

王道！ 時系列データで学ぶ6種の特徴抽出と異常検知

MathWorks Japan

アプリケーションエンジニアリング部

王 晓星 博士（理学）

xwang@mathworks.com



A



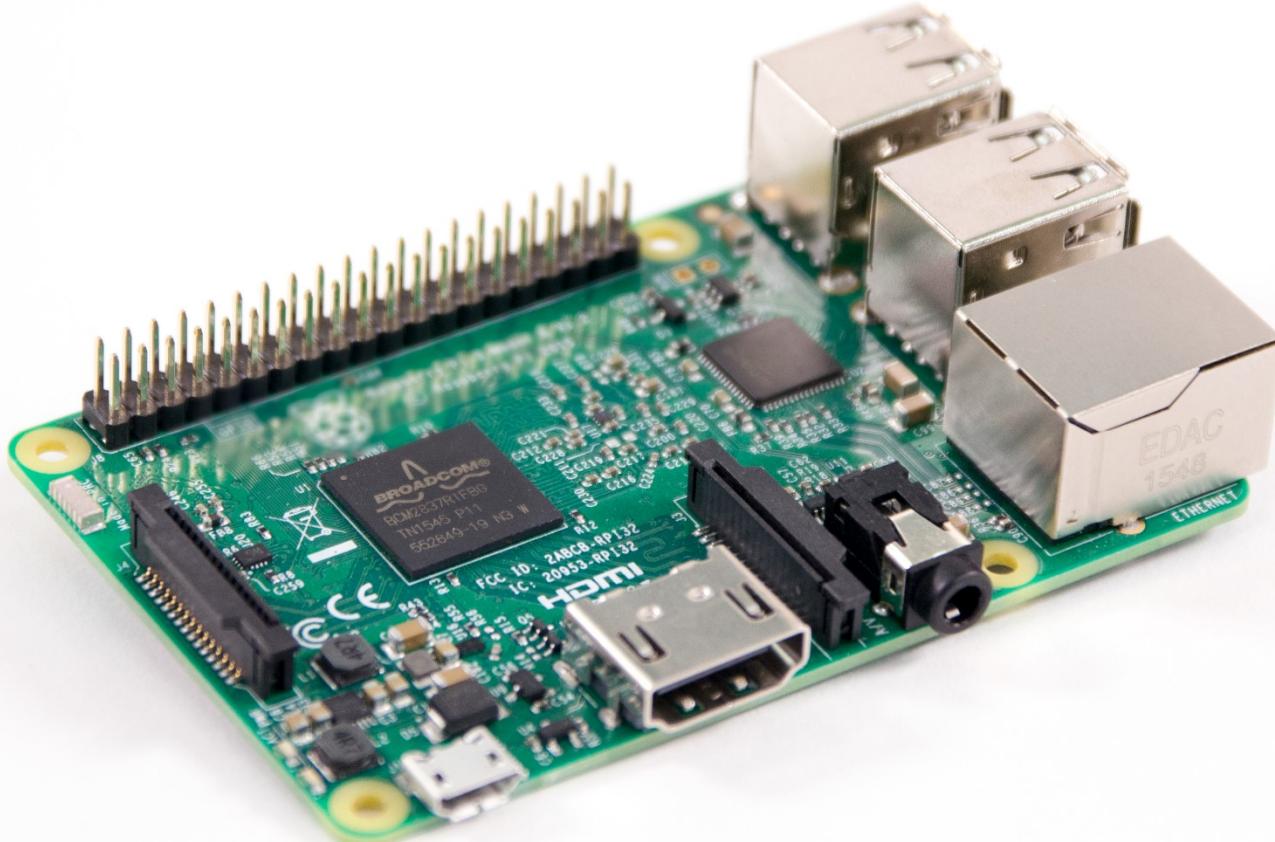
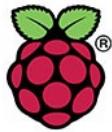
B



C



Raspberry Pi x 異常検知



```
pi@raspberrypi-mn2M91tFd1:~/airCompressor_spectrogram/MATLAB_ws/R2021a/C/... - pi@raspberrypi-mn2M91tFd1:~/airCompressor_spectrogram/ws/R2021a/C/Users/xwang/OneDrive - MathWorks/06_Pres... 05_Demos/01_PdM/03_audio/01_airCompressor/deployRaspbe 5_PdMwebinar_20210521 $
```



ライブエディター - C:\Users\xwang\OneDrive - MathWorks\06_Presentation\05_Demos\01_PdM\03_audio\01_airCompressor\deployRaspbe 5_PdMwebinar_20210521 \$

ライブエディター 挿入 ビュー

単一表示 左/右 タブ位置▼ 行番号 データヒント 全画面 すべての出力をクリア 出力

上/下 カスタム▼ 収まるようにタブを縮小 アルファベット順 ドキュメントタブ 表示

並べて表示

run_processAirCompressor.mlx +

Raspberry Pi 上で異音判定アルゴリズムを実行

初期化

```
1 clear; close all;
2 format short; format compact;
3 % 乱数群の制御
4 rng('default');
```

Raspberry Piへの接続

```
5 r = raspi
```

実行可能ファイルのパスを取得

```
6 applicationDirPaths = raspi.utils.getRemoteBuildDirectory('applicationName','prc
7 targetDirPath = applicationDirPaths{1}.directory;
```

データフォルダ

06_Presentation > 05_Demos > 00_dataset > 03_airCompressor > AirCompressorDataset > AirCompressorDataset > 99_csvfiles > Healthy

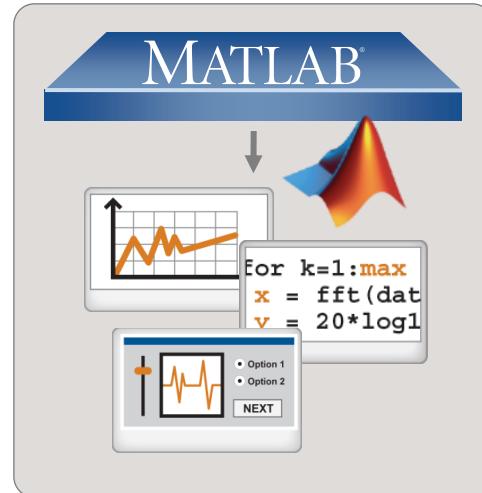
Healthy001.csv	Healthy002.csv	Healthy003.csv
Healthy004.csv	Healthy005.csv	Healthy006.csv

UTF-8 LF スクリプト 行 4 列 16

システムの構築

様々な環境への統合

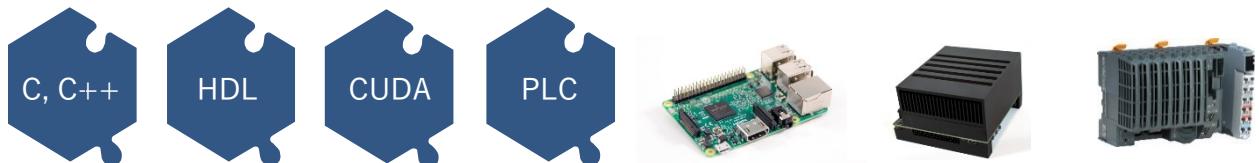
目的に応じた統合をサポート
MATLABライセンスがない環境でも利用可



MATLABで
アルゴリズム開発



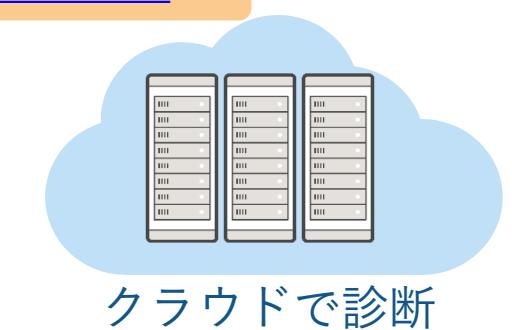
MATLAB Coder / GPU Coder / Embedded Coder



MATLAB Compiler / MATLAB Compiler SDK



MATLAB Production Server



データ解析のワークフロー

データアクセス



センサー



ファイル



データベース

データ解析



データ探索



前処理

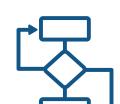


特徴抽出

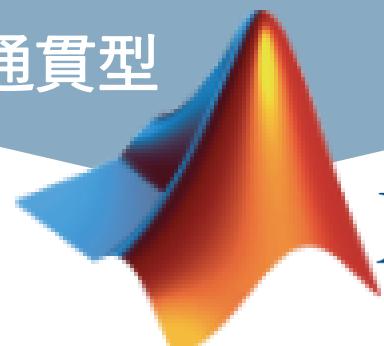
開発



AI モデル



アルゴリズム
開発



MATLAB® & SIMULINK®

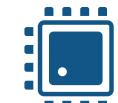
実装 & 展開



デスクトップ
アプリ



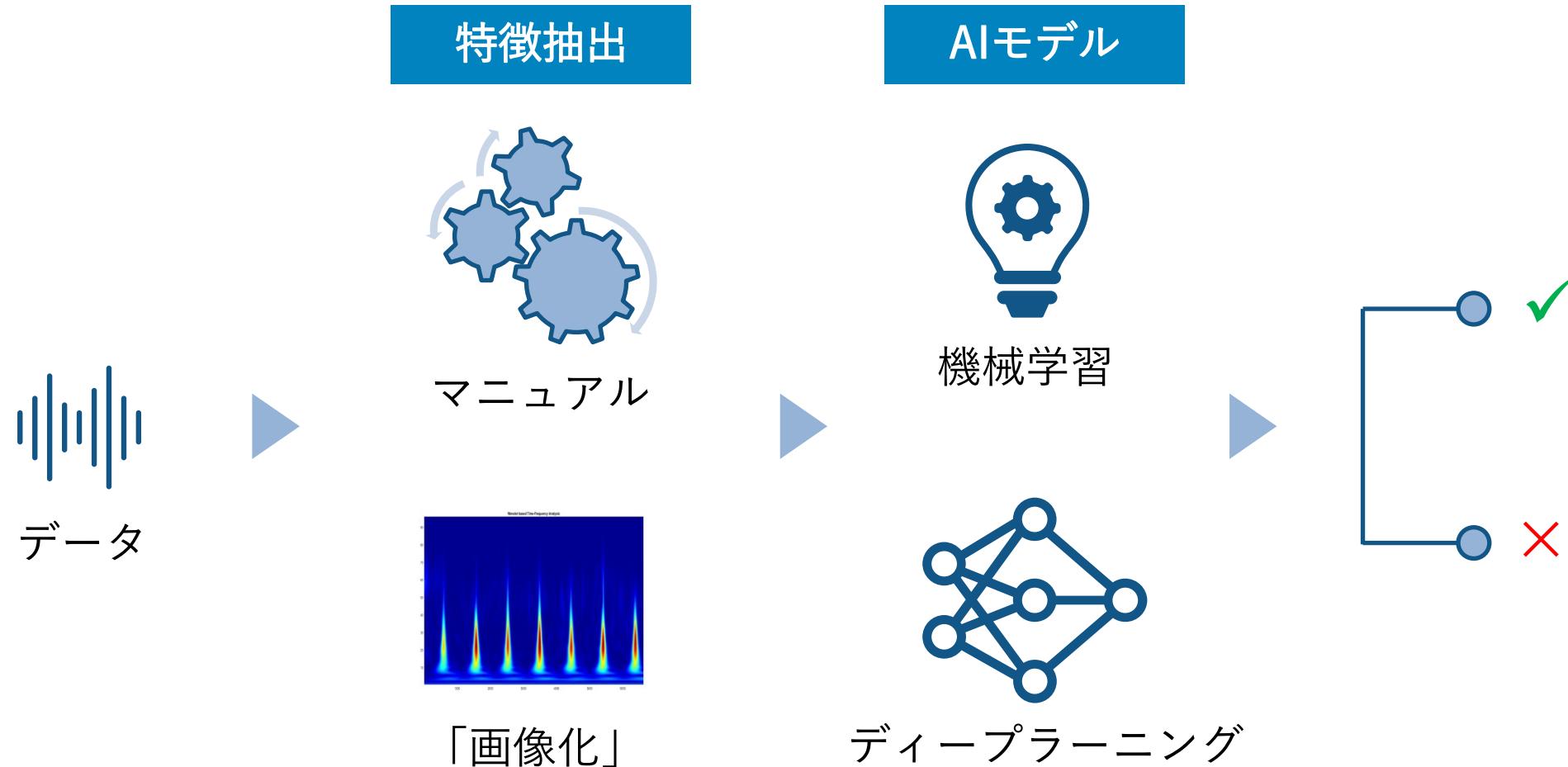
エンタープライ
ズシステム



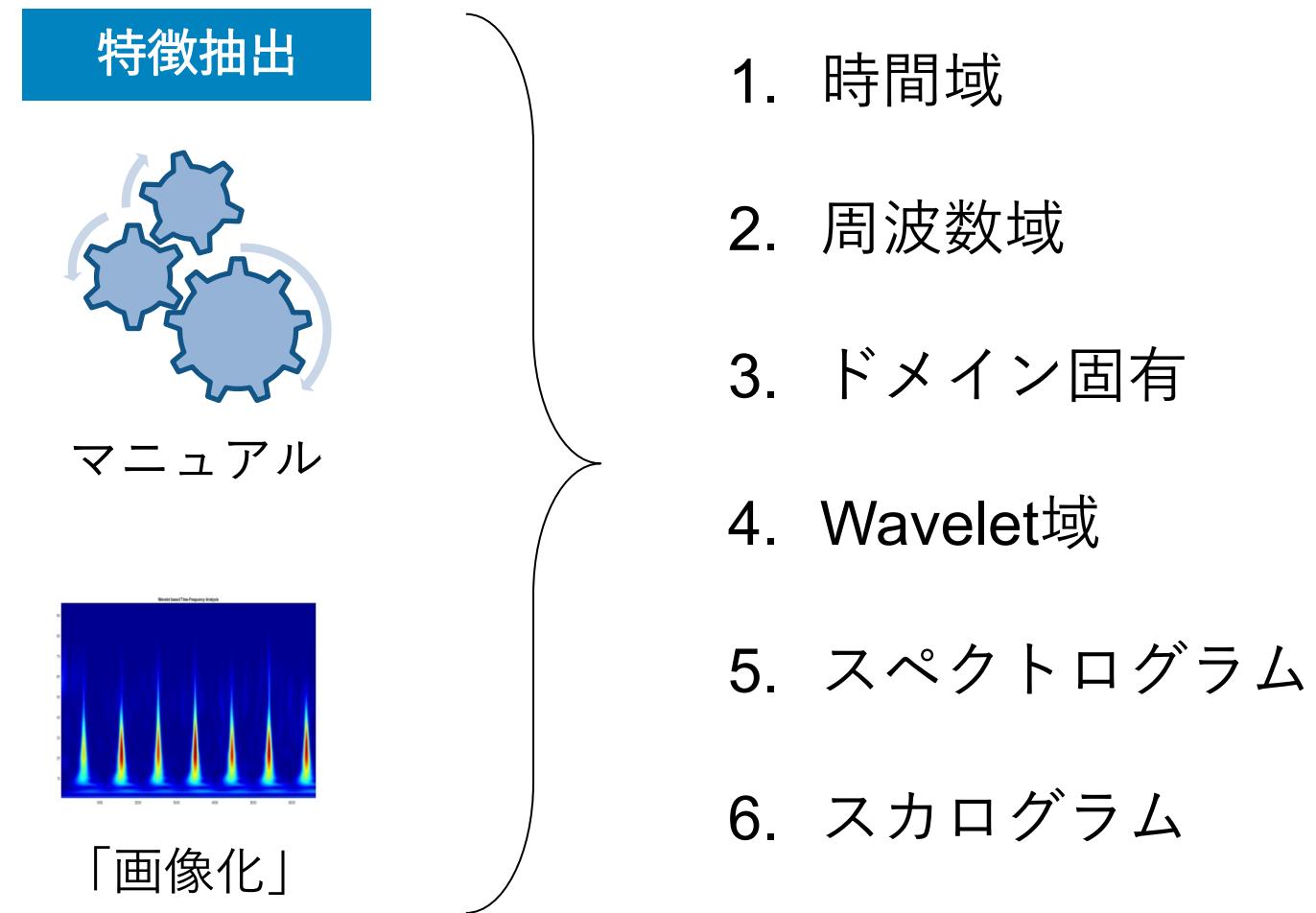
組み込み
デバイス

一気通貫型

異常検知・異常分類に役立つ2つのアプローチ



主な特徴抽出の方法



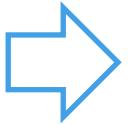
様々な時系列データ

No.	種類	特徴	クラス	データサイズ (合計)	データ点/ ファイル	ファイル数
1	振動データ	時間とともに徐々に変化	2	20 MB	10万	50個
2	AEセンサー	不定期な突発的な変化	2	1.5G	10.5万	400個
3	音データ	シーケンス長が異なる信号	2	150 MB	0.7万－1.6万	4720個
4	心電図	周期変化 & 異常値	3	70 MB	6.6万	162個

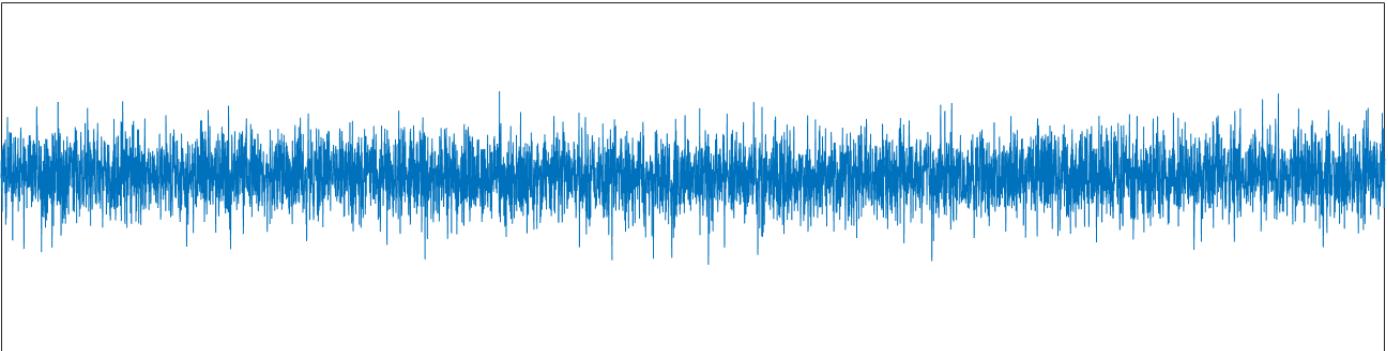
様々な時系列データ

No.	種類	特徴	クラス	データサイズ (合計)	データ点/ ファイル	ファイル数
1	振動データ	時間とともに徐々に変化	2	20 MB	10万	50個
2	AEセンサー	不定期な突発的な変化	2	1.5G	10.5万	400個
3	音データ	シーケンス長が異なる信号	2	150 MB	0.7万－1.6万	4720個
4	心電図	周期変化 & 異常値	3	70 MB	6.6万	162個

振動データ



時系列データ



- 周波数：約 100 kHz
- 計測時間：1秒 / 日
- 正常：25日、異常：25日

ゴール

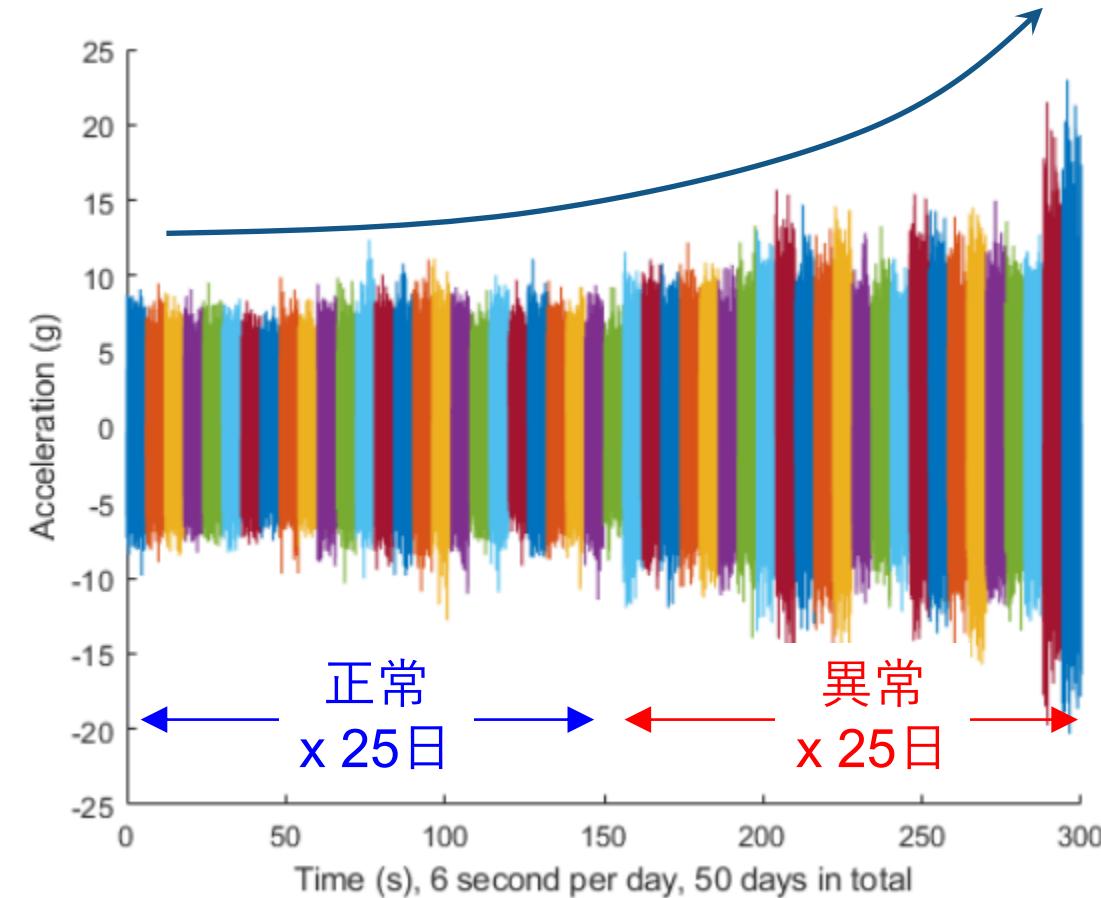
振動データから
劣化のヒントを示す**特徴量を特定**

データ元：

Bechhoefer, Eric, Brandon Van Hecke, and David He. "Processing for improved spectral analysis." *Annual Conference of the Prognostics and Health Management Society, New Orleans, LA, Oct. 2013.*

振動データの時間推移

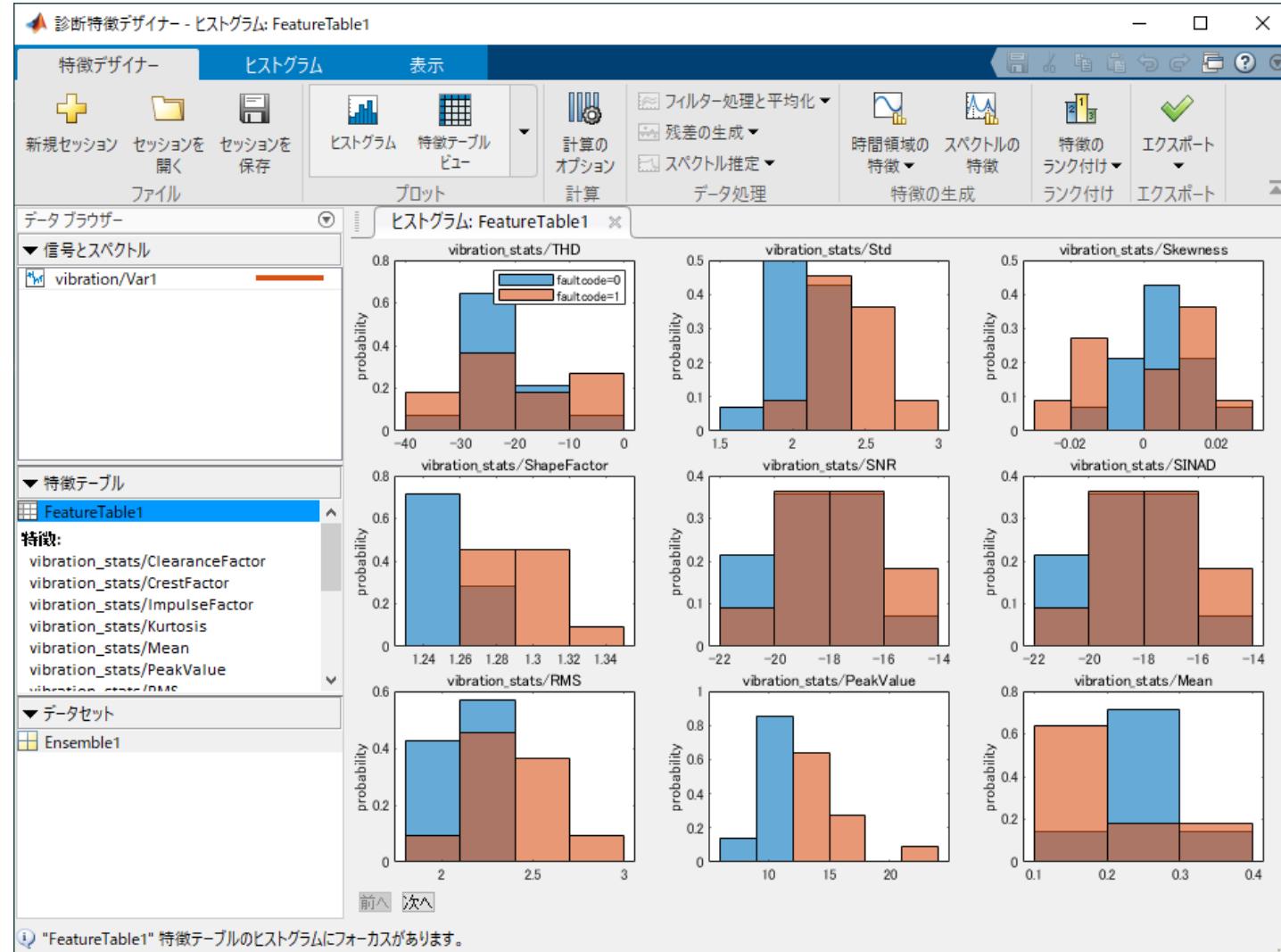
時間経過とともに振幅が増大



診断特徴デザイナー 特徴量選択を簡単に実施するアプリ

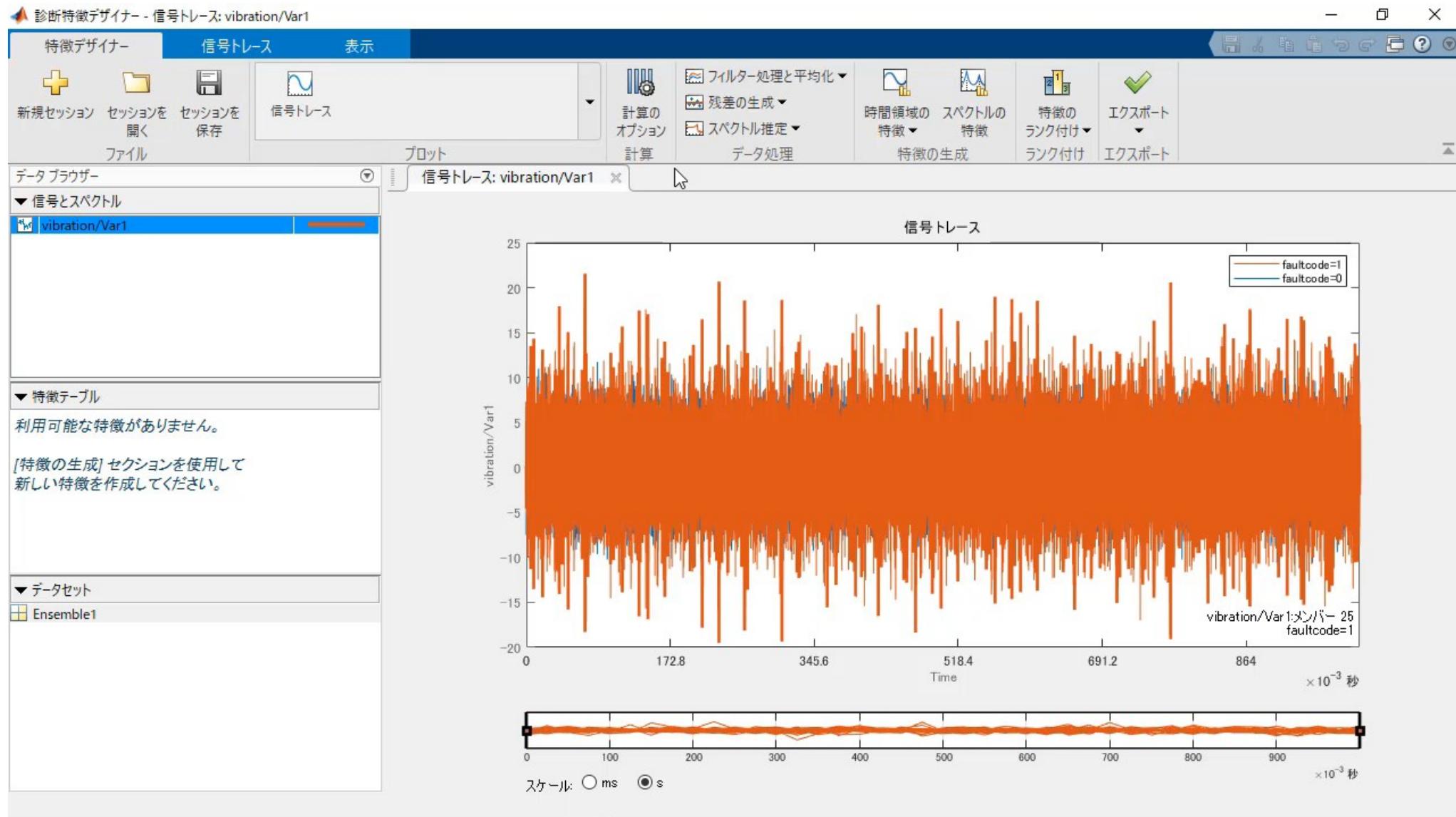


Predictive Maintenance Toolbox



複数の特徴量を算出・可視化し、
「正常/異常」を区別できる
特徴量の特定を容易に

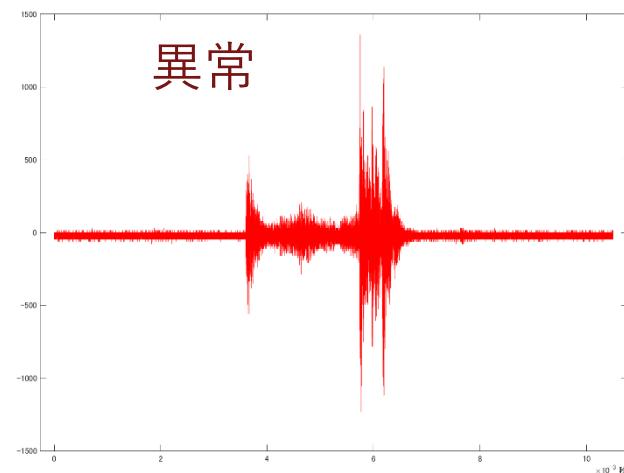
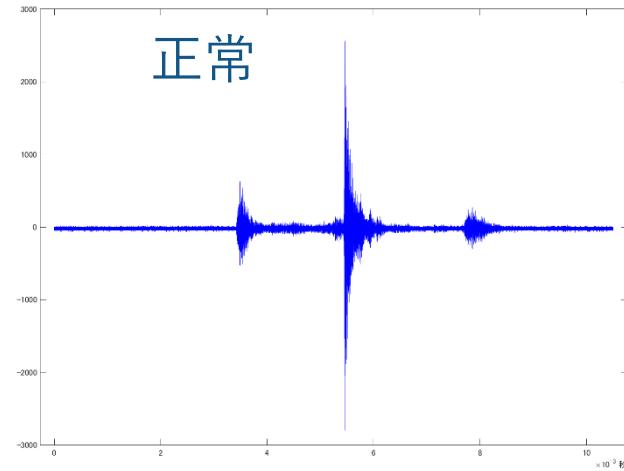
診断特徴デザイナーを用いた特徴抽出と異常検知



様々な時系列データ

No.	種類	特徴	クラス	データサイズ (合計)	データ点/ ファイル	ファイル数
1	振動データ	時間とともに徐々に変化	2	20 MB	10万	50個
2	AEセンサー	不定期な突発的な変化	2	1.5G	10.5万	400個
3	音データ	シーケンス長が異なる信号	2	150 MB	0.7万－1.6万	4720個
4	心電図	周期変化 & 異常値	3	70 MB	6.6万	162個

(株) ミスズ工業：AEセンサーによる異常検知



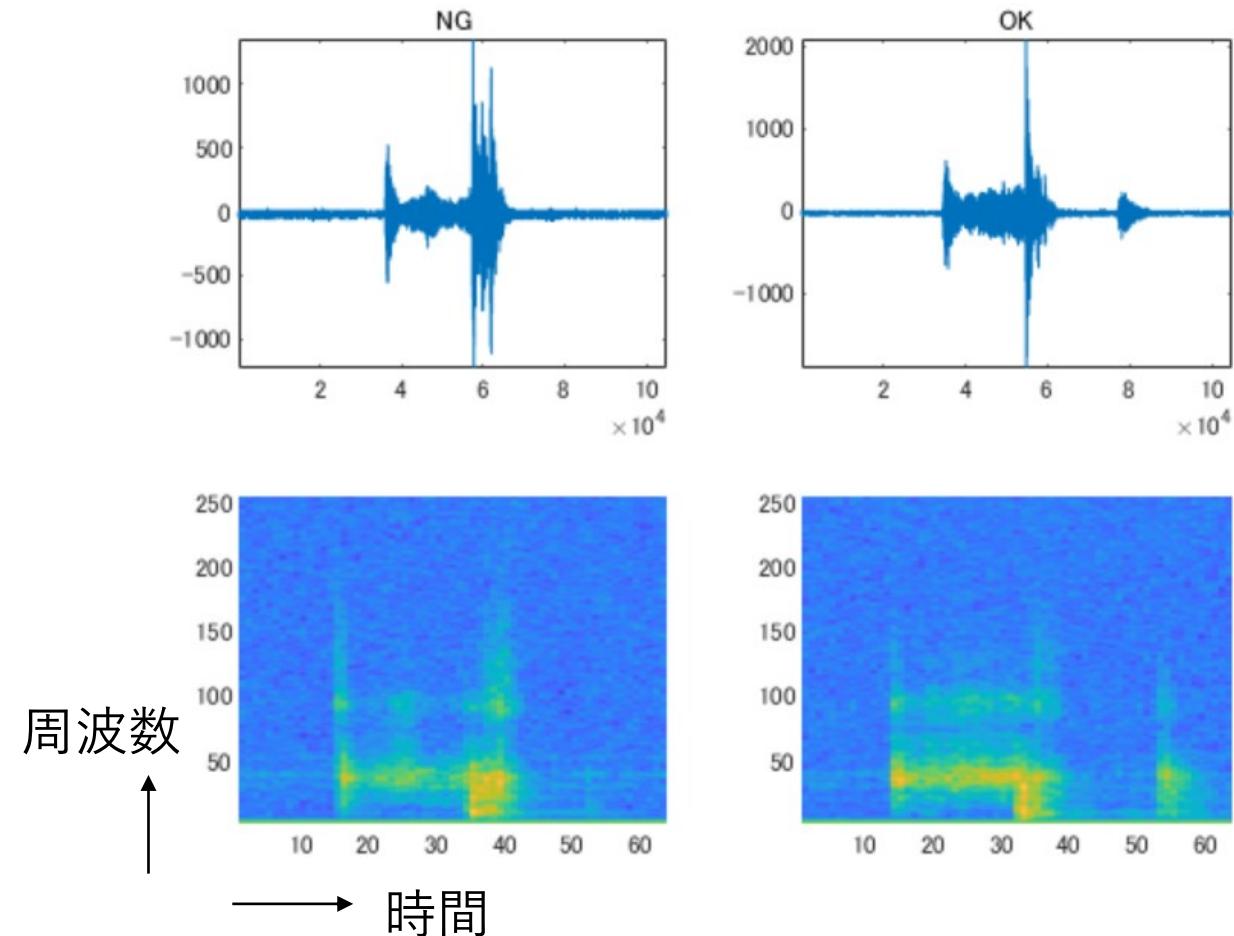
計測周波数：10 MHz

謝辞：

現場データをご提供くださいました [\(株\) ミスズ工業](#) および [Industrial Value Chain Initiative](#) の皆様に心よりお礼申し上げます。

スペクトログラム+ディープラーニングを使用した異常検知

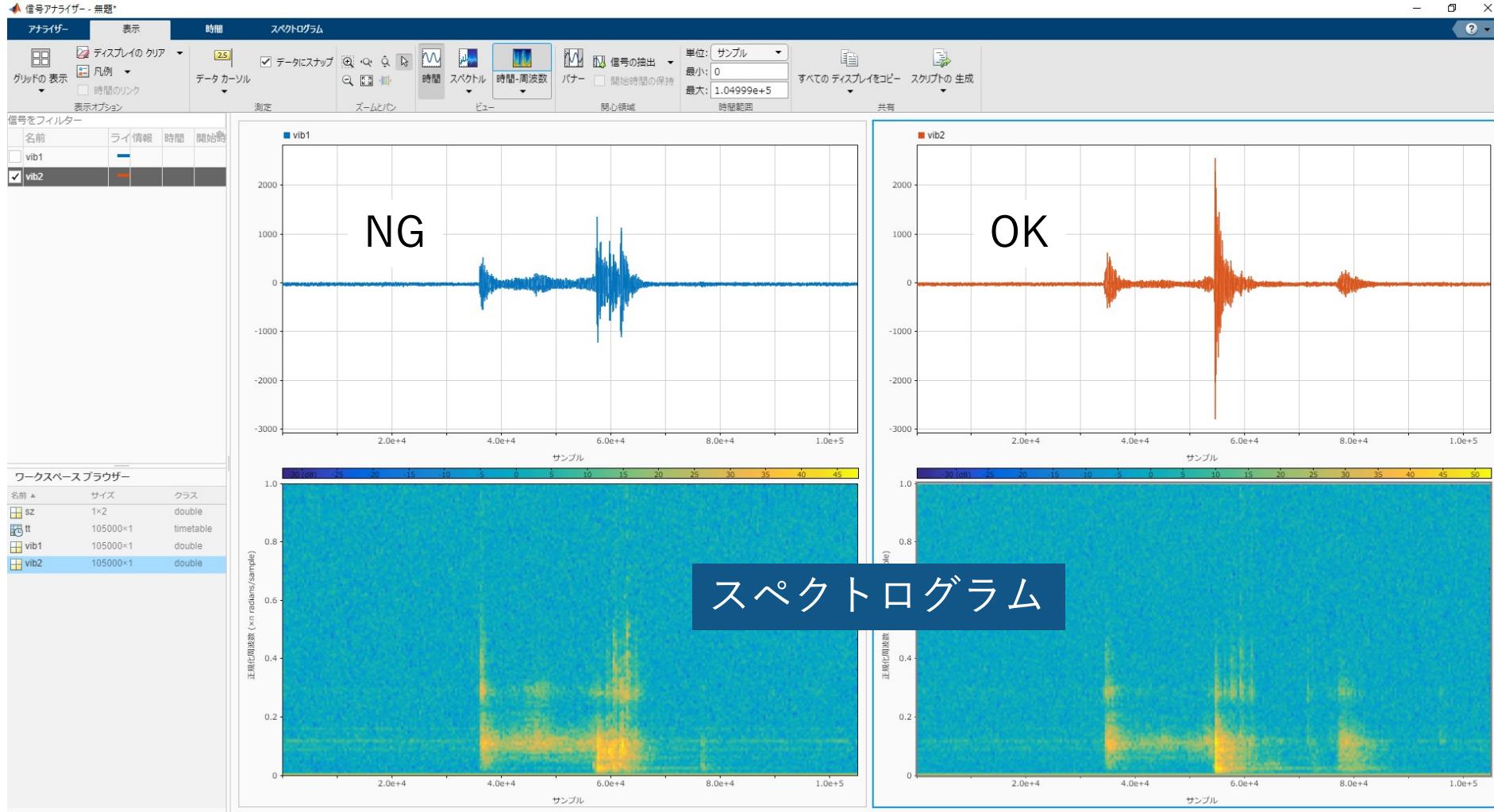
- 時間周波数解析
- スペクトログラム
(信号を画像化)
- ニューラルネットワーク
による画像識別



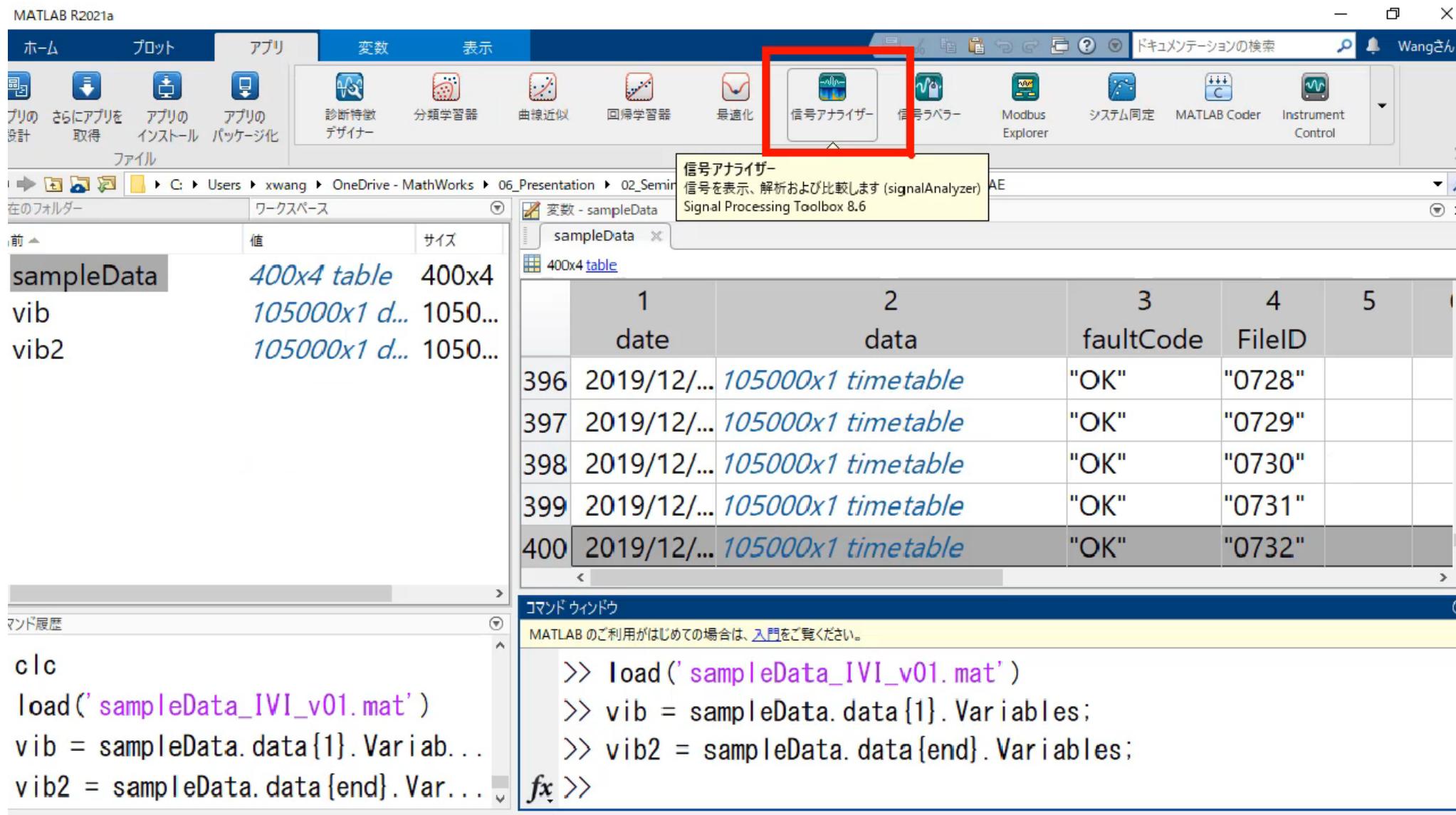
時系列データの信号処理アプリ



Signal Processing Toolbox



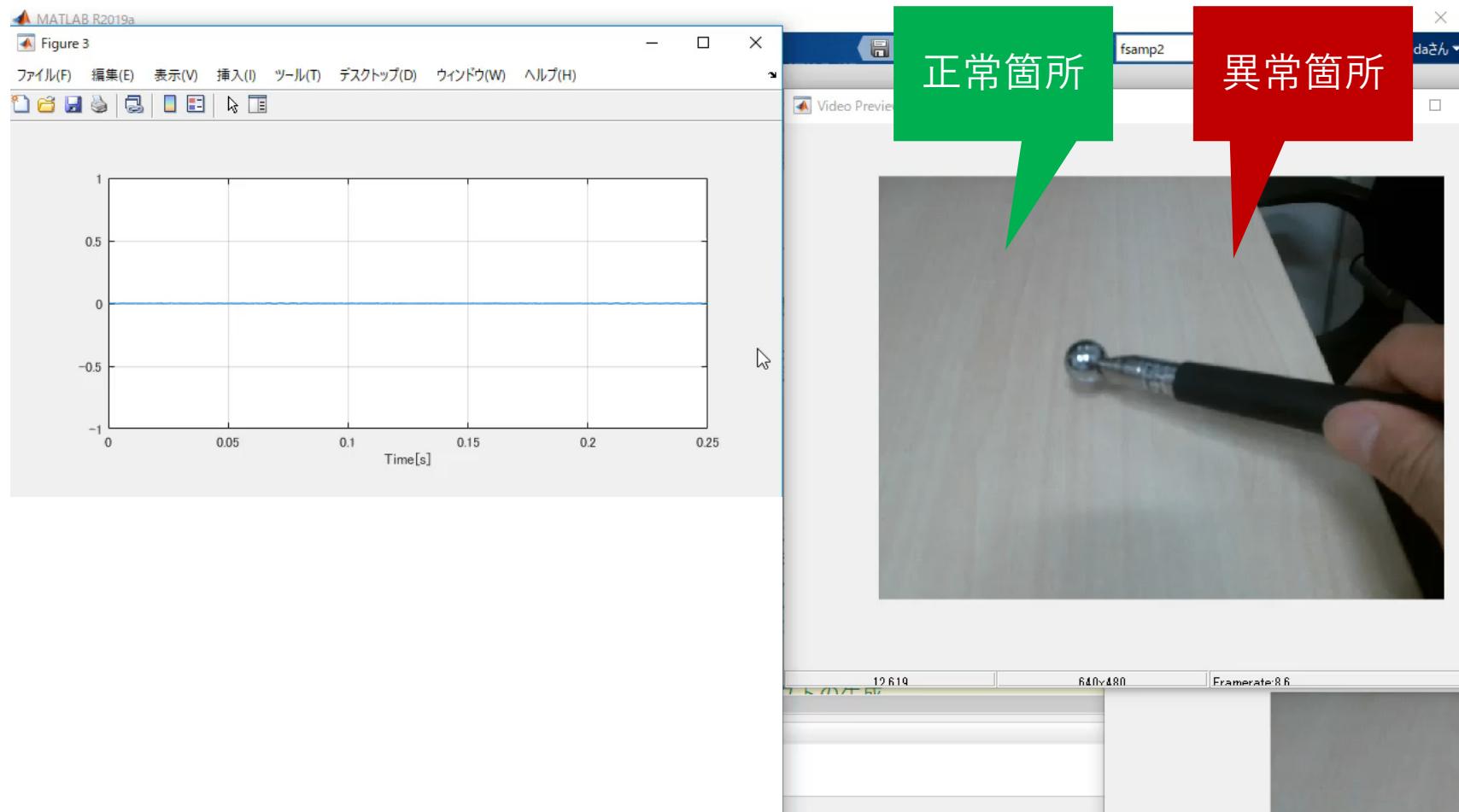
スペクトログラム + ディープラーニング



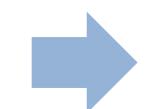
様々な時系列データ

No.	種類	特徴	クラス	データサイズ (合計)	データ点/ ファイル	ファイル数
1	振動データ	時間とともに徐々に変化	2	20 MB	10万	50個
2	AEセンサー	不定期な突発的な変化	2	1.5G	10.5万	400個
3	音データ	シーケンス長が異なる信号	2	150 MB	0.7万－1.6万	4720個
4	心電図	周期変化 & 異常値	3	70 MB	6.6万	162個

打音検査



詳細は [こちら](#)

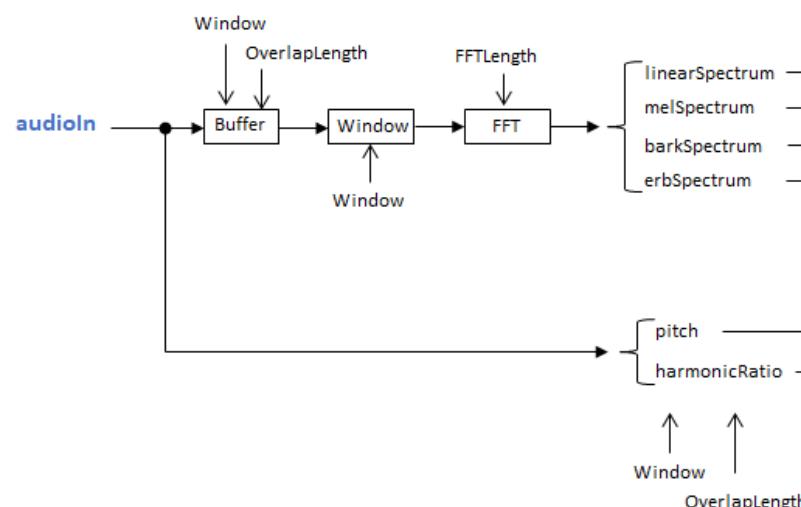


音の特徴抽出を試みる

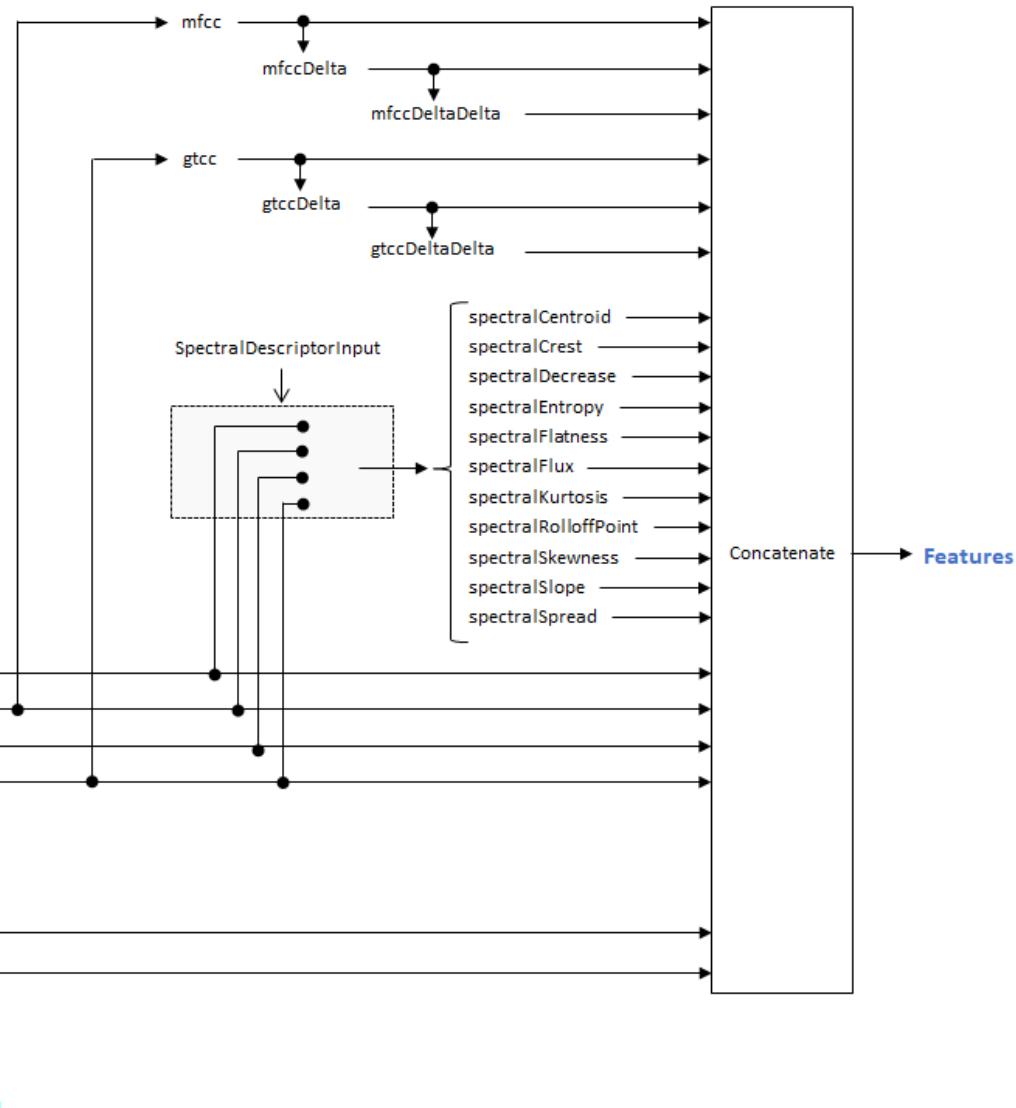
音データの特徴抽出 audioFeatureExtractor

```
>>aFE = audioFeatureExtractor( ...
    "SampleRate", fs, ...
    "Window", hamming(round(0.03*fs)), "periodic", ...
    "OverlapLength", round(0.02*fs), ...
    "mfcc", true, ...
    "mfccDelta", true, ...
    "mfccDeltaDelta", true, ...
    "pitch", true, ...
    "spectralCentroid", true);

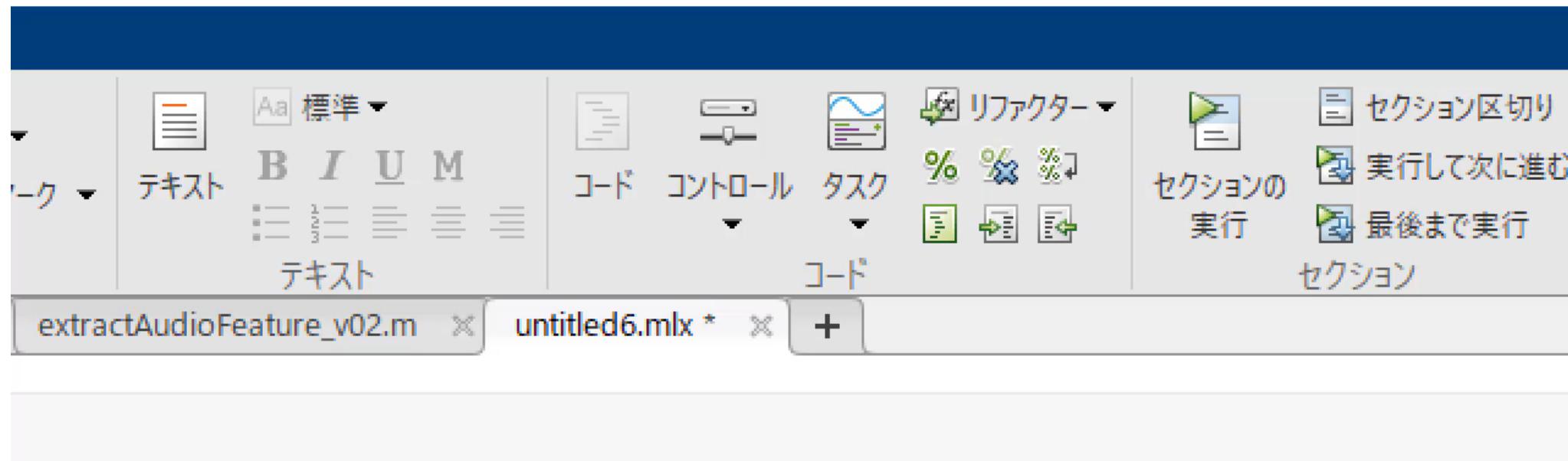
>>features = extract(aFE, audioIn);
```



Audio Toolbox



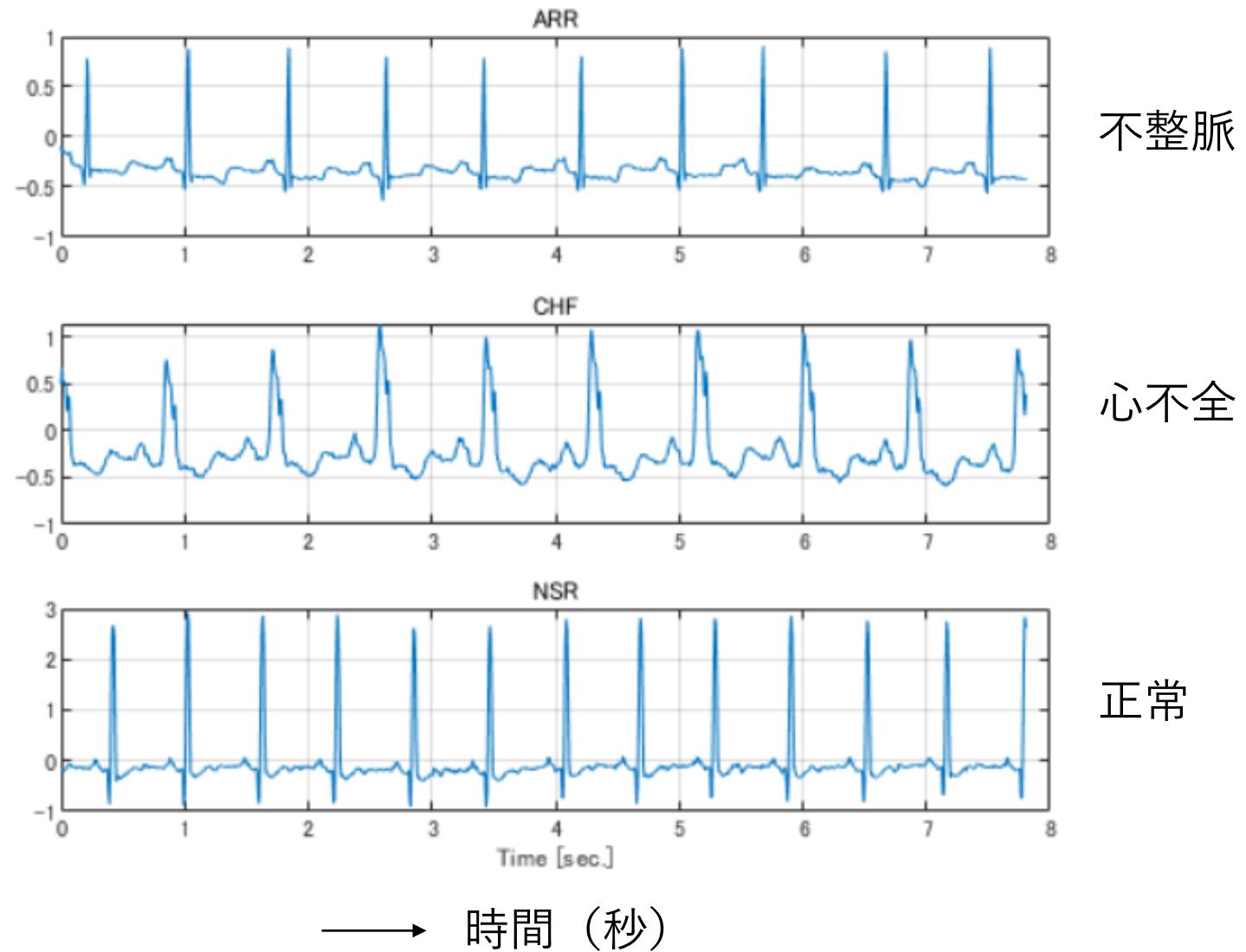
音の特徴量 + LSTM



様々な時系列データ

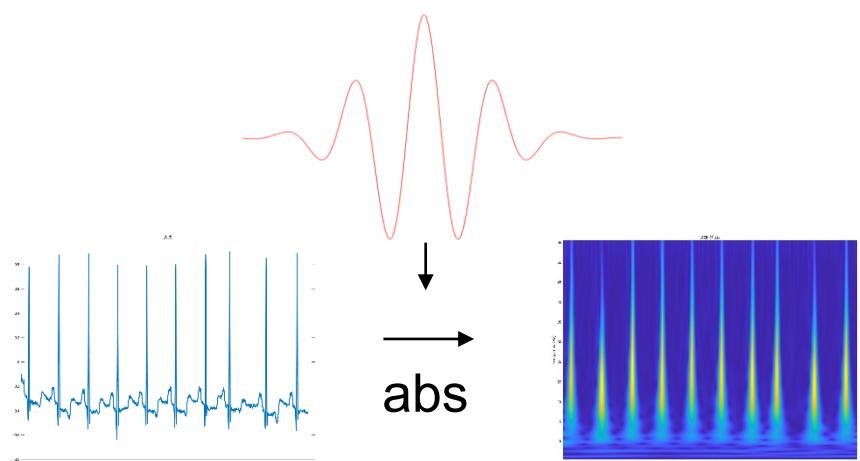
No.	種類	特徴	クラス	データサイズ (合計)	データ点/ ファイル	ファイル数
1	振動データ	時間とともに徐々に変化	2	20 MB	10万	50個
2	AEセンサー	不定期な突発的な変化	2	1.5G	10.5万	400個
3	音データ	シーケンス長が異なる信号	2	150 MB	0.7万－1.6万	4720個
4	心電図	周期変化 & 異常値	3	70 MB	6.6万	162個

生体信号

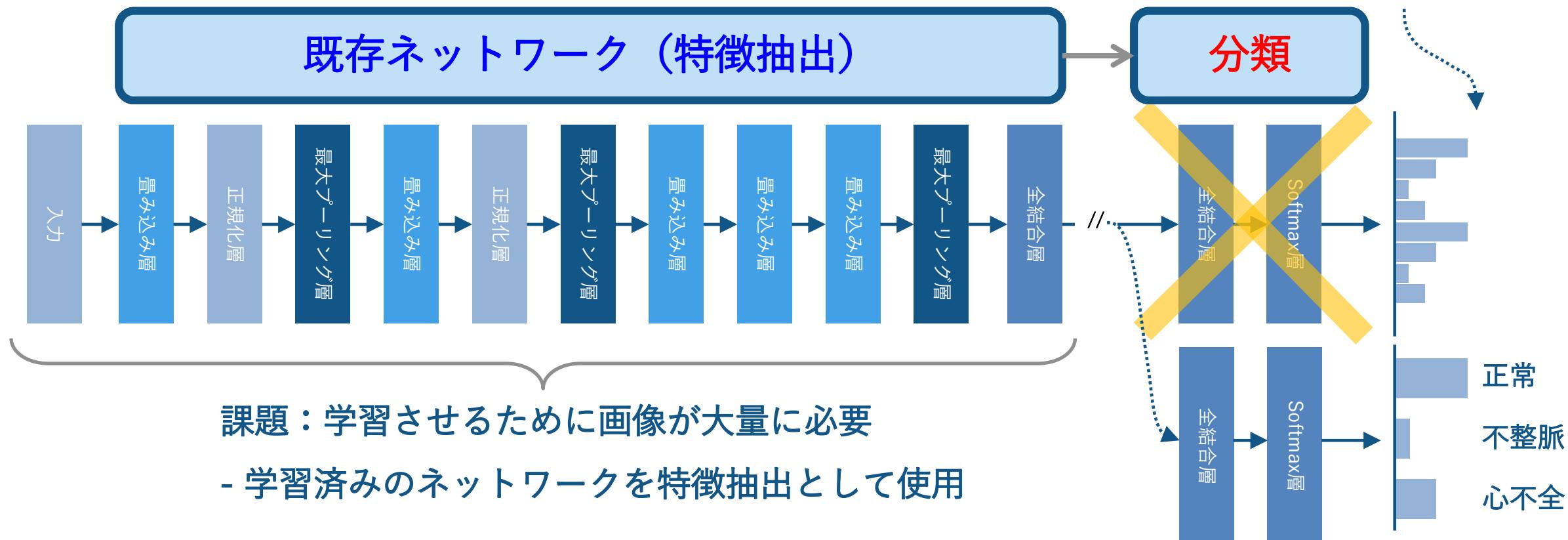


連続ウェーブレット変換とスカログラム

- スペクトル成分の変化を時間の関数として表現
- 激しい過渡状態とゆっくりと変化する振動を同時に「見つけ出す」ことが可能 ⇒ 局所的な周波数成分の探索を手助けする
- パターンは画像として保存できる



転移学習



学習済みネットワーク + 独自分類器

10行ができる転移学習

<https://www.youtube.com/watch?v=Nrx3WssLMHY>

```
1 alex = alexnet; % ネットワークを呼び出し  
2 layers = alex.Layers  
3  
4 layers(23) = fullyConnectedLayer(2); % 独自の分類をするためにネットワークを変更  
5 layers(25) = classificationLayer  
6  
7 allImages = imageDatastore('myImages', 'IncludeSubfolders', true, 'LabelSource','foldernames');  
8 [trainingImages, testImages] = splitEachLabel(allImages, 0.7, 'randomize');  
9  
10 %% ネットワークを再学習  
11 opts = trainingOptions('sgdm', 'InitialLearnRate', 0.001, 'MaxEpochs', 5, 'MiniBatchSize', 28);  
12 myNet = trainNetwork(trainingImages, layers, opts);  
13  
14 %% 精度の評価  
15 predictedLabels = classify(myNet, testImages);  
16 accuracy = mean(predictedLabels == testImages.Labels)
```

分類する個数

フォルダ名

70% 学習用

データセット利用の繰り返し

学習

スカラグラムと転移学習

Deep Network Designerの活用

The screenshot shows the MATLAB Deep Network Designer interface. The top menu bar includes 'ファイル', 'ナビゲート', 'テキスト', 'コード', 'セクション', and '実行' menus. Below the menu bar, two files are open: 'dr_ECG_scalogramTL_v03_PdMwebinar20210521 mlx' and 'dr_ECG_scalogramTL_v02_bk mlx'. The code editor window displays the following code:

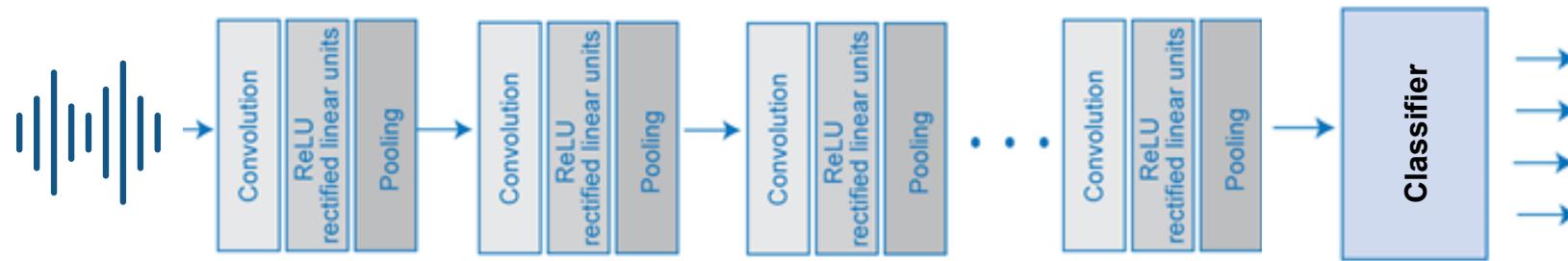
```
28 [adsTrain,adsValidation,adsTest] = splitEachLabel(ads,0.7,0.2,0.1,'randomized');

29 uniqueLabels = unique(ads.Labels);
30 tblTrain = countEachLabel(adsTrain);
31 tblValidation = countEachLabel(adsValidation);
32 tblTest = countEachLabel(adsTest);
33 figure; clf;
34 H = bar(uniqueLabels,[tblTrain.Count, tblValidation.Count, tblTest.Count],'stacked');
35 legend(H',{'学習','検証','テスト'},'Location','NorthEastOutside','FontSize',12)
```

The code uses the `splitEachLabel` function to partition the dataset into training, validation, and test sets. It then counts the number of samples for each class in each set and creates a stacked bar chart titled "データの構成を確認: 学習用 + 検証用 + テスト用". The chart shows the distribution of classes across the three sets. The legend indicates: 学習 (blue), 検証 (orange), and テスト (yellow). The y-axis ranges from 40 to 100.

セット	学習	検証	テスト
合計	68	30	2

ウェーブレット散乱変換*



◆ 比較表

	ウェーブレット散乱変換	畠み込みニューラルネットワーク
重み	固定	学習
目的	特徴抽出	分類・回帰
出力	各層の出力を使用	最終層の出力を使用

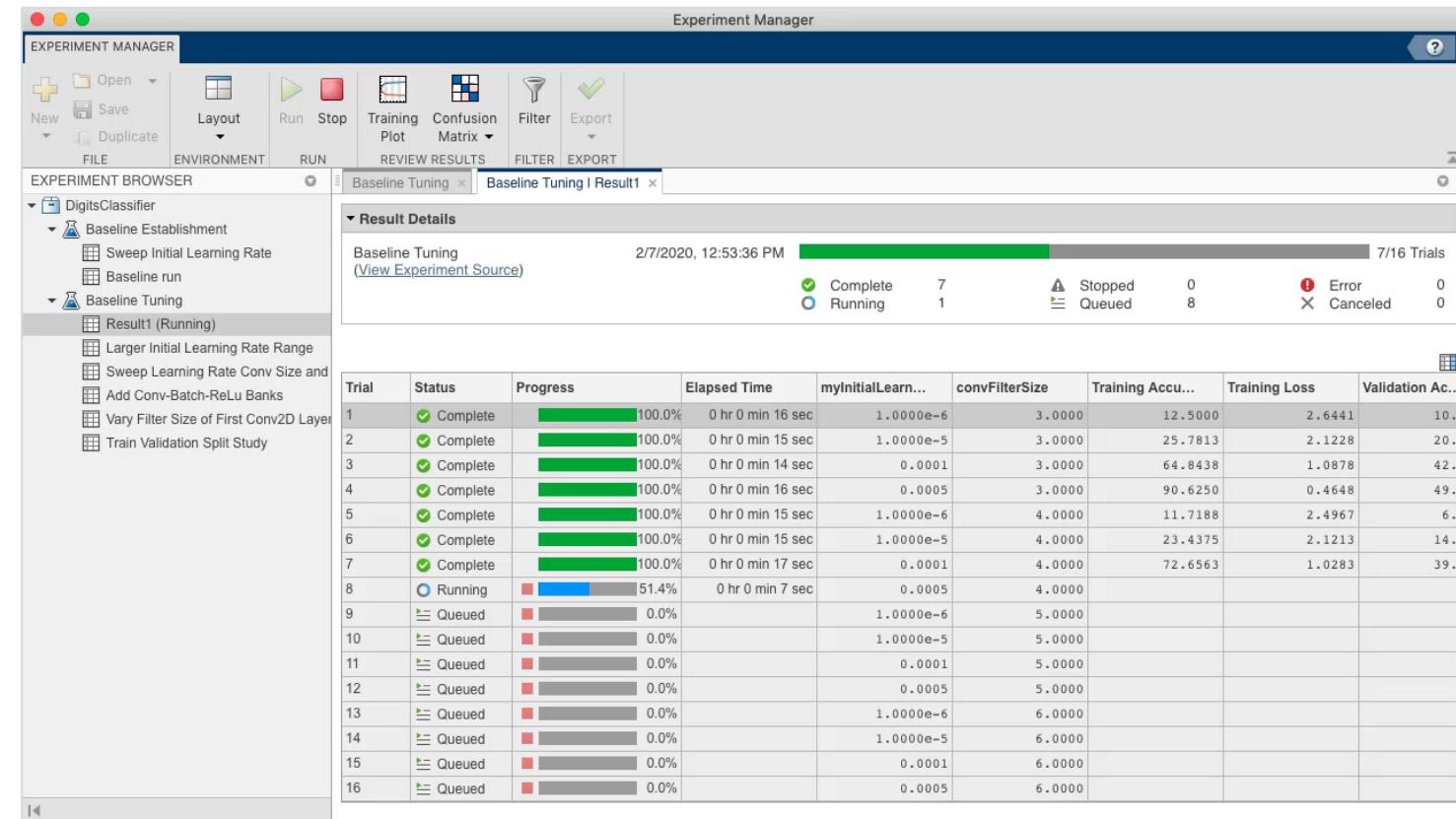
*. J. Bruna and S. Mallat, "[Invariant Scattering Convolution Networks](#)," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1872-1886, Aug. 2013.

実験マネージャー R2020a

Deep Learning Toolbox

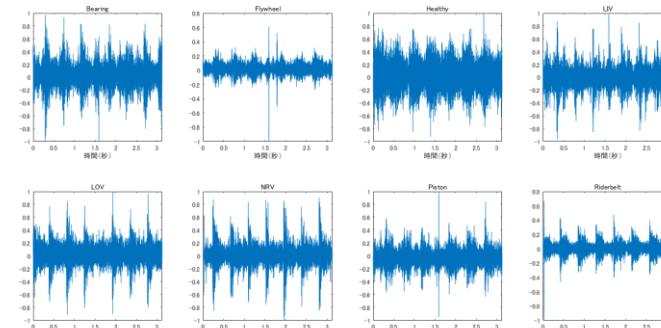
- ディープラーニングの結果管理
 - ベイズ最適化
 - ハイパーパラメータの探索
 - テンプレートあり
 - 並列計算対応

ハイパーパラメータ	
手法:	網羅的なスイープ
名前	値
Threshold	[250 200 150]
LSTMDepth	[1 2 3]
探索パラメータの指定	



ディープラーニングの解析モデルについて、ハイパーパラメータのチューニングを行い、より良いモデルを探索することができます。また、データを一元管理するためのGUIを用意しています。

応用：コンプレッサーの異音判定

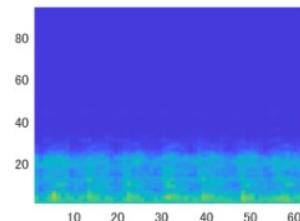


正常と7種の異音の計8クラス
225 個 / クラス

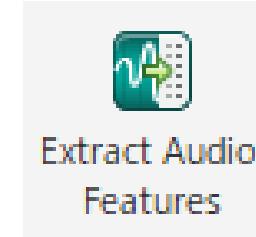
時間 / 周波数



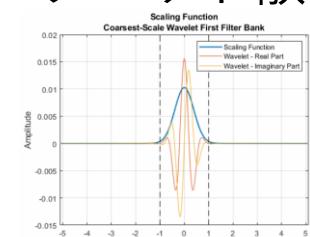
時間-周波数



ドメイン固有



ウェーブレット散乱変換

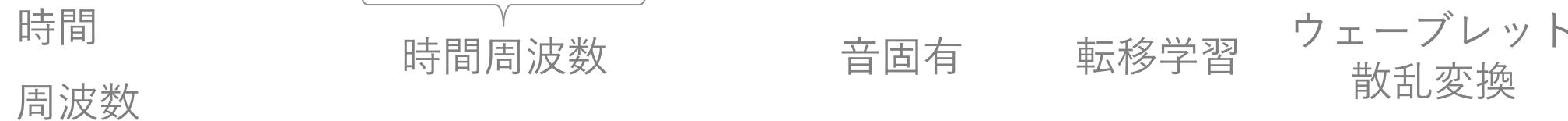


比較結果

単一CPU @ 1.9GHz

以下は前述のコンプレッサーの異音判定データに対する比較表です。
各手法のパフォーマンスを保証するものではありません。
異なるデータセットでは結果が変わる場合があります。
1つ目のPdM+MLでは29種類の機械学習モデルが生成されます。

	PdM+ ML	Spec.+ DCNN	Scalo.+ DCNN	AFE + LSTM	AFE + YAMNet	WST + LSTM
学習精度	85.2%	100%	98.9%	98.6%	98.8%	99.4%
予測精度	85.8%	99.4%	97.7%	98.3%	98.8%	98.9%
作業時間*	15 min.	17 min.	25 min.	5 min.	10 min.	3 min.



*データ加工とモデルの学習を含む目安の時間です。詳細はスクリプトをご覧ください。

比較結果

MATLABは多様な特徴抽出とそれらを実装するプラットフォーム



MathWorksは、ガートナー社の2021年の
データ・サイエンスと機械学習プラットフォーム部門についての
マジック・クアドラン트においてリーダーの1社と評価されました

MathWorksは、リーダー・クアドラントにおいて
2年連続でリーダーの1社に位置付けられ、ビジョンの完全性を
基に高い評価をされました

この評価は、以下の実行能力を実証するものと考えています

- AIおよびデータサイエンスの経験が限られているチームを強化
- データの準備、AIモデリング、システム設計、量産のための包括的なワークフローの提供
- AIモデルを組み込み機器、AIモデリング、エッジデバイス、エンタープライズシステム、クラウドに実装
- インテグレーションの課題に取り組み、AIドリブンシステムの開発におけるリスクを軽減するためのSimulinkの活用

Gartner Magic Quadrant for Data Science and Machine Learning Platforms, Peter Krensky, Carlie Idoine, Erick Brethenoux, Pieter den Hamer, Farhan Choudhary, Afraz Jaffri, Shubhangi Vashisth, 1st March 2021.

This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from MathWorks.

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner research organization and should not be construed as statements of fact. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



ご紹介内容のふりかえり

No.	種類	特徴	アプローチ	アプリ
1	振動データ	時間とともに徐々に変化	時間・周波数+ML	診断特徴デザイナー 分類学習器
2	AEセンサー	不定期な突発的な変化	スペクトログラム+DL	信号アナライザ
3	音データ	シーケンス長が異なる信号	MFCC+LSTM	audioFeatureExtractor
4	心電図	周期変化 & 異常値	スカラグラム+TL & ウェーブレット散乱変換 +LSTM	ディープネットワークデザイナー 実験マネージャー
5	異音データ	8クラスの異音判定	上記すべて	-

まとめ

- 特徴抽出には様々な方法があり、上手に活用することで異常検知を可能にする。
- MATLABには多彩な特徴抽出と機械学習・ディープラーニングとのコンボを達成できる既存の例題が豊富に提供されている。
- MATLABは一気通貫型のデータ分析のツールであり、解析から実装までのワークフローを可能にする。

【オンライントレーニングのご紹介】

AI4days トレーニングより

- ① 『機械学習アルゴリズム・理論入門』
- ② 『回帰分析 - 統計・機械学習・ディープラーニング-』

■次回開催スケジュール

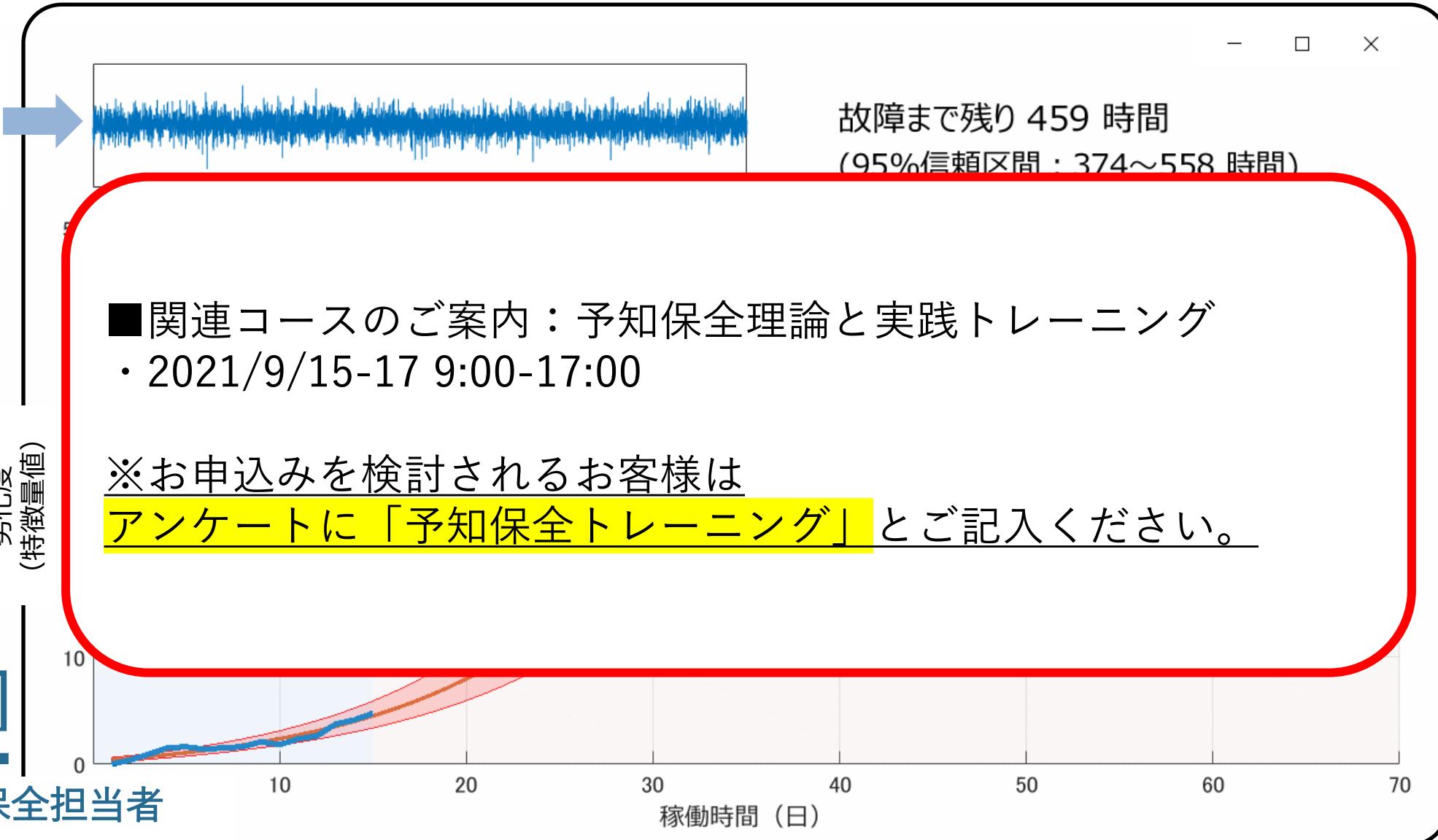
- ①2021/7/15-16 9:00-17:00
- ②2021/7/20-21 9:00-17:00

※お申込みを検討されるお客様はアンケートに「AI4days」と
ご記入ください。

Training Services,
MathWorks Japan

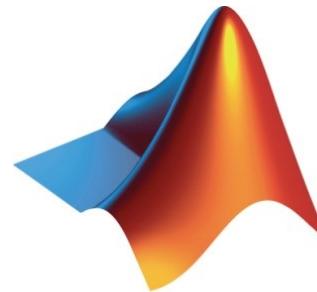
故障予測・予知保全

Predictive Maintenance Toolbox



保全担当者

項目	紹介製品
診断特徴デザイナー	<ul style="list-style-type: none">Predictive Maintenance Toolbox
機械学習 & 分類学習器	<ul style="list-style-type: none">Statistics and Machine Learning Toolbox
信号アナライザー	<ul style="list-style-type: none">Signal Processing Toolbox
ディープラーニング & 実験マネージャー	<ul style="list-style-type: none">Deep Learning Toolbox
audioFeatureExtractor 音の特徴抽出	<ul style="list-style-type: none">Audio Toolbox
スカラグラムなど Wavelet解析	<ul style="list-style-type: none">Wavelet Toolbox
コード生成 Raspberry Pi	<ul style="list-style-type: none">MATLAB Coder



MathWorks®

Accelerating the pace of engineering and science

© 2021 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See 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.