

# **Master's Thesis: Time-Sequential Wind Dancing Signal Anomaly Detection**

Xiaonan Wang

University of Electronic Science and Technology of China

2018.06

(Reorganized and Translated in 2023.04)



# Table of Contents

**1** Introduction

**2** Wind Dancing Monitoring System

**3** Benchmark Dataset

**4** Feature Engineering

**5** Time-Sequential Detection and Prediction

**6** Conclusion and Future Work

# 1 Introduction

**Research Objective** Design and build methods for time-sequential anomaly detection

Traditional Approaches:

- Mechanics Methods
- Traditional data mining methods

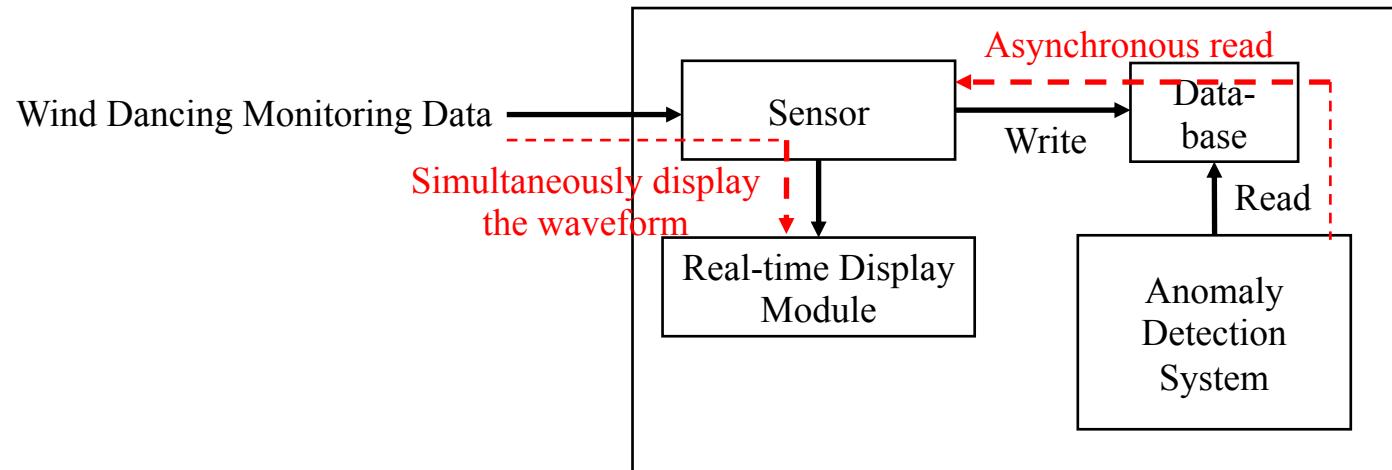
Research Objectives:

- Mining useful information from time-sequential wind dancing signal
- Anomaly Detection

Proposed:

- A Benchmark contains sequential token data
- Integrate time-domain and frequency-domain feature to 2D fusion feature
- Batch Normalization with Sliding-Window Preprocessing (BNSWP)
- CNN classifier and RNN predictor

## 2 Wind Dancing Monitoring System



# 3 Benchmark Dataset

Original wind dancing signal of a month:

$$\{sample_1, sample_2, \dots, sample_i, \dots\}$$



Framed time-sequential tokens:

$$\{token_1: \{sample_1, \dots, sample_n\}, \dots, token_i: \{sample_1, \dots, sample_n\}, \dots\}$$

Consists of  $n$  samples within  $frame_i$

$$len(frame_i) = 50s$$

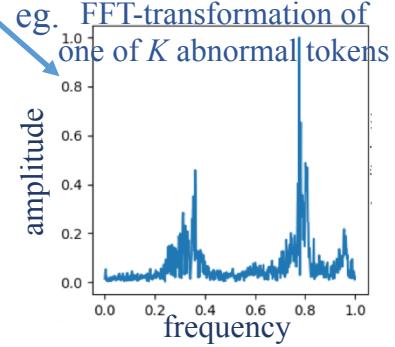
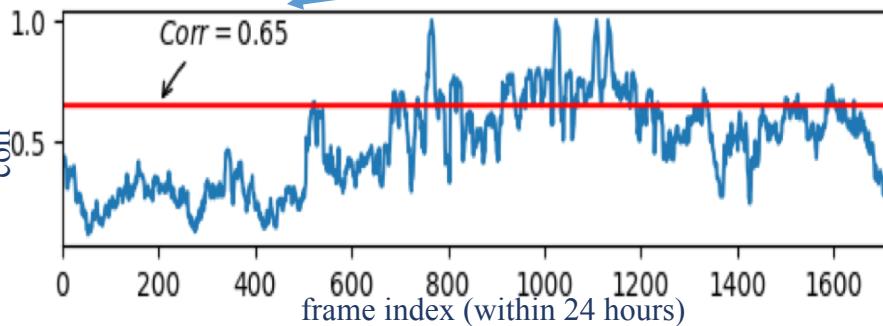
$$\frac{60s/min \times 60min/hour \times 24hour/day}{50s/frame} = 1728 \text{ frames/day}$$

Remaining Part



- 1. Relatively large dancing amplitude  
(empirical observation + box-plot statistics)
- 2. Has low frequency harmonics:

$$corr_i > corr_{th}$$
$$corr_i = \max_{1 \leq k \leq K} PearsonCorrealtion(FFT(token_i), FFT(token\_anomaly_k))$$



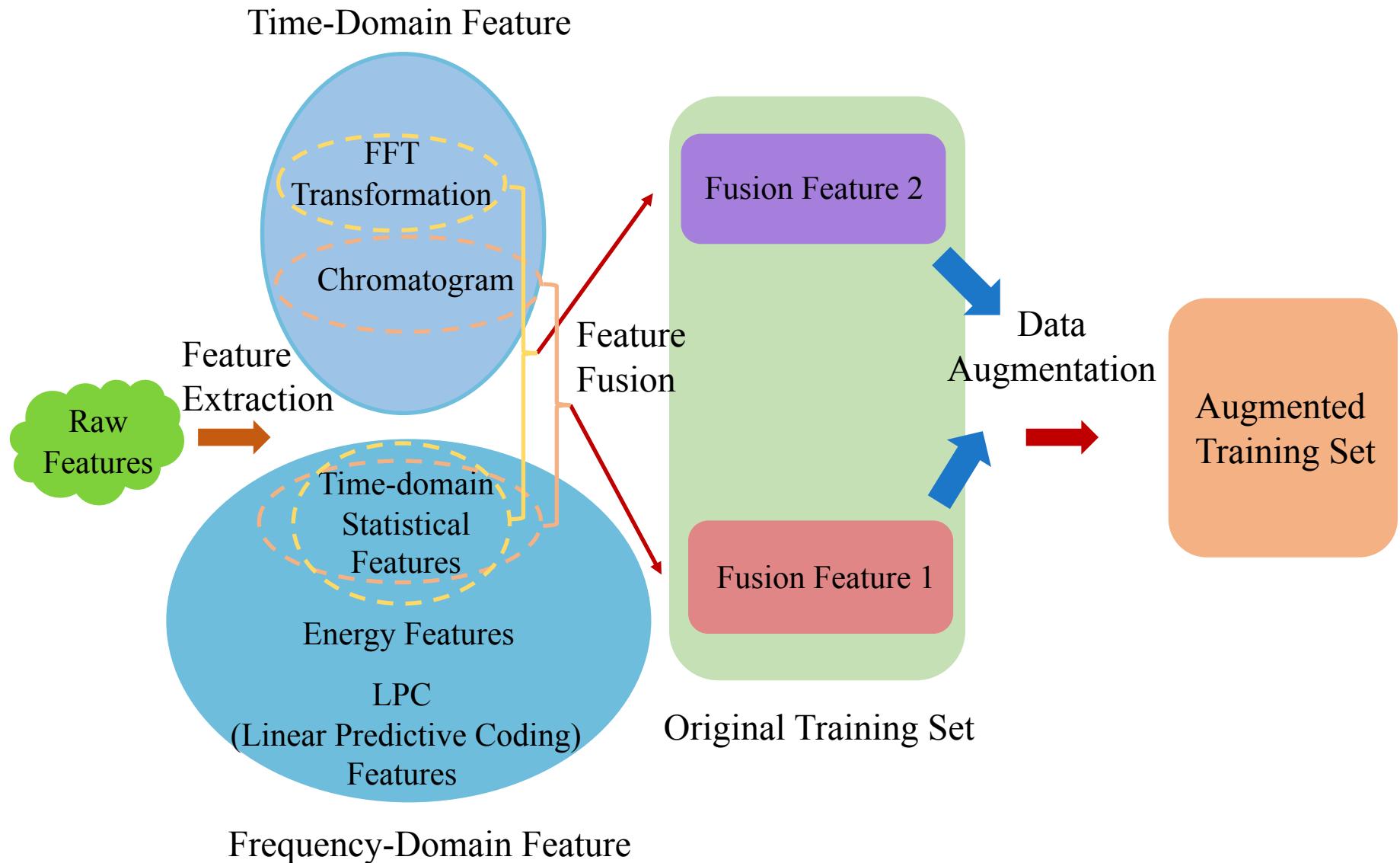
normal tokens  
*label: 0*

abnormal tokens  
*label: 1*

(933120 tokens) (42660 tokens)  
Benchmark Dataset

# 4 Feature Engineering

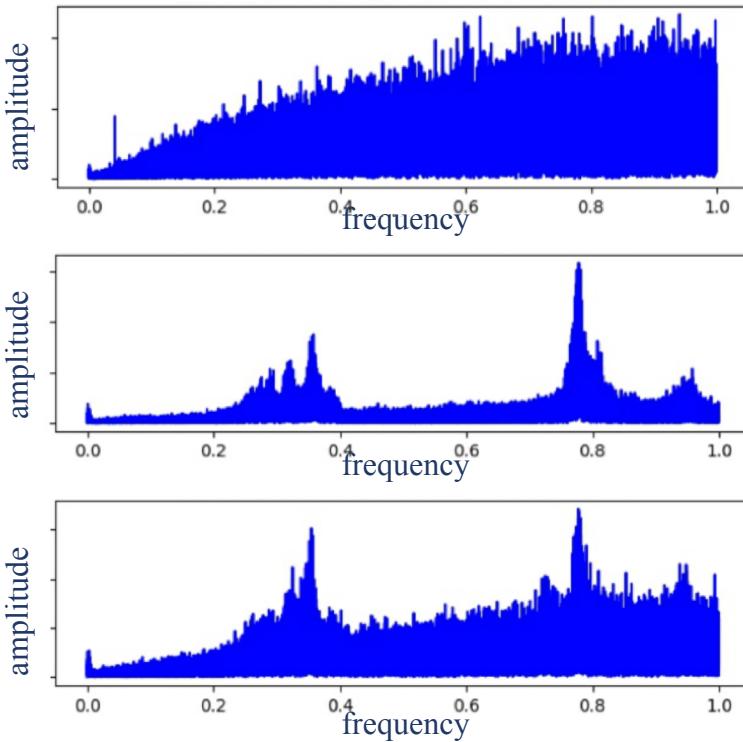
## Overview of Feature Engineering



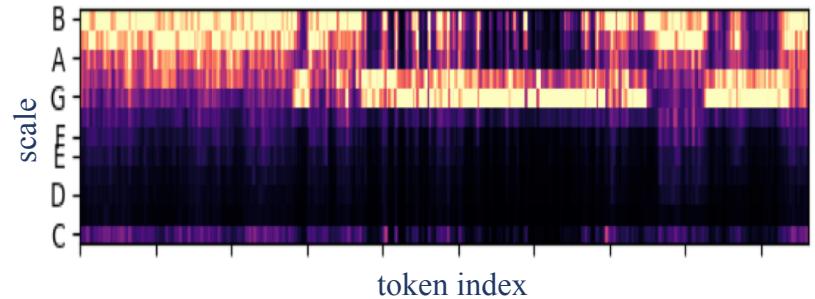
# 4 Feature Engineering

## Frequency-Domain Feature

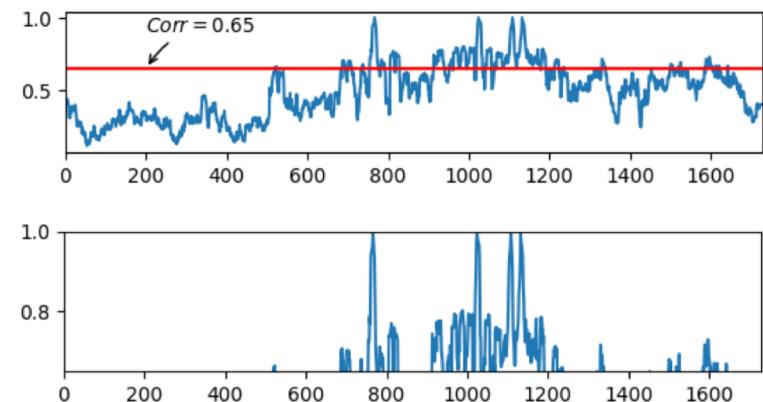
FFT Transform of Certain Token



Chromatogram Changes as Token Varies

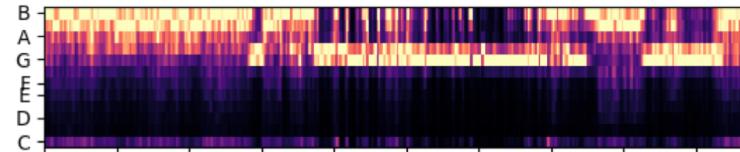


Corr vs Chromatogram



Chromatogram Reference:

D. P. W. Ellis, G. E. Poliner. Identifying cover songs' with chroma features and dynamic programming beat tracking[C]. Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on. IEEE, 2007, 4: IV-1429-IV-1432



# 4 Feature Engineering

## Time-Domain Feature

### Step1: Calculate statistical values

$$timeMax_i = \max\{abs(sample_j), sample_j \in token_i\}$$

$$timeMaxMin_i = \max\{sample_j \in token_i\} - \min\{sample_j \in token_i\}$$

$$timeMean_i = \text{mean}\{abs(sample_j), sample_j \in token_i\}$$

$$timeStd_i = \text{std}\{sample_j \in token_i\}$$
 std: standard deviation

$$timeRMS_i = \text{RMS}\{sample_j \in token_i\} = \sqrt{\text{mean}\{sample_j^2, sample_j \in token_i\}}$$

### Step2: Batch Normalization with Sliding-Window Preprocessing (BNSWP)

**Input:** Values of  $x$  over a sliding-window:  $\mathcal{SW} = \{x_{l \dots r}\}$ ;

Size of sliding-window  $m = r - l + 1$ ;

For  $i < m - 1$ :  $l = 0, r = m - 1$ ;

Else:  $\begin{cases} l = i - \frac{m-1}{2}, r = i + \frac{m-1}{2}; & \text{when } m \text{ is odd} \\ l = i - \frac{m}{2} + 1, r = i + \frac{m}{2}; & \text{when } m \text{ is even} \end{cases}$

Parameters to be learned:  $\gamma_1, \beta_1$

**Output:**  $\{y_i = \text{SWP}_{\gamma_1, \beta_1}(x_i)\}$

$$\hat{x}_i \leftarrow \frac{x_i}{\max_{\mathcal{SW}}\{x_j\}} + 1 \quad // \text{normalize}$$

$$y_i \leftarrow \gamma_1 \hat{x}_i + \beta_1 \equiv \text{SWP}_{\gamma_1, \beta_1}(x_i) \quad // \text{scale and shift}$$

**Algorithm 1:** Sliding-Window Preprocess, applied to  $x = fea_i$  (e.g.  $timeMax_i$ ) over a sliding-window in time-ordered training set.

**Note:**  $m$  is decided by the length of sliding-window, we choose 24 hours as the sliding-window's length, so  $m = 1728$  tokens,

# 4 Feature Engineering

## Time-Domain Feature

### Step2: Batch Normalization with Sliding-Window Preprocessing (BNSWP)

**Input:** Values of  $x$  over a mini-batch  $\mathcal{B} = \{x_1 \dots n\}$ ;

Parameters to be learned:  $\gamma_2, \beta_2$

**Output:**  $\{z_i = \text{BN}_{\gamma_2, \beta_2}(y_i) = \text{BN}_{\gamma_2, \beta_2}(\text{SWP}_{\gamma_1, \beta_1}(x_i))\}$

**Algorithm 2:** Batch Normalization with Sliding-Window Preprocess (BNSWP), applied to  $x$  over a mini-batch in shuffled training set.

**Input:** Network  $N$  with trainable parameters  $\Theta$ ;  
subset of shuffled inputs  $\{x\}$

**Output:** BNSWP network for inference,  $N_{\text{BNSWP}}^{\text{inf}}$

- 1:  $N_{\text{BNSWP}}^{\text{tr}} \leftarrow N$  // Training BNSWP network
- 2: Add  $z = \text{BNSWP}_{\gamma_1, \beta_1, \gamma_2, \beta_2}(x)$  to  $N_{\text{BNSWP}}^{\text{tr}}$  (Alg. 2)
- 3: Train  $N_{\text{BNSWP}}^{\text{tr}}$  to optimize  $\Theta \cup \{\gamma_1, \beta_1, \gamma_2, \beta_2\}$
- 4:  $N_{\text{BNSWP}}^{\text{inf}} \leftarrow N_{\text{BNSWP}}^{\text{tr}}$  // Inference BNSWP network  
// with frozen parameters
- 5: Process multiple training sliding-window  $\mathcal{SW}$ ,  
each of size  $m$ , and average over them:

$$\hat{x} \leftarrow \frac{x_i}{\mathbb{E}[\max_{\mathcal{SW}}\{x_j\}]} + 1$$

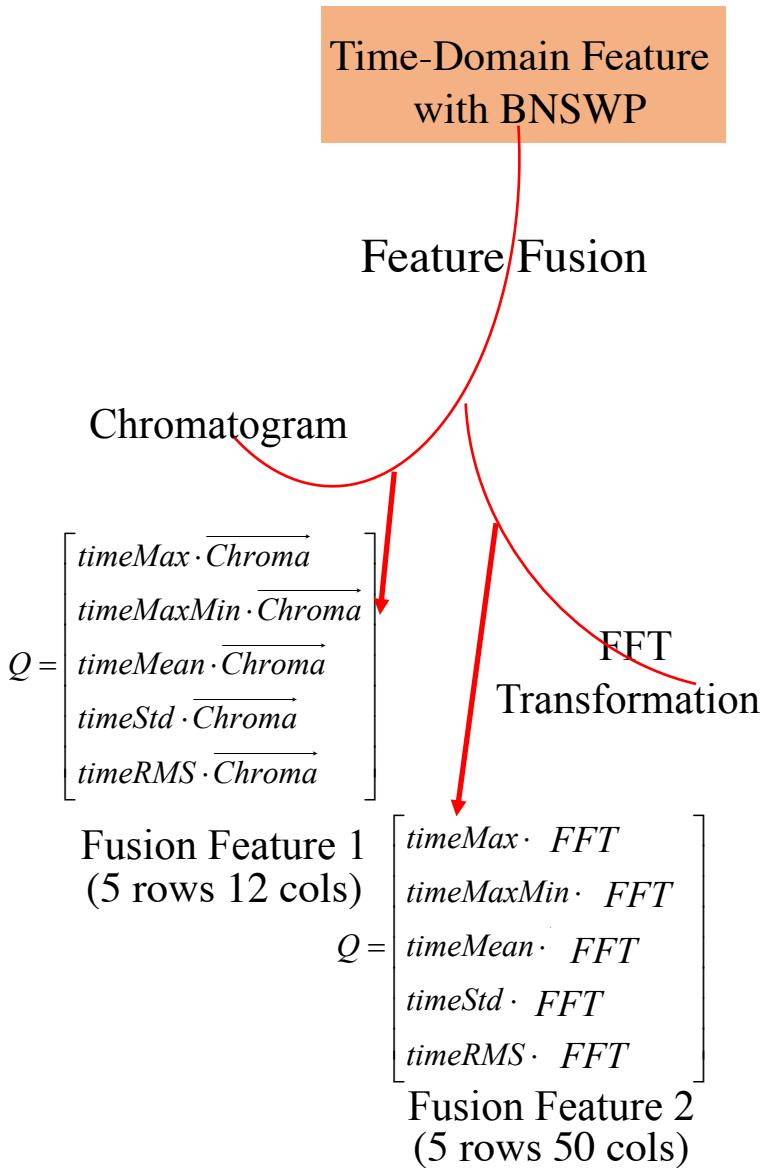
$$y \leftarrow \gamma_1 \hat{x} + \beta_1$$

- 6: Process multiple training mini-batches  $\mathcal{B}$ , each of size  $n$ , average over  $\mathcal{B}$ ; replace... // see original BN

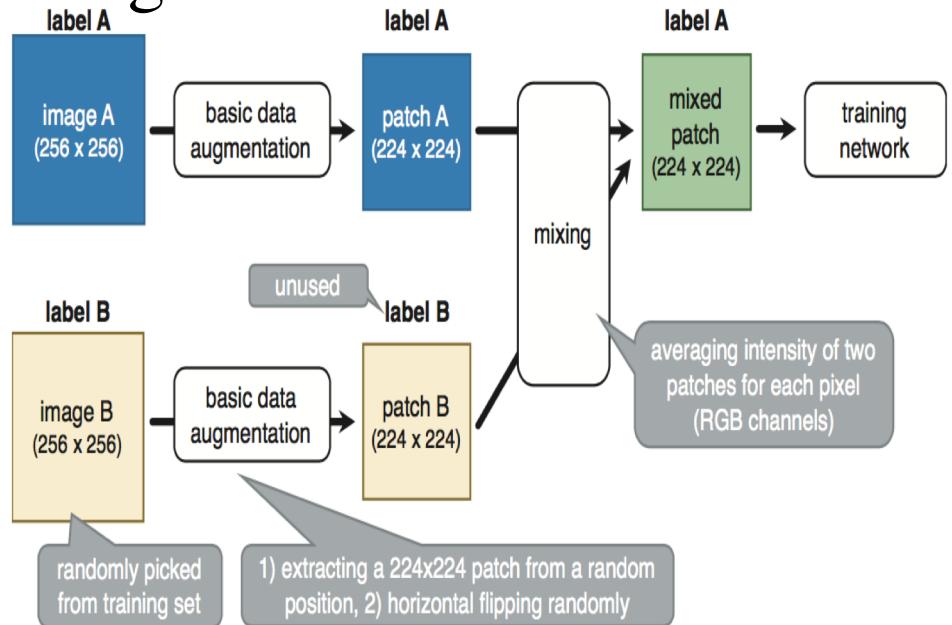
**Algorithm 3:** Training a BNSWP Network

# 4 Feature Engineering

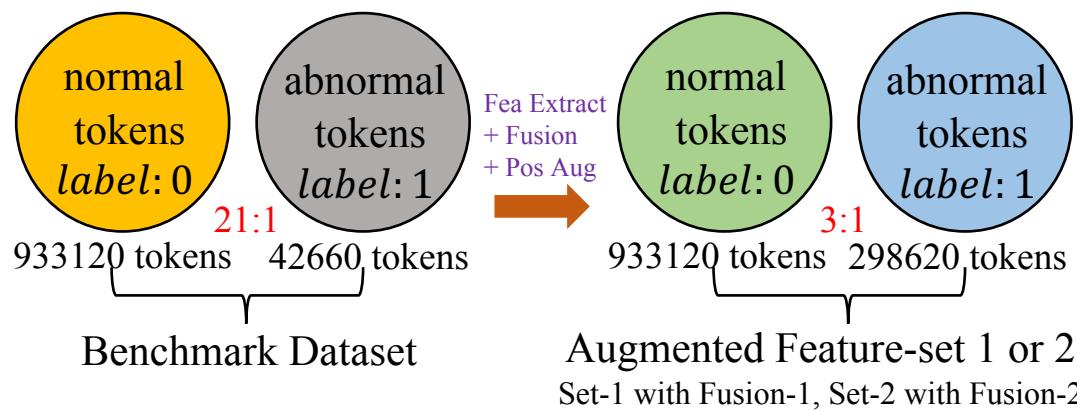
## Feature Fusion



## Data Augmentation



Reference: H. Inoue. Data Augmentation by Pairing Samples for Images Classification[J]. arXiv preprint arXiv:1801.02929, 2018



# 5 Time-Sequential Detection and Prediction

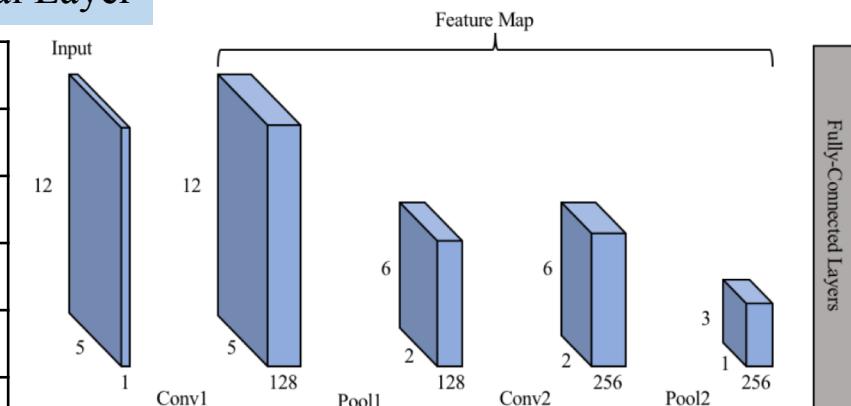
## Detection: CNN Classifier

### CNN Classifier for Feature-set 1 (Fusion 1)

$P = \text{Padding\_Width} = \text{Padding\_Height}$

#### Convolutional Layer

	Conv-Kernel Size	Stride	Padding	Output Channels
Conv-Layer 1	$5 \times 5$	1	2	128
	$3 \times 3$	1	1	128
Pool-Layer 1	$2 \times 2$	2	0	128
Conv-Layer 2	$3 \times 3$	1	1	256
Pool-Layer 2	$2 \times 2$	2	0	256

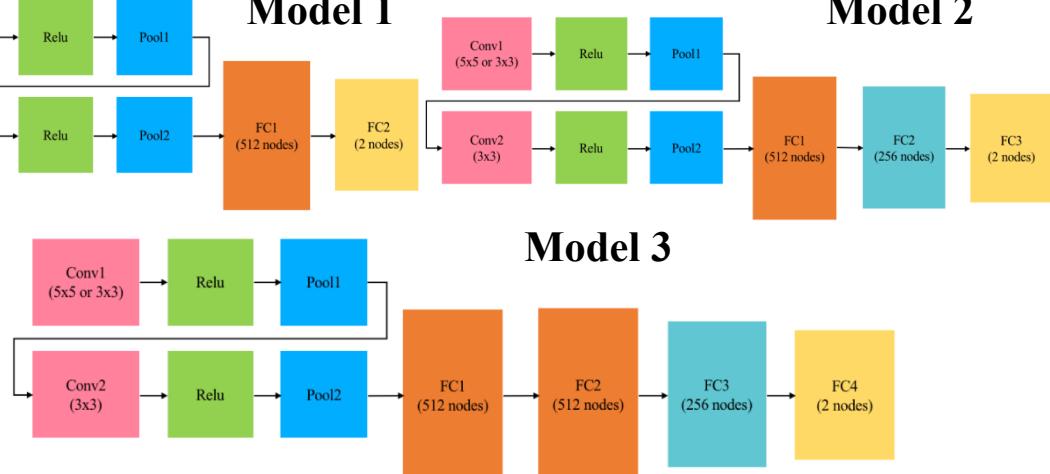


Reshape  
to  
 $1 \times 768$

#### Fully Connected Layer

	Model 1	Model 2	Model 3
FC-Layer 1	512 units	512 units	512 units
FC-Layer 2	2 units	256 units	512 units
FC-Layer 3	--	2 units	256 units
FC-Layer 4	--	--	2 units

#### Model 1



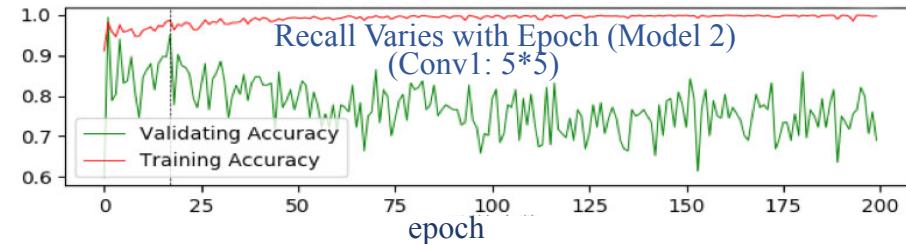
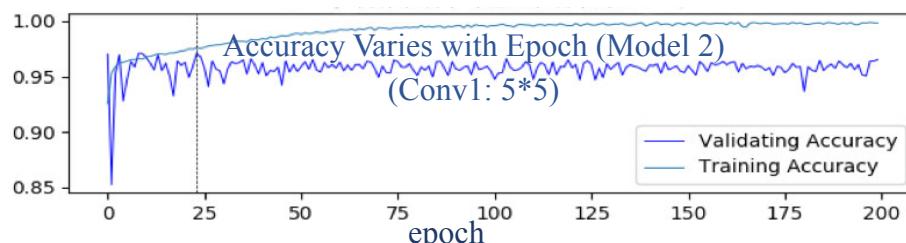
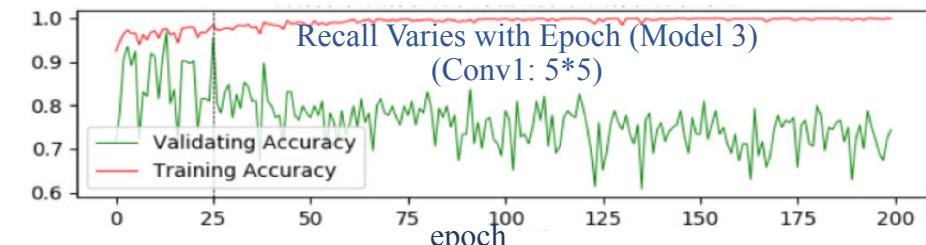
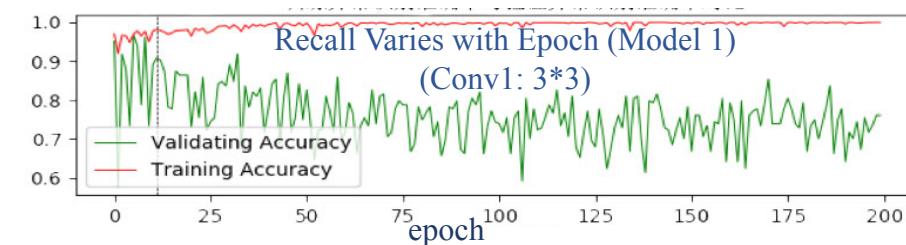
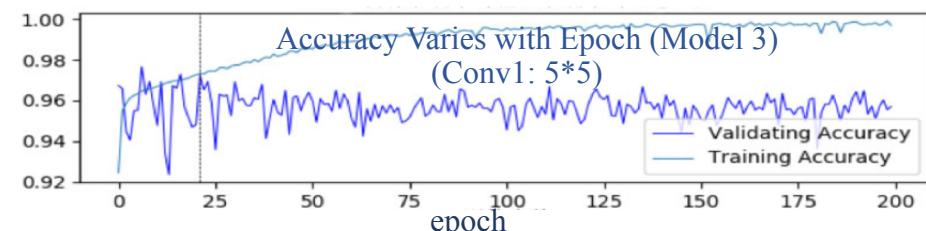
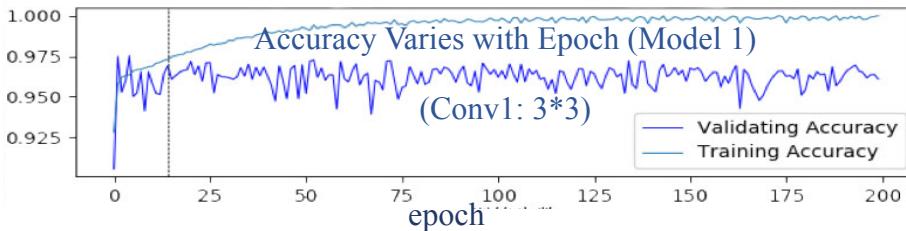
# 5 Time-Sequential Detection and Prediction

## Detection: CNN Classifier

### CNN Classifier for Feature-set 1 (Fusion 1) —— Experiments

Train

1. Regularization: L2 Norm + Early Stop + Dropout;
2. Threshold: 0.5



Test

	Conv1 Size	Epoch	Accuracy	Recall	Precision
Model 1	3 × 3	17	94.39%	82.3%	94.11%
		18	92.63%	85.9%	92.36%
	5 × 5	20	94.31%	82.8%	94.04%
		21	94.17%	83.3%	93.91%
Model 2	5 × 5	22	91.65%	88.5%	91.49%
		25	92.94%	85.4%	92.66%
Model 3	5 × 5	16	90.26%	94.3%	90.5%
		26	94.27%	84.9%	94.05%

Note: 1. Recall and Precision for Anomaly; 2. Threshold: 0.5

# 5 Time-Sequential Detection and Prediction

## Detection: CNN Classifier

### CNN Classifier for Feature-set 2 (Fusion 2)

#### Model

	Conv-Kernel Size	Stride	Padding	Output Channels
Conv-Layer 1	$5 \times 5$	1	2	128
Pool-Layer 1	$2 \times 2$	2	0	128
Conv-Layer 2	$5 \times 5$	1	2	256
Pool-Layer 2	$2 \times 2$	2	0	256
Conv-Layer 3	$3 \times 3$	1	1	512
Pool-Layer 3	$2 \times 2$	2	0	512
FC-Layer 1	1024 units			
FC-Layer 2	512 units			
FC-Layer 3	2 units			

Output of Pool 3  $\rightarrow$  Reshape to  $1 \times 1536 \rightarrow$  Input of FC 1

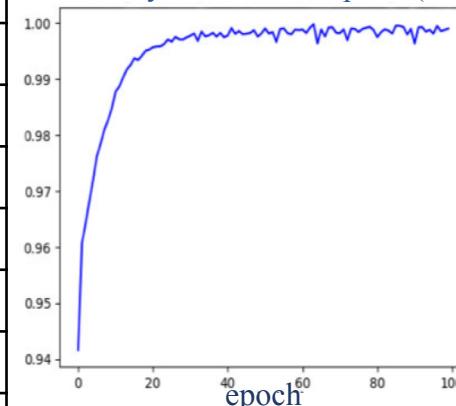
Much Weaker than  
Using Feature-set 1

#### Experiment

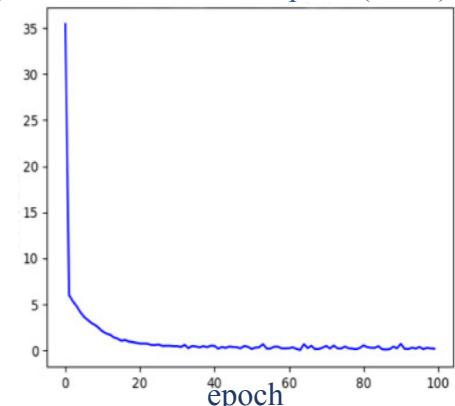
1. Regularization in Training: L2 Norm + Dropout

2. Threshold: 0.5

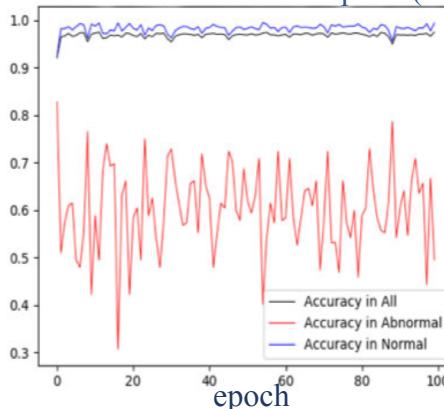
Accuracy Varies with Epoch (Train)



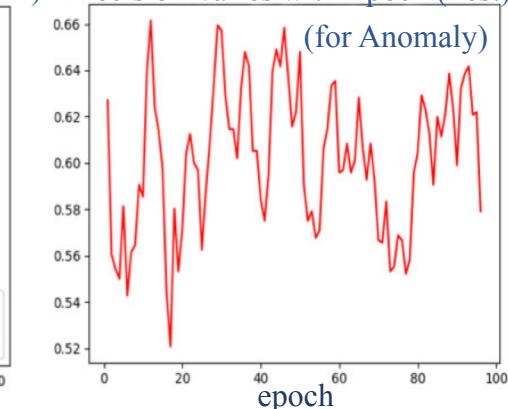
Loss Varies with Epoch (Train)



Acc/Recall Varies with Epoch (Test)



Precision Varies with Epoch (Test)



# 5 Time-Sequential Detection and Prediction

## Detection: CNN Classifier

Comparison with other models

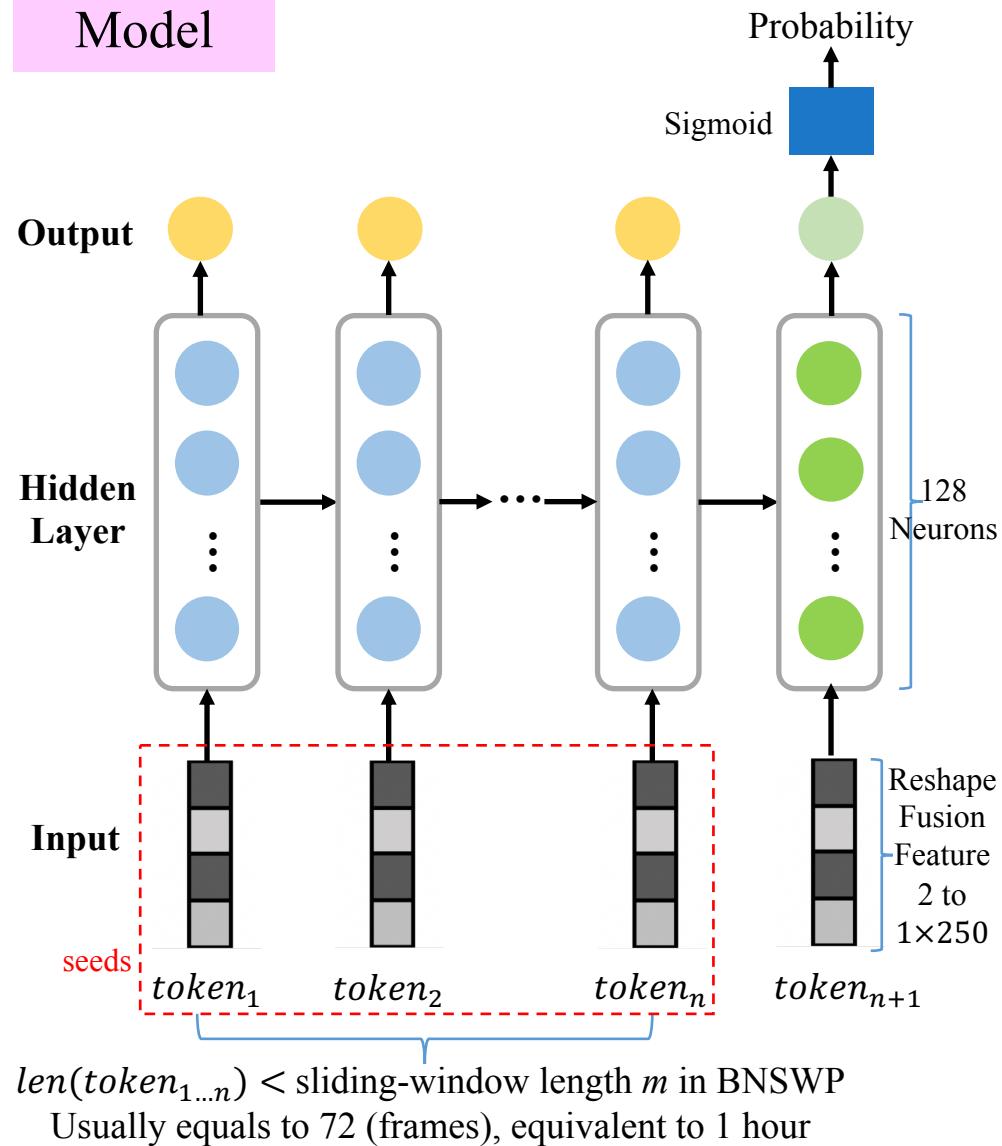
Model	kNN	LR	SVM	CNN
Training Set	Feature-set 2	Feature-set 2	Feature-set 2	Feature-set 1
Test Accuracy	90.28%	93.98%	91.67%	96%~98%
Test Recall (Anomaly)	84.90%	66.67%	70.31%	85%~90%
Test Precision (Anomaly)	89.92%	93.06%	90.35%	90%~95%
Other Instructions		SAGA Optimizer	SVM Kernel: 2nd Order Polynomial Kernel	<ol style="list-style-type: none"><li>1. Training epochs about 10 to 30</li><li>2. 3 or 4 fully connected layers</li></ol>

CNN Classifier with Feature-set 2 as inputs has a better classification effect comparing with other models

# 5 Time-Sequential Detection and Prediction

## Prediction: RNN Predictor

Model



Experiment

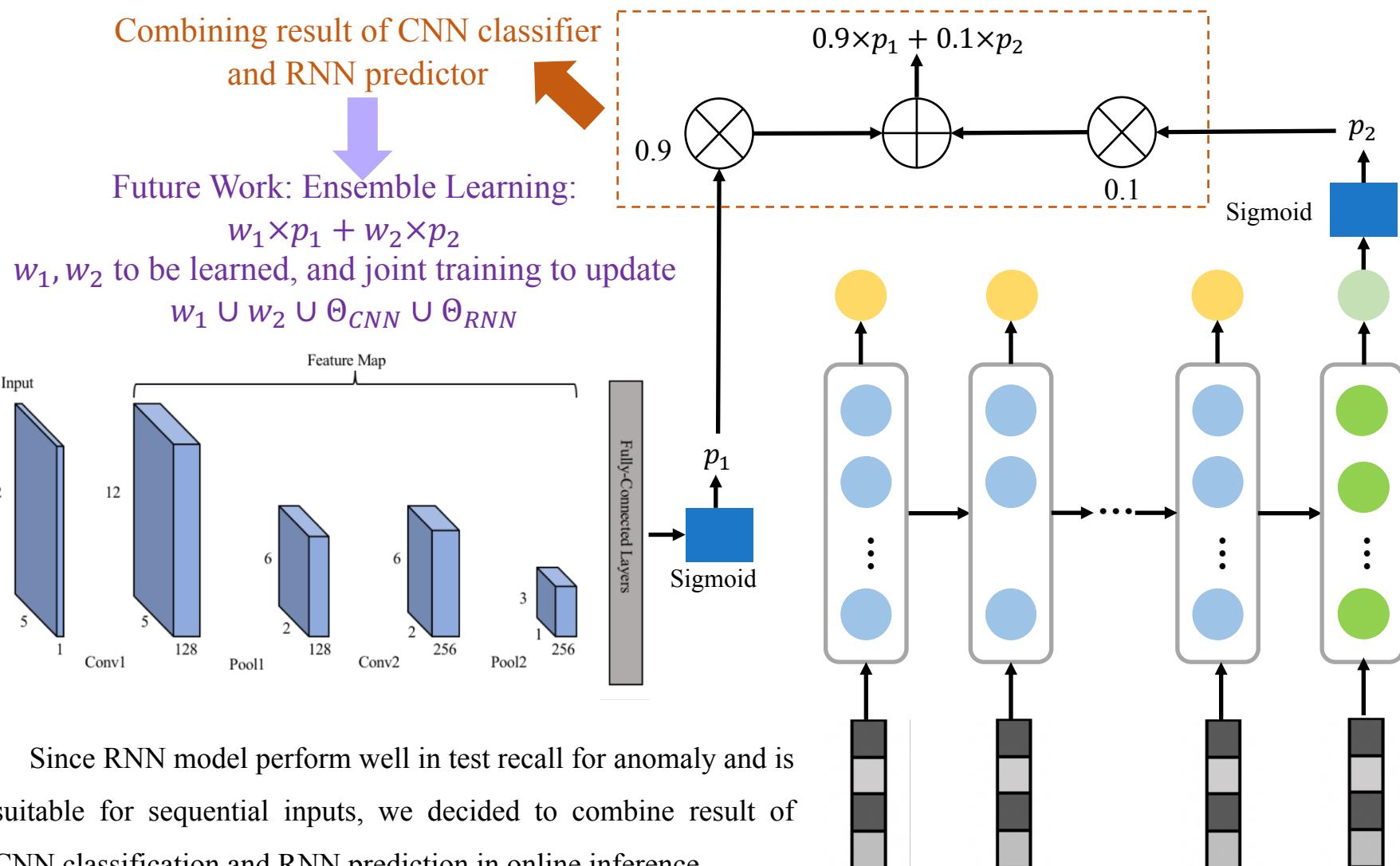
	Training Data	Stride	Test Accuracy	Test Recall	Test Precision
Model 1	1 month data (origin)	1	92.36%	73.2%	64.11%
		3	92.63%	76.9%	62.46%
Model 2	1.5 month data (augmented)	1	91.50%	78.8%	60.49%
		3	91.94%	73.4%	62.67%
Model 3	2 month data (augmented)	1	89.26%	74.3%	61.52%
		3	92.42%	77.9%	64.05%

Note:

1. For training data: *Origin* as original feature-set 2, and *Augmented* as augmented (both for positive and negative samples) feature-set 2
2.  $\text{len}(\text{seeds}) = \text{seg\_len} = n = 72$
3. Threshold: 0.5
4. Test Recall and Precision for Anomaly

# 5 Time-Sequential Detection and Prediction

## Combine Classification and Prediction in Online Inference



# 6 Conclusion and Future Work

## Conclusion

- Build a benchmark which contains sequential token data for training and evaluating models for wind dancing signal anomaly detection.
- Combine time-domain and frequency-domain feature to 2D-feature so that models such as CNN can be used. Propose Batch Normalization with Sliding-Window Preprocessing (BNSWP) method for normalization.
- Design and train CNN model with 2D fusion feature as inputs for anomaly classification and RNN model for prediction. Experiments show that CNN model performed well both in test recall and precision comparing with other models and RNN performed well in test recall. Combine result of CNN and RNN together when in inference.

## Innovation

- BNSWP: Add time-sequential information when normalizing.
- Data Fusion (1D->2D): Combine time-domain and frequency-domain information, build 2D time-sequential feature.
- Combine Classification and Prediction: Use both CNN and RNN model for anomaly detection and prediction, take advantages of both CNN and RNN.

# 6 Conclusion and Future Work

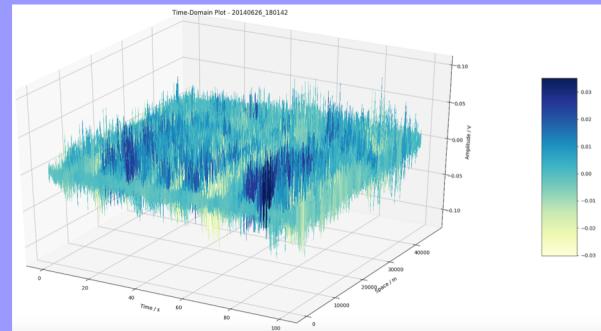
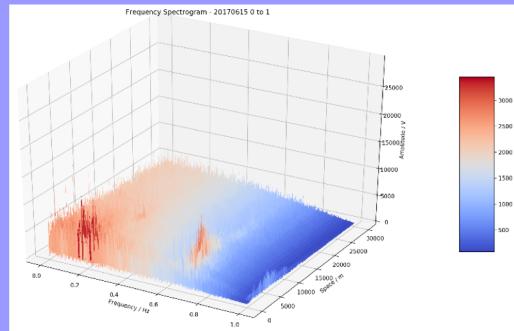
## Discussion

For CNN classifier, using fusion feature 1 get a much better result than using fusion feature 2: We suspect that it introduced more noises than information when applying fusion feature 2 to convolutional situation.

## Future Work

Dealing with problems about noise and sampling, utilize prior/posterior information when building models.

Try to use continuous time-space/frequency-space data for further more complex modeling.



Ensemble Learning: Assemble CNN and RNN as a whole model and joint training.

Consider about clustering algorithms.



# Thanks!