MATLAB EXPO

理解和验证AI模型

马文辉，Principal Application Engineer, MathWorks China

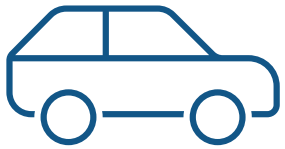
MathWorks®

AI技术在生产中的应用

- AI在很多任务中取得了瞩目的效果
- 越来越多的关键领域应用AI模型
- AI模型解释、验证、以及识别偏差的测试越来越受到关注

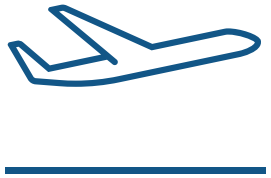
各个行业在验证AI模型方面正在取得进展

发布白皮书、标准和规划



Automotive

New WIP ISO PAS 8800 (Road Vehicles — Safety and artificial intelligence)



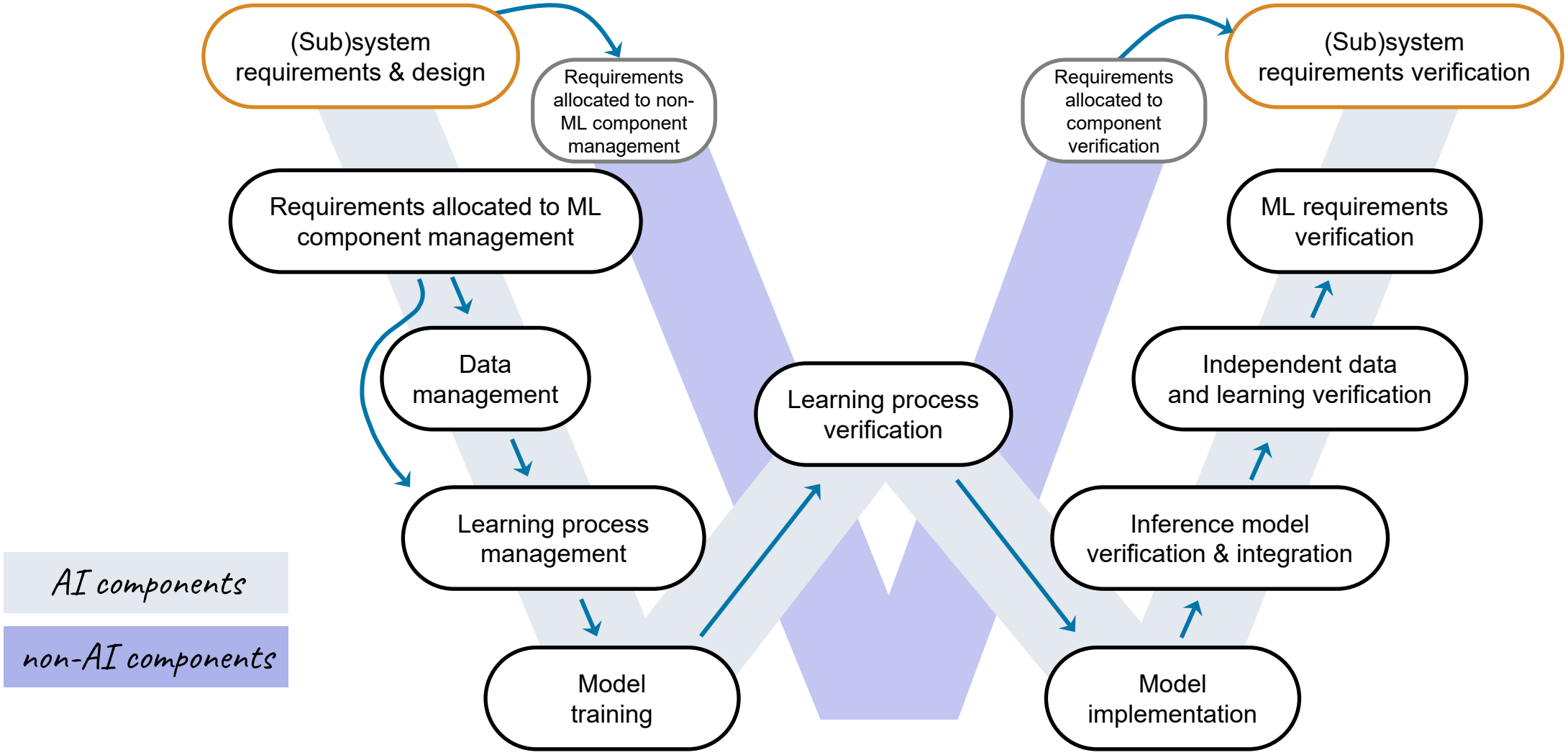
Aerospace

New standard (AS6983) from EUROCAE WG-114 / SAE G-34 is expected in 2024

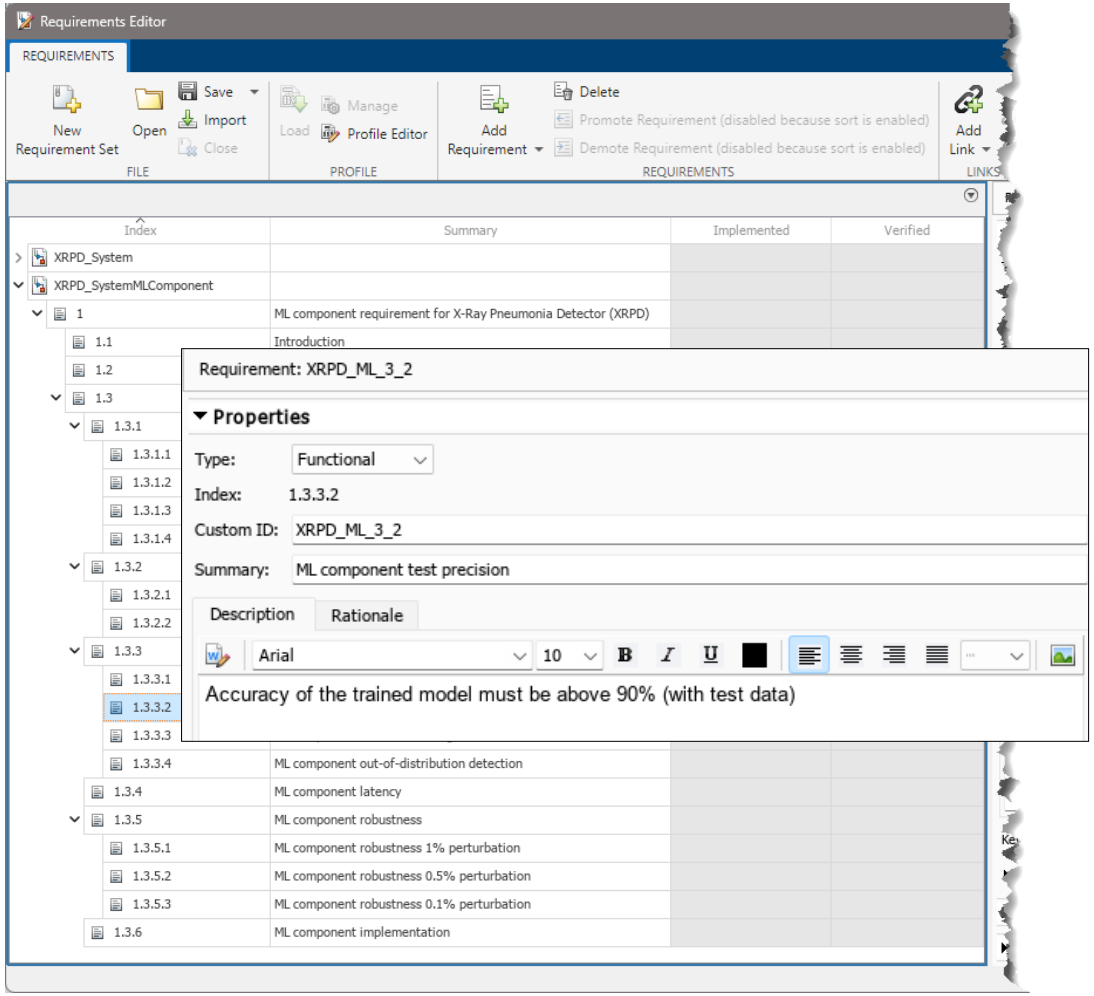
Medical Devices

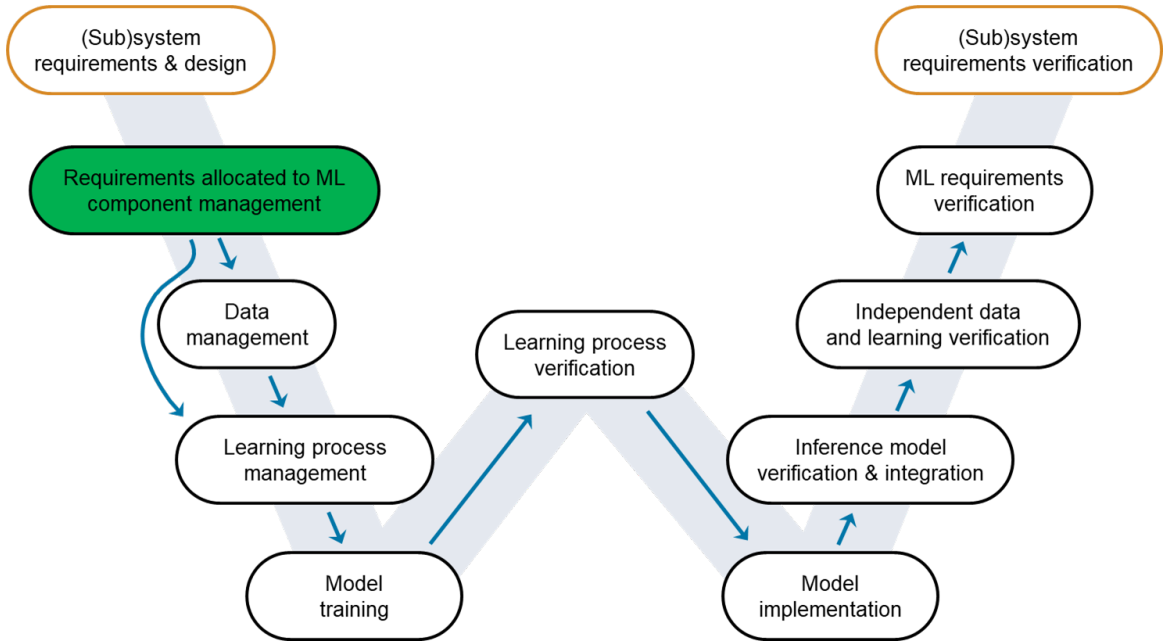
FDA released its first AI/ML-Based Software as a Medical Device (SaMD) Action Plan

W-shape: 经典V型方法应用于AI项目开发



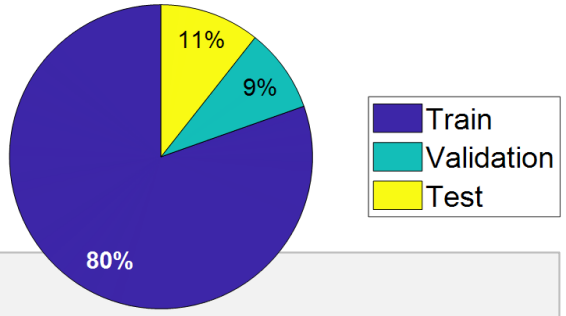
MATLAB对W-Shape方法的支持 - 需求定义与管理



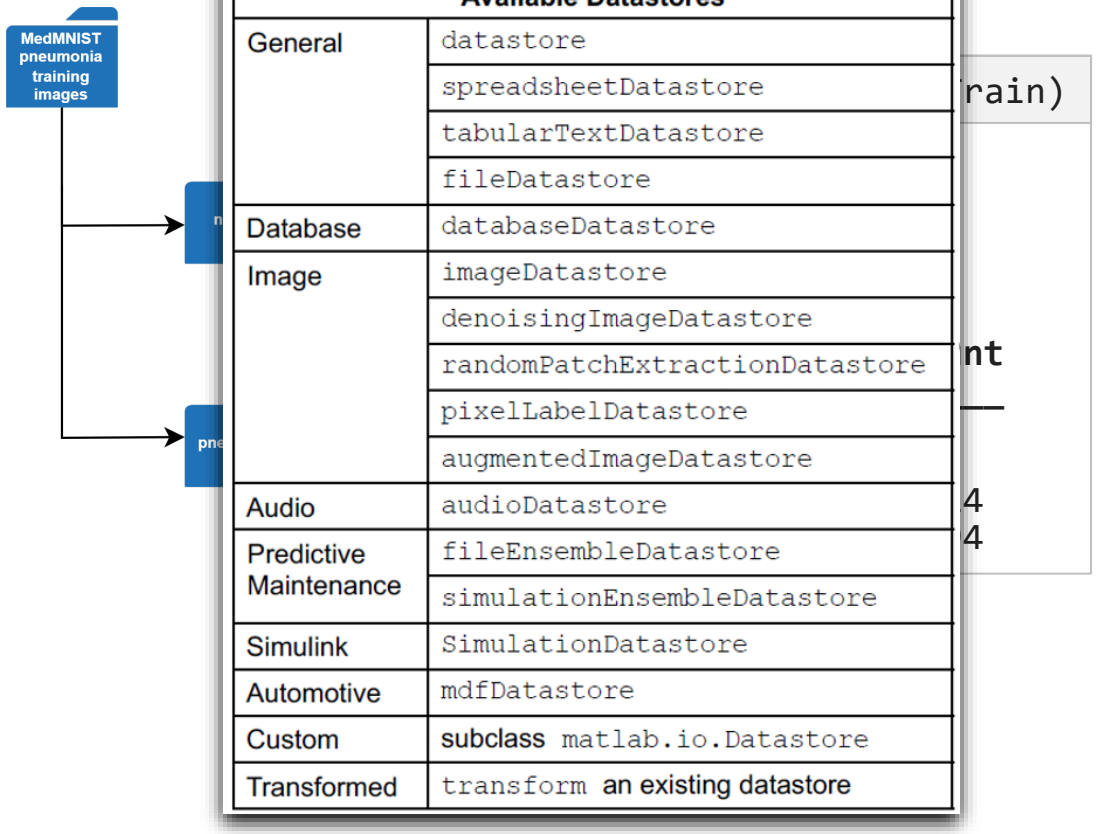


[Requirement Toolbox]

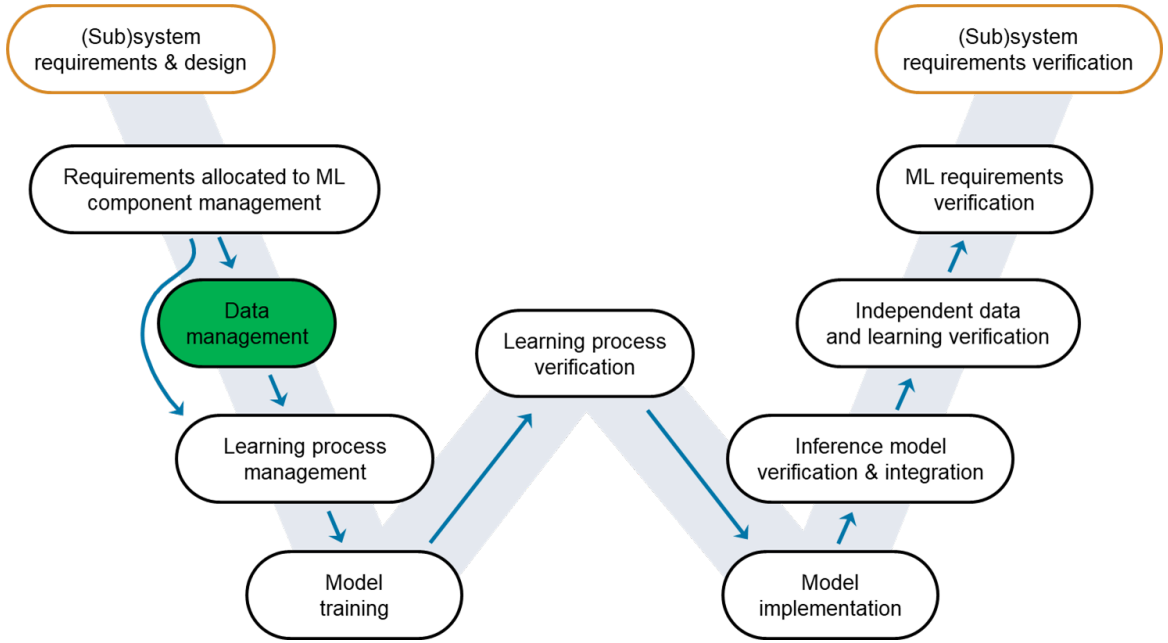
大数据管理



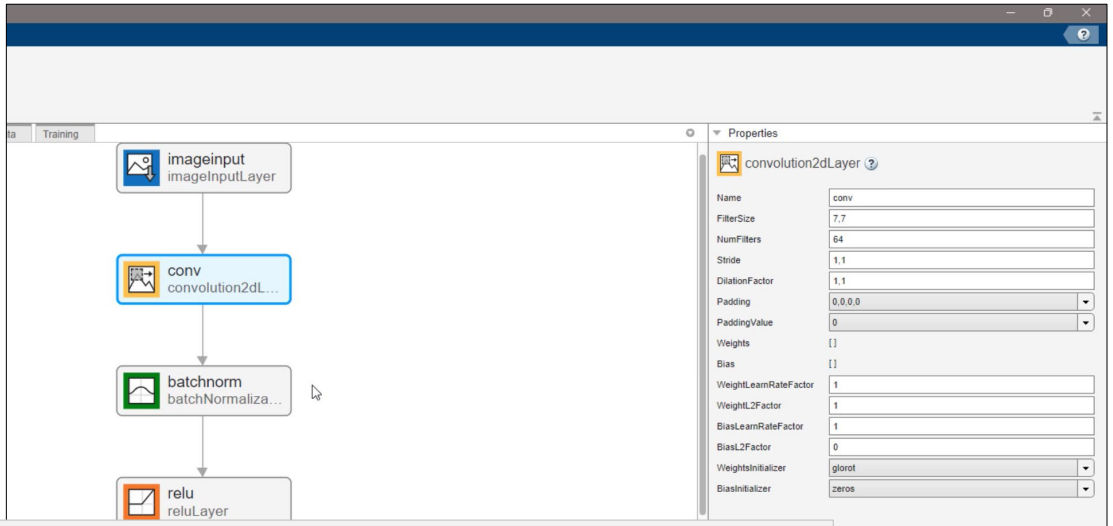

```
trainingDataFolder = "pneumoniamnist\Train";
imdsTrain = imageDatastore(trainingDataFolder,IncludeSubfolders=true,LabelSource="foldernames");
```



Available Datastores	
General	datastore
	spreadsheetDatastore
	tabularTextDatastore
	fileDatastore
Database	databaseDatastore
Image	imageDatastore
	denoisingImageDatastore
	randomPatchExtractionDatastore
	pixelLabelDatastore
	augmentedImageDatastore
Audio	audioDatastore
Predictive Maintenance	fileEnsembleDatastore
	simulationEnsembleDatastore
Simulink	SimulationDatastore
Automotive	mdfDatastore
Custom	subclass matlab.io.Datastore
Transformed	transform an existing datastore

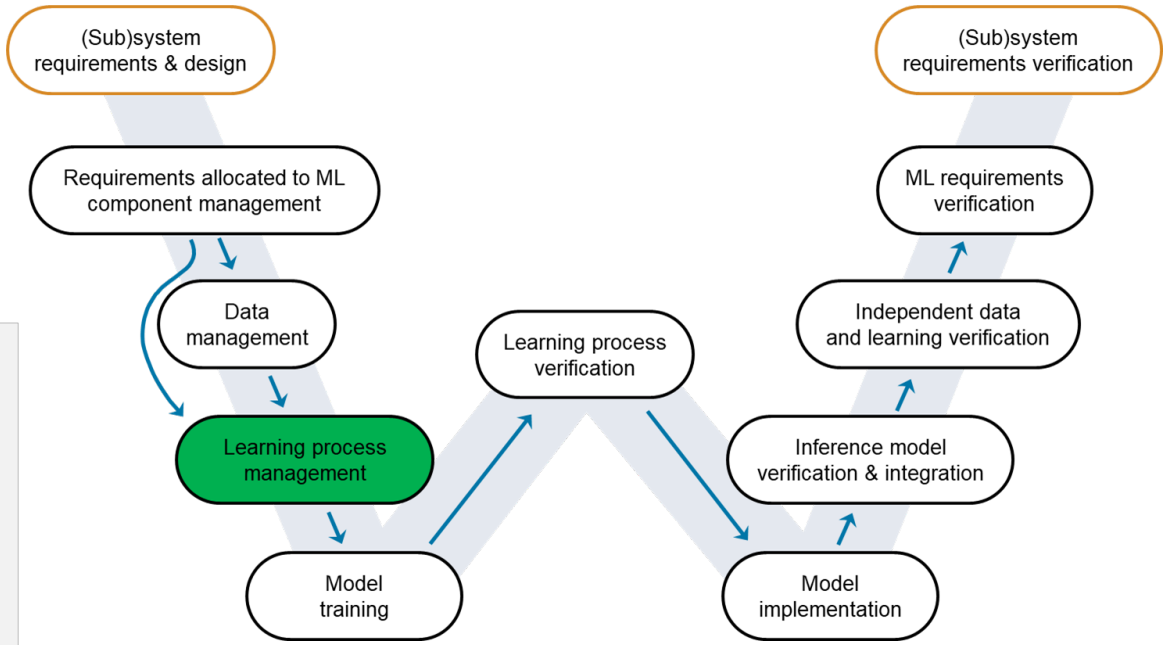


可视化模型创建

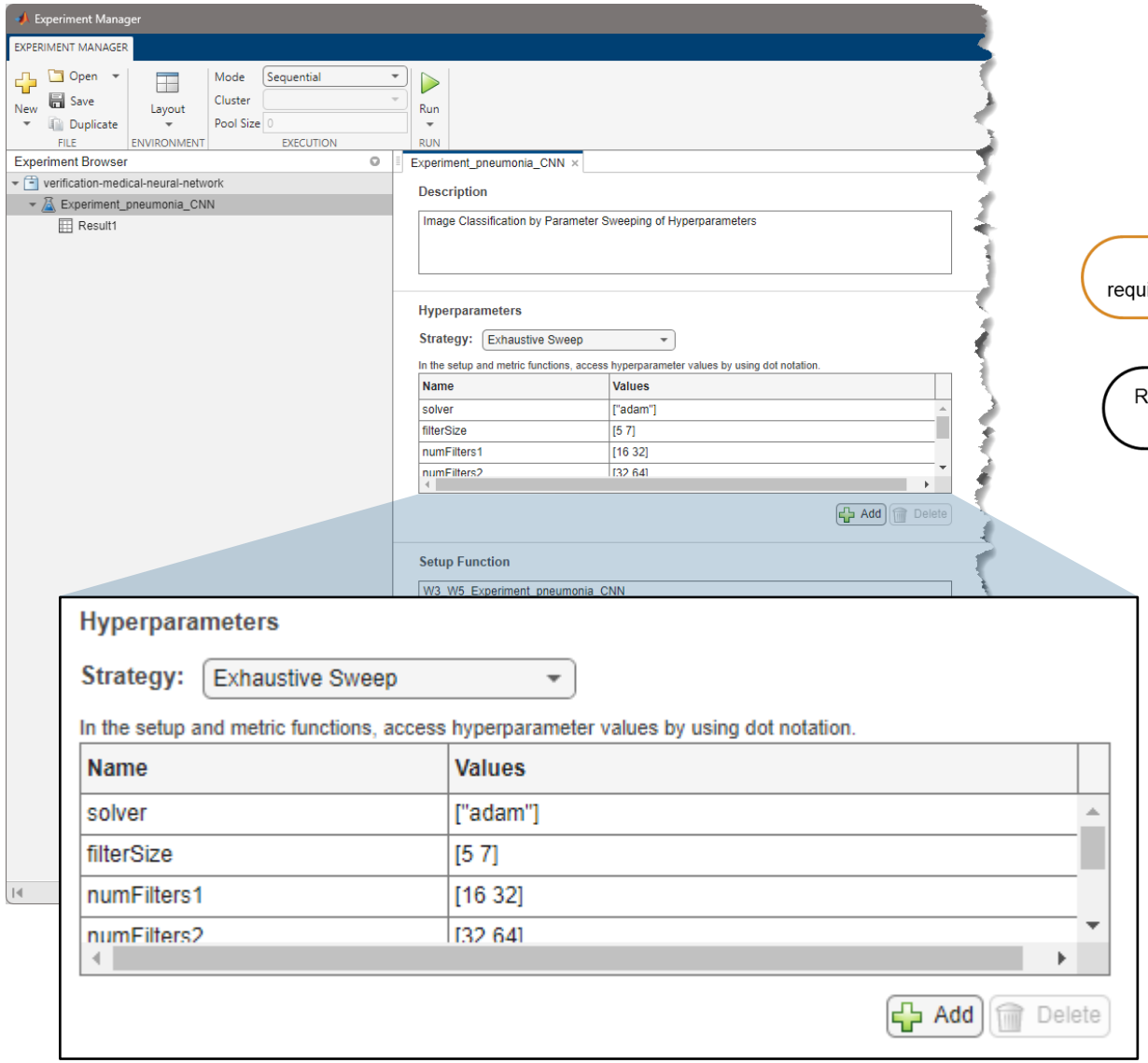


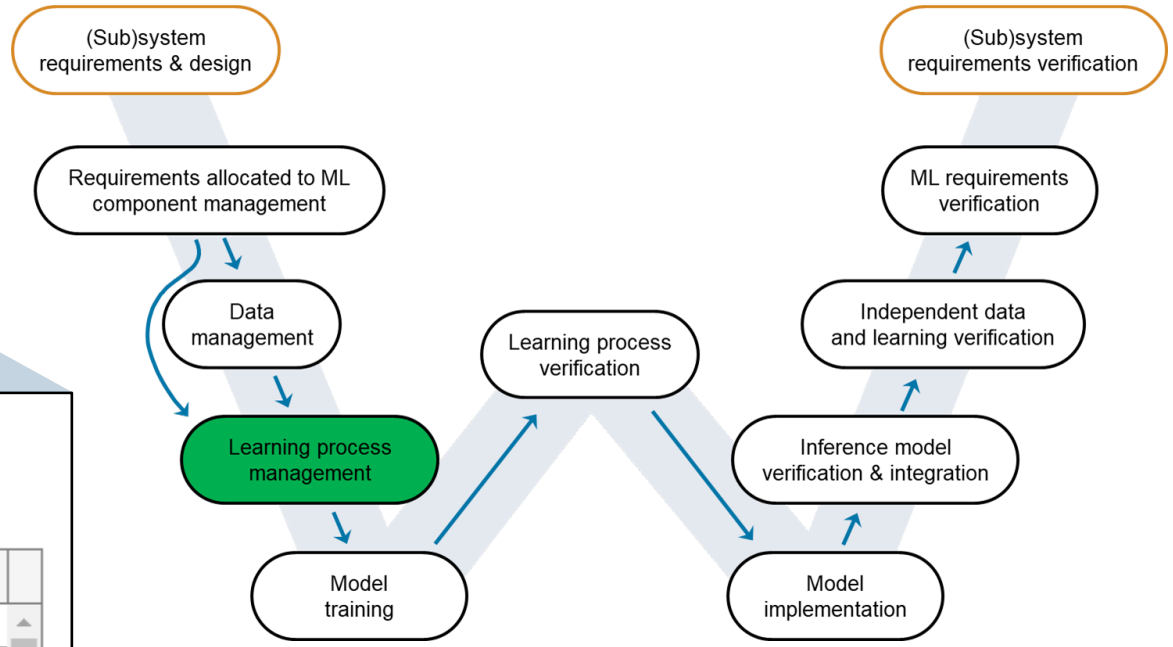

```
numClasses = numel(classNames);
layers = [
    imageInputLayer(imageSize,Normalization="none")
    convolution2dLayer(7,64,Padding=0)
    batchNormalizationLayer()
    reluLayer()
    dropoutLayer(0.5)
    averagePooling2dLayer(2,Stride=2)
    convolution2dLayer(7,128,Padding=0)
    batchNormalizationLayer()
    reluLayer()
    dropoutLayer(0.5)
    averagePooling2dLayer(2,Stride=2)
    fullyConnectedLayer(numClasses)
    softmaxLayer
    classificationLayer(Classes=classNames,ClassWeights=classWeights)];
```

```
options = trainingOptions("adam", ...
    ExecutionEnvironment="auto", ...
    InitialLearnRate=0.001, ...
    MaxEpochs=50, ...
    MiniBatchSize=256, ...
    Shuffle="every-epoch",...
    LearnRateSchedule="piecewise", ...
    LearnRateDropPeriod=30, ...
    LearnRateDropFactor=0.1, ...
    Plots="training-progress", ...
    ValidationData={XVal,TVal}, ...
    ValidationPatience=10, ...
    OutputNetwork="best-validation-loss");
```



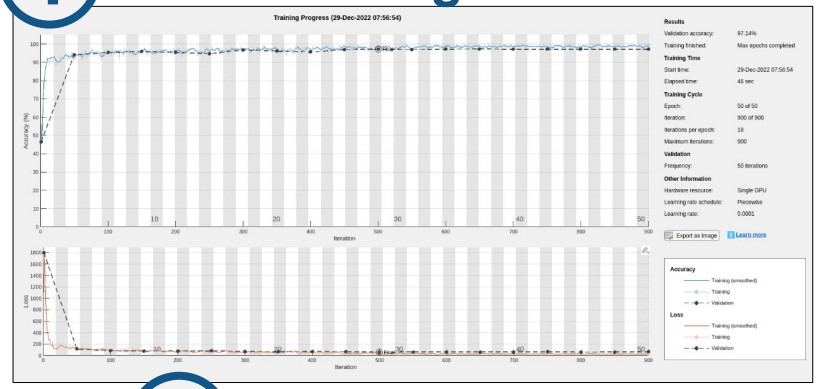
模型训练管理和超参数调优





迭代训练与过程可视化

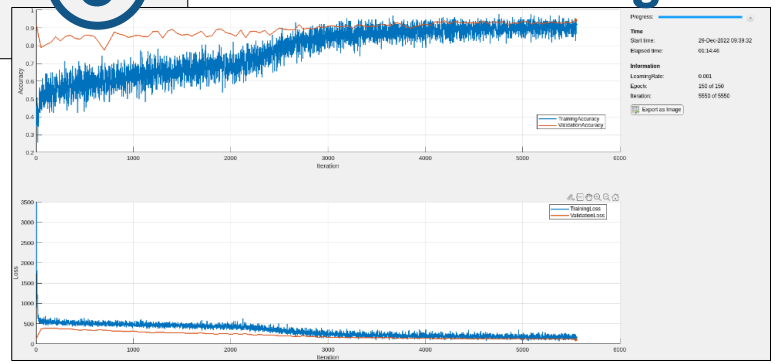
1 Initial training

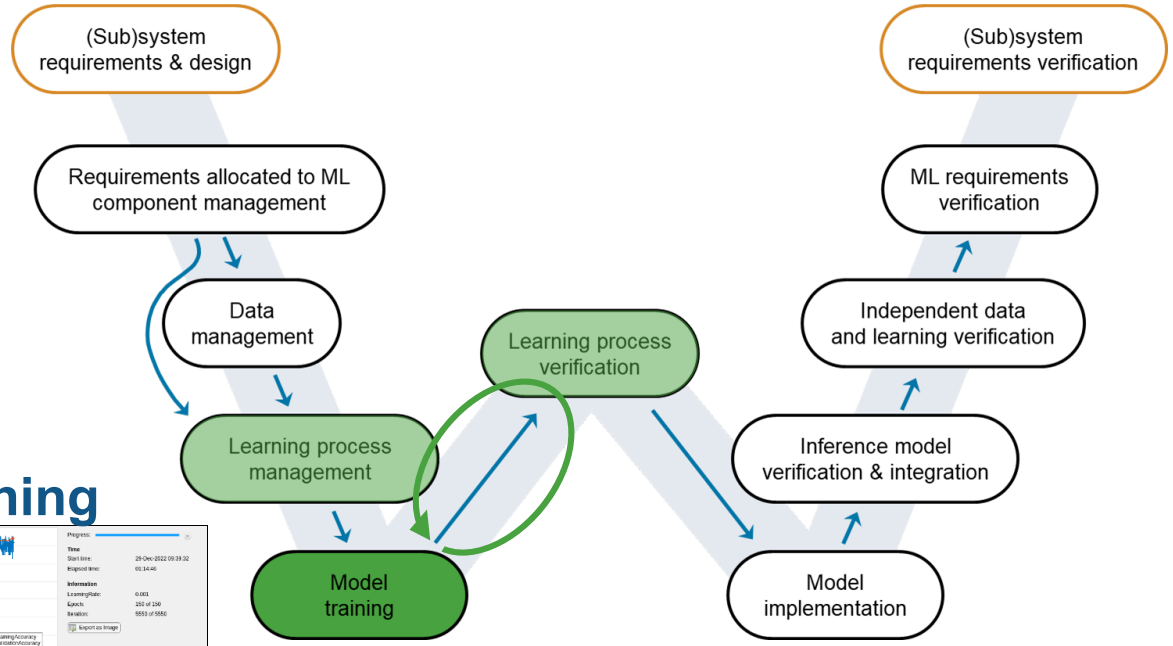


2 Data-augmented training

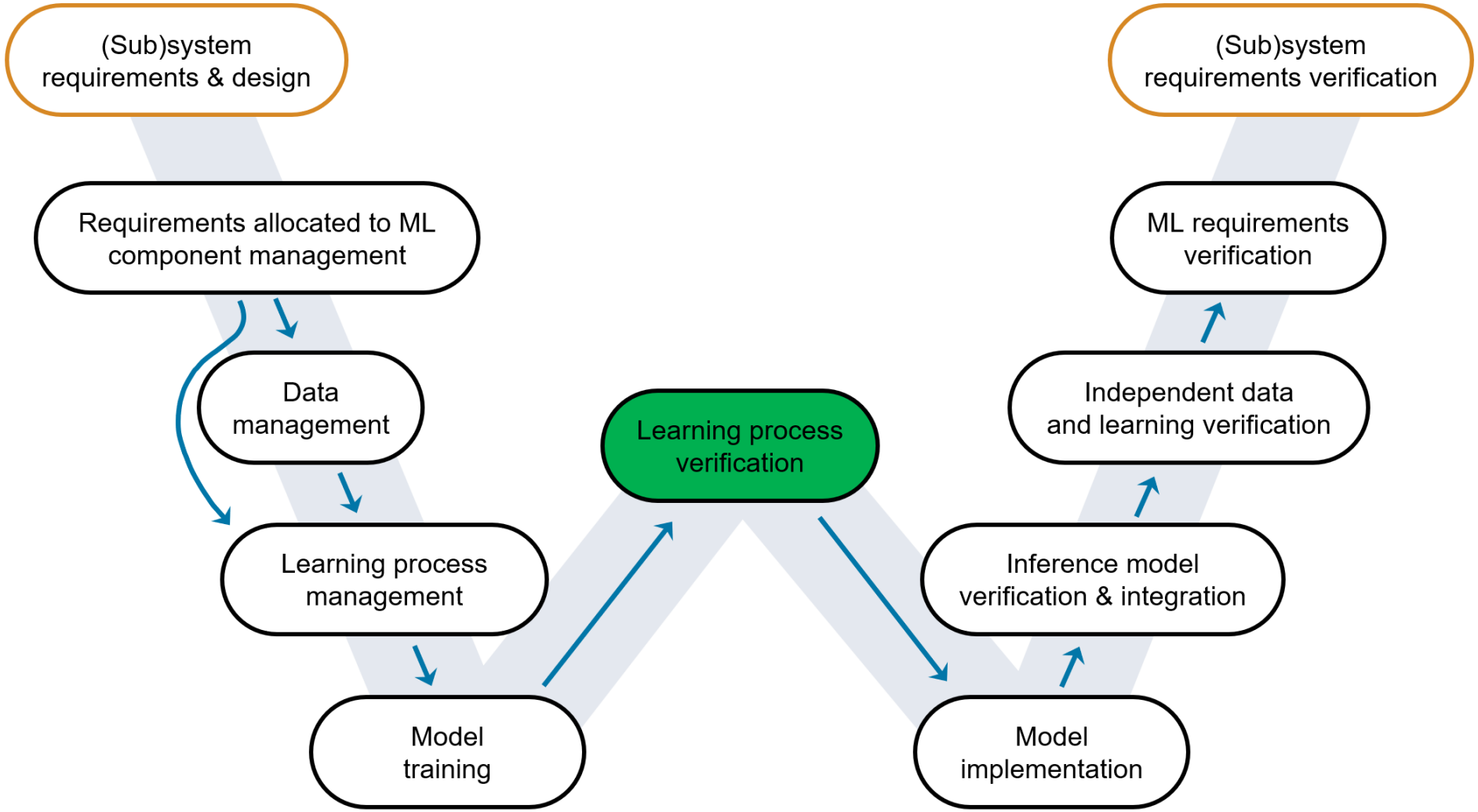
```
imageAugmenter = imageDataAugmenter(...
    FillValue=mean(XTrain(:)), ...
    RandXReflection=true, ...
    RandXTranslation=[-2,2], ...
    RandYTranslation=[-2,2], ...
    RandRotation=[-10,10],...
    RandScale=[1,1.25], ...
    RandXShear=[-5,5], ...
    RandYShear=[-5,5]);
```

3 Adversarial training

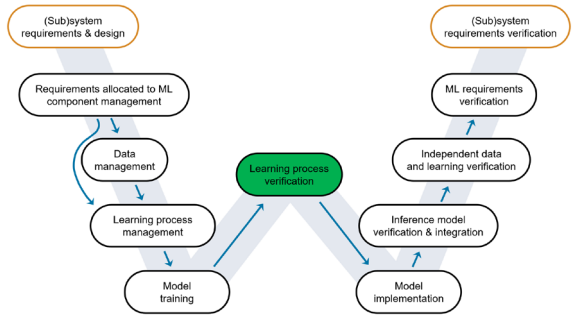




模型验证

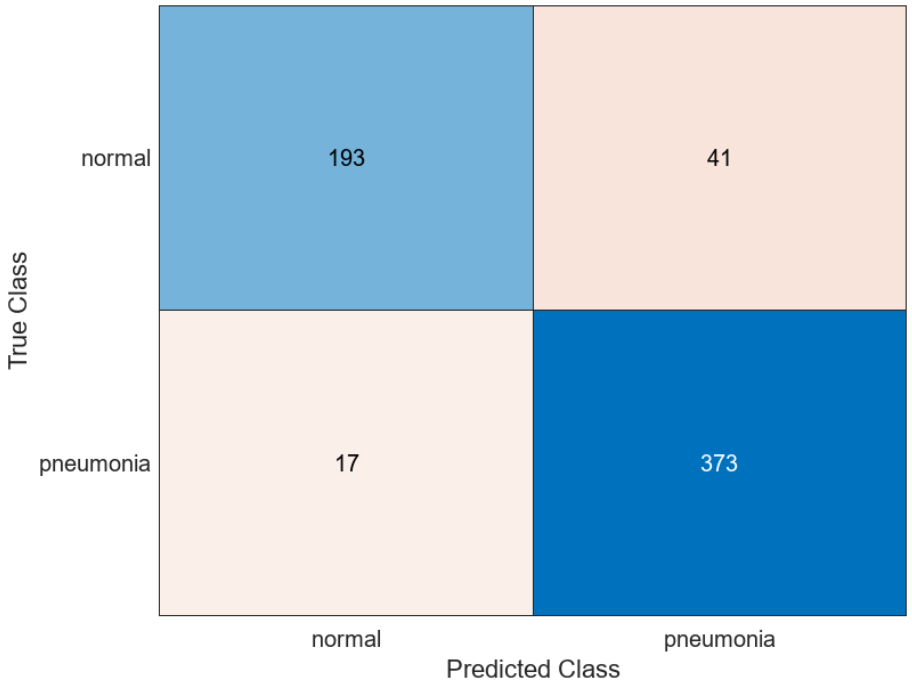

(Sub)system requirements & design
Requirements allocated to ML component management
Data management
Learning process management
Model training
Learning process verification
Model implementation
Inference model verification & integration
Independent data and learning verification
ML requirements verification
(Sub)system requirements verification

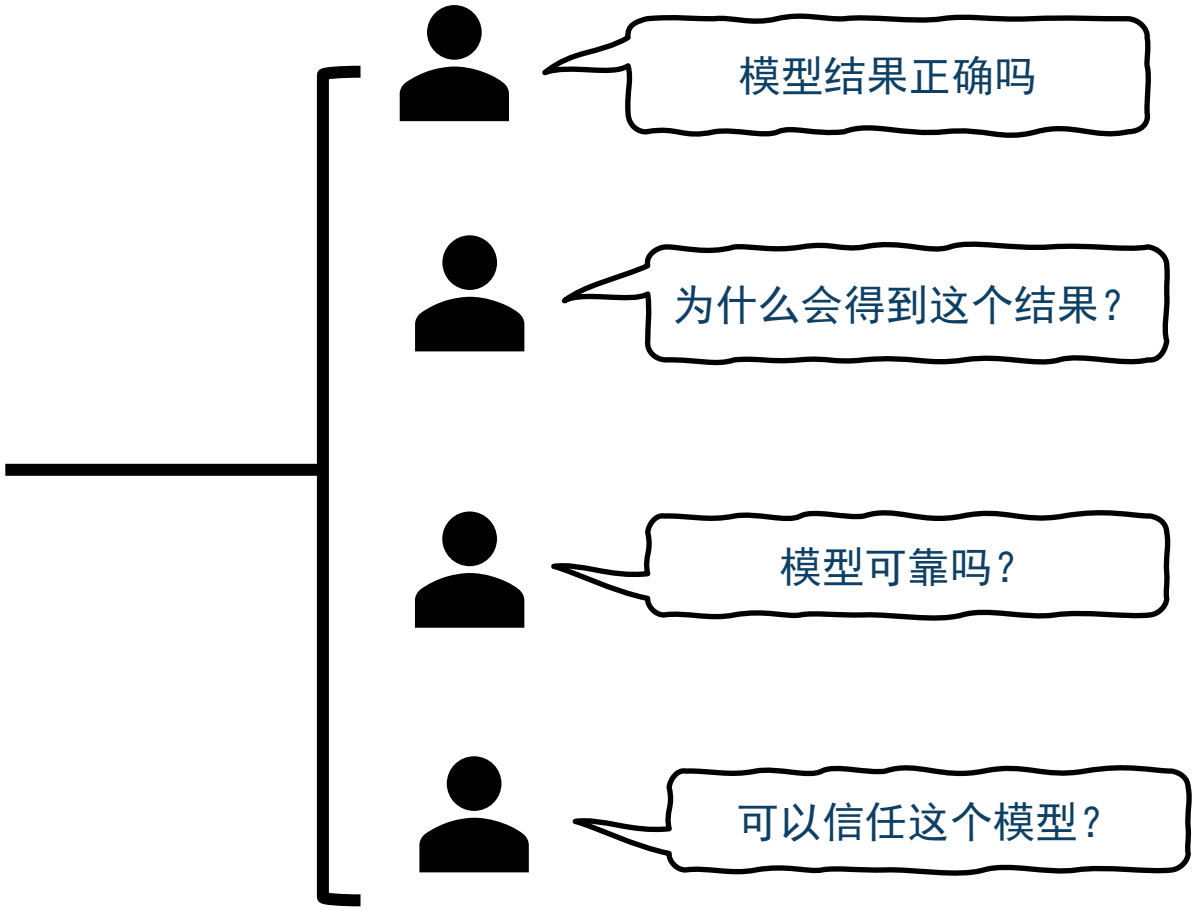
理解和测试模型



Accuracy: 90.71%

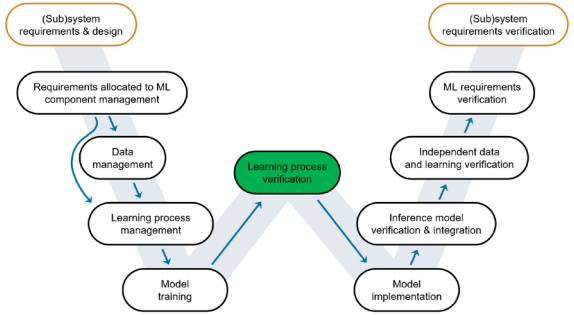
```
confusionchart(T,Y)
```

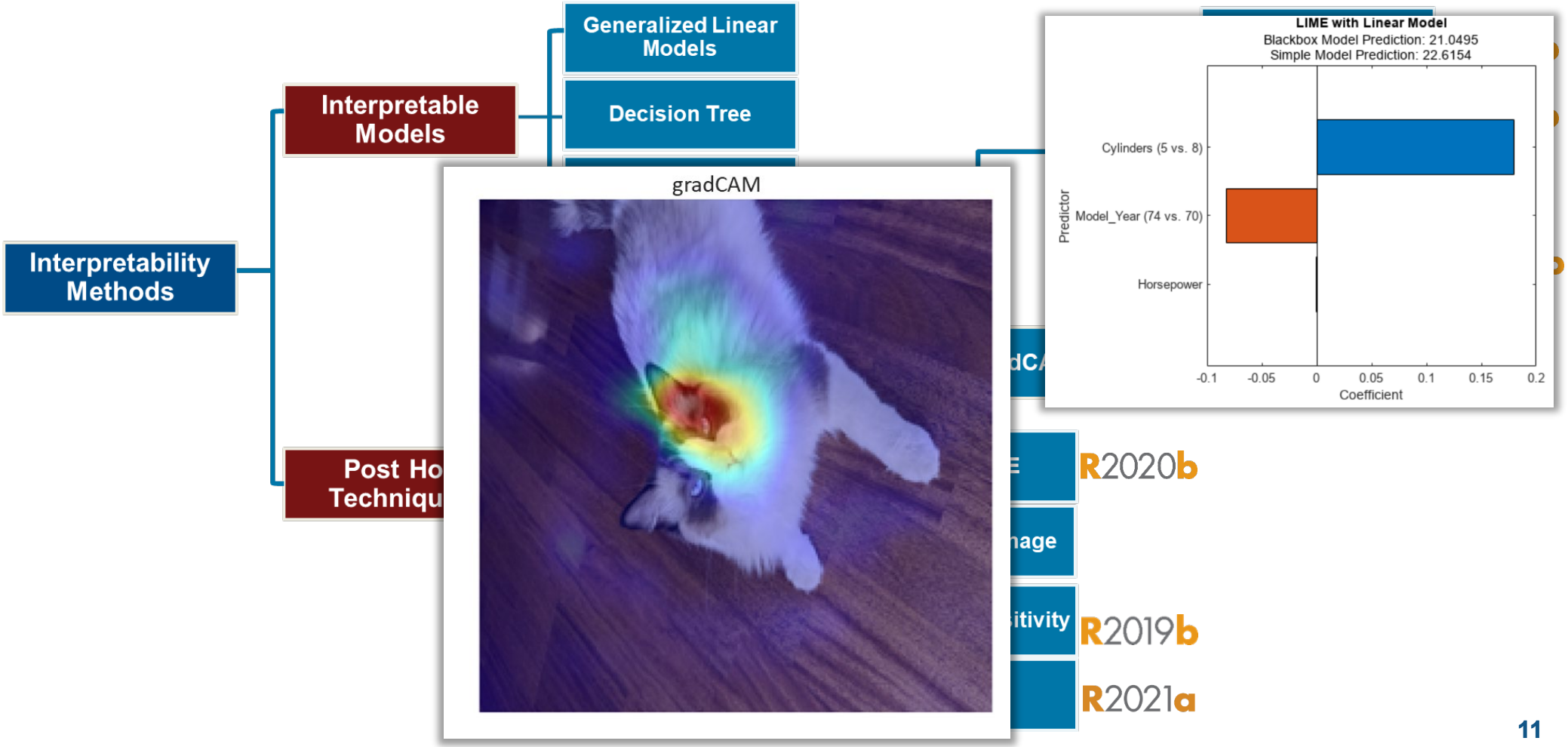




模型可解释性

- MATLAB提供多种模型解释方法





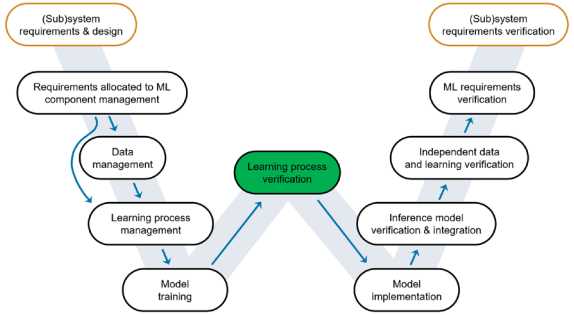
MATLAB Deep Learning Toolbox Verification

Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team STAFF

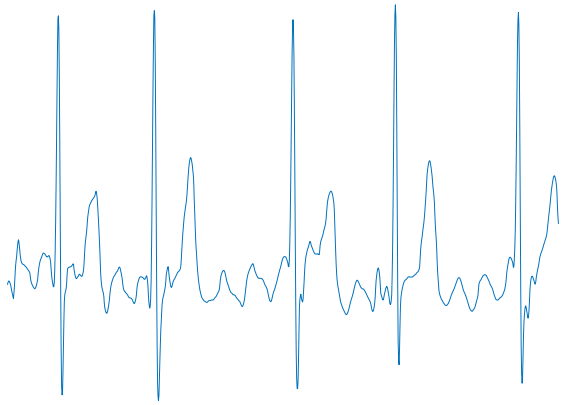
Verify and test robustness of deep learning networks

R2023a



- 验证神经网络对抗性样本的健壮性 (Adversarial Examples)
- 评估网络对输入扰动的敏感程度
- 创建数据分布识别器，划分分布内和分布外数据
- 检测网络输入的分布外数据(Detect out-of-distribution ,OOD)

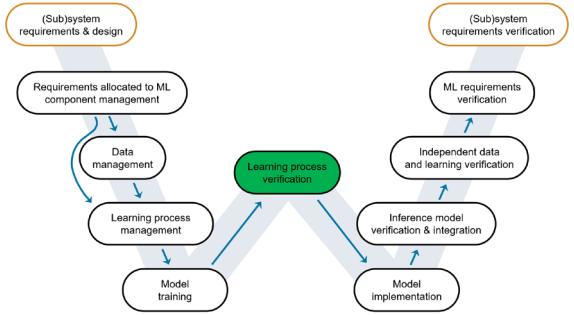
模型鲁棒性



$+ \delta$

Neural Network

Arrhythmia
Normal



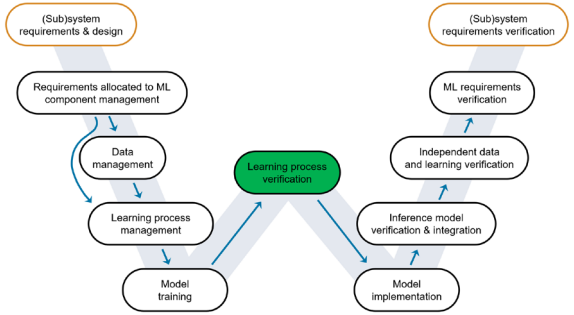
Formal Verification

violated

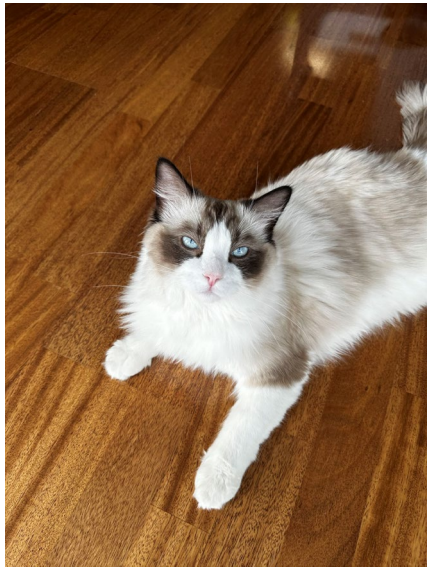
violated

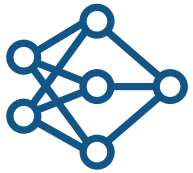
unproven

利用对抗样本(Adversarial Examples)验证模型鲁棒性



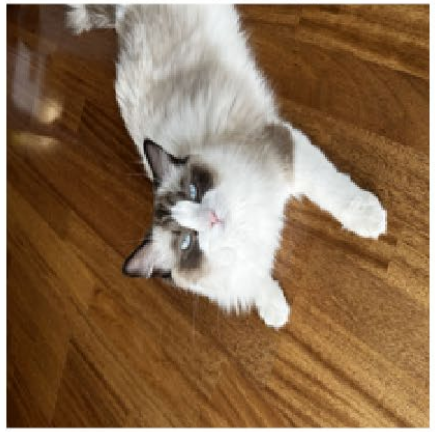
通过将生成的微小扰动加入到输入数据，以验证模型预测的鲁棒性





squeezenet

Original Image
Class: Persian cat



Perturba

$$X_{\text{adv}} = X + \epsilon.\,\text{sign}(\nabla_X L(X, T)).$$

fast gradient sign method (FGSM)

Adversarial Image (Epsilon = 1)
Class: Tibetan terrier

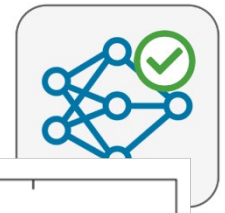
西藏梗犬

Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. “Intriguing Properties of Neural Networks.” Preprint, submitted February 19, 2014. https://arxiv.org/abs/1312.6199.

https://ww2.mathworks.cn/help/deeplearning/ug/generate-adversarial-examples.html?s_tid=srchtitle_Generate%20Untargeted%20and%20Targeted%20Adversarial%20Examples%20for%20Image%20Classification_1

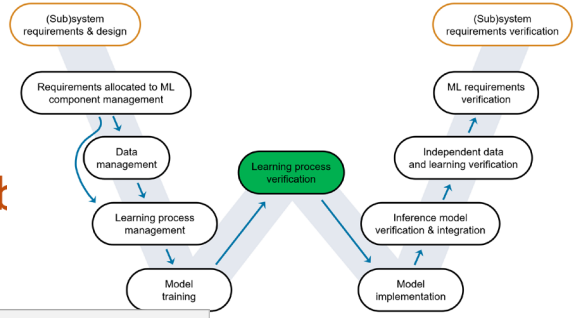
MATLAB Deep Learning Toolbox Verification

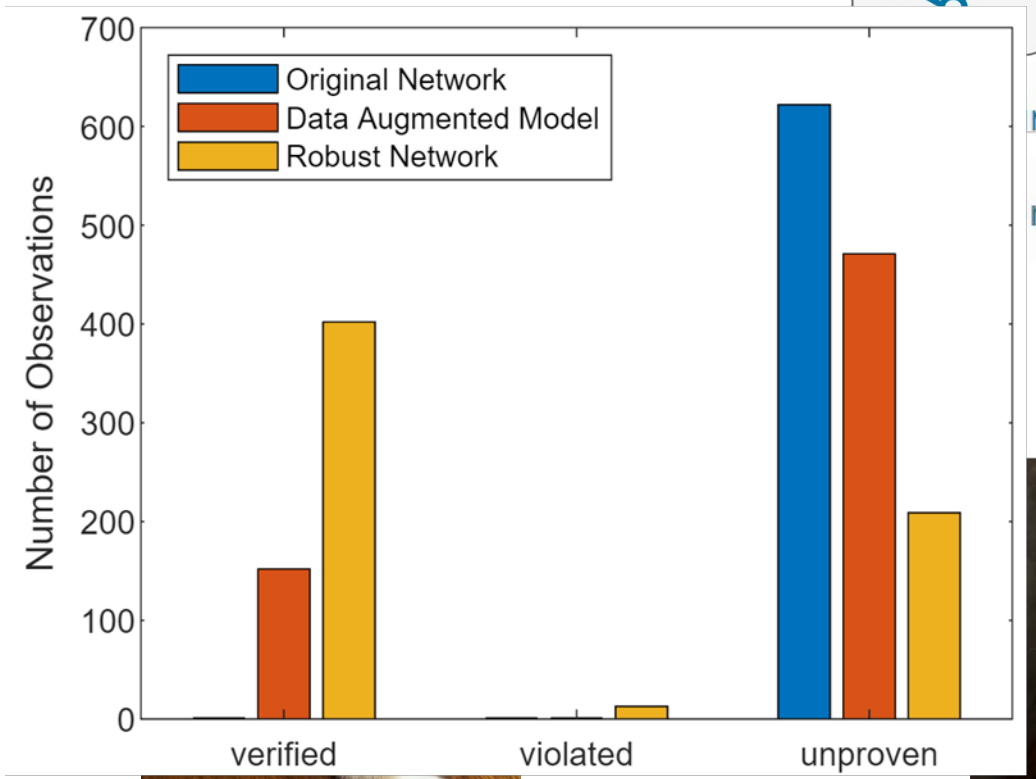
- 验证和评估网络鲁棒性



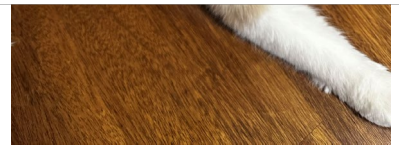
Deep Learning Toolbox Verification Lib

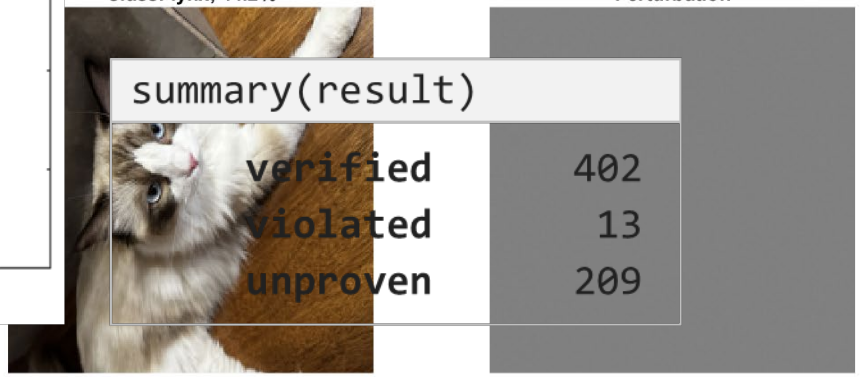
by MathWorks Deep Learning Toolbox Team STAFF





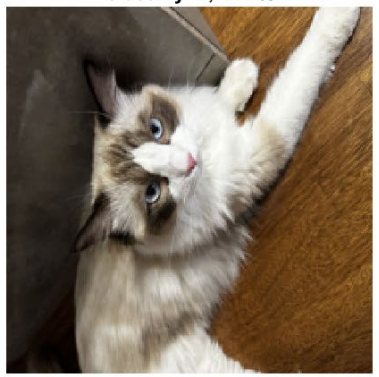

```
perturbation = 0.01;
XLower = XTest - perturbation;
XUpper = XTest + perturbation;
XLower = dlarray(XLower,"SSCB");
XUpper = dlarray(XUpper,"SSCB");
result = verifyNetworkRobustness(net,...
    XLower,XUpper,TTest);
```





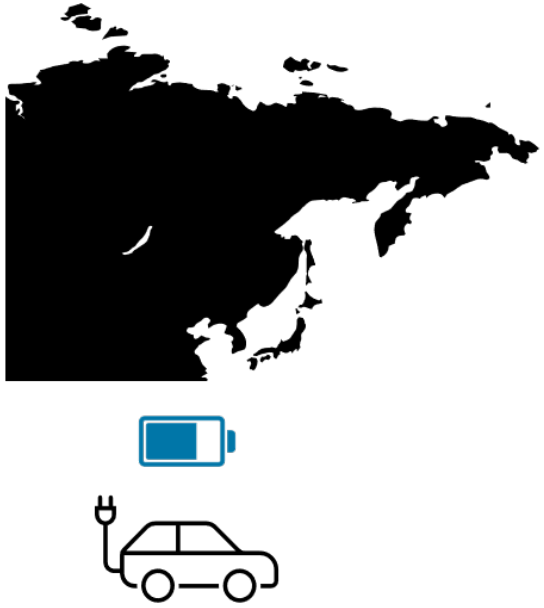
```
summary(result)
```

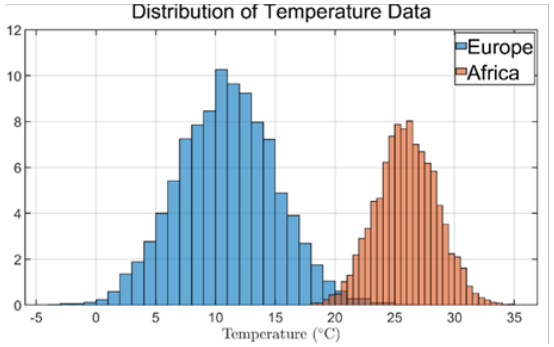
verified	402
violated	13
unproven	209

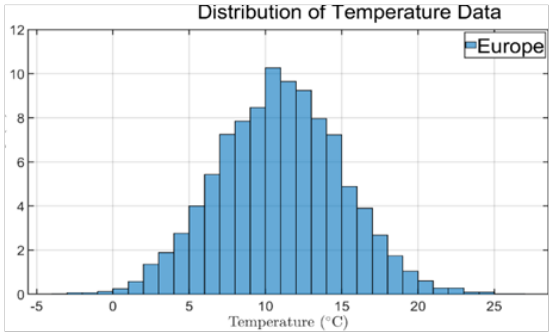


MATLAB Deep Learning Toolbox Verification

- **Out-of-Distribution Detection**
- In-distribution (ID)： 用来构建和训练模型的数据。
- Out-of-distribution (OOD)：不同于训练数据的数据。例如，数据是在不同的方式、时间、条件下采集。





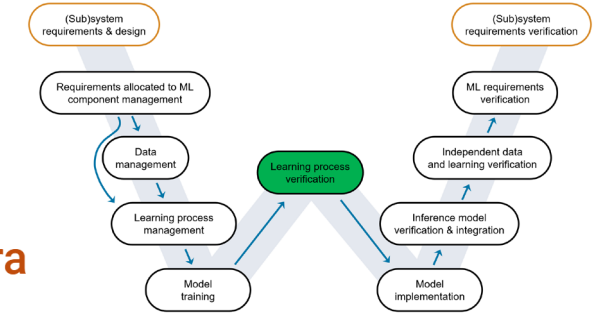


未知样本的识别与处理

对未知情况的处理

1. 拒绝

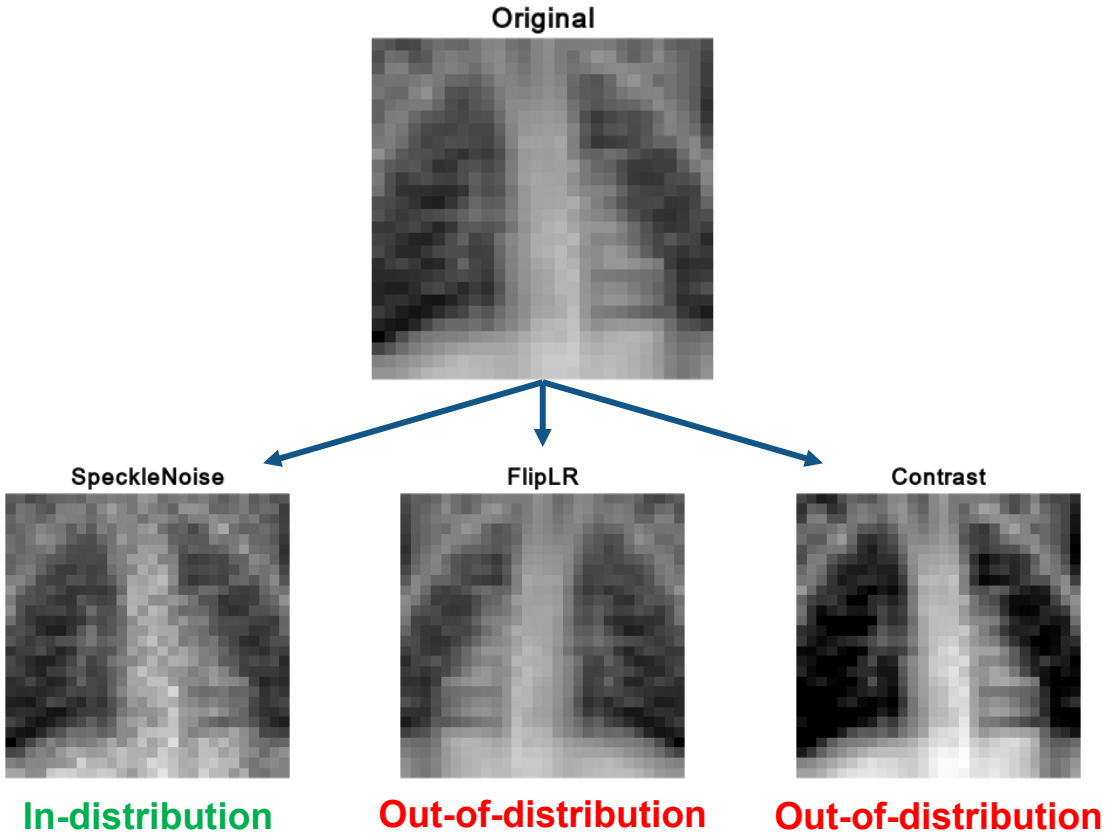
2. 人工处理

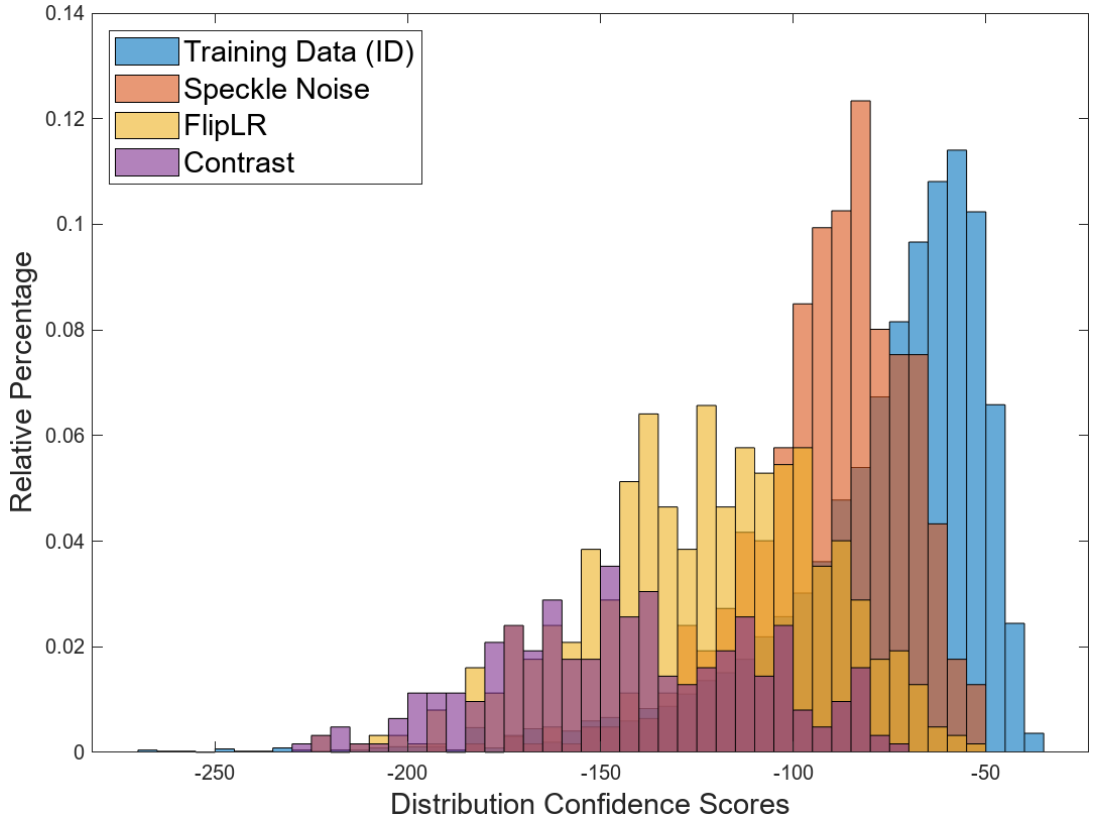


Deep Learning Toolbox Verification Libra

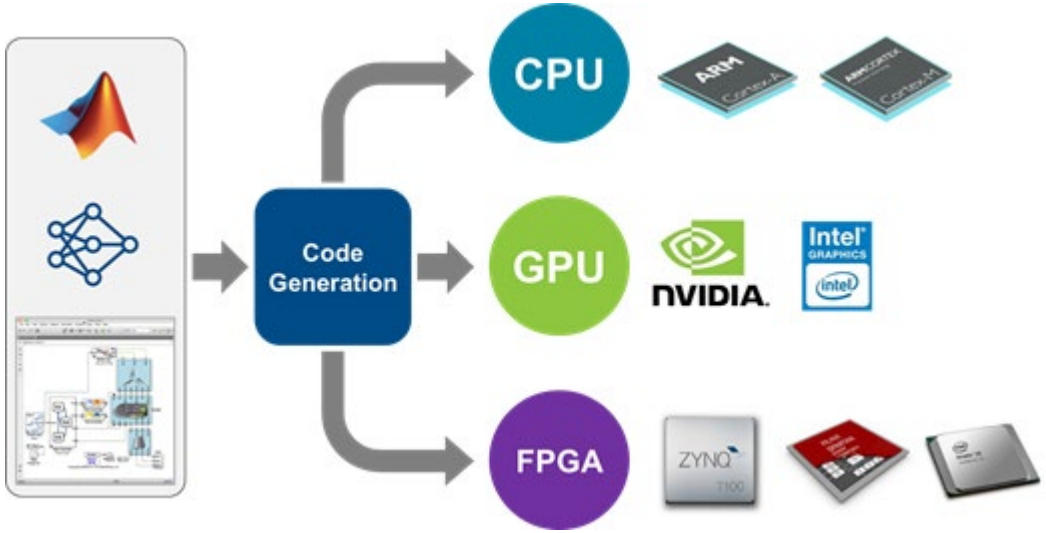
by MathWorks Deep Learning Toolbox Team STAFF

Verify and test robustness of deep learning networks

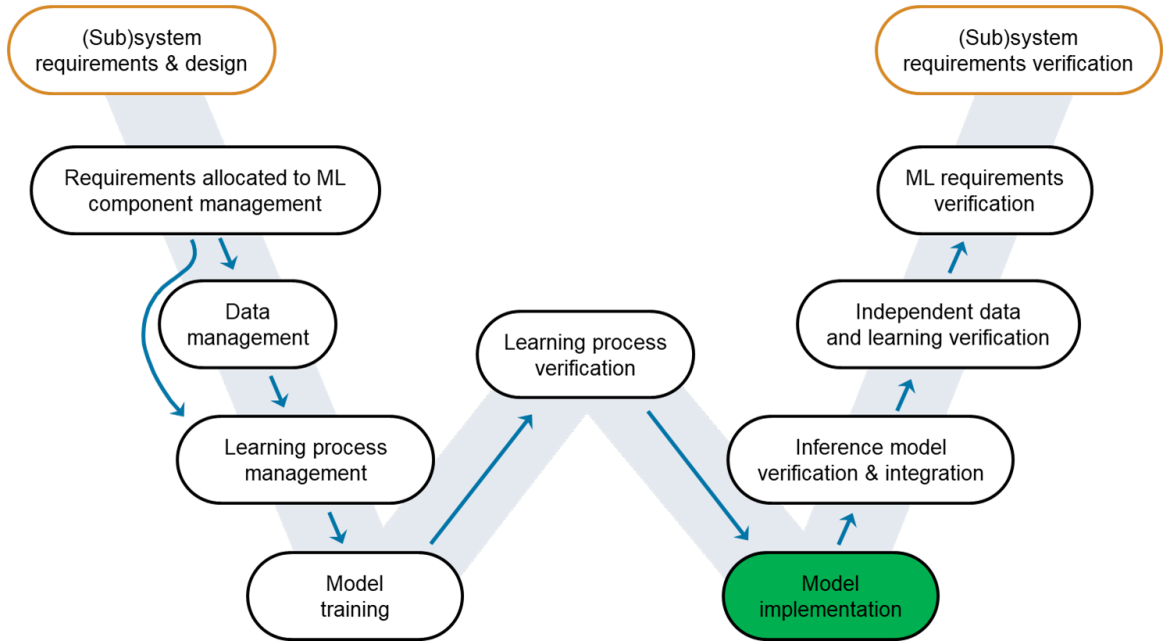




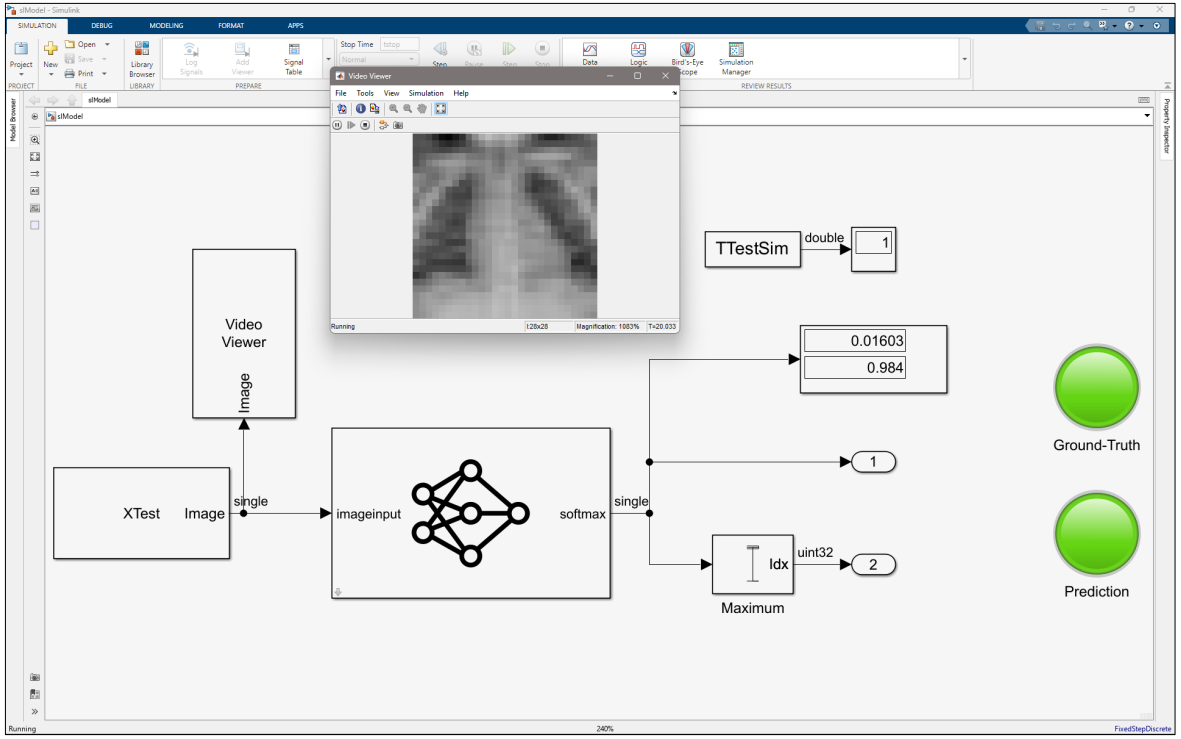
模型部署

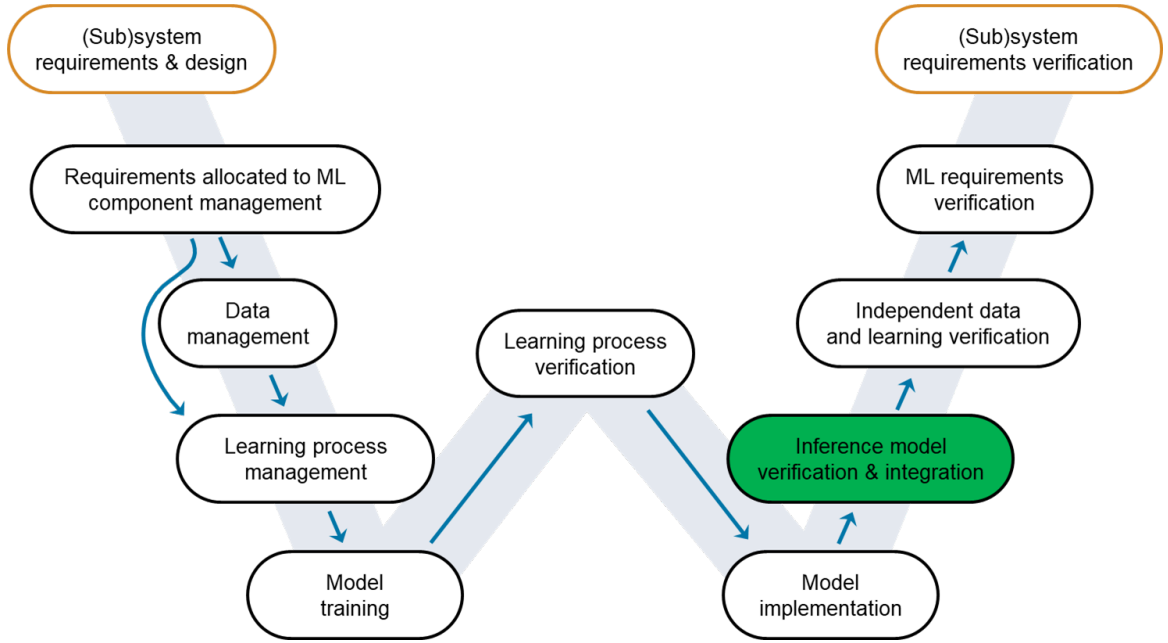



```
analyzeNetworkForCodegen(net)

                  Supported
                  _________

    none            "Yes"
    arm-compute     "Yes"
    mkldnn          "Yes"
    cudnn           "Yes"
    tensorrt        "Yes"
```

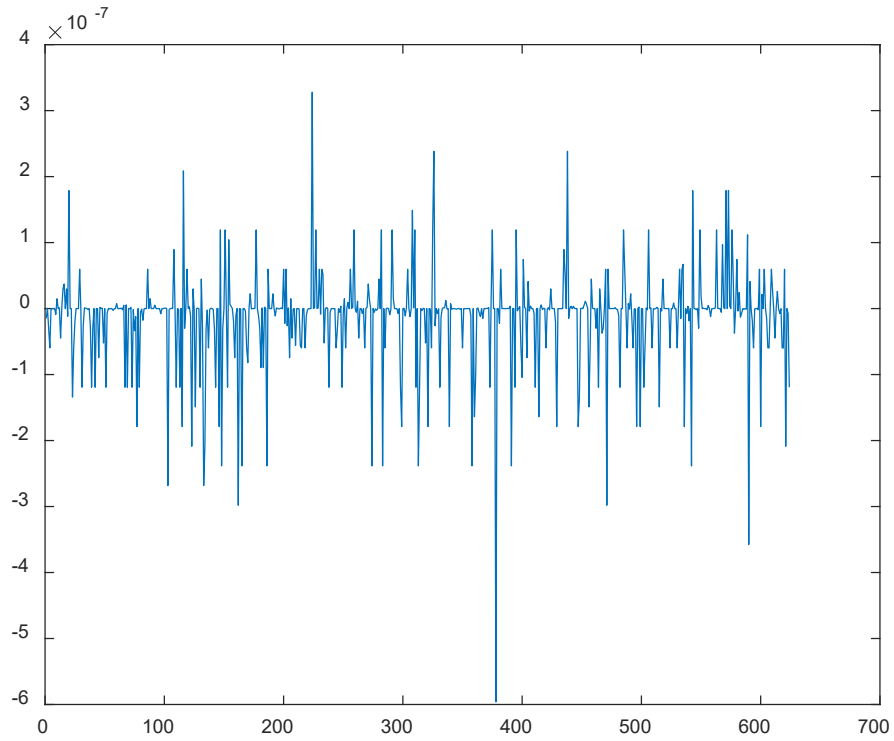


与Simulink集成实现系统级仿真



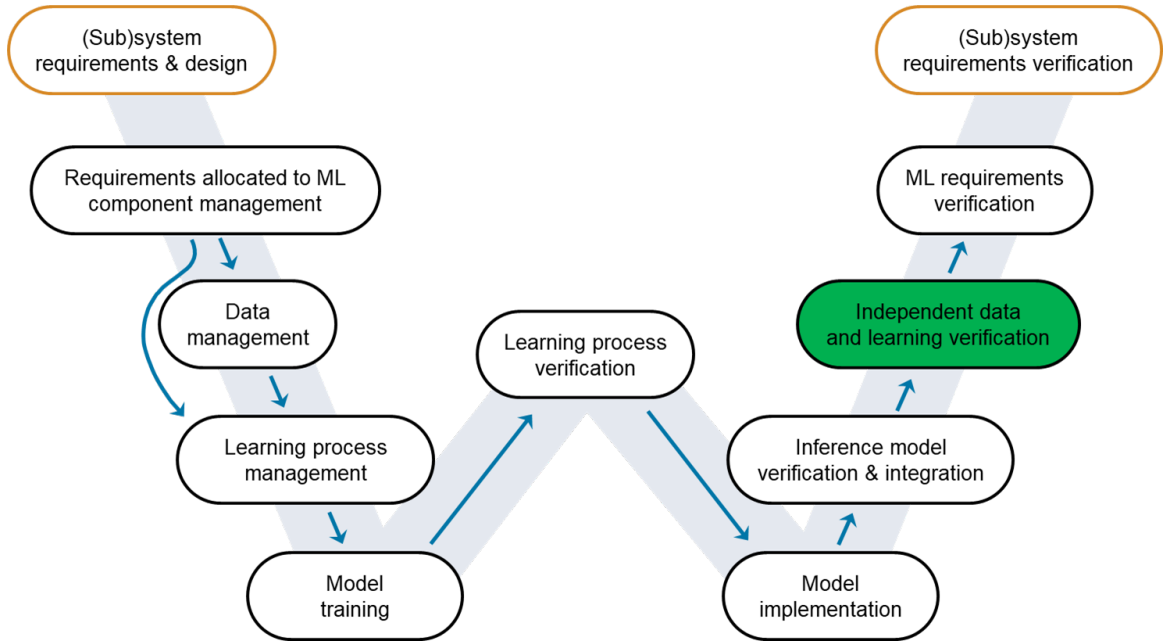


模型泛化能力验证

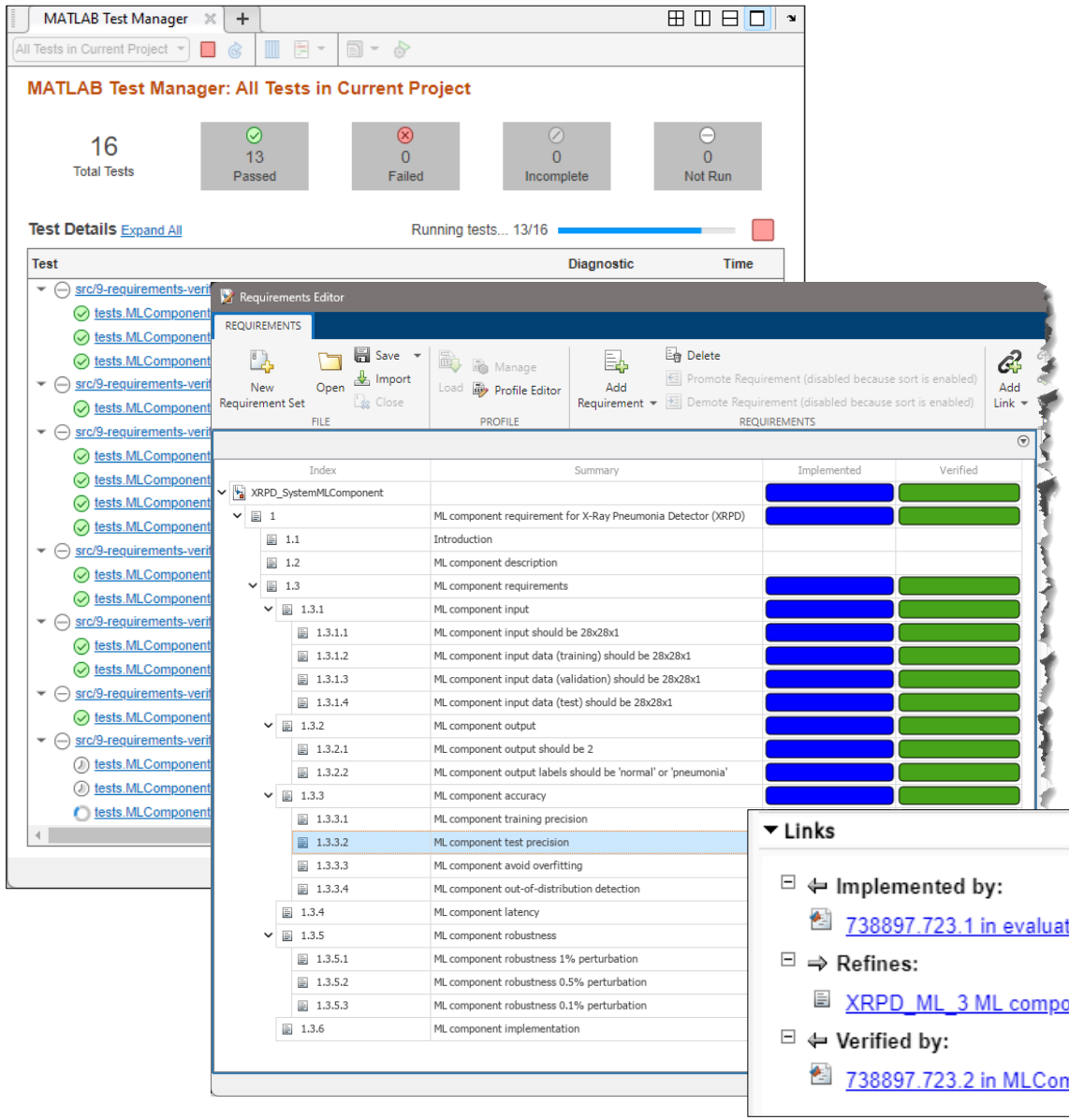


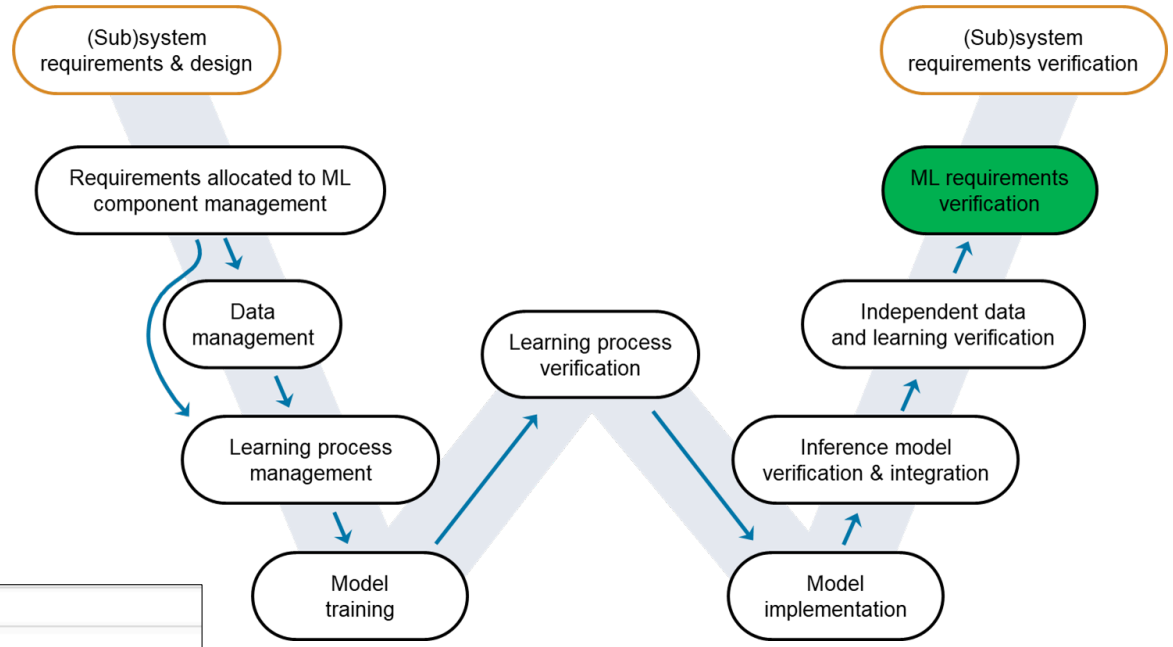

```
max(abs(differences))
```

```
ans = single
5.9605e-07
```



需求完整度验证


MATLAB Test Manager
MATLAB Test Manager: All Tests in Current Project
16
Total Tests
13
Passed
0
Failed
0
Incomplete
0
Not Run
Test Details Expand All
Running tests... 13/16
Test
Diagnostic
Time
Requirements Editor
REQUIREMENTS
New Requirement Set
Open
Save
Import
Close
FILE
Load
Manage
Profile Editor
PROFILE
Add Requirement
Delete
Promote Requirement (disabled because sort is enabled)
Demote Requirement (disabled because sort is enabled)
REQUIREMENTS
Add Link
Index
Summary
Implemented
Verified
XRPD_SystemMLComponent
1 ML component requirement for X-Ray Pneumonia Detector (XRPD)
1.1 Introduction
1.2 ML component description
1.3 ML component requirements
1.3.1 ML component input
1.3.1.1 ML component input should be 28x28x1
1.3.1.2 ML component input data (training) should be 28x28x1
1.3.1.3 ML component input data (validation) should be 28x28x1
1.3.1.4 ML component input data (test) should be 28x28x1
1.3.2 ML component output
1.3.2.1 ML component output should be 2
1.3.2.2 ML component output labels should be 'normal' or 'pneumonia'
1.3.3 ML component accuracy
1.3.3.1 ML component training precision
1.3.3.2 ML component test precision
1.3.3.3 ML component avoid overfitting
1.3.3.4 ML component out-of-distribution detection
1.3.4 ML component latency
1.3.5 ML component robustness
1.3.5.1 ML component robustness 1% perturbation
1.3.5.2 ML component robustness 0.5% perturbation
1.3.5.3 ML component robustness 0.1% perturbation
1.3.6 ML component implementation
Links
Implemented by:
738897.723.1 in evaluateModelAccuracy.m
Refines:
XRPD_ML_3 ML component accuracy
Verified by:
738897.723.2 in MLComponent_Accuracy.m


(Sub)system requirements & design
Requirements allocated to ML component management
Data management
Learning process management
Model training
Learning process verification
Model implementation
Inference model verification & integration
Independent data and learning verification
ML requirements verification
(Sub)system requirements verification

MATLAB EXPO

谢谢