

# MATLAB EXPO

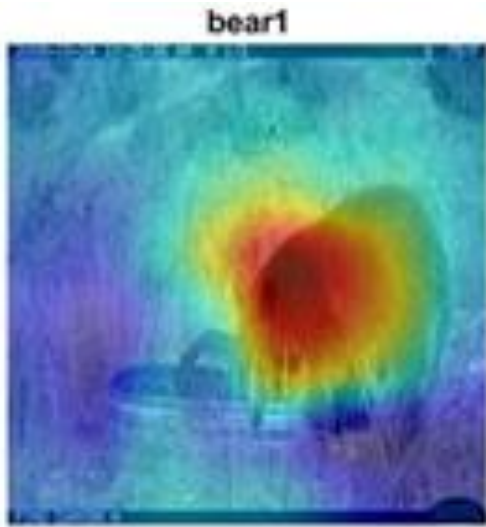
## AI的高安全性应用与可靠性验证 ——医学影像分析与可解释性

MathWorks

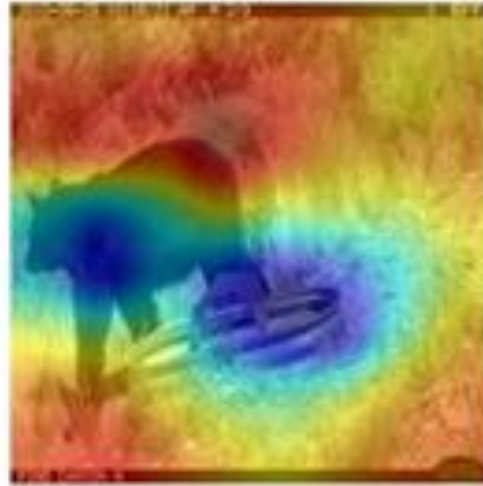
中国区医疗行业市场经理  
单博



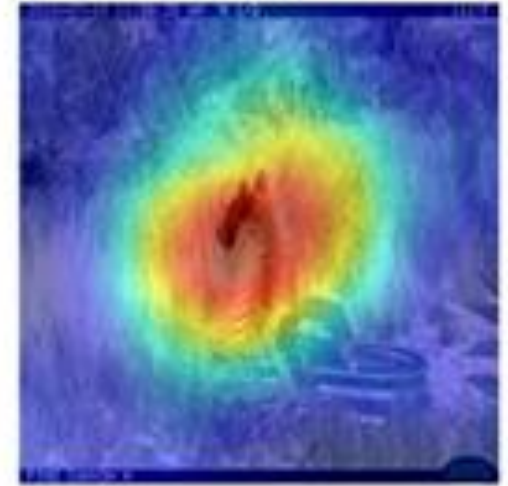
net1



bear2



bear3

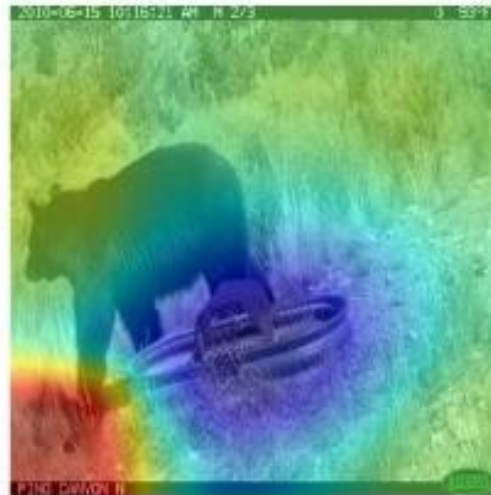


net2

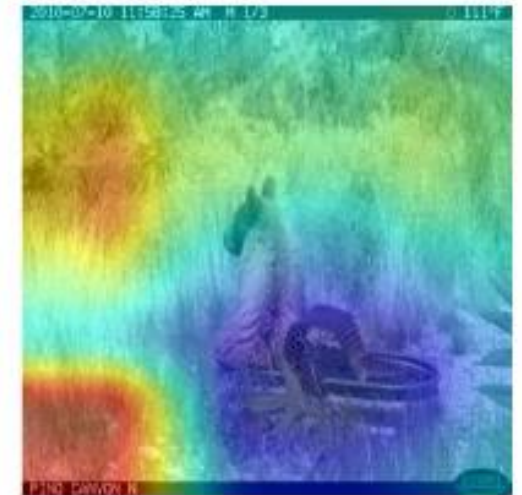
bear1



bear2



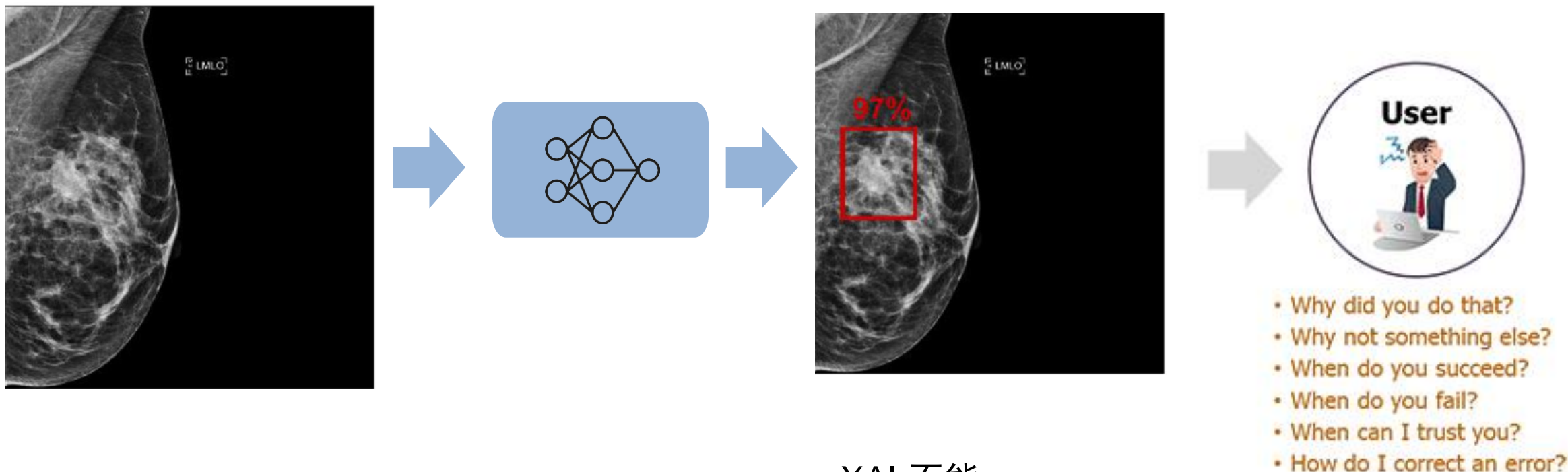
bear3



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：

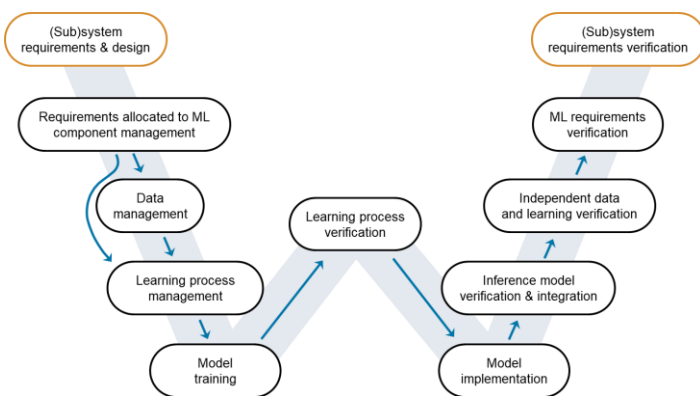
- 解析方式解释AI模型
- 代替对于高安全模型具有关键意义的Good Machine Learning Practices (GMLP),

# 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

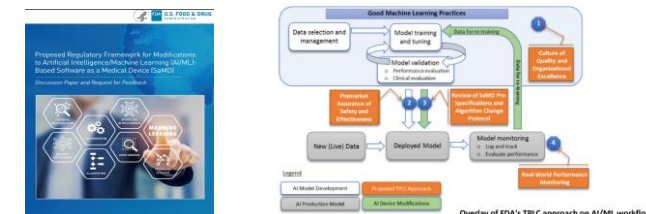
高安全性验证的经验助力推动全新AI标准



## Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team **STAFF**

Verify and test robustness of deep learning networks



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

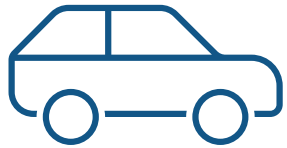
随AI的使用量在快速增加，迫切需要在高安全性领域，对其解释、确认和验证。



# 对包含AI组件系统的确认与验证的挑战

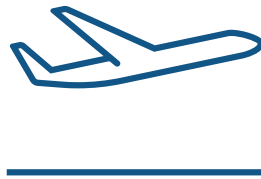


# 工业界正在努力推动高安全性系统中AI模型的验证 提供白皮书, 标准及计划



汽车

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



航空

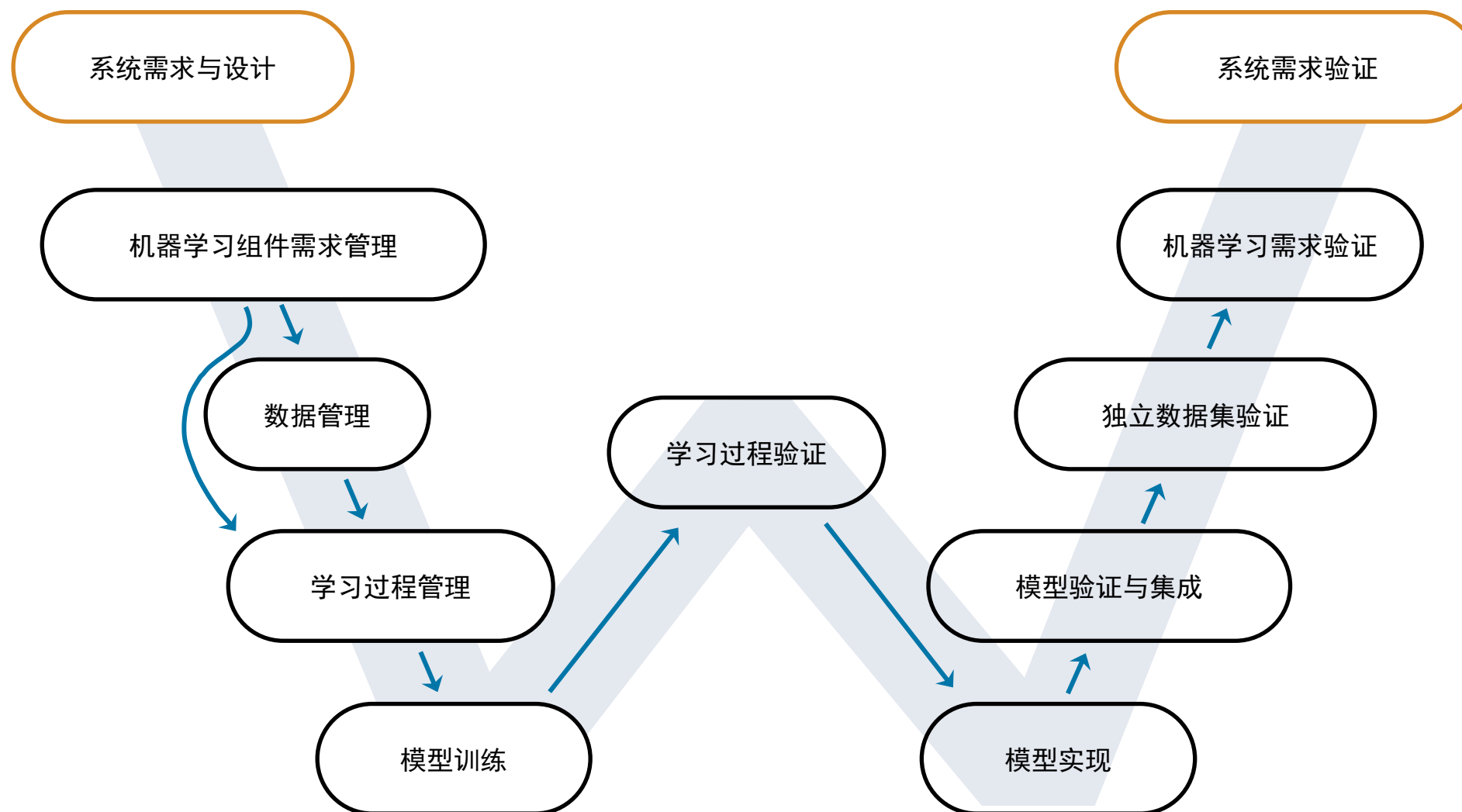
全新标准([AS6983](#)) from  
[EUROCAE WG-114 / SAE G-34](#)  
预计2024发布



医疗

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

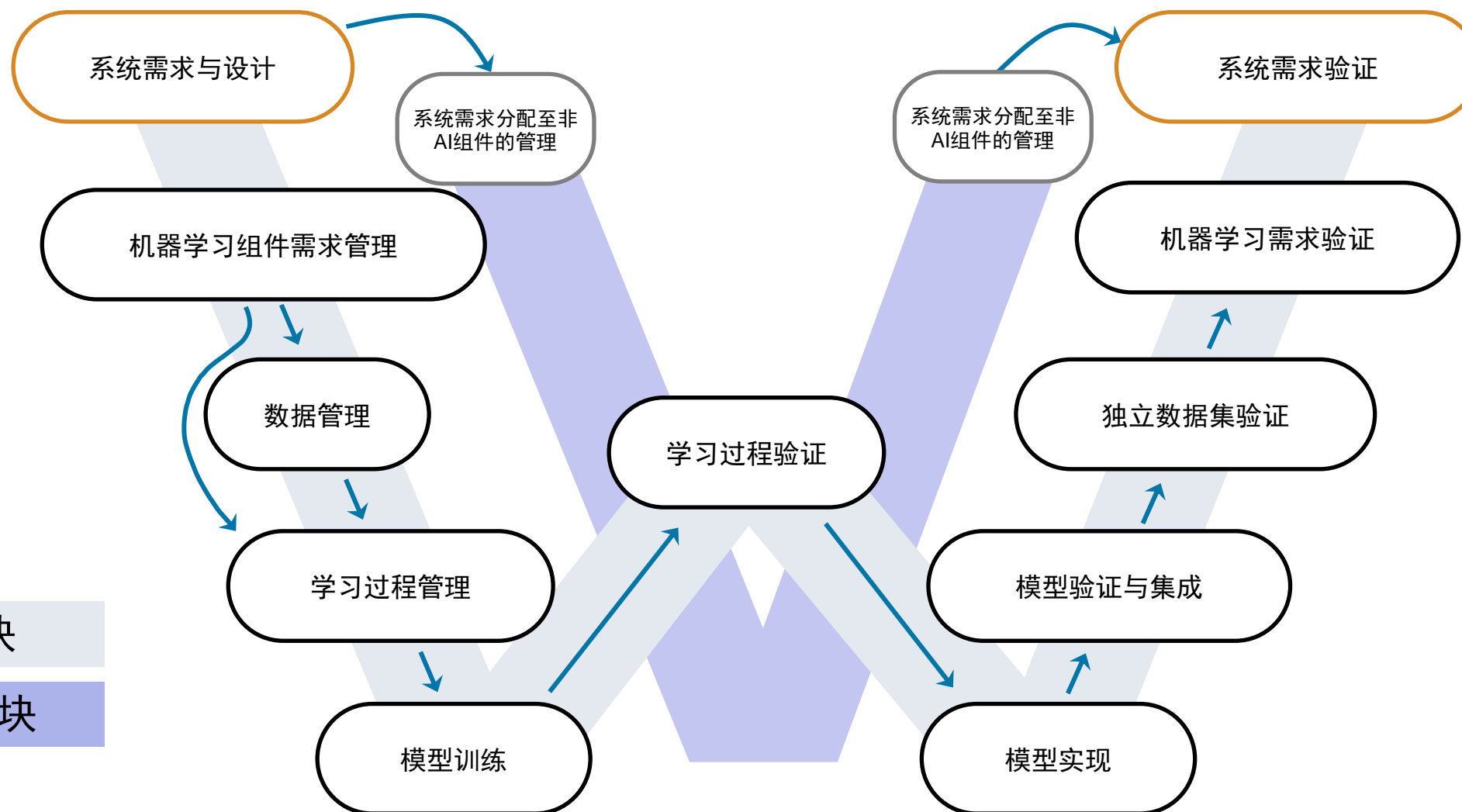
# W型流程将V流程的应用延伸到AI的应用



Credit: EASA, Daedalean



# W型流程可与非AI模块的V流程共存



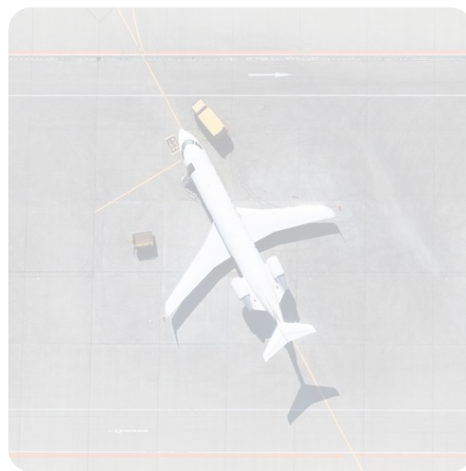
Credit: EASA, Daedalean

# Task: 验证一个图像分类网络

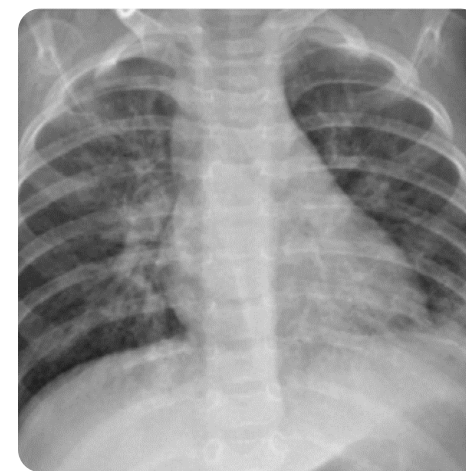
汽车



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

<sup>1</sup> Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup> Boston College, Chestnut Hill, MA

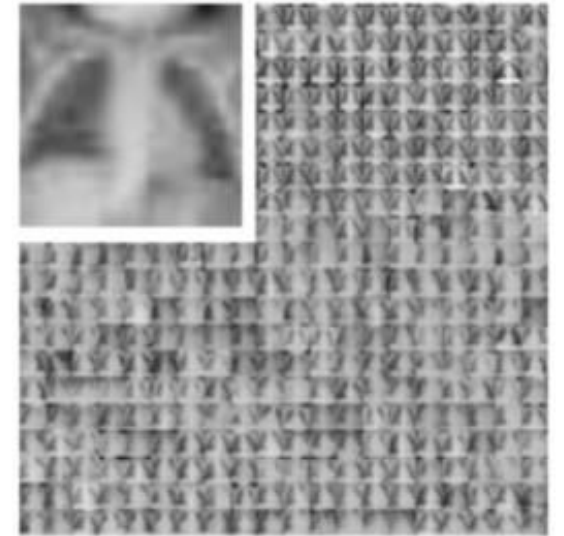
<sup>3</sup> RWTH Aachen University, Aachen, Germany

<sup>4</sup> Fudan Institute of Metabolic Diseases, Zhongshan Hospital, Fudan University, Shanghai, China

<sup>5</sup> Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

<sup>6</sup> Harvard University, Cambridge, MA

## PneumoniaMNIST



# 始于与机器学习相关的需求收集

**Requirements Editor**

REQUIREMENTS

FILE PROFILE REQUIREMENTS LINKS

Index	Summary	Implemented	Verified
> XRPD_System			
> 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		

**Requirement: XRPD\_ML\_3\_2**

**Properties**

Type: Functional

Index: 1.3.3.2

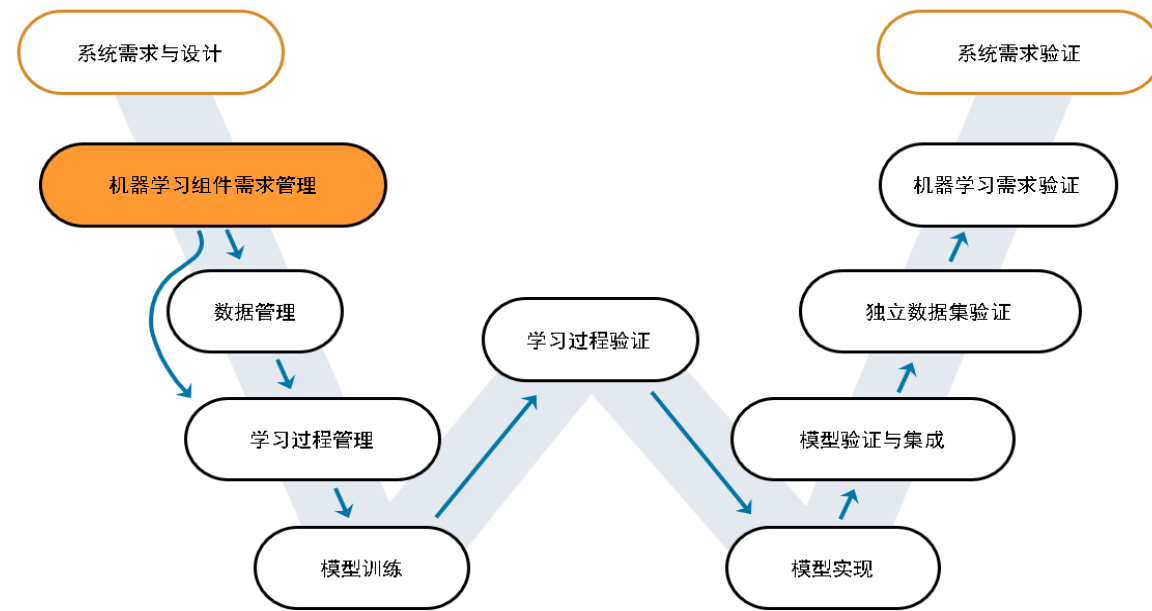
Custom ID: XRPD\_ML\_3\_2

Summary: ML component test precision

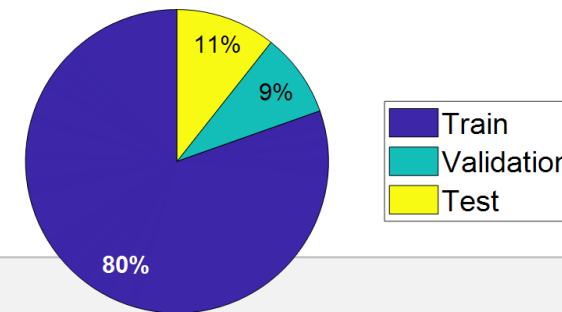
Description Rationale

Arial 10

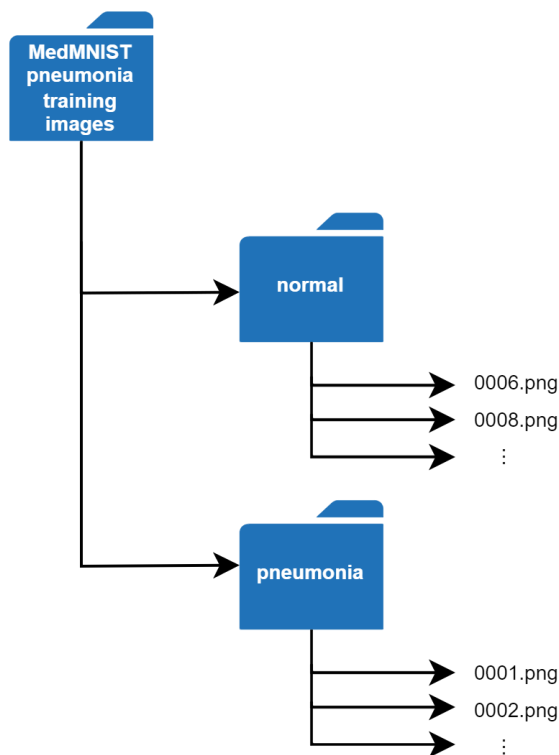
Accuracy of the trained model must be above 90% (with test data)



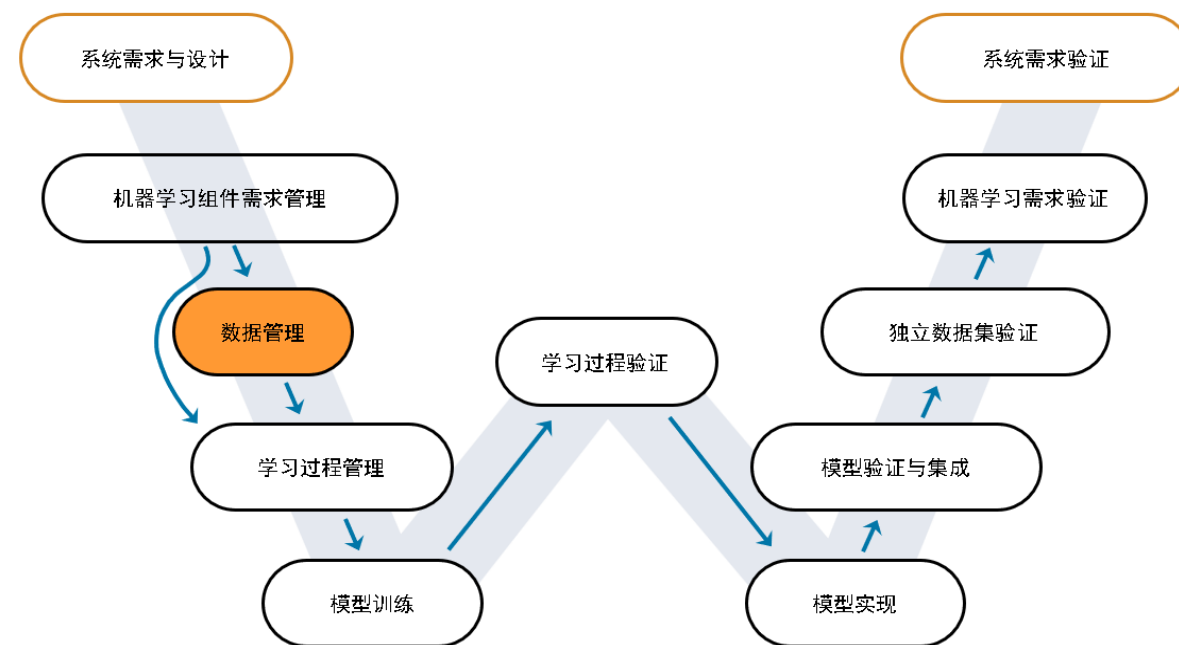
# 海量数据的便捷管理



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



```
countEachLabel(imdsTrain)
ans =
  2x2 table
    Label    Count
    _____  _____
    normal    1214
    pneumonia  3494
```

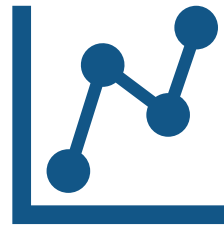


# AI模型的偏见（Bias）

## Bias的来源



数据不足，Bias因所选  
数据集而来



继承性偏见



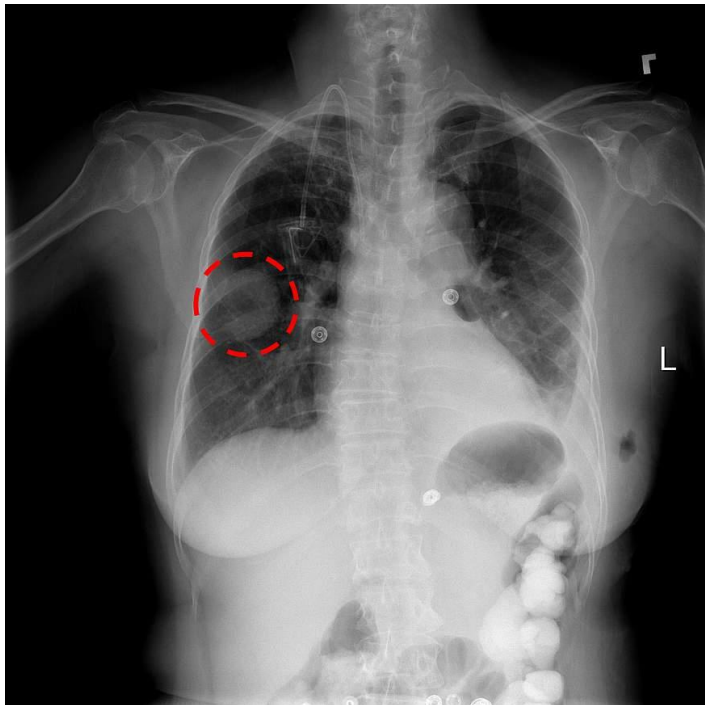
模型问题

**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<sup>a,1</sup>, Nicolás Nieto<sup>a,b,1</sup>, Victoria Peterson<sup>b,c</sup>, Diego H. Milone<sup>a</sup>, and Enzo Ferrante<sup>a,2</sup>



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*

BRIEF REPORT | APPLIED MATHEMATICS |

f t in ✉

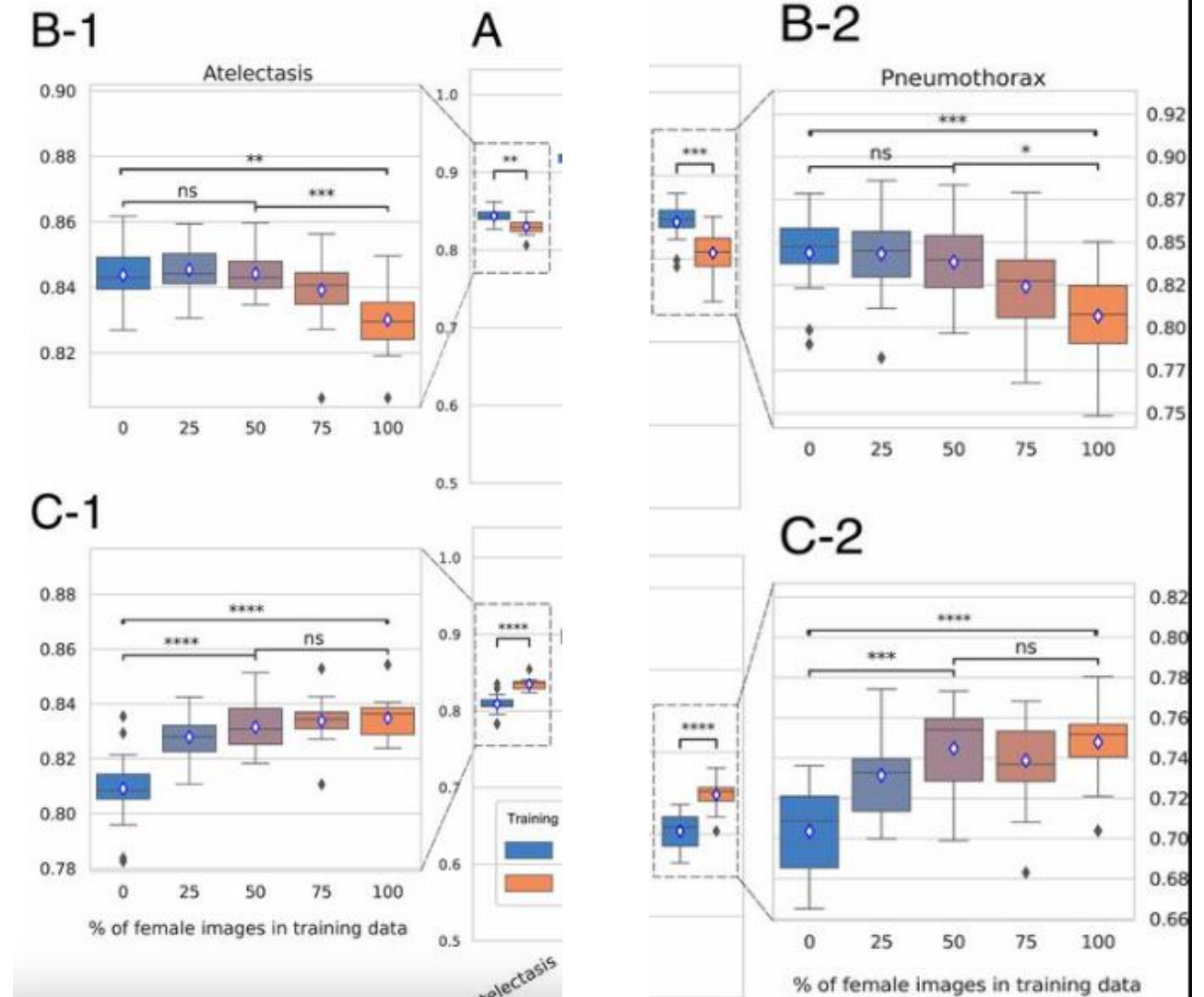
# Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis

Agostina J. Larrazabal, Nicolás Nieto, Victoria Peterson , , 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

$$\text{Disparate Impact} = \frac{\# \left( \begin{array}{c} \text{Female} \\ \text{Smoker} \end{array} \right)}{\# \left( \begin{array}{c} \text{Female} \\ \text{Nonsmoker} \end{array} \right)} \div \frac{\# \left( \begin{array}{c} \text{Male} \\ \text{Smoker} \end{array} \right)}{\# \left( \begin{array}{c} \text{Male} \\ \text{Nonsmoker} \end{array} \right)}$$

Disparate Impact < 1 for females indicating bias

VisualizeFairnessWeightsExample.mlx

```

6 numSmoker = sum(tblstats.GroupCount([2 4]));
7 numTotal = sum(tblstats.GroupCount);
8 numFemale = sum(tblstats.GroupCount([1 2]));
9 numFemaleSmoker = tblstats.GroupCount(1);
10
11 pIdealFemaleSmoker = (numSmoker/numTotal)*(numFemale/numTotal);
12 pObservedFemaleSmoker = numFemaleSmoker/numTotal;

```

This result indicates bias against the smoker class for female patients in the original data set.  
We need the proportion of female nonsmokers to female patients is the same as the proportion of male nonsmokers to male patients.

Compute fairness weights with respect to the sensitive attribute Gender and the binary response variable Smoker.

```

13 tbl.Weights = fairnessWeights(tbl,"Gender","Smoker")

```

**Compute by Group**  
tblstats = Compute counts for each group in tbl

Select groups and data to compute on  
Group by: tbl, Gender, Smoker, Weights  
Group by unique values

Compute on: All non-grouping variables

Select computation for groups  
Compute stats by group, Transform by group, Filter by group

Computations per group: Counts  
 Include empty groups  
Display results

	Gender	Smoker	GroupCount
1	Female	Nonsmoker	40
2	Female	Smoker	13
3	Male	Nonsmoker	26
4	Male	Smoker	21

pIdealFemaleSmoker = 0.1802  
pObservedFemaleSmoker = 0.4000

	Diastolic	Gender	Smoker	Systolic	Weights
1	93	Male	Smoker	124	0.7610
2	77	Male	Nonsmoker	109	1.1931
3	83	Female	Nonsmoker	125	0.8745
4	75	Female	Nonsmoker	117	0.8745
5	80	Female	Nonsmoker	122	0.8745
6	70	Female	Nonsmoker	121	0.8745
7	88	Female	Smoker	130	1.3862
8	82	Male	Nonsmoker	115	1.1931
9	78	Male	Nonsmoker	115	1.1931

	Gender	Smoker	Weights	GroupCount
1	Female	Nonsmoker	0.8745	40
2	Female	Smoker	1.3862	13
3	Male	Nonsmoker	1.1931	26
4	Male	Smoker	0.7610	21

Visualize the fairness weights using grouped scatter plots. Without the fairness weights, all observations have the same weight by default.

```

17 scatterPlotFair(tbl, tblstats);

```

To understand how fairness weights affect the observations, find the statistical parity difference (SPD) for each group in Gender after applying the fairness weights. This measure must be equal to 0 to be fair. Use the `fairnessMetrics` function, which computes bias and group metrics for a data set or binary classification model with respect to sensitive attributes.

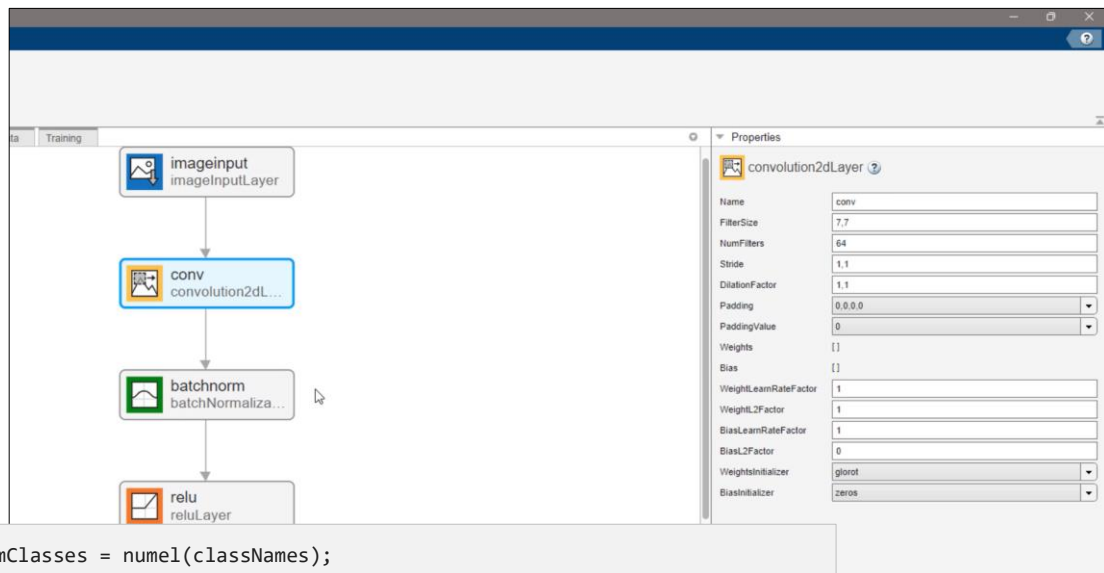
Original Observations

Weighted Observations

# Bias 检测与降低

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

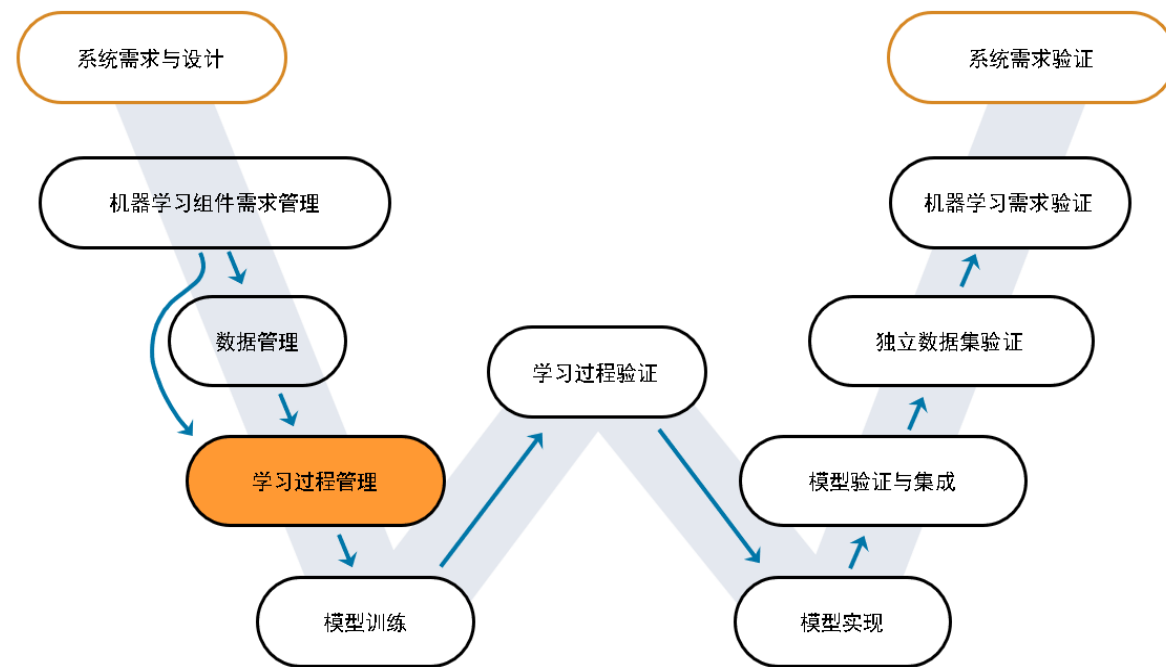
# 模块化快速搭建网络模型



```
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

Mode: Sequential

Cluster: [ ]

Pool Size: 0

Run

Experiment Browser

verification-medical-neural-network

Experiment\_pneumonia\_CNN

Result1

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]

Setup Function

W3\_W5\_Experiment\_pneumonia\_CNN

---

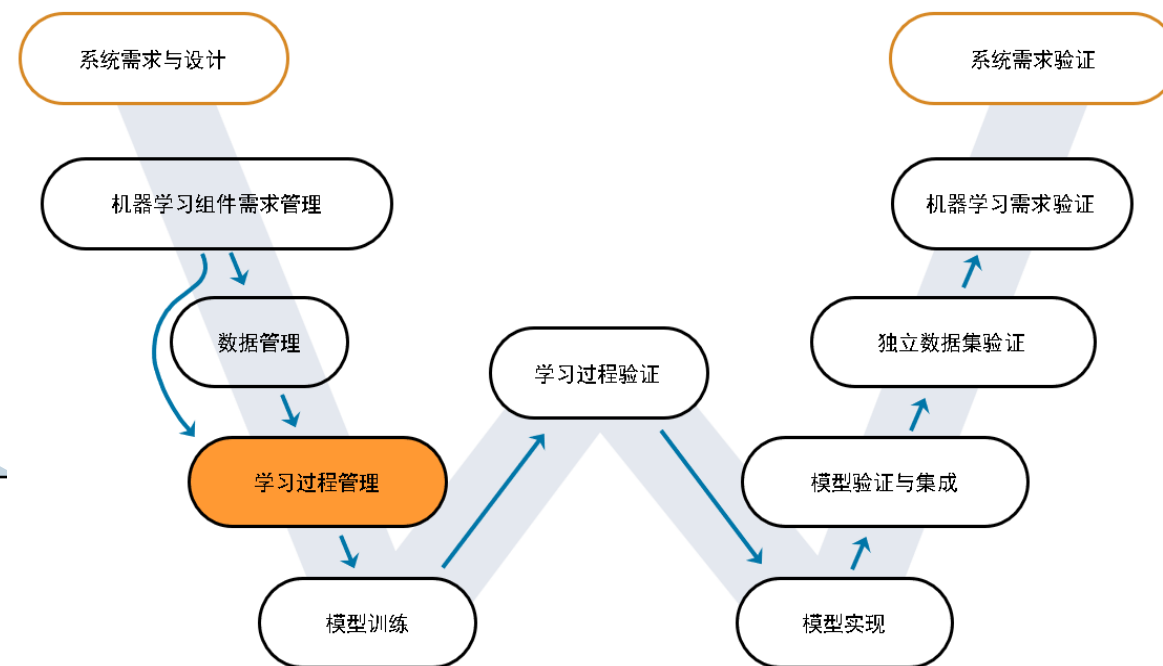
**Hyperparameters**

Strategy: Exhaustive Sweep

In the setup and metric functions, access hyperparameter values by using dot notation.

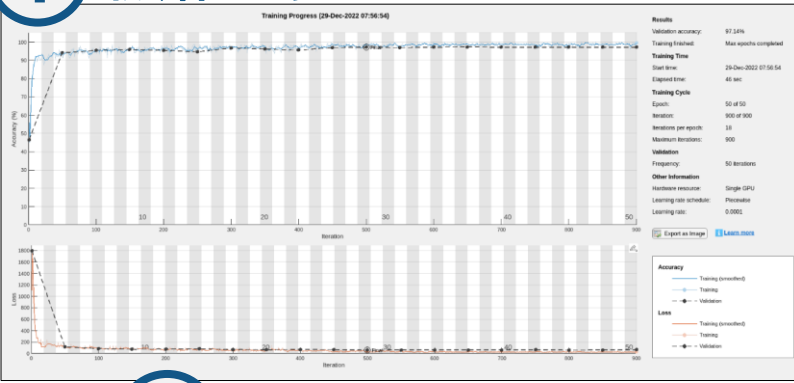
Name	Values
solver	["adam"]
filterSize	[5 7]
numFilters1	[16 32]
numFilters2	[32 64]

+ Add - Delete



# 渐进式优化迭代得到高精度、高可靠性模型

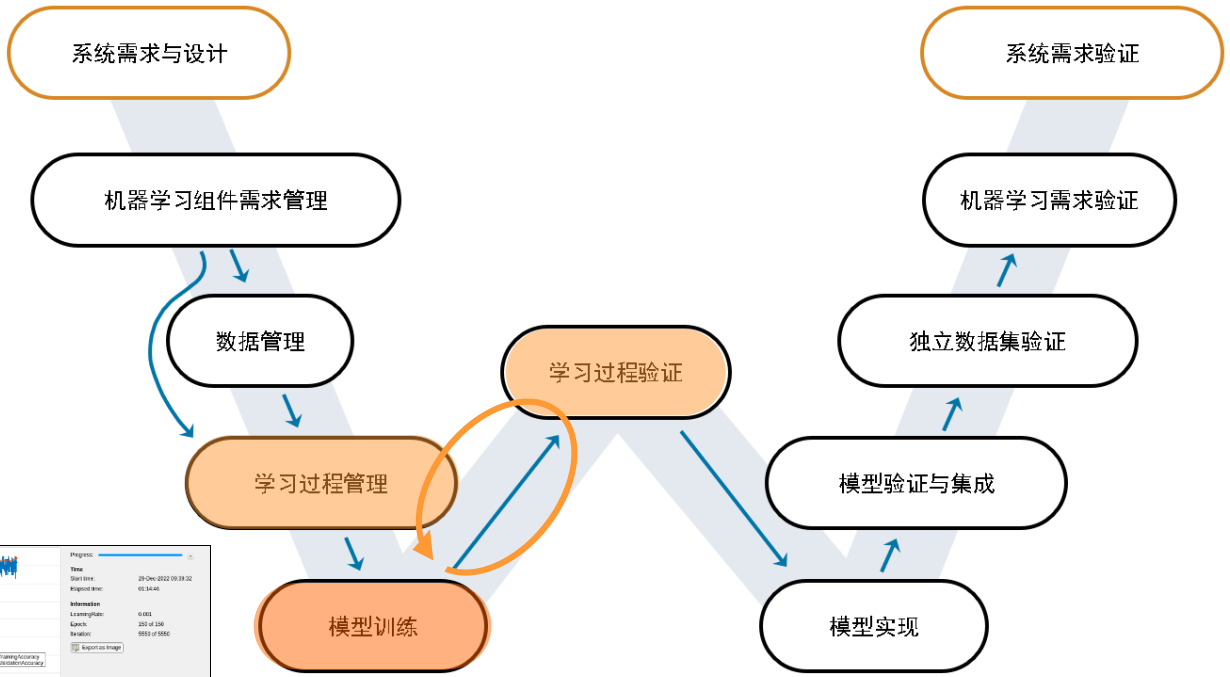
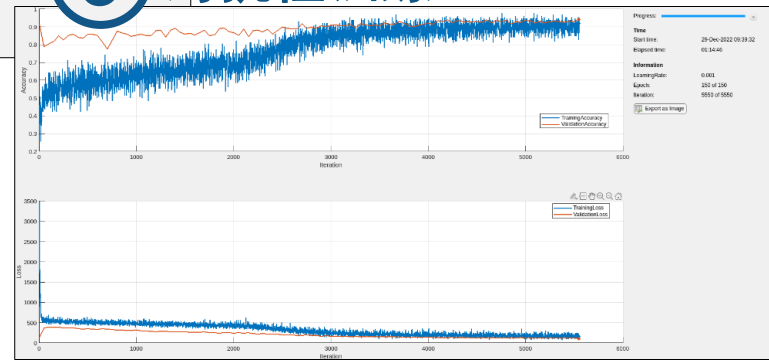
## 1 初始训练



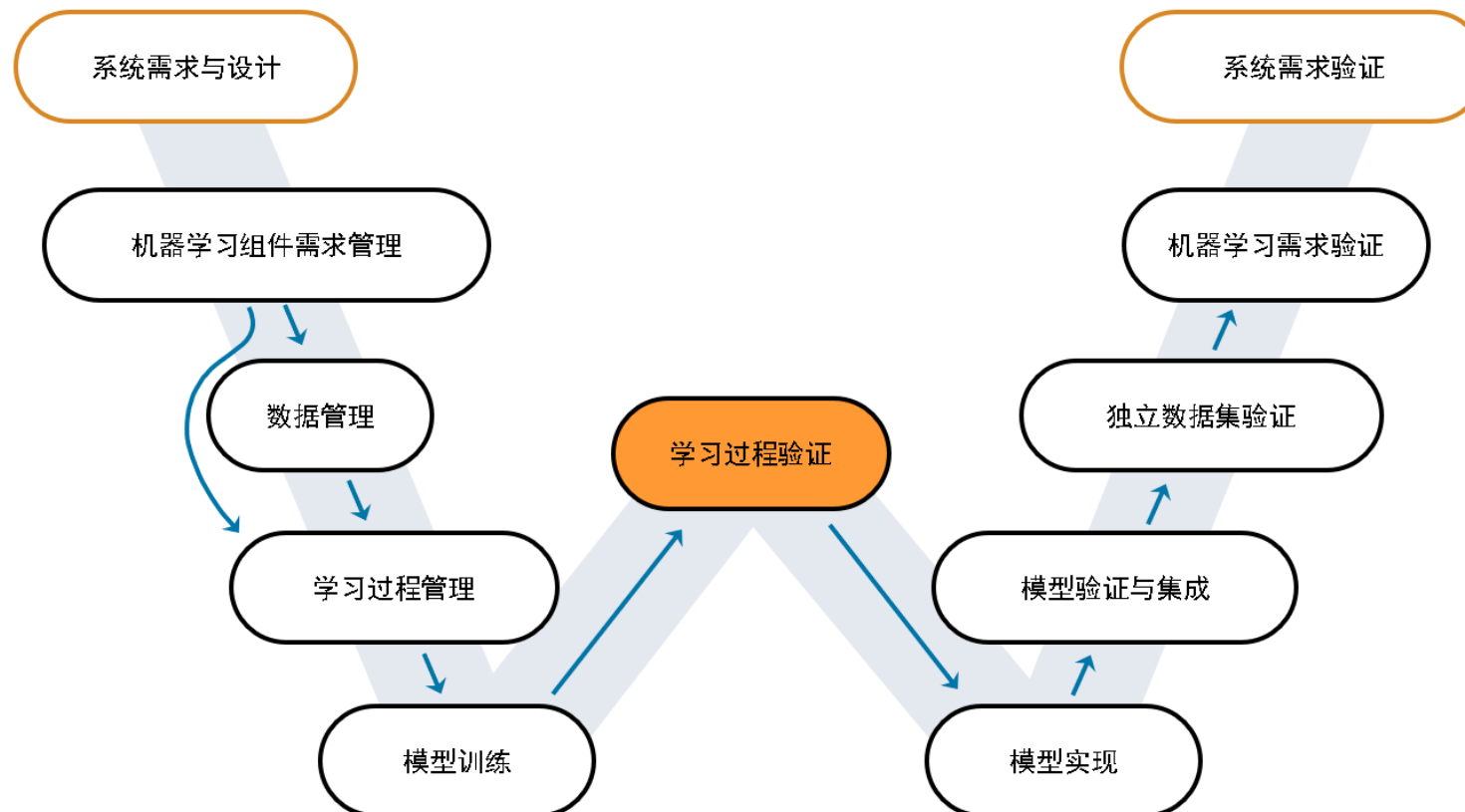
## 2 数据增强训练

```
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



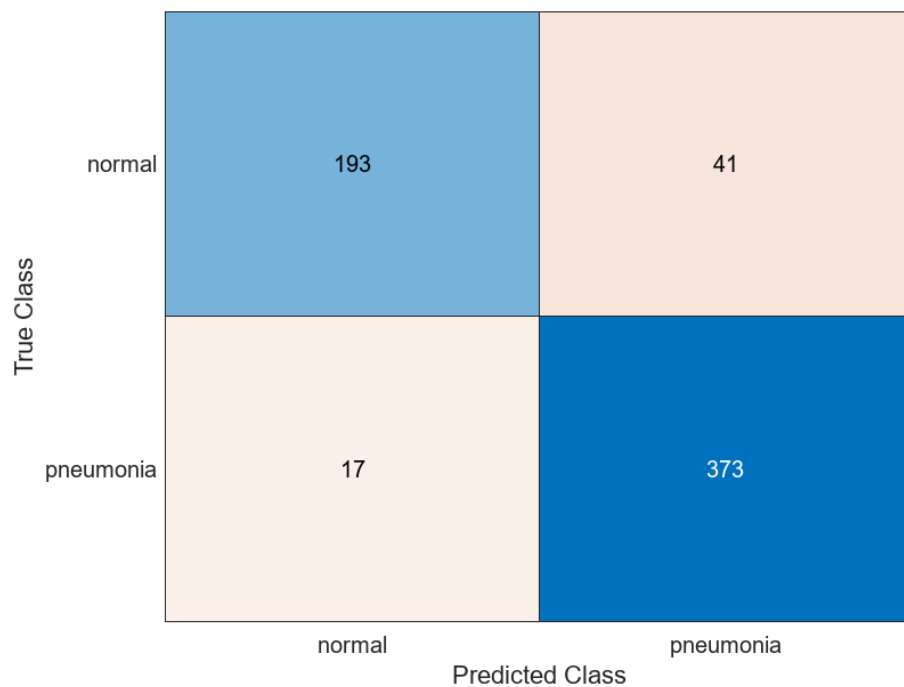
# 学习过程验证



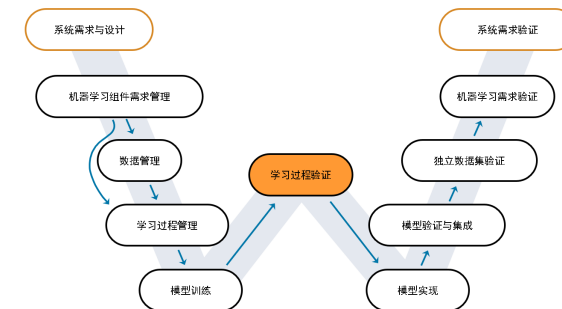
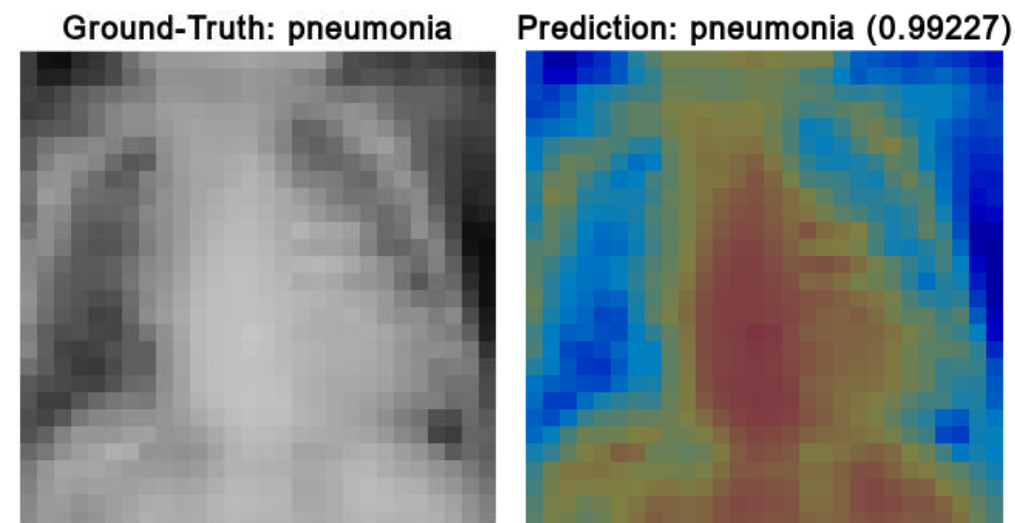
# 基于独立数据集的模型性能测试与理解

Accuracy: 90.71%

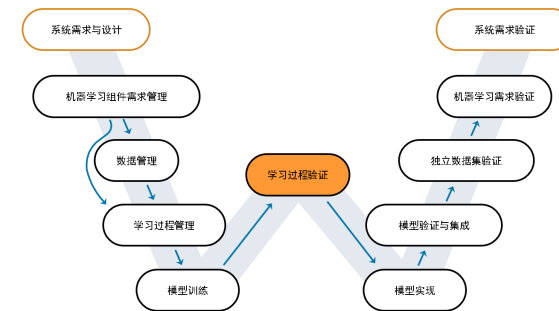
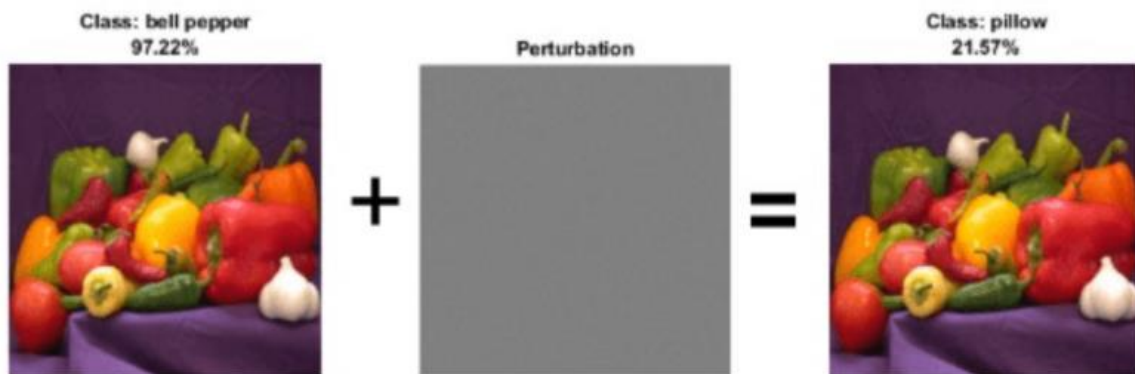
```
confusionchart(T,Y)
```



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



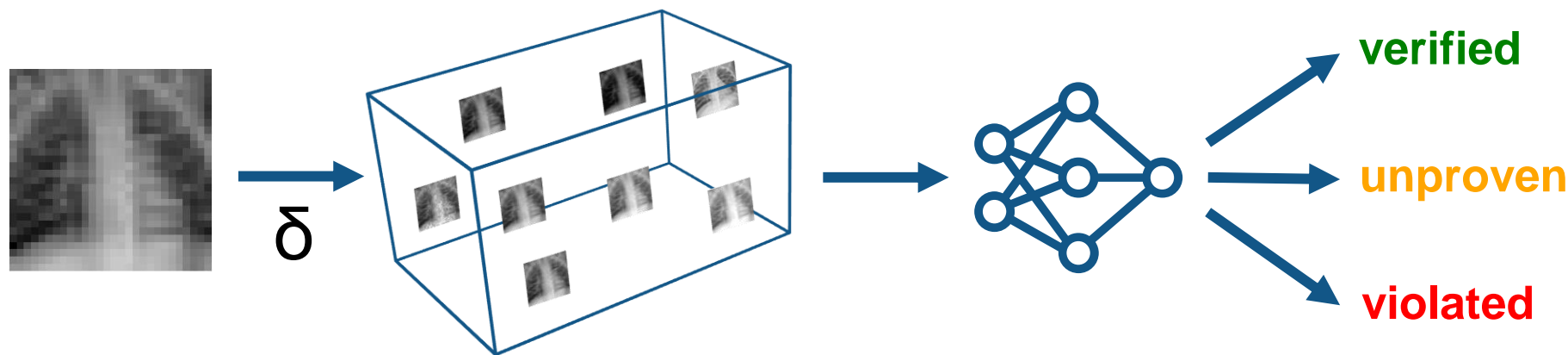
# 神经网络的鲁棒性验证



## 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



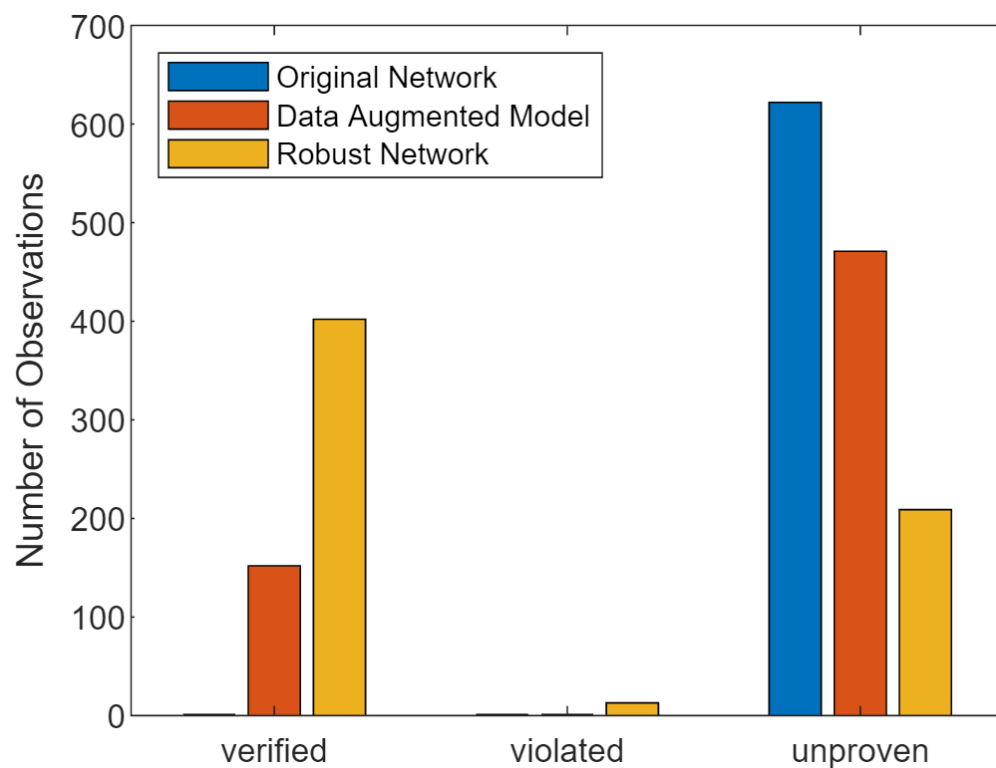
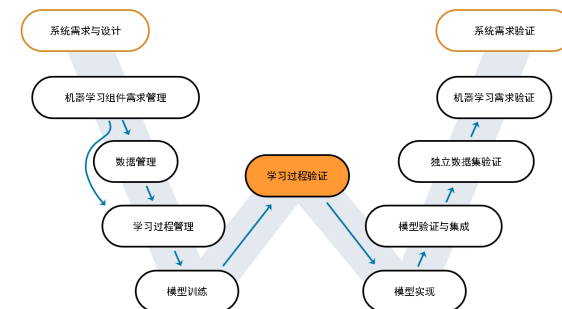
# 神经网络的鲁棒性验证



## 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)
```

<b>verified</b>	402
<b>violated</b>	13
<b>unproven</b>	209

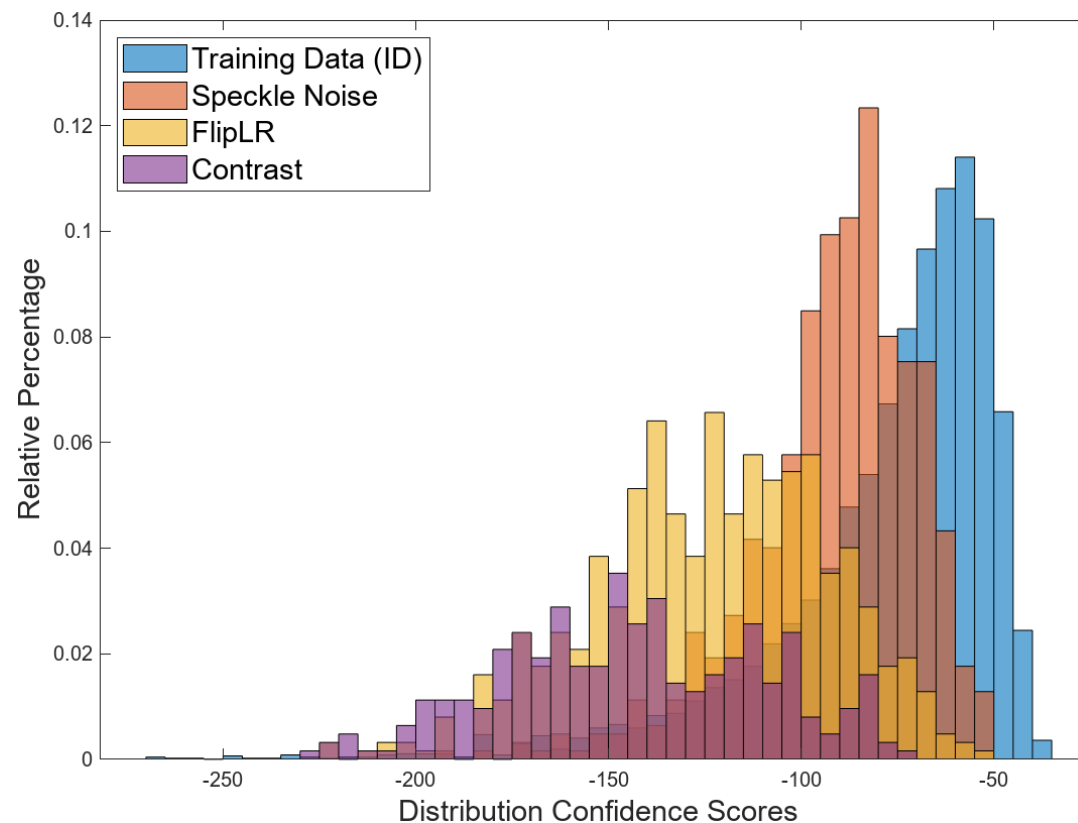
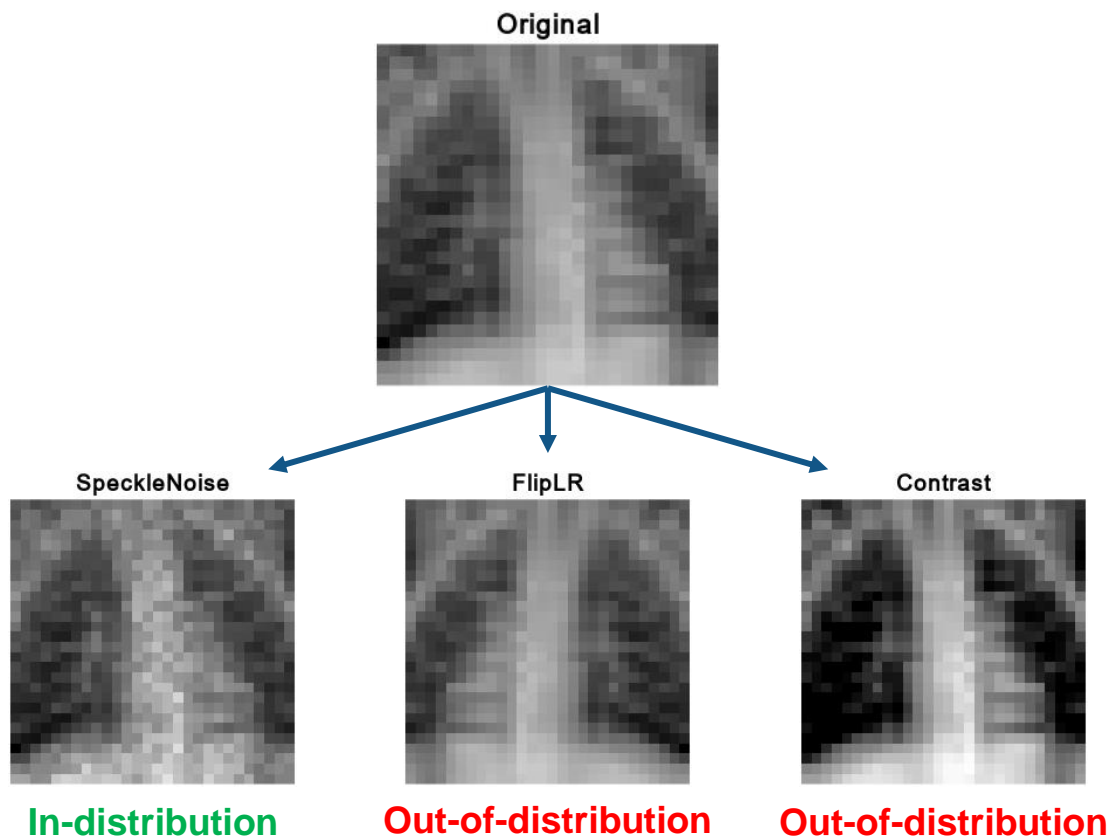
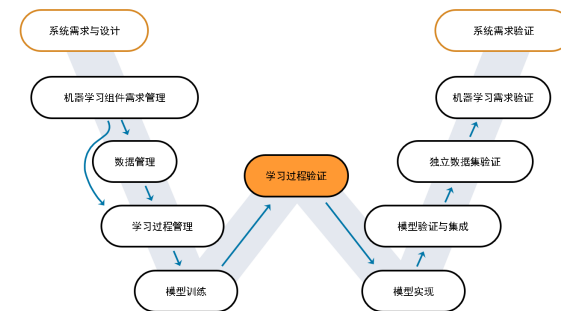
# 检测分布外样本，拒收或转给专家复核



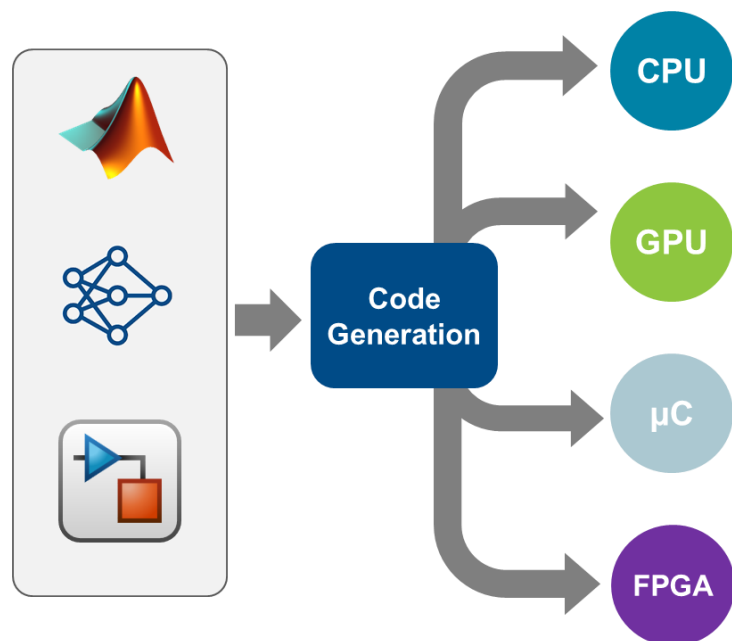
## Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team **STAFF**

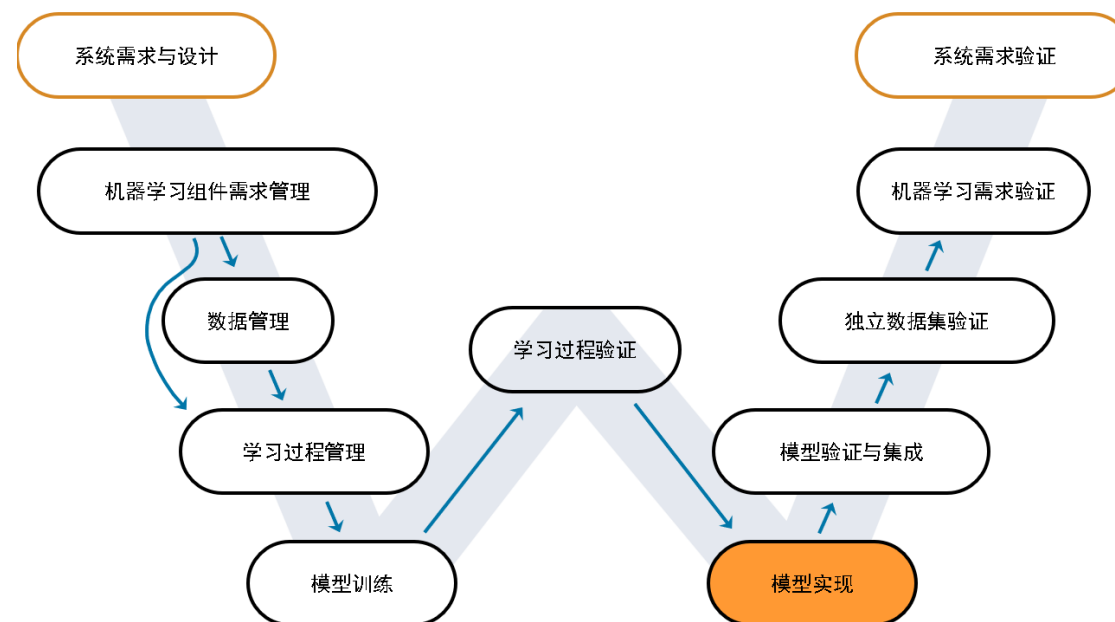
Verify and test robustness of deep learning networks



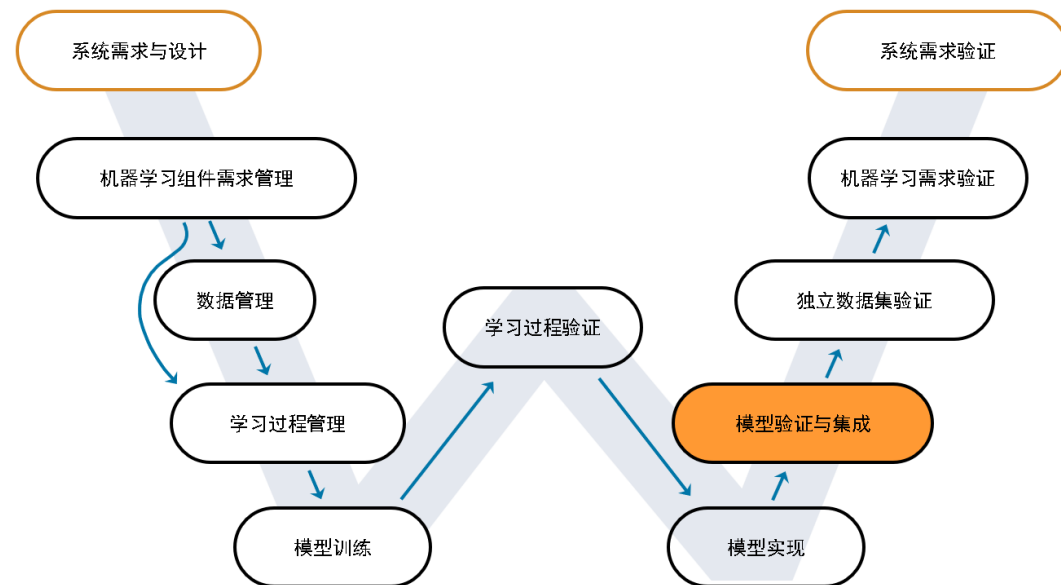
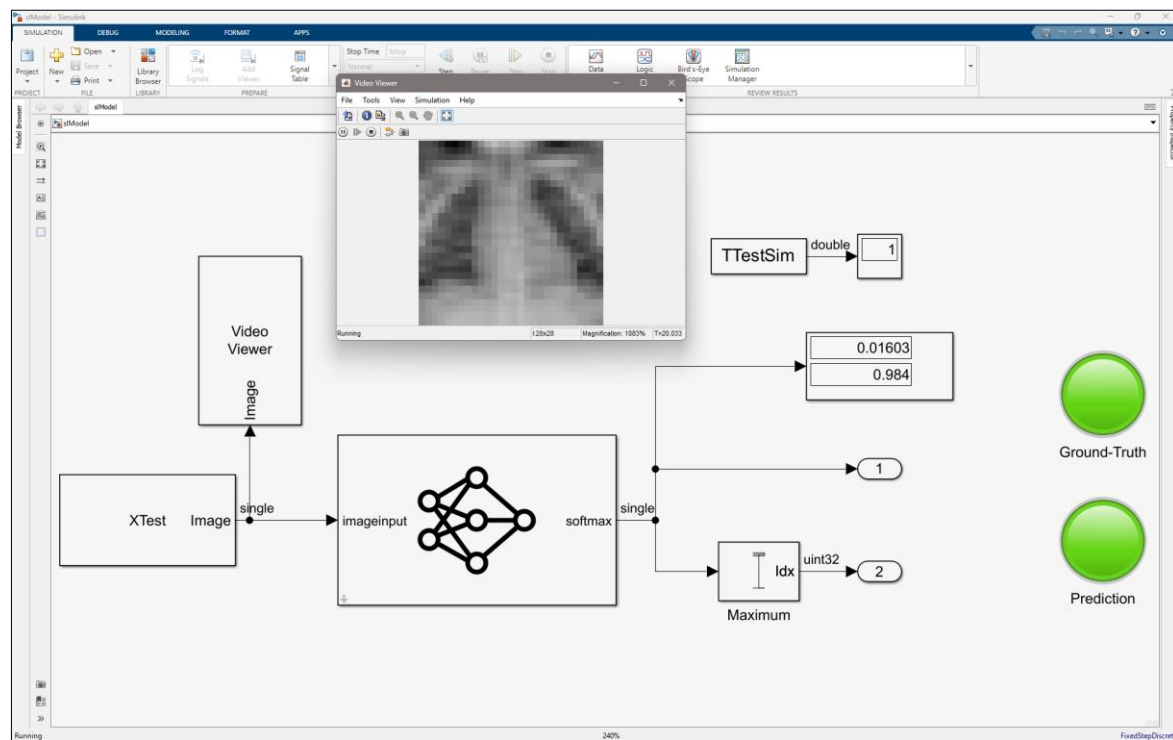
# 无bug自动部署至目标硬件



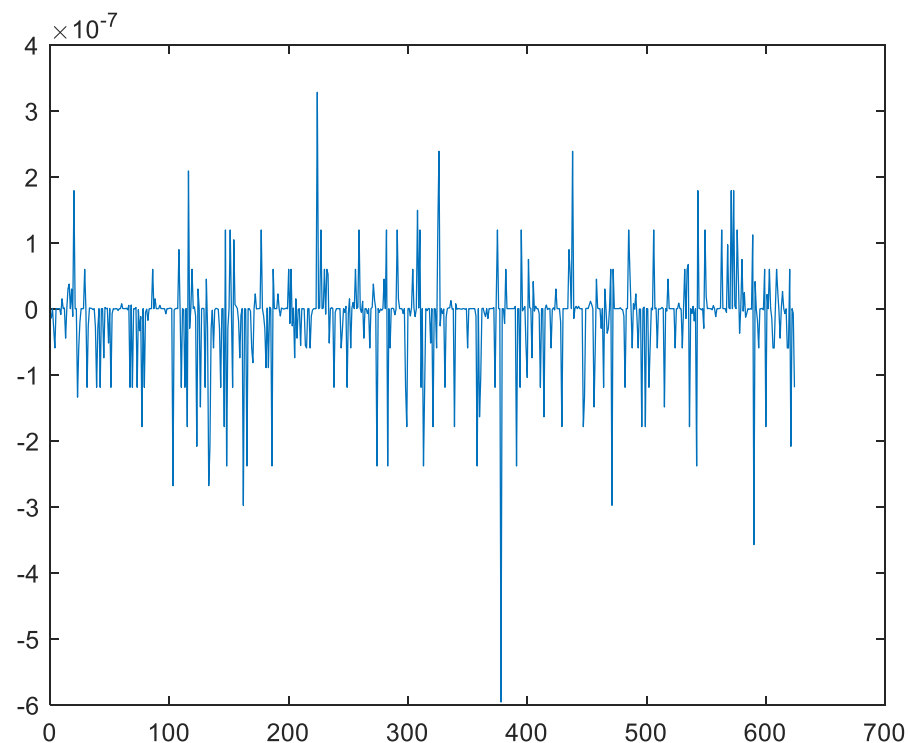
analyzeNetworkForCodegen(net)	
	Supported
none	"Yes"
arm-compute	"Yes"
mkldnn	"Yes"
cuda	"Yes"
tensorrt	"Yes"



# 将AI模块集成在Simulink中，进行系统级仿真测试

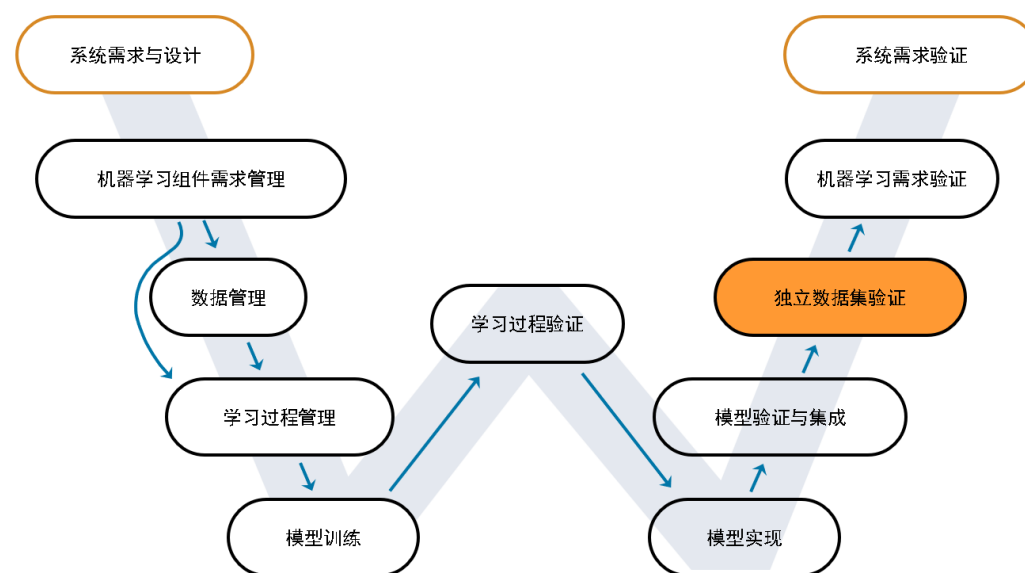


# 开发与推理模型完全无差别



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

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



# Verifying requirements have been fully tested

**MATLAB Test Manager: All Tests in Current Project**

16 Total Tests  
 13 Passed  
 0 Failed  
 0 Incomplete  
 0 Not Run

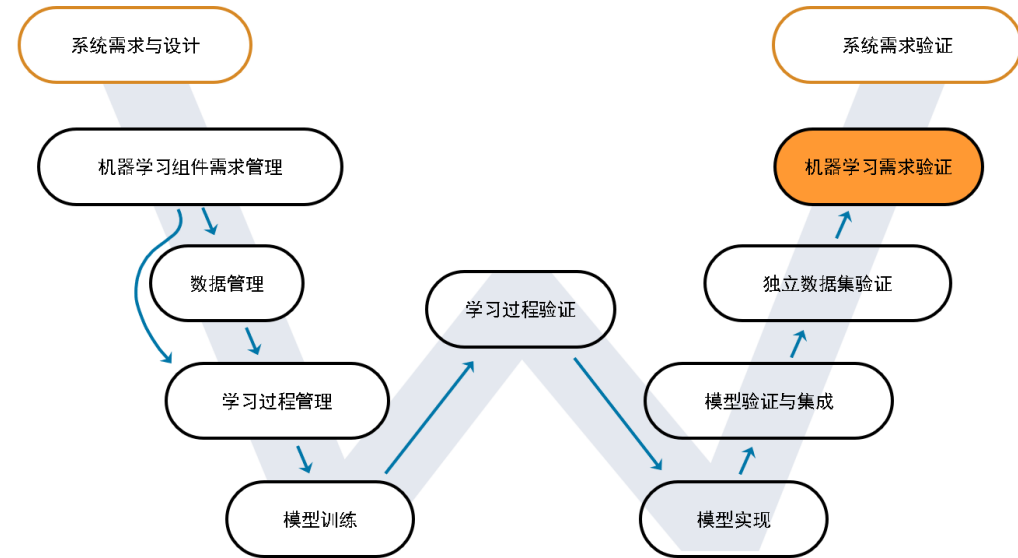
Running tests... 13/16

**Requirements Editor**

Index	Summary	Implemented	Verified
1	ML component requirement for X-Ray Pneumonia Detector (XRPD)	Implemented	Verified
1.1	Introduction		
1.2	ML component description		
1.3	ML component requirements	Implemented	Verified
1.3.1	ML component input	Implemented	Verified
1.3.1.1	ML component input should be 28x28x1	Implemented	Verified
1.3.1.2	ML component input data (training) should be 28x28x1	Implemented	Verified
1.3.1.3	ML component input data (validation) should be 28x28x1	Implemented	Verified
1.3.1.4	ML component input data (test) should be 28x28x1	Implemented	Verified
1.3.2	ML component output	Implemented	Verified
1.3.2.1	ML component output should be 2	Implemented	Verified
1.3.2.2	ML component output labels should be 'normal' or 'pneumonia'	Implemented	Verified
1.3.3	ML component accuracy	Implemented	Verified
1.3.3.1	ML component training precision	Implemented	Verified
1.3.3.2	ML component test precision	Implemented	Verified
1.3.3.3	ML component avoid overfitting	Implemented	Verified
1.3.3.4	ML component out-of-distribution detection	Implemented	Verified
1.3.4	ML component latency	Implemented	Verified
1.3.5	ML component robustness	Implemented	Verified
1.3.5.1	ML component robustness 1% perturbation	Implemented	Verified
1.3.5.2	ML component robustness 0.5% perturbation	Implemented	Verified
1.3.5.3	ML component robustness 0.1% perturbation	Implemented	Verified
1.3.6	ML component implementation	Implemented	Verified

**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](#) ✓

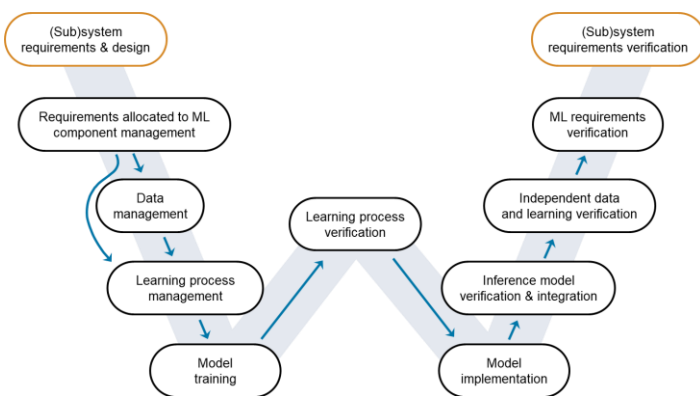


# 要点

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

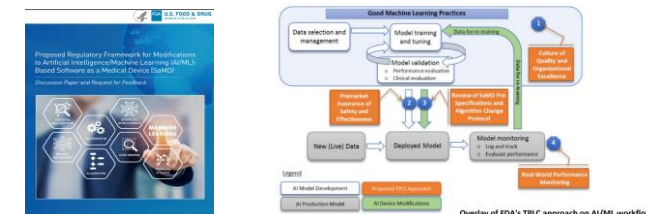
高安全性验证的经验助力推动全新AI标准



## Deep Learning Toolbox Verification Library

by MathWorks Deep Learning Toolbox Team **STAFF**

Verify and test robustness of deep learning networks



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

# MATLAB EXPO

Thank you



© 2023 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See [mathworks.com/trademarks](https://www.mathworks.com/trademarks) for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.