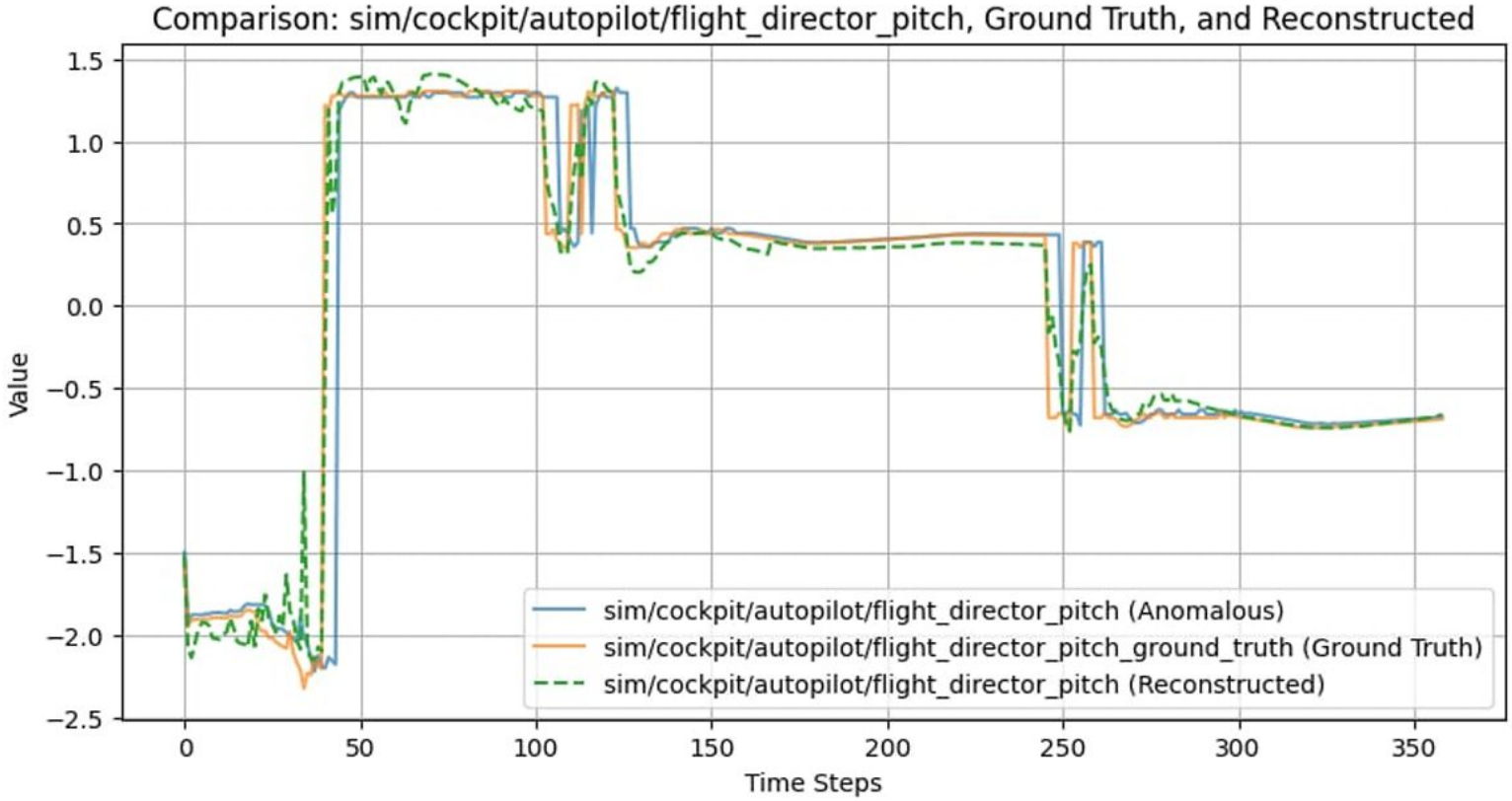
Results for LSTM Autoencoder



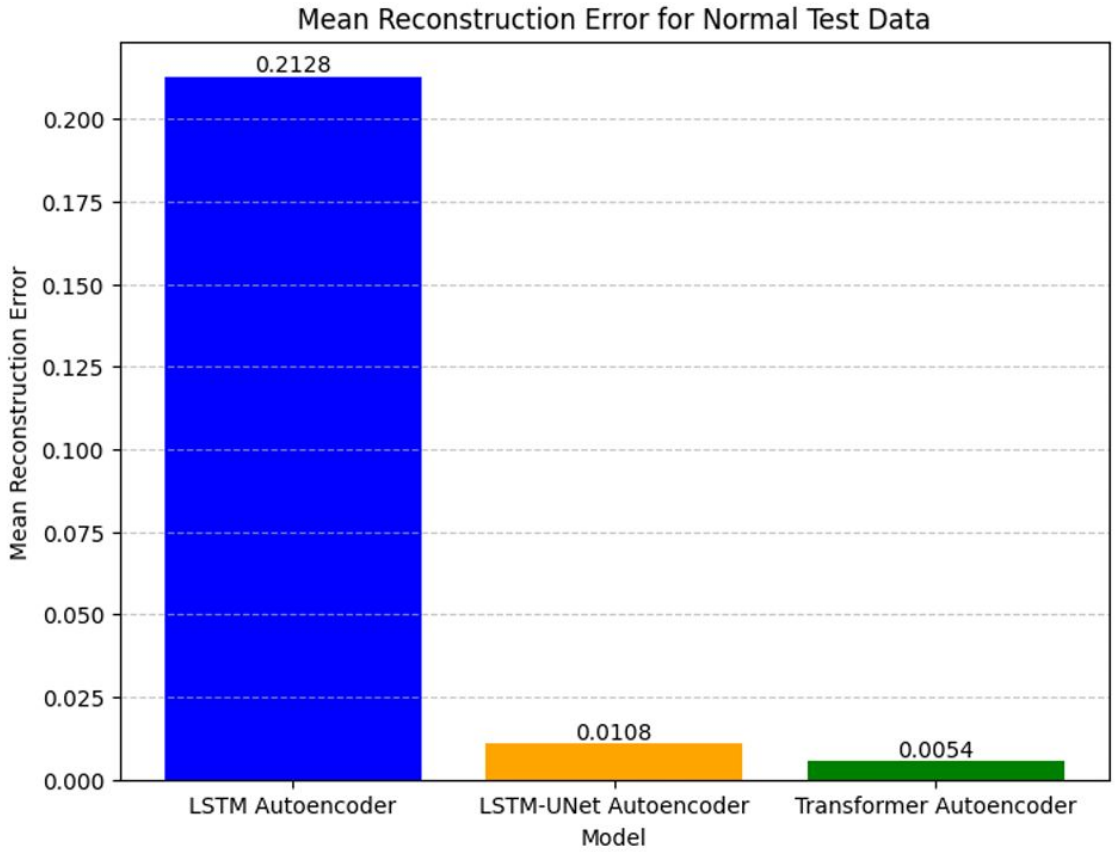
Results for LSTM-UNet Autoencoder



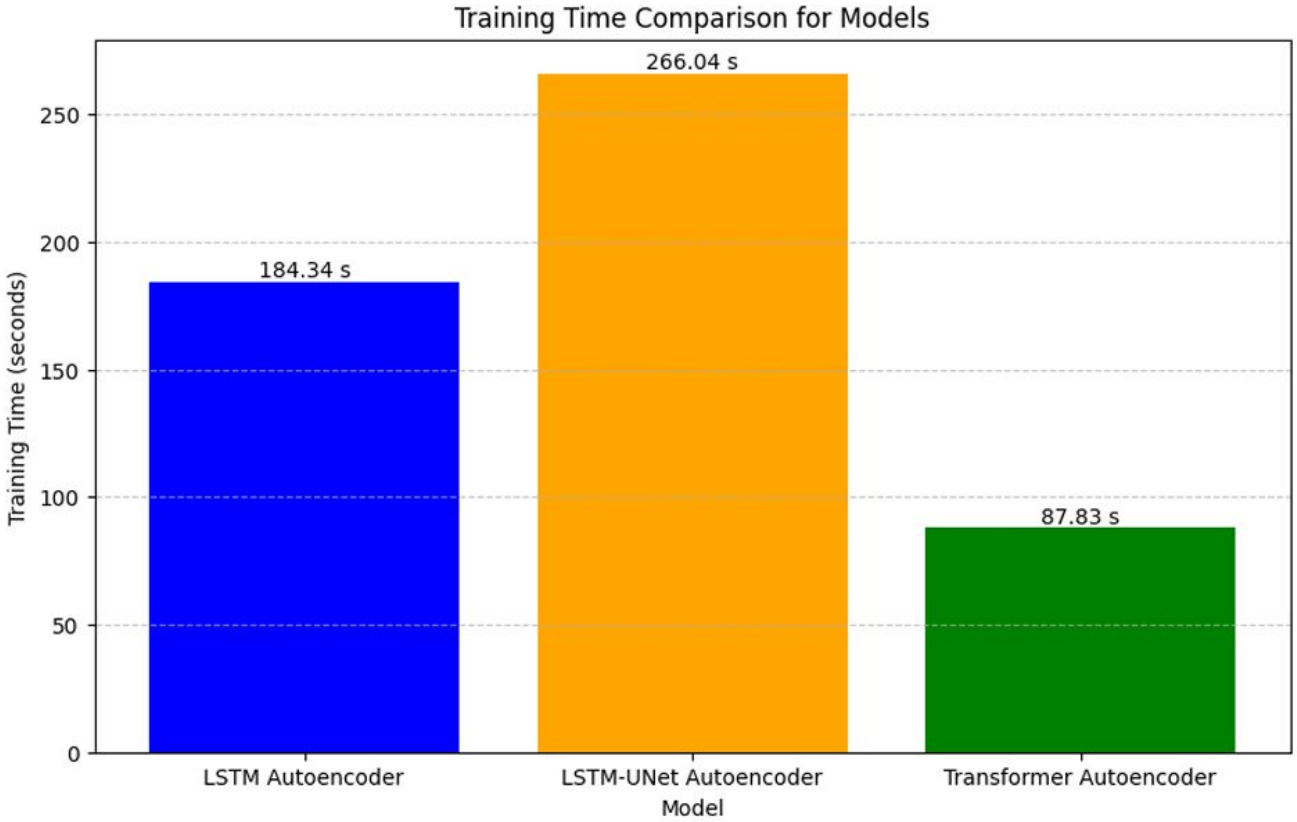
Results for Transformers



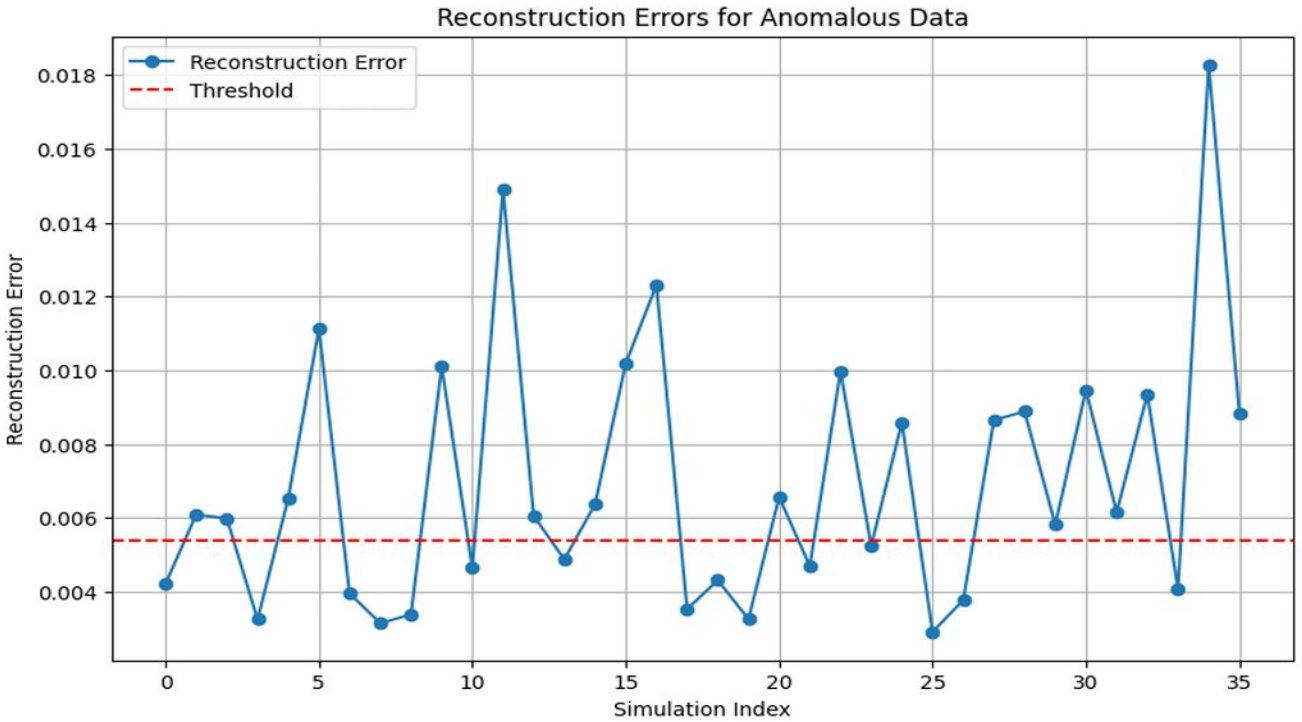
Comparison of Average MSE on Normal Simulation Test Data



Comparison of Training Time



Detecting Simulation Anomalies



← 80th percentile

21 out of 36 anomalous simulations being detected

V. Conclusion and perspectives

Conclusion

✅ Analyzed the dataset and prepared it for model development.

✅ Successfully reconstructed data from anomalous simulations to closely resemble the ground truth values.

⇨ Determine how to use the reconstructed data to identify whether a given simulation is an anomaly

Perspectives

Use the ROC curve to determine a more suitable threshold

Evaluate the average error on the two anomalous variables and the remaining 133 variables.

THANK YOU FOR LISTENING !