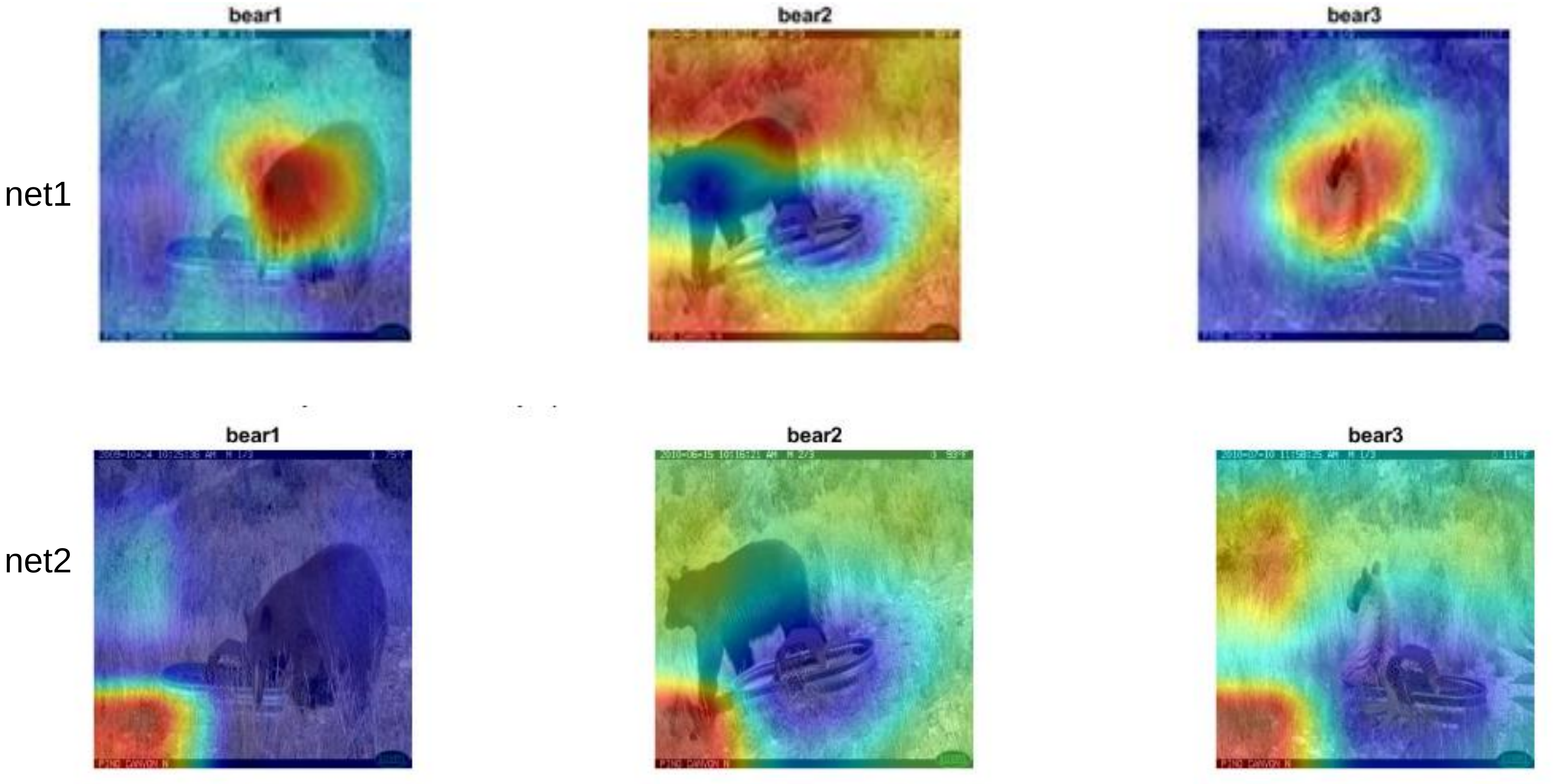
AI的高安全性应用与可靠性验证

——医学影像分析与可解释性

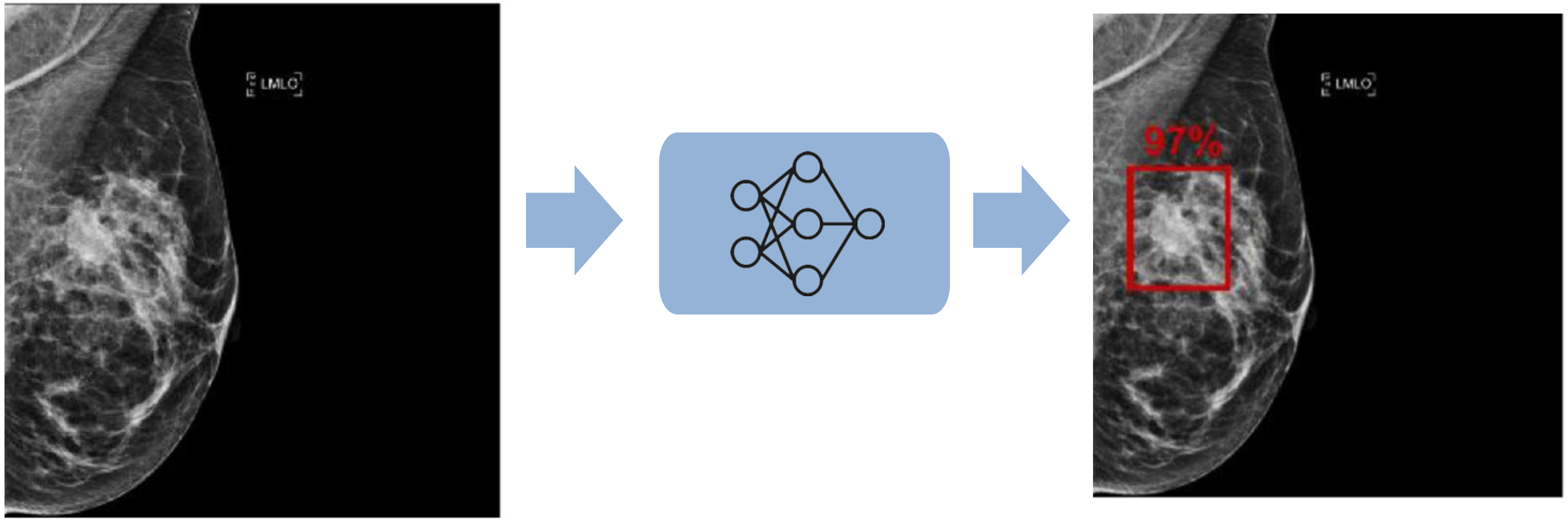
MathWorks
中国区医疗行业市场经理
单博



Results from Grad-CAM identifying portions of images which influence the classification.

可解释的AI 模型

仅提供预测/推理就够了吗？

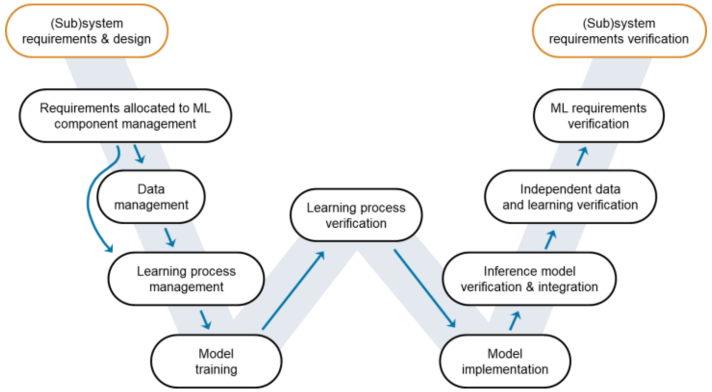


XAI 能：
- 提示什么对得到结论贡献最大.
- 提示AI的缺陷/弱点.

XAI 不能:
- 解析方式解释AI模型
- 代替对于高安全模型具有关键意义的Good Machine Learning Practices (GMLP),

要点

MathWorks 提供高安全性AI开发W流程各阶段的支持



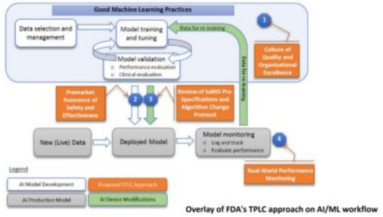
神经网络模型
鲁棒性测试与验证专用库

Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team STAFF

Verify and test robustness of deep learning networks

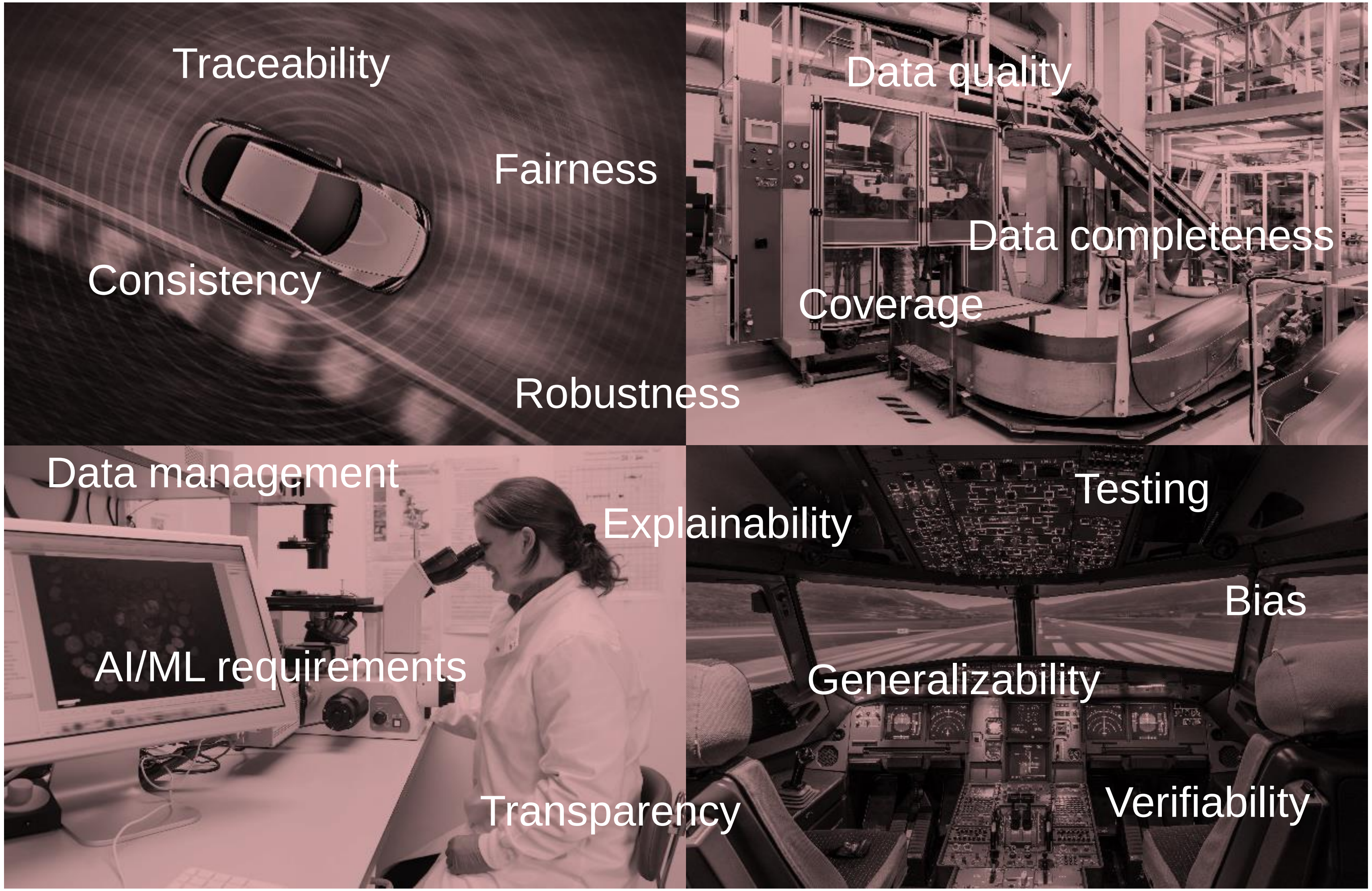
高安全性验证的经验助力推动全新AI标准



EUROCAE WG-114 / SAE G-34
Standardization Working Group
“Artificial Intelligence in Aviation”

随AI的使用量在快速增加，迫切需要在高安全性领域，对其解释、确认和验证。

对包含AI组件系统的确认与验证的挑战



工业界正在努力推动高安全性系统中AI模型的验证

提供白皮书, 标准及计划

汽车

全新 WIP ISO PAS 8800 (Road Vehicles — 安全性与 AI)

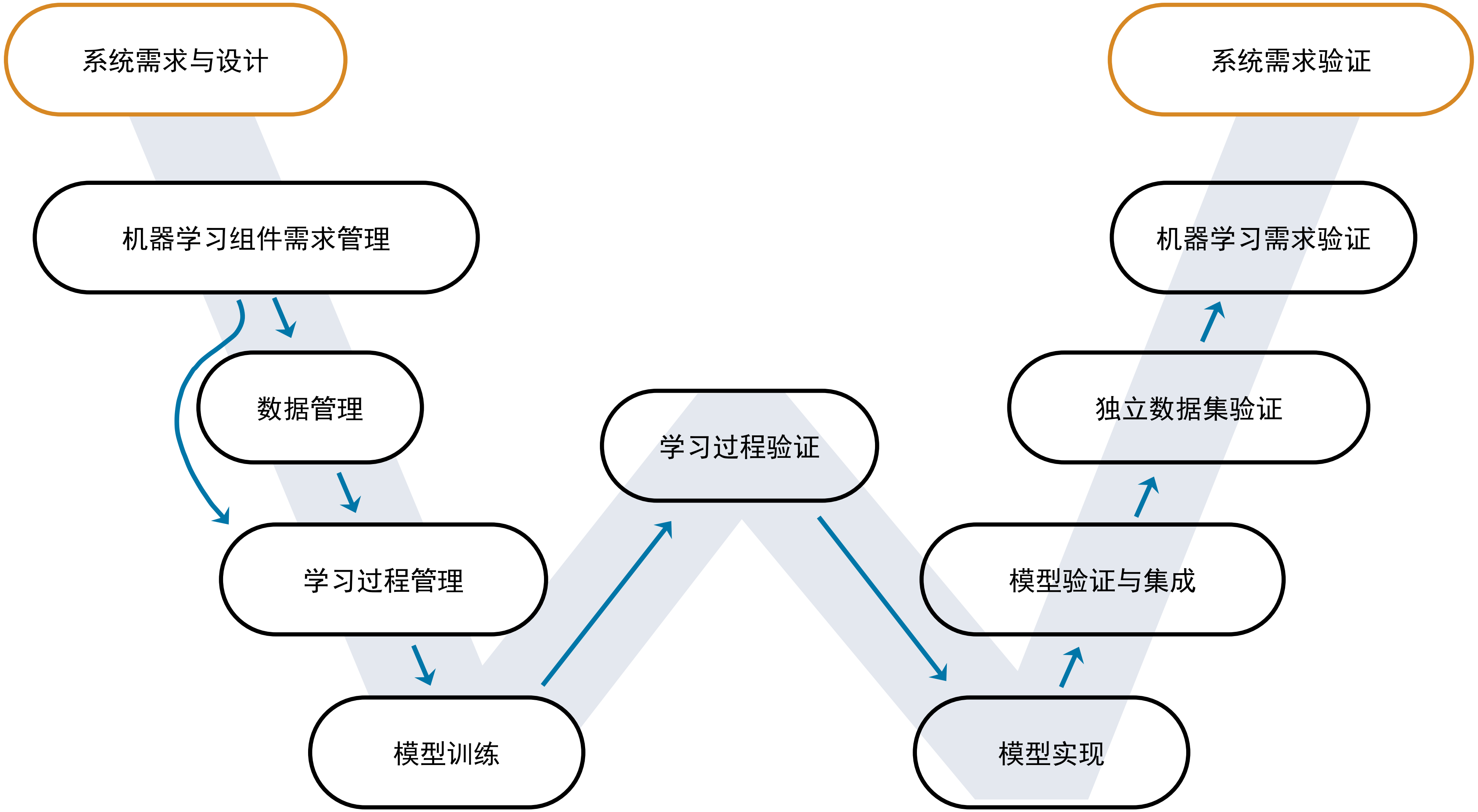
航空

全新标准(AS6983) from EUROCAE WG-114 / SAE G-34 预计2024发布

医疗

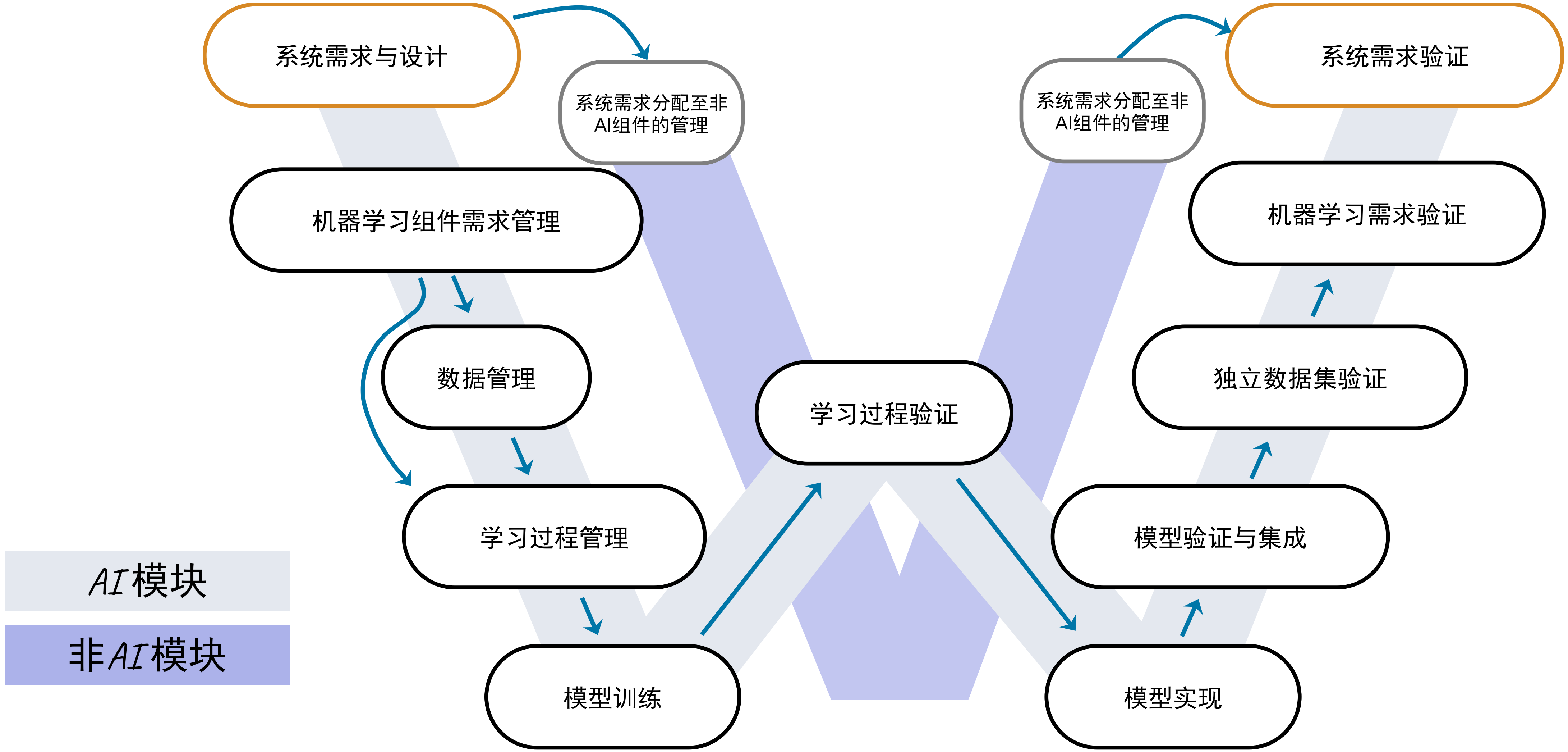
FDA 发布了第一份规范 AI/ML-Based Software as a Medical Device (SaMD) Action Plan

W型流程将V流程的应用延伸到AI的应用



Credit: EASA, Daedalean

W型流程可与非AI模块的V流程共存



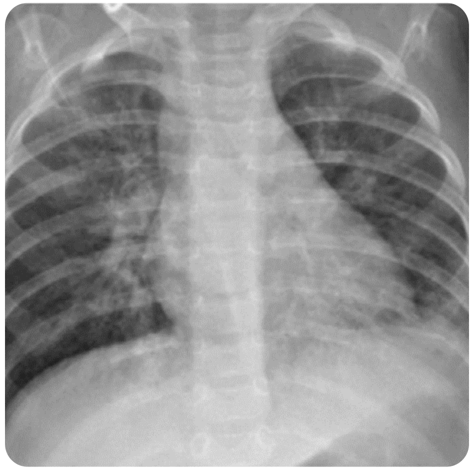
Credit: EASA, Daedalean

Task: 验证一个图像分类网络

汽车

航空

医疗



MedMNIST v2 数据集

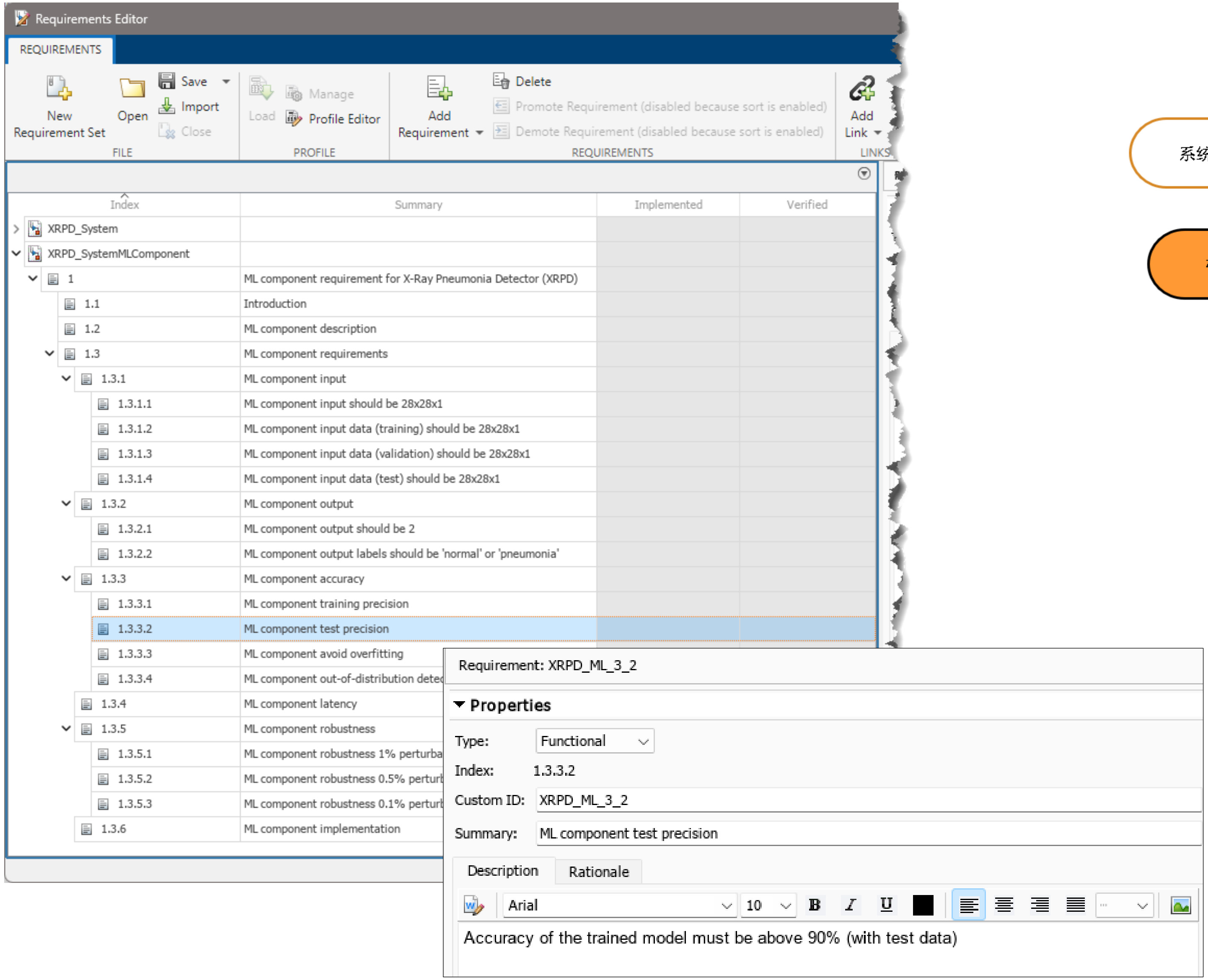
MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification

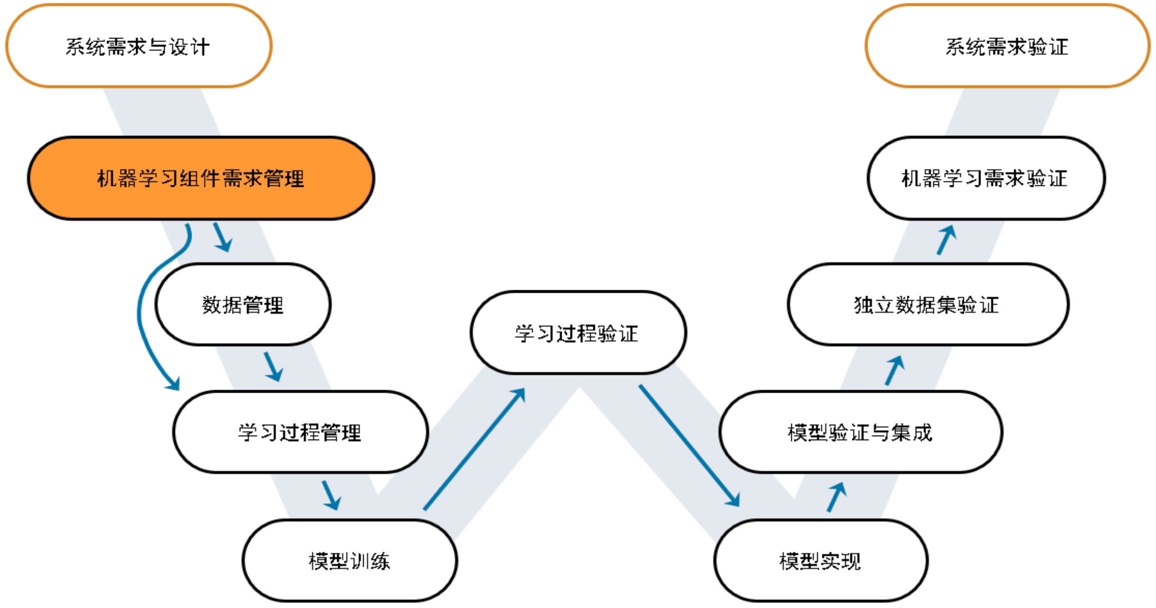
Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, Bingbing Ni

[1] Shanghai Jiao Tong University, Shanghai, China
[2] Boston College, Chestnut Hill, MA
[3] RWTH Aachen University, Aachen, Germany
[4] Fudan Institute of Metabolic Diseases, Zhongshan Hospital, Fudan University, Shanghai, China
[5] Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[6] Harvard University, Cambridge, MA

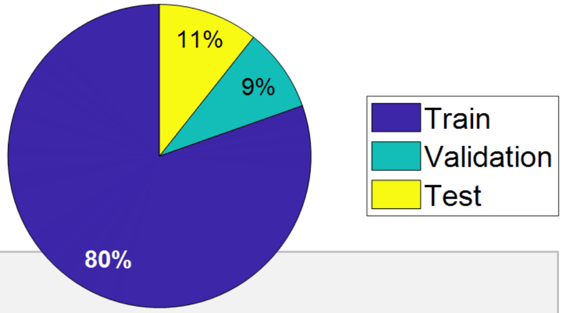
PneumoniaMNIST

始于与机器学习相关的需求收集

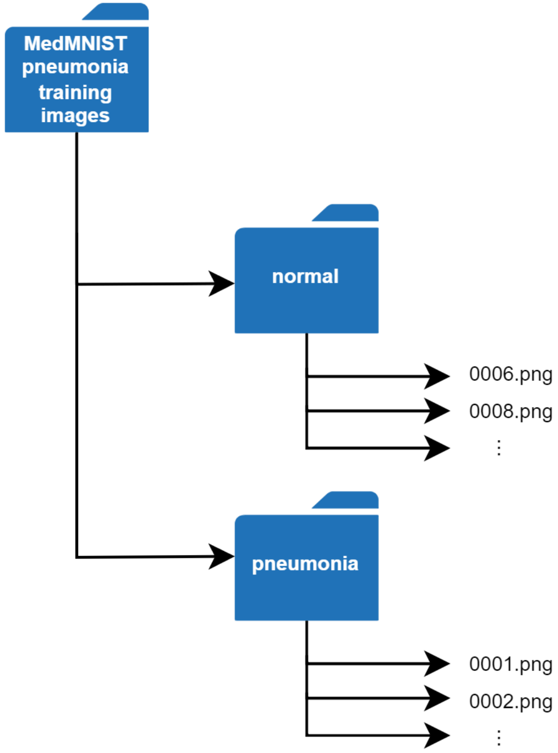




海量数据的便捷管理




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

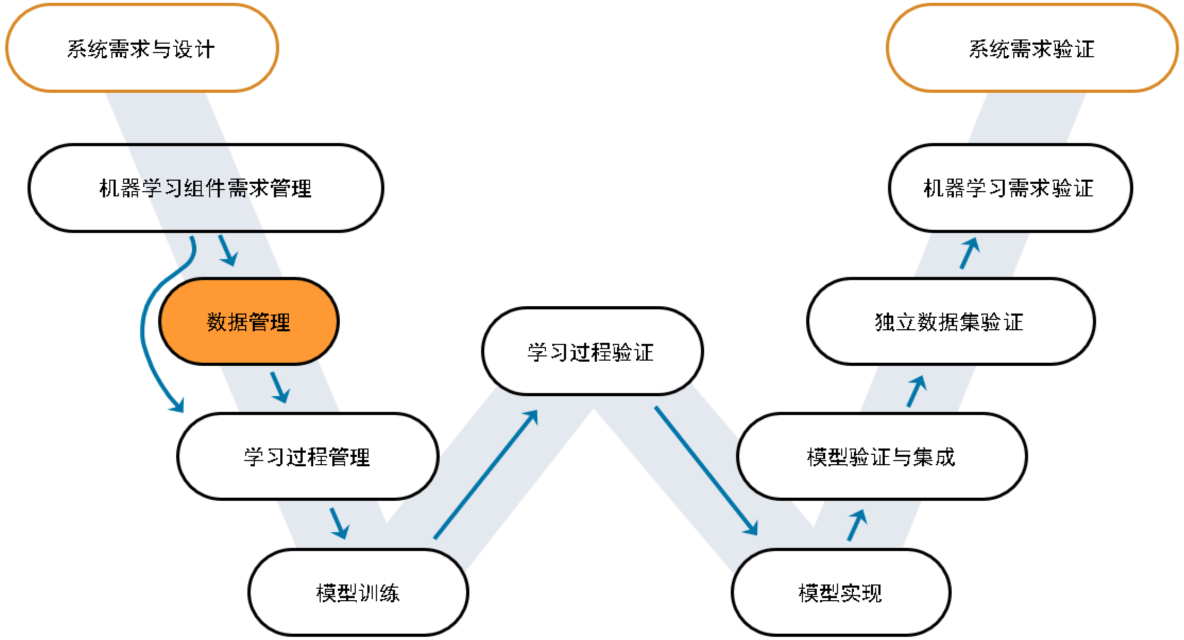



```
countEachLabel(imdsTrain)
```

ans =

2×2 table

Label	Count
normal	1214
pneumonia	3494



AI模型的偏见（Bias）

Bias的来源

数据不足，Bias因所选数据集而来

继承性偏见

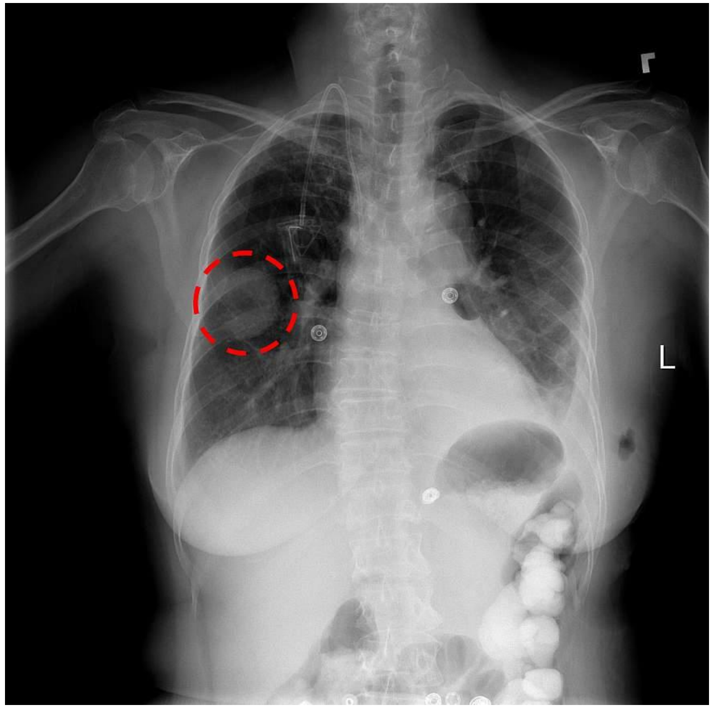
模型问题

Fairness in Responsible AI: Detecting and mitigating bias against unprivileged groups in ML modeling

身边的医疗设备可能有偏见!

Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis

Agostina J. Larrazabal[a,1], Nicolás Nieto[a,b,1], Victoria Peterson[b,c], Diego H. Milone[a], and Enzo Ferrante[a,2]



Courtesy : *PNAS*

From oximeters to AI, where bias in medical devices may lurk

Analysis: issues with some gadgets could contribute to poorer outcomes for women and people of colour

Some research suggest that oximeters work less well for patients with darker skin. Photograph: Grace Cary/Getty Images

Courtesy : *The Guardian*

Fixing Medical Devices That Are Biased against Race or Gender

Designers should show how well instruments perform across different populations

Courtesy : *Scientific American*

美国科学院院刊 PNAS

PNAS ARTICLES FRONT MATTER AUTHORS TOPICS +

BRIEF REPORT | APPLIED MATHEMATICS

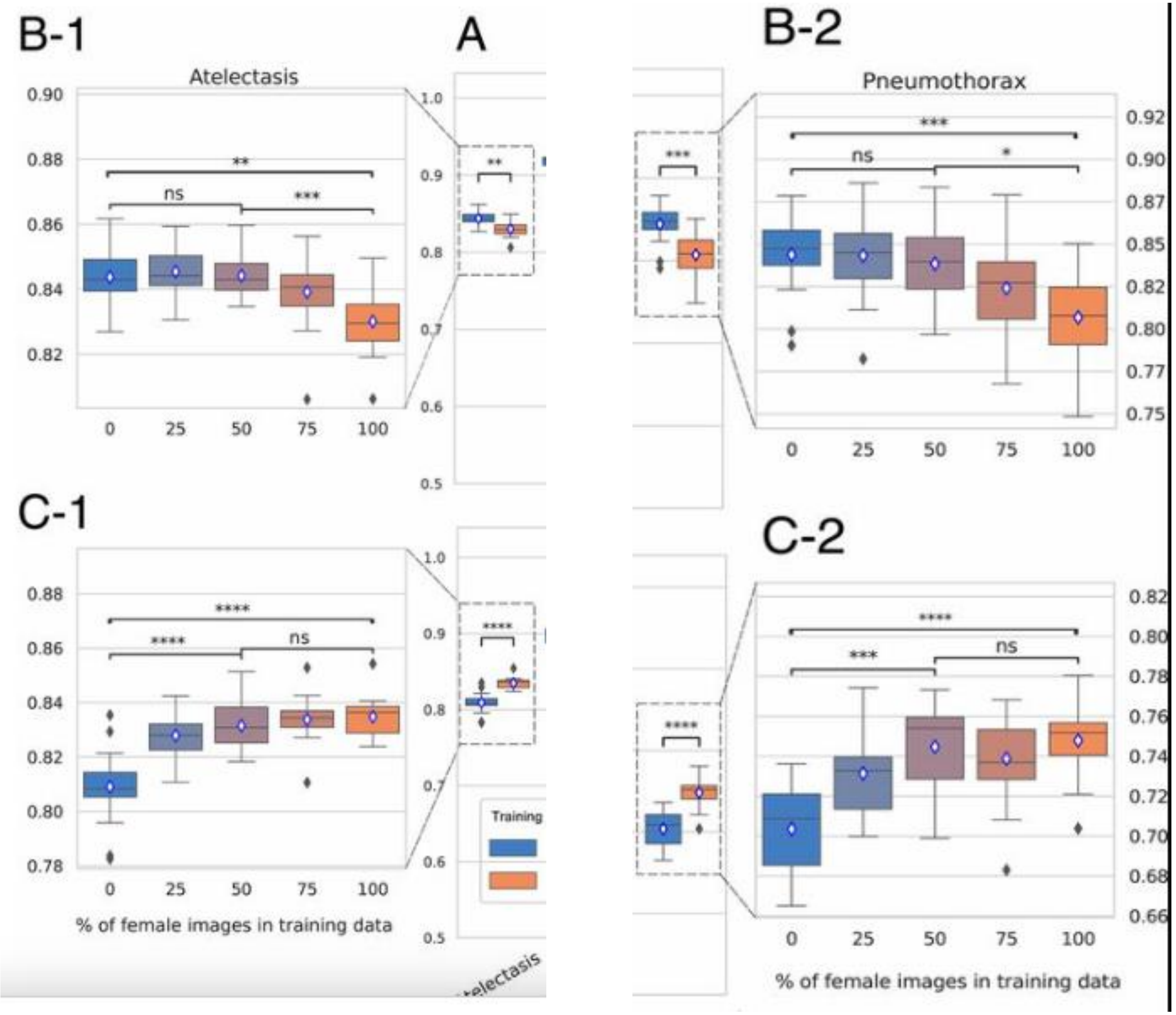
Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis

Agostina J. Larrazabal, Nicolás Nieto, Victoria Peterson, +1, and Enzo Ferrante Authors Info & Affiliations

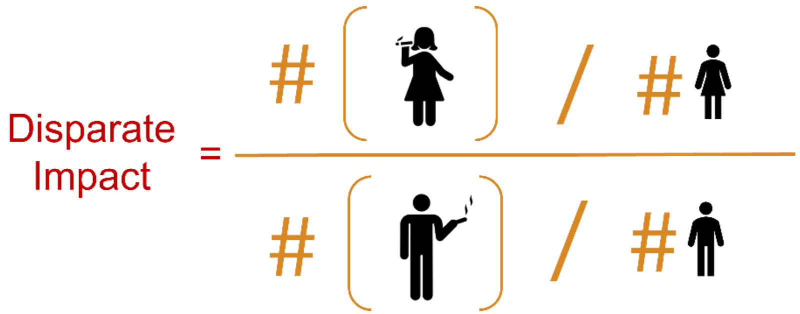
Edited by David L. Donoho, Stanford University, Stanford, CA, and approved April 30, 2020 (received for review October 30, 2019)

May 26, 2020 | 117 (23) 12592-12594 | https://doi.org/10.1073/pnas.1919012117

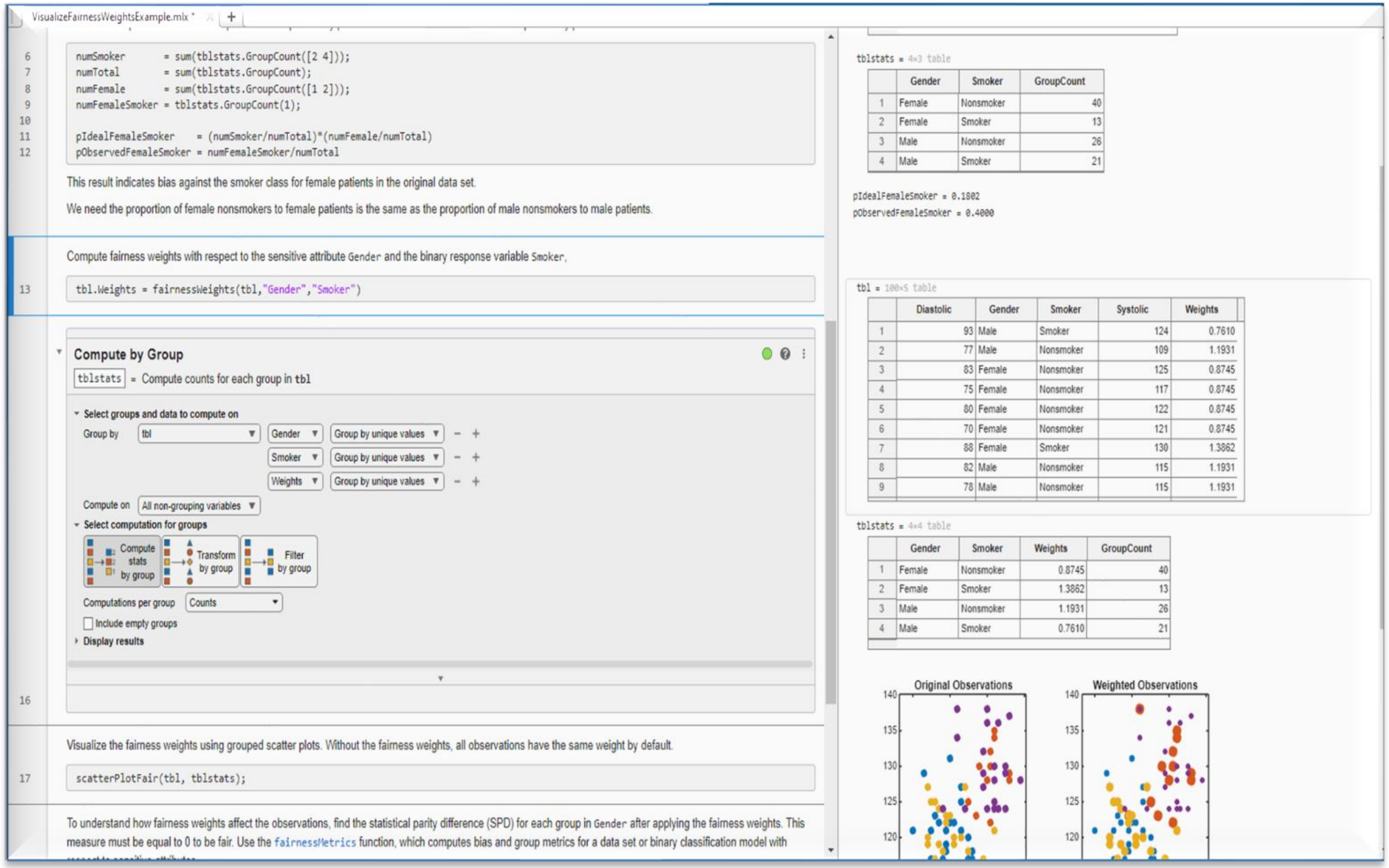
for models trained only with male images, while orange boxes indicate training with female-only images. Both models are evaluated over male-only (Fig. 1 *A*, *Top*) and female-only (Fig. 1 *A*, *Bottom*) test images. A consistent decrease in performance is observed when using male patients for training and female for testing (and vice-versa). The same tendency was confirmed when evaluating three different deep learning architectures in two X-ray datasets with different pathologies.



均衡性度量- Detect bias



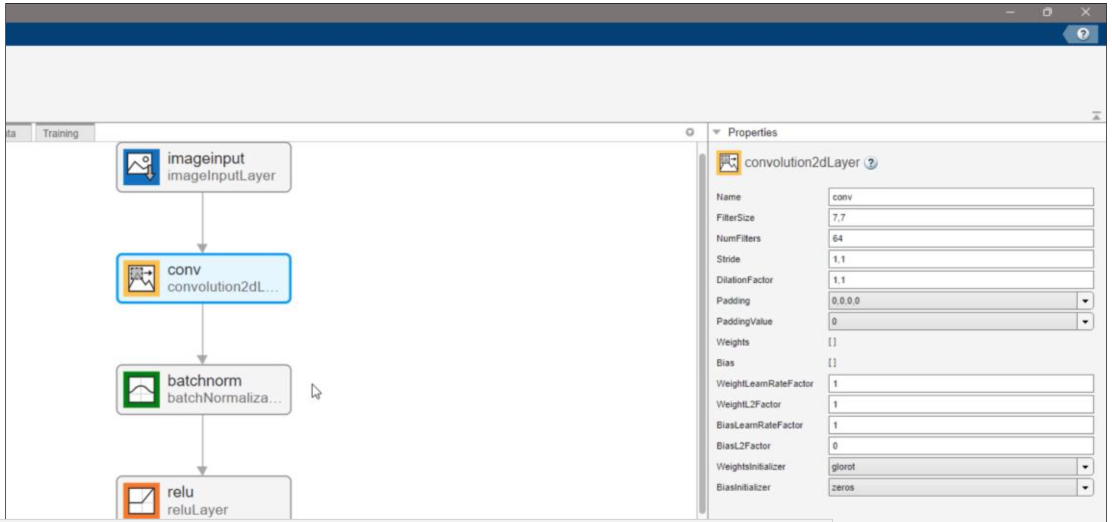
Disparate Impact < 1 for females indicating bias



Bias 检测与降低

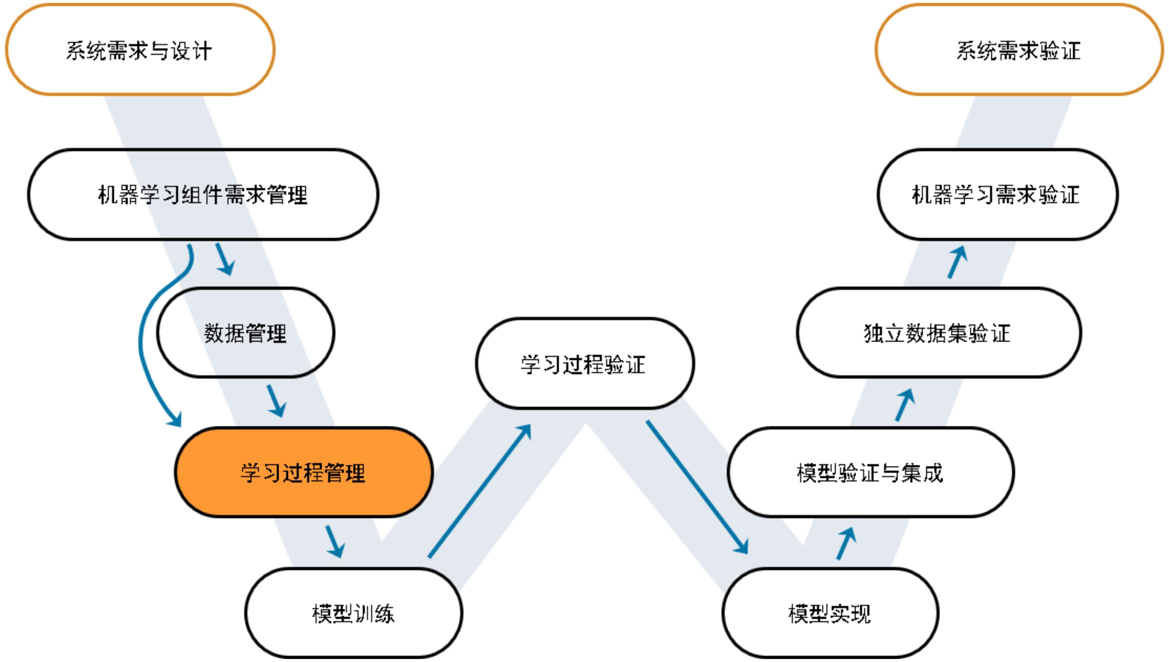
Stage	Description
Pre-processing	Removes the information correlated to the sensitive attribute
In-processing	Add constraint or regularization term to the objective, Adversarial models
Post-Processing	Edit posteriors to satisfy fairness constraints

模块化快速搭建网络模型

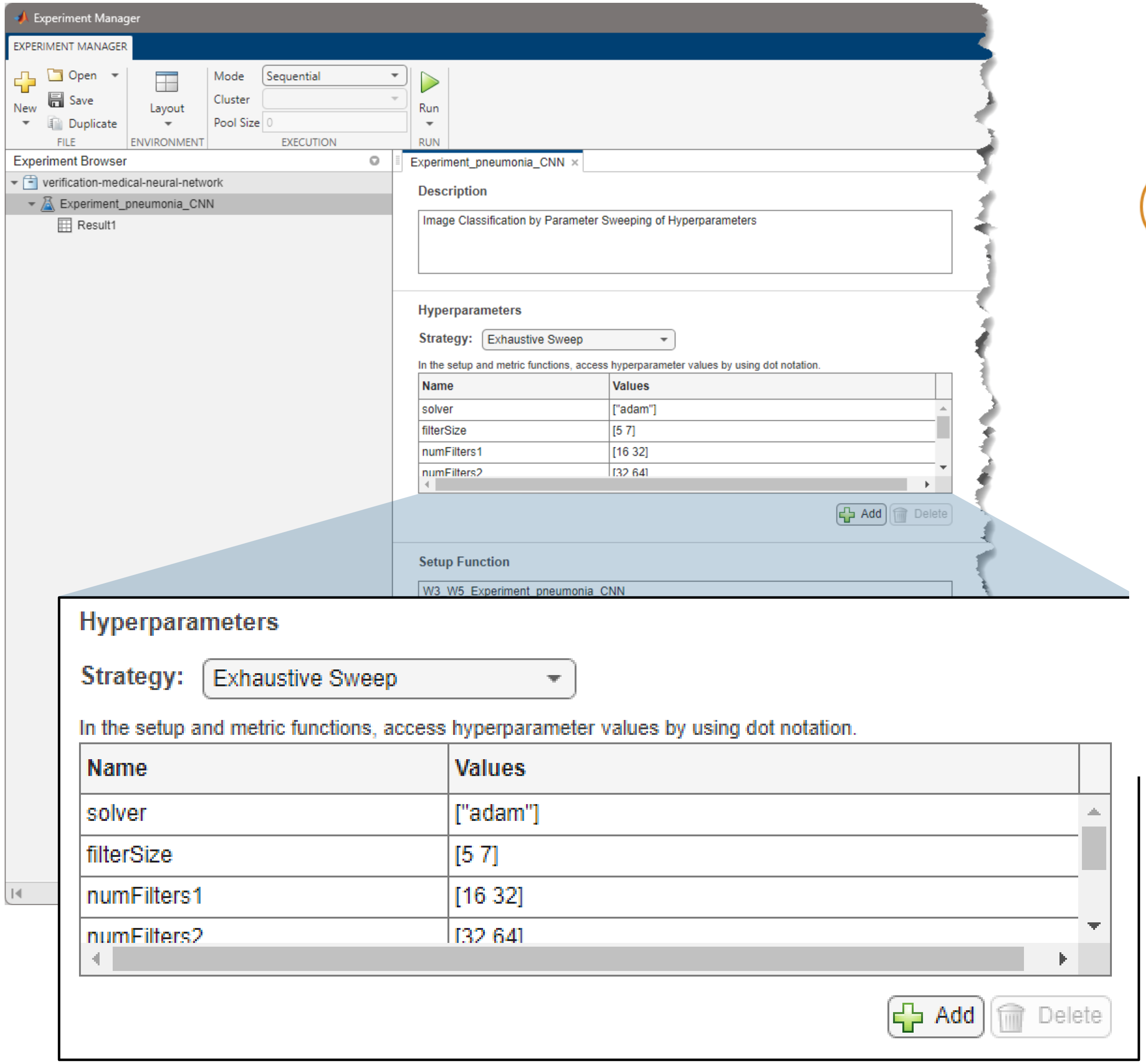


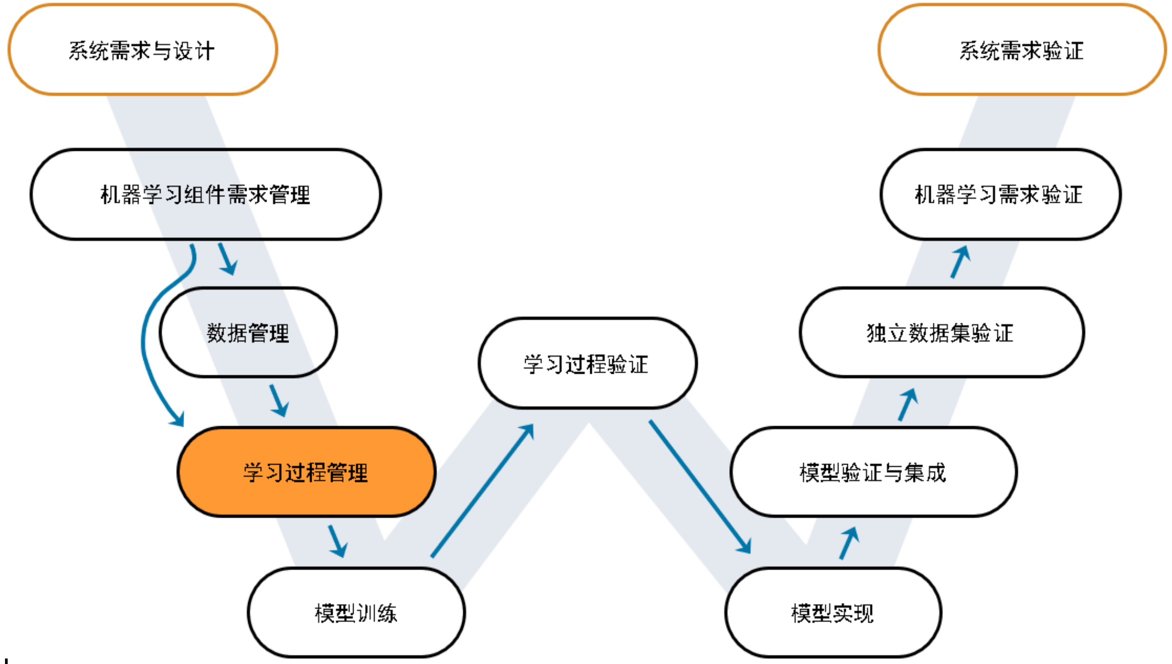

```
numClasses = numel(classNames);
layers = [
    imageInputLayer(imageSize,Normalization="none")
    convolution2dLayer(7,64,Padding=0)
    batchNormalizationLayer()
    reluLayer()
    dropoutLayer(0.5)
    averagePooling2dLayer(2,Stride=2)
    convolution2dLayer(7,128,Padding=0)
    batchNormalizationLayer()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



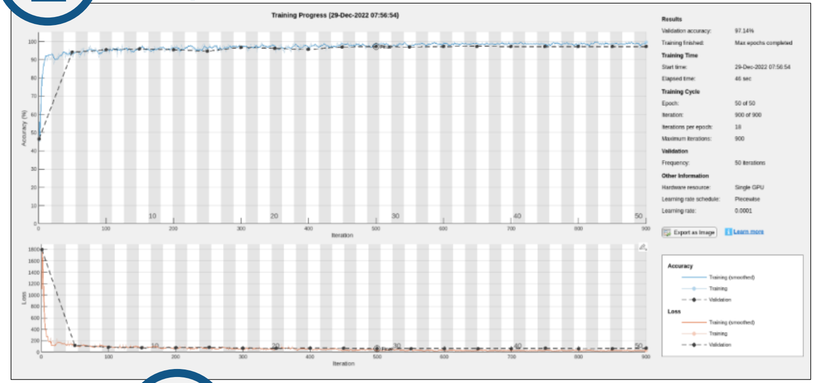
超参调优


Experiment Manager
EXPERIMENT MANAGER
New
Open
Save
Duplicate
FILE
Layout
ENVIRONMENT
Mode
Sequential
Cluster
Pool Size
EXECUTION
Run
RUN
Experiment Browser
verification-medical-neural-network
Experiment_pneumonia_CNN
Result1
Experiment_pneumonia_CNN
Description
Image Classification by Parameter Sweeping of Hyperparameters
Hyperparameters
Strategy:
Exhaustive Sweep
In the setup and metric functions, access hyperparameter values by using dot notation.
Name
Values
solver
["adam"]
filterSize
[5 7]
numFilters1
[16 32]
numFilters2
[32 64]
Add
Delete
Setup Function
W3_W5_Experiment_pneumonia_CNN


系统需求与设计
机器学习组件需求管理
数据管理
学习过程管理
模型训练
学习过程验证
模型实现
模型验证与集成
独立数据集验证
机器学习需求验证
系统需求验证

渐进式优化迭代得到高精度、高可靠性模型

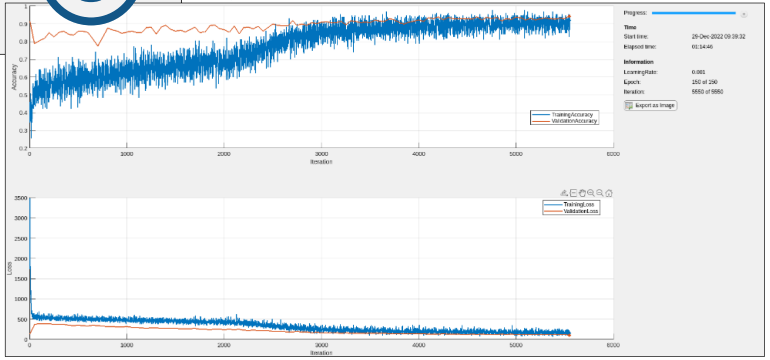
1 初始训练

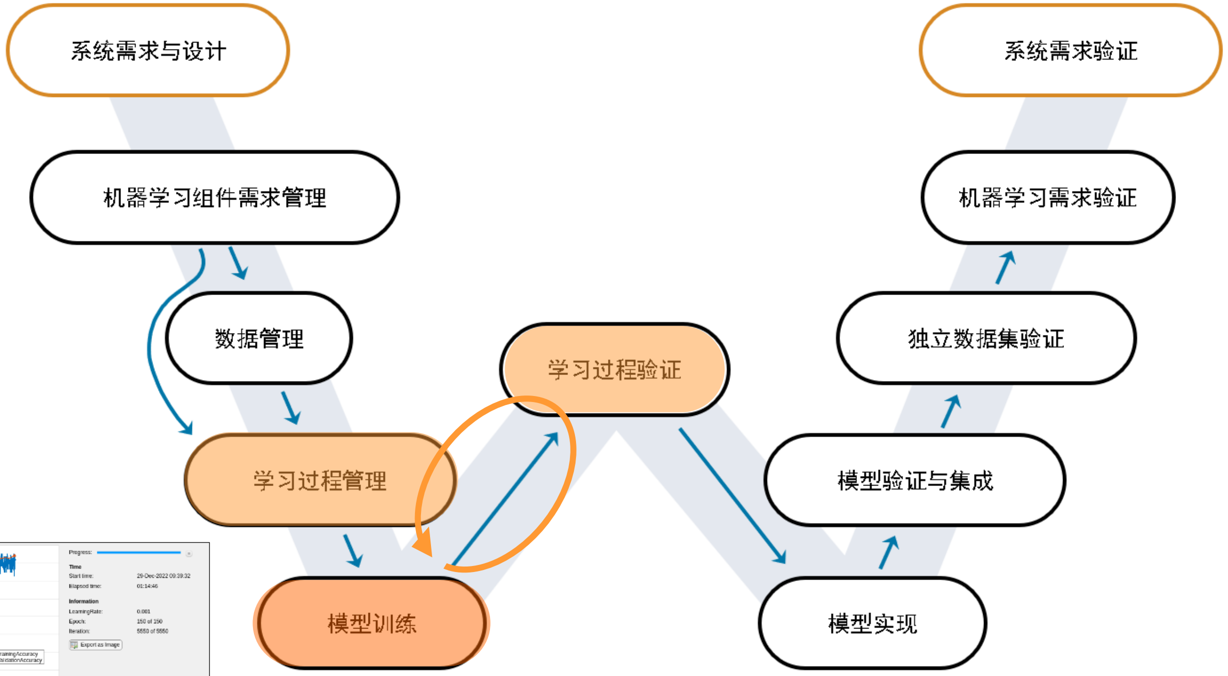


2 数据增强训练

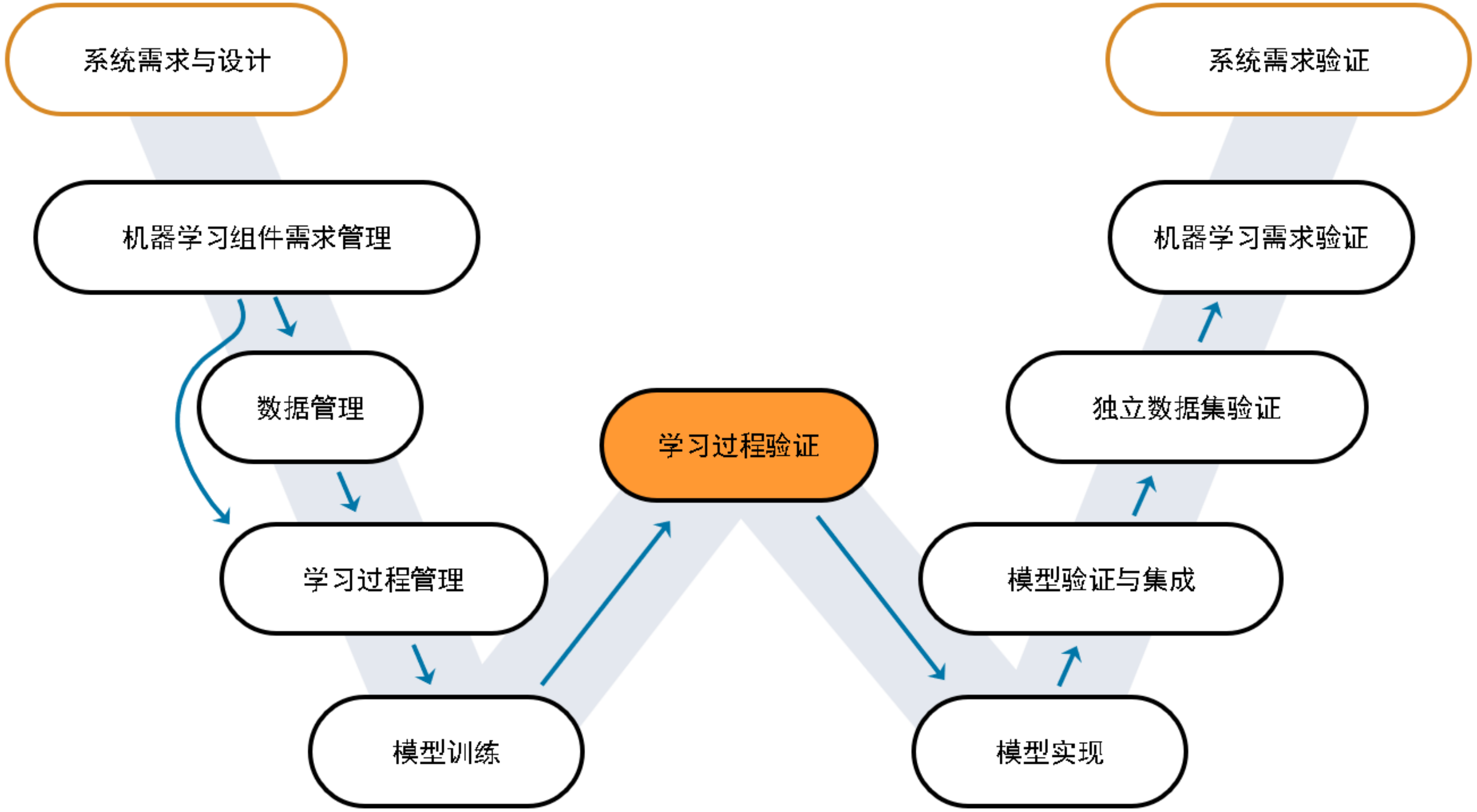
```
imageAugmenter = imageDataAugmenter(...
    FillValue=mean(XTrain(:)), ...
    RandXReflection=true, ...
    RandXTranslation=[-2,2], ...
    RandYTranslation=[-2,2], ...
    RandRotation=[-10,10],...
    RandScale=[1,1.25], ...
    RandXShear=[-5,5], ...
    RandYShear=[-5,5]);
```

3 对抗性训练

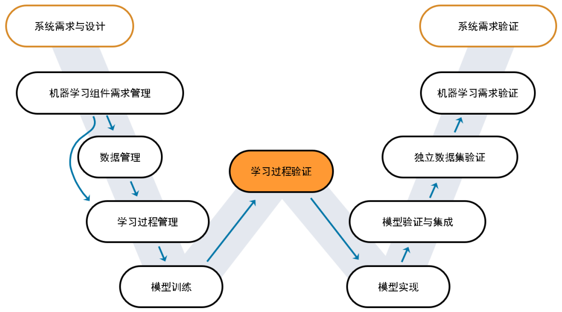




学习过程验证

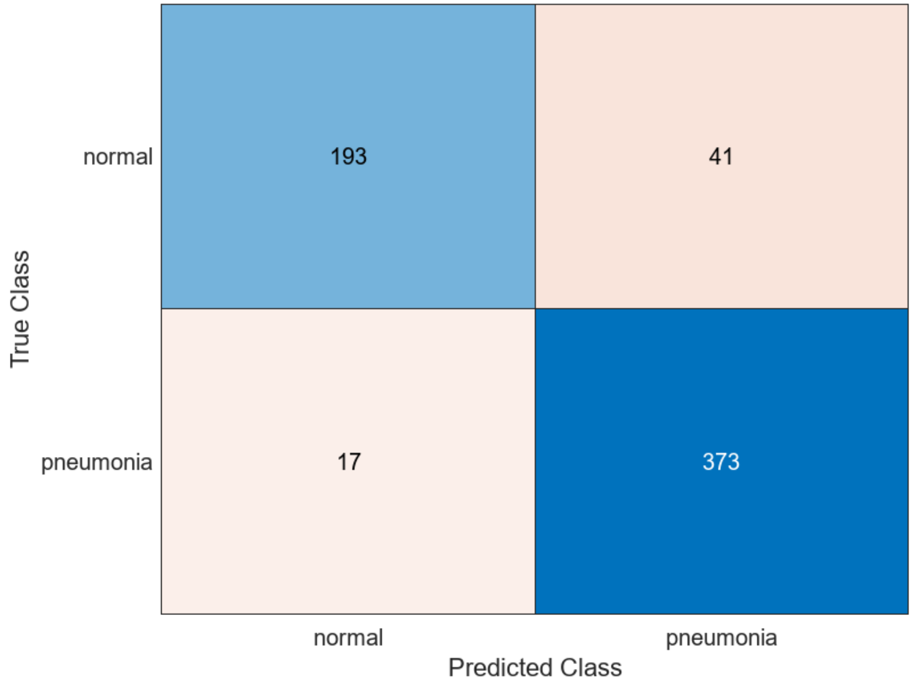

系统需求与设计
机器学习组件需求管理
数据管理
学习过程管理
模型训练
学习过程验证
模型实现
模型验证与集成
独立数据集验证
机器学习需求验证
系统需求验证

基于独立数据集的模型性能测试与理解

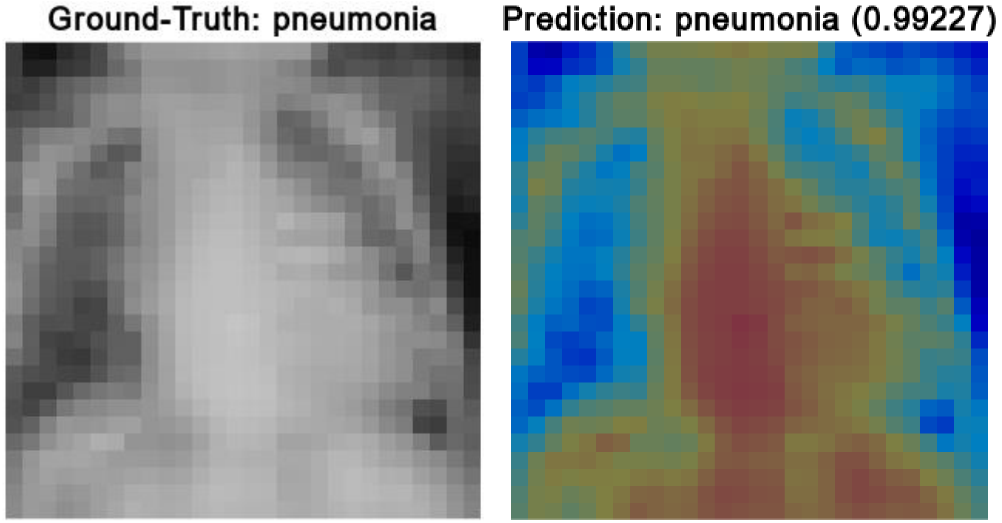



```
Accuracy: 90.71%
```

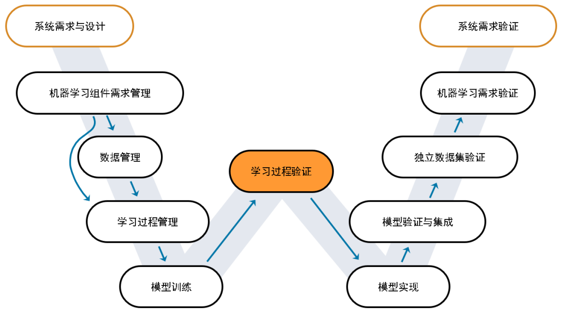
```
confusionchart(T,Y)
```

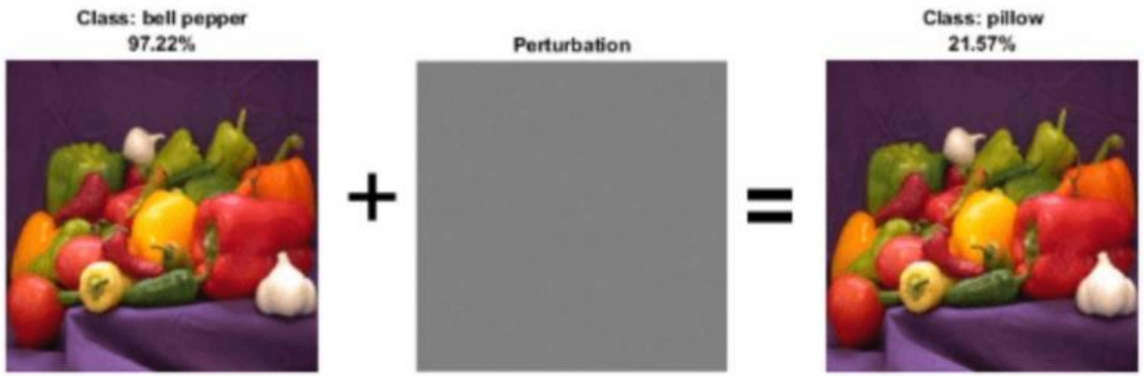



```
scoreMap = gradCAM(net,X,label)
```



神经网络的鲁棒性验证

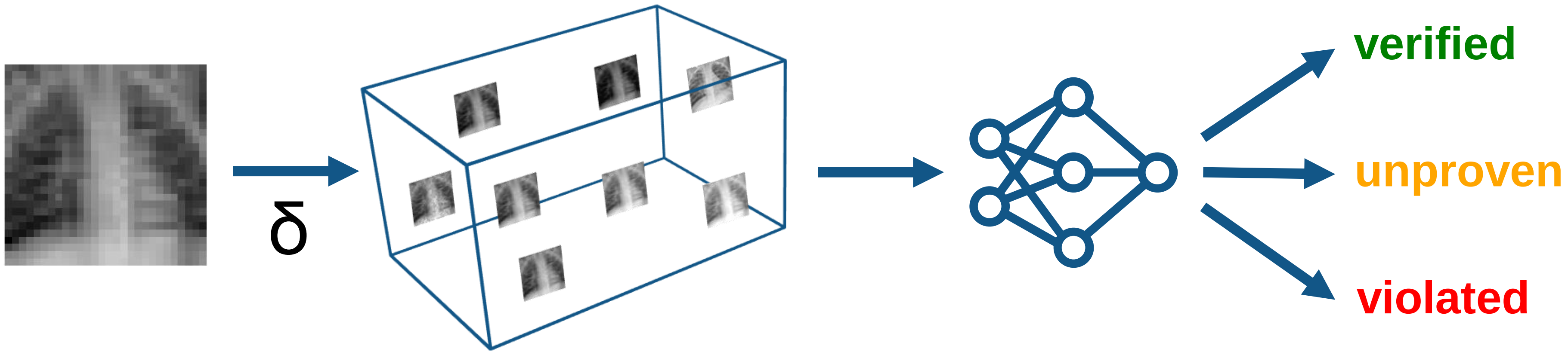




Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team STAFF

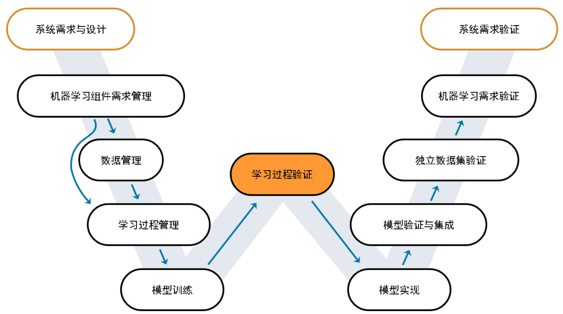
Verify and test robustness of deep learning networks



形式化验证

MATLAB EXPO

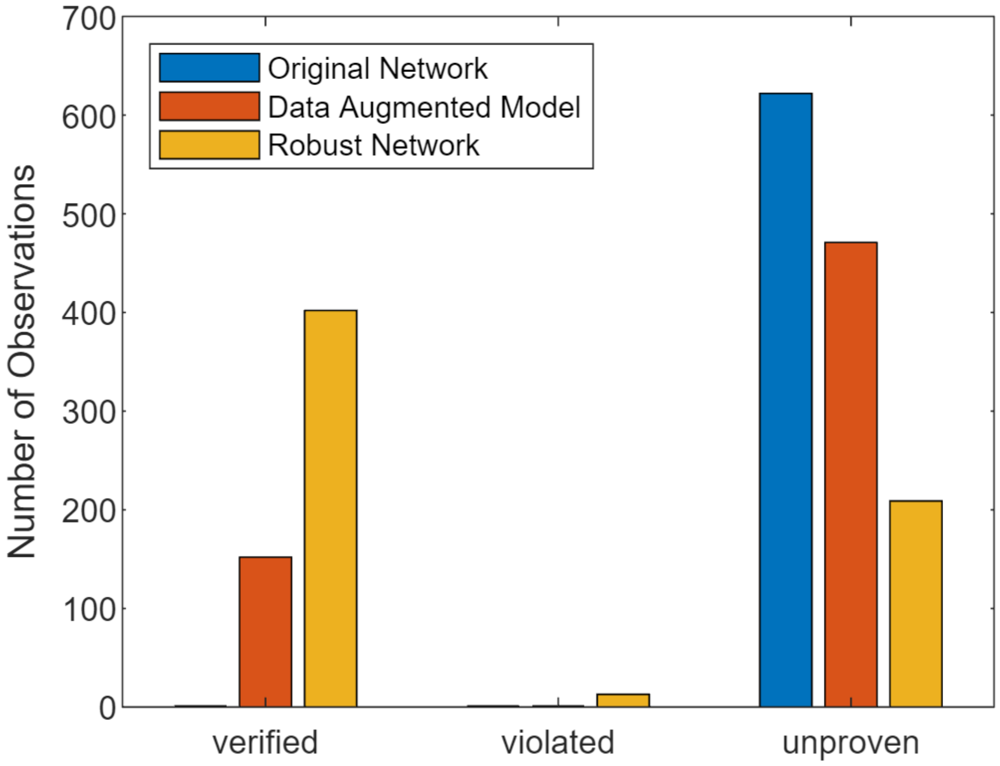
神经网络的鲁棒性验证



Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team STAFF

Verify and test robustness of deep learning networks

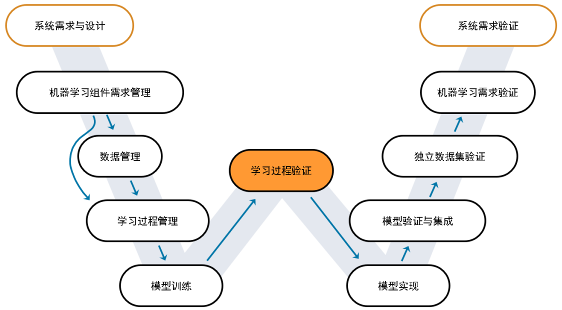



```
perturbation = 0.01;
XLower = XTest - perturbation;
XUpper = XTest + perturbation;
XLower = dlarray(XLower,"SSCB");
XUpper = dlarray(XUpper,"SSCB");
result = verifyNetworkRobustness(net,...
    XLower,XUpper,TTest);
```

```
summary(result)
```

verified	402
violated	13
unproven	209

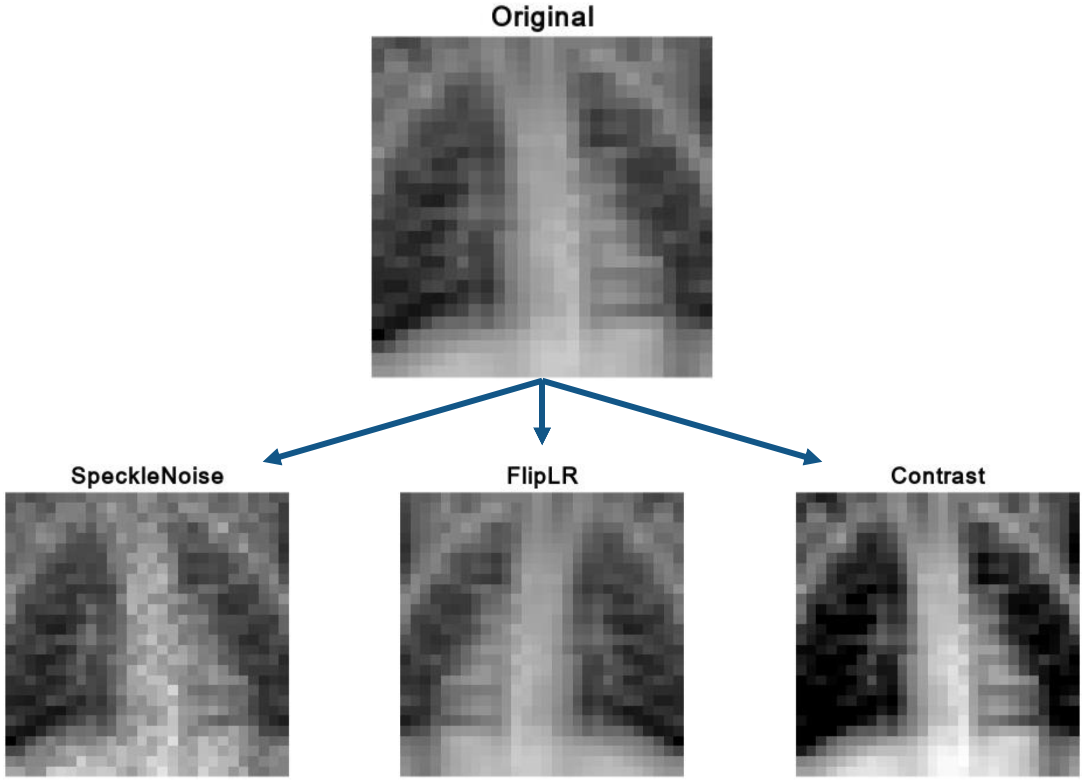
检测分布外样本，拒收或转给专家复核



Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team STAFF

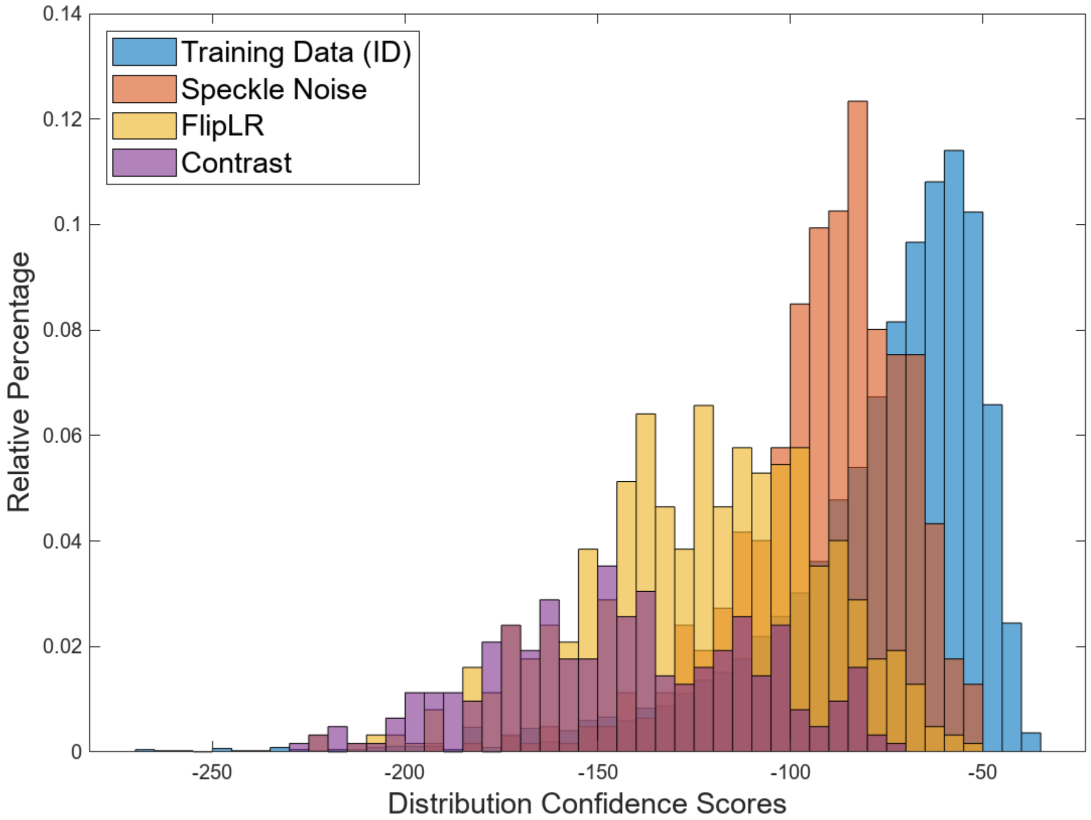
Verify and test robustness of deep learning networks



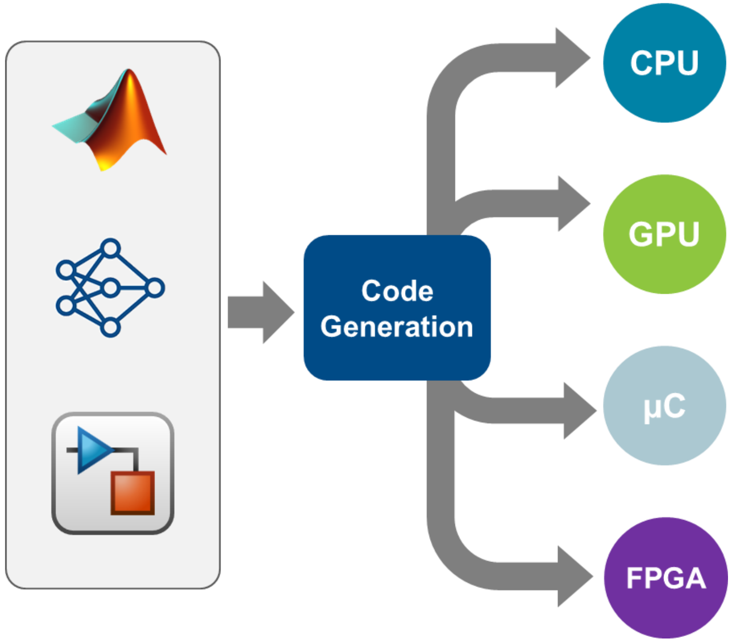
In-distribution

Out-of-distribution

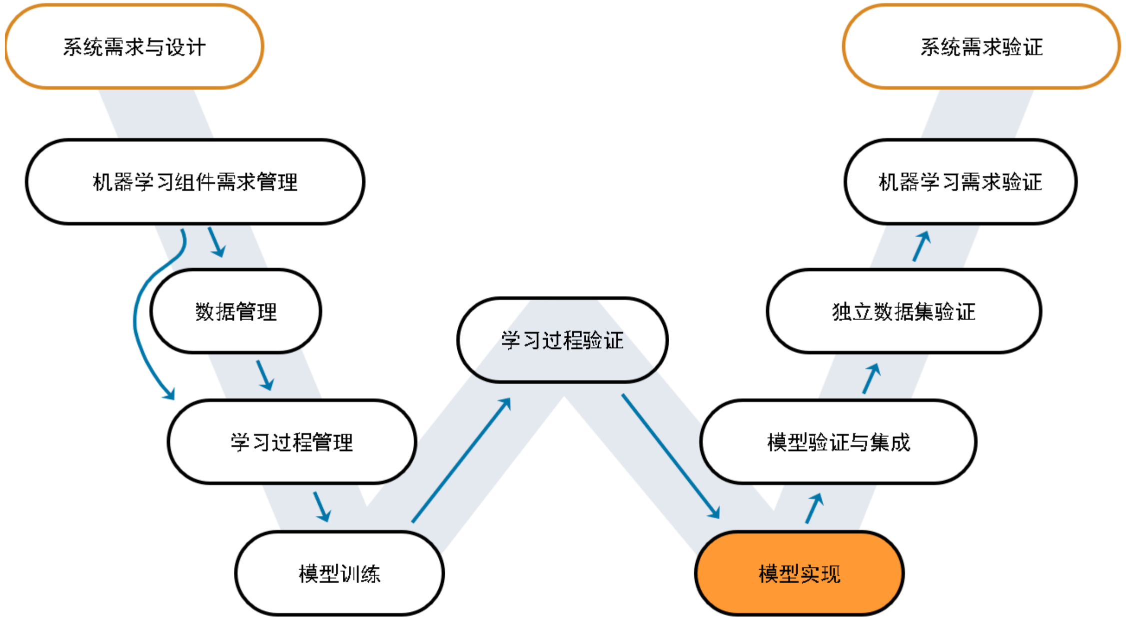
Out-of-distribution



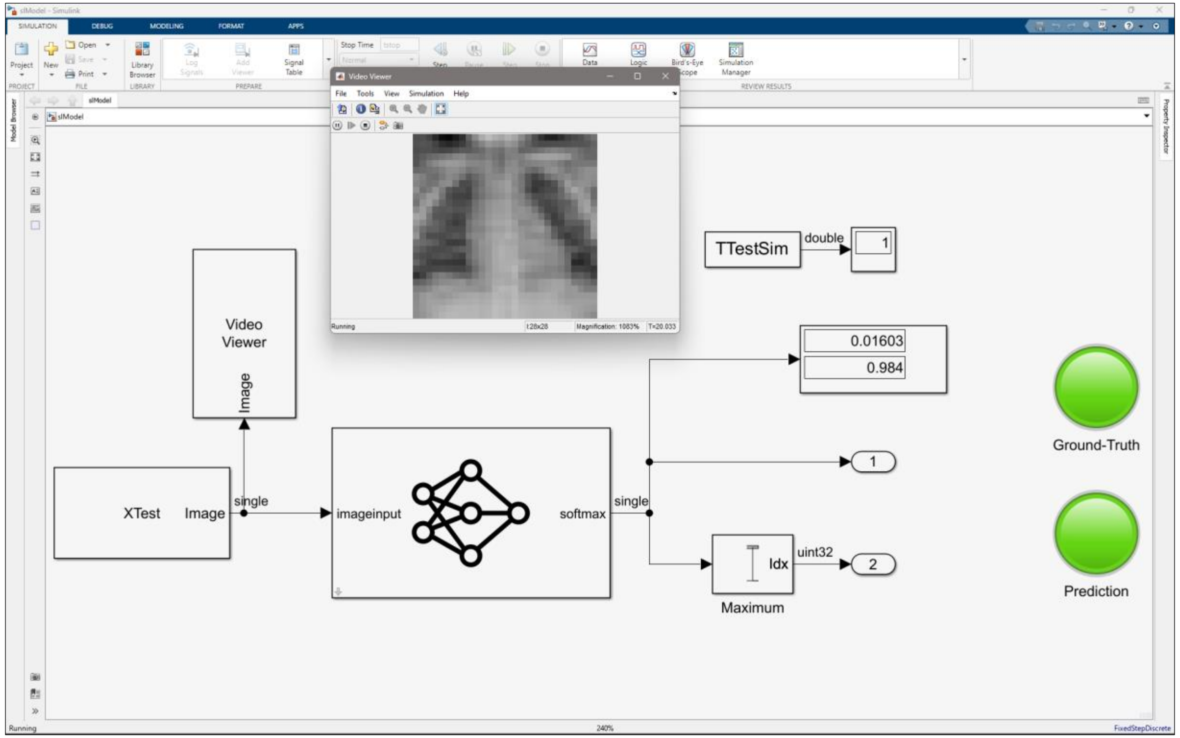
无bug自动部署至目标硬件

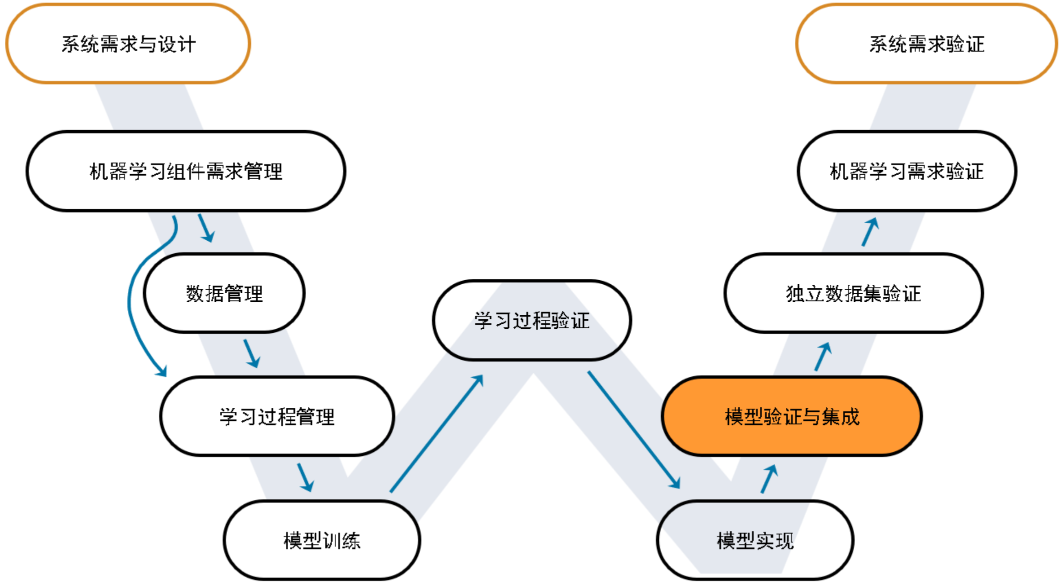



```
analyzeNetworkForCodegen(net)

                    Supported
                    _________

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

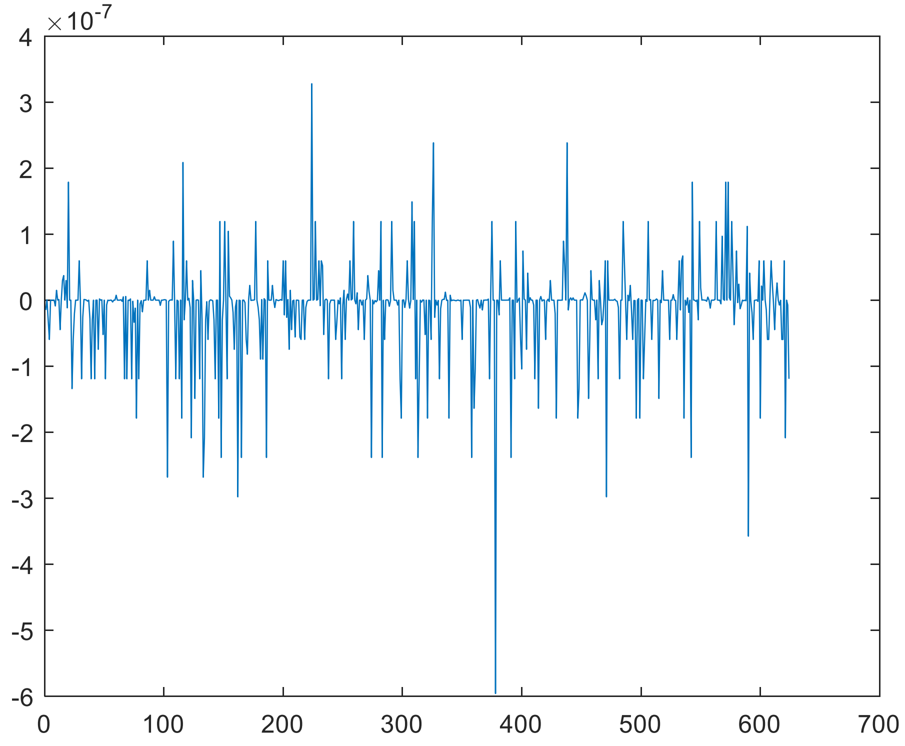


将AI模块集成在Simulink中，进行系统级仿真测试

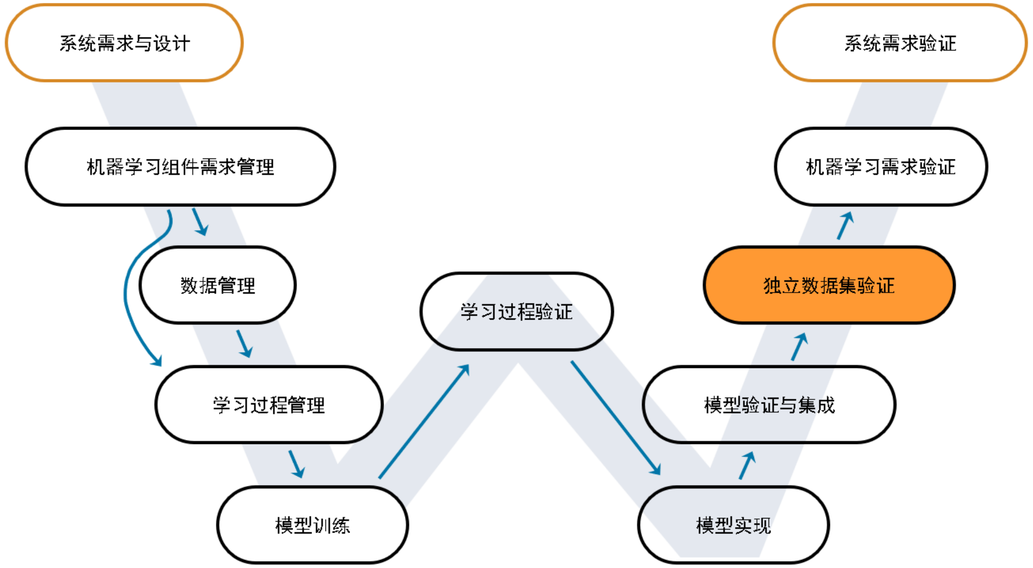




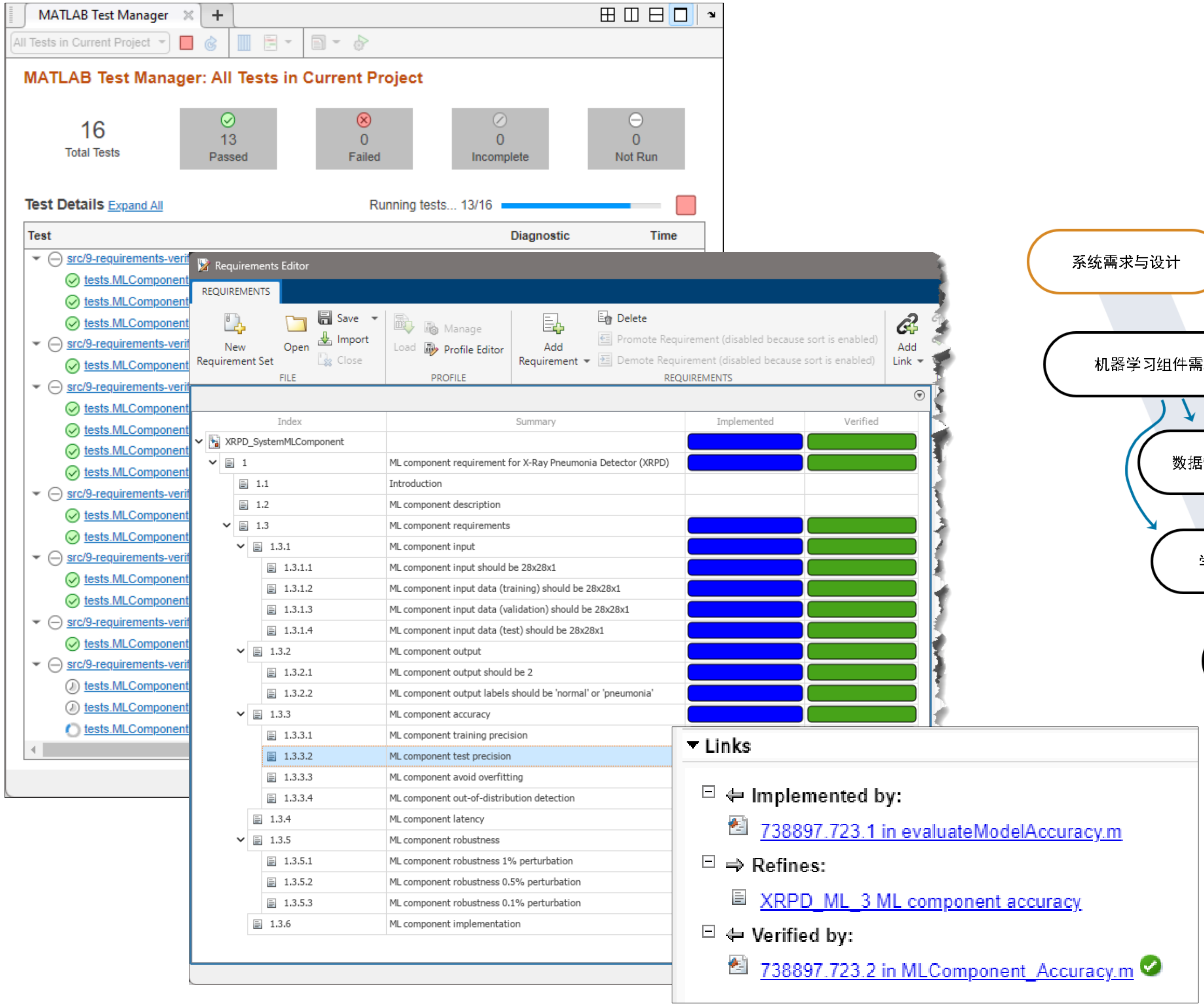
开发与推理模型完全无差别

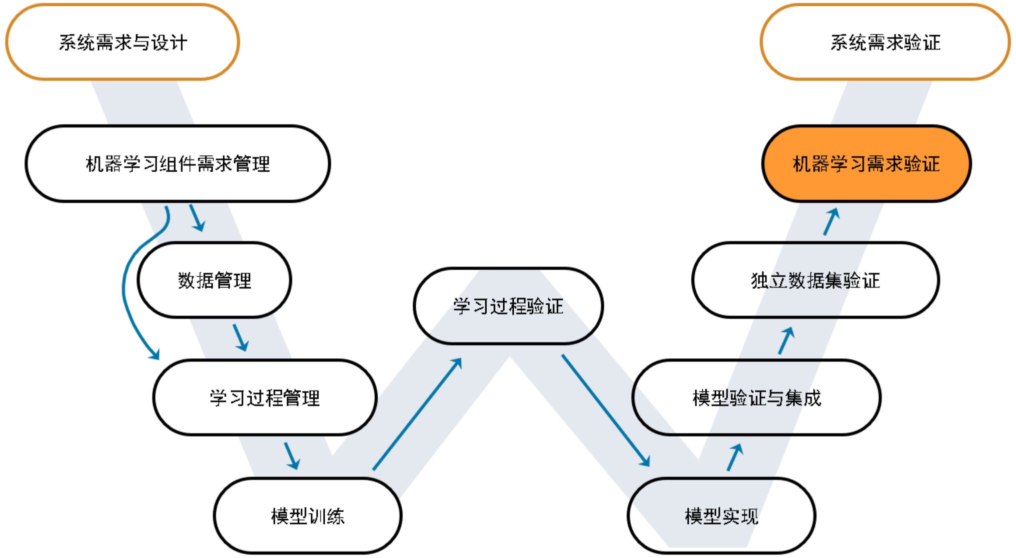



```
max(abs(differences))

ans = single
5.9605e-07
```



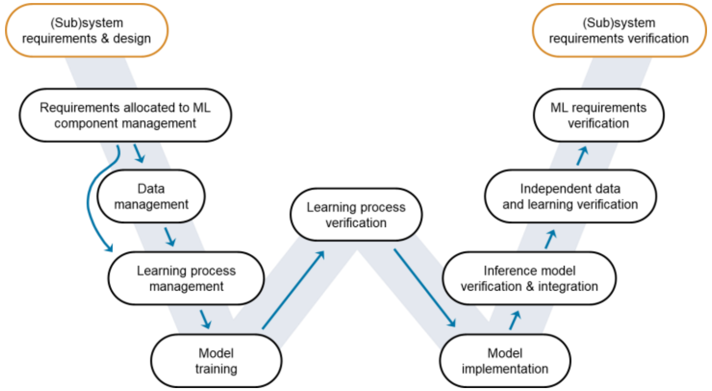
Verifying requirements have been fully tested





要点

MathWorks 提供高安全性AI开发W流程各阶段的支持



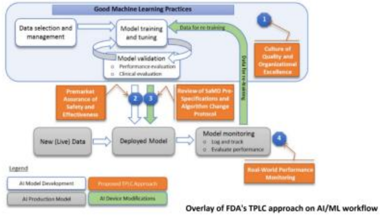
神经网络模型
鲁棒性测试与验证专用库

Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team **STAFF**

Verify and test robustness of deep learning networks

高安全性验证的经验助力推动全新AI标准



Overlay of FDA's TPLC approach on AI/ML workflow

EUROCAE WG-114 / SAE G-34
Standardization Working Group
"Artificial Intelligence in Aviation"

MATLAB EXPO

Thank you

MathWorks