 doi:10.20944/preprints202001.0028.v1

Article

Anomaly Detection in Particulate Matter Sensor using Hypothesis Pruning Generative Adversarial Network

YeongHyeon Park [1,*], Won Seok Park [1], Yeong Beom Kim[1], and Seok Woong Chang [1]

[1] SK Planet Co.,Ltd., Seongnam, 13487, Korea;
yeonghyeon@sk.com, pwswonder@sk.com, ybkim@sk.com, and swchang@sk.com
* Correspondence: swchang@sk.com

Abstract: World Health Organization (WHO) provides the guideline for managing the Particulate Matter (PM) level because when the PM level is higher, it threats the human health. For managing PM level, the procedure for measuring PM value is needed firstly. The Beta Attenuation Monitor (BAM)-based PM sensor can be used for measuring PM value precisely. However, BAM-based sensor occurs not only high cost for maintaining but also cause of lower spatial resolution for monitoring PM level. We use Tapered Element Oscillating Microbalance (TEOM)-based sensors, which needs lower cost than BAM-based sensor, as a way to increase spatial resolution for monitoring PM level. The disadvantage of TEOM-based sensor is higher probability of malfunctioning than BAM-based sensor. In this paper, we aim to detect malfunctions for the maintenance of these cost-effective sensors. In this paper, we call many kinds of malfunctions from sensor as anomaly, and our purpose is detecting anomalies in PM sensor. We propose a novel architecture named with Hypothesis Pruning Generative Adversarial Network (HP-GAN) for anomaly detection. We present the performance comparison with other anomaly detection models with experiments. The results show that proposed architecture, HP-GAN, achieves cutting-edge performance at anomaly detection.

Keywords: anomaly detection; Generative Adversarial Network; multiple hypothesis; particulate matter

1. Introduction

The smaller particle can infiltrate into more the deeper site of respiratory organ. When the diameter of Particulate Matter (PM) is same or less than $2.5\mu g/m^3$ it named $PM_{2.5}$, and when diameter is between 2.5 and 10 it named PM_{10}. Thus, the smaller particle, $PM_{2.5}$, has the probability to make more bad effect to human health than PM_{10} when exposed to the same amount [1].

The PM can trigger not only respiratory diseases [2] but also cardiovascular disease [3], lung cancer [4], or some other diseases. This is the reason that World Health Organization (WHO) provides their guideline to recommend managing the PM level as shown in Table 1 [5].

Table 1. Air Quality Guideline (AQG) for Particulate Matter (PM) level management from World Health Organization (WHO) [5]. The unit of each value is $\mu g/m^3$.

Model	Annual mean	24-hour mean
$PM_{2.5}$	10	25
PM_{10}	20	50

Before managing PM level as above table, the process for measuring the air condition should be preceded. In the Republic of Korea, Air Korea that operated by Korea Environment Corporation

 doi:10.20944/preprints202001.0028.v1

provides measured values of SO_2, CO, O_3, NO_2, $PM_{2.5}$ and PM_{10} with unit of one hour. However, the spatial resolution is relatively low when monitoring PM levels using information provided by Air Korea.

For measuring PM value, two kinds of the method-based sensor can be used respectively [6,7]. One of methods is Beta Attenuation Monitor (BAM), and the other one is Tapered Element Oscillating Microbalance (TEOM). BAM-based sensor is precise and less affected by the environment, but disadvantage of them is sensor cost and maintaining cost. Also, the spatial resolution of the PM monitoring via BAM-based sensor is lower because of above disadvantage. On the other hand, TEOM-based sensor needs lower cost than BAM-based sensor. Altough, TEOM-based sensor slightly influenced by the external environment, it performs as reasonable level with providing right information.

The two coefficients are already measured between BAM-based sensor and TEOM-based sensor for proving the ability of TEOM-based sensor. One of them is Pearson's correlation coefficient and the other is coefficient of determination for 1-hour averages. They are measured as 0.91 and 0.81 respectively, and it represents the TEOM-based sensor can be used for monitoring the PM level [6].

For informing the air pollution to the public, the higher spatial resolution may be more effective than lower. In above purpose and based on the cost-effectiveness, we have installed TEOM-based sensor at the several regions as a trial. However, the TEOM-based sensor has a more probability of measuring error or malfunctioning, in the information collecting process than the BAM-based sensor, because it affected by the external environment that already mentioned above; we call these as an anomaly. Therefore, it is necessary to increase the reliability of the collected information by anomaly detection and maintaining the overall of the sensor.

The organization of this paper is described as follows. In Section 2, we summarize previous studies that have efforted for anomaly detection. We present the proposed architecture and experimental results in Section 3 and Section 4 respectively, and conclude the whole content in final section. We only deal with anomaly detection task of functioning sensors in this paper. Thus, the task after anomaly detection, can be categorized to correcting the collected values or maintain the sensor via engineer, will be handled in future study.

2. Related Works

Several previous studies have conducted the anomaly detection via classification approach with various methods [8–12]. However, there are some problem for using classification method such as difficulty of collecting diverse abnormal cases and labeling cost from collected data to specific category.

For detecting anomaly, the above problems can be eased by regression-based method such as One-Class SVM [13], Auto-Encoder (AE) [14,15], or Long-Short Term Memory (LSTM) [16,17]. The idea of regression-based method is such simple, and the cost of preparation for training conventional machine learning or deep learning-based anomaly detection algorithm is not high. Because we need only the normal case (healthy) data for training and it does not need the labeling or categorizing process. Labeling is the essential thing of validation process for assess quantitative performance of anomaly detection model, but labeling task is simple in regression concept because each data needs only checked whether normal or abnormal.

There are several neural network based anomaly detection model are already published [18–24]. The generative neural network which trained with variational bound can make the user desired data from the random noise [18,25]. However, the process of generating data from noise is not needed, moreover it is not not essential in anomaly detection procedure. Also, above deep learning architecture may generating the blurred data because of using variational bound with distribution assumption.

The Generative Adversarial Network (GAN)-based anomaly detection model is published with naming AnoGAN [19]. The AnoGAN can generate more sharped data because it does not use the variational bound. However, the one procedure, finding the most cloest data to the input data that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

generated from random noise, is essentially needed for using AnoGAN. This procedure makes the lower throughput than typical and simple AE-based anomaly detection model.

Some of the recent research, such as BiGAN [20] and GANomaly [21], eased the above limitation. These two models do not use the random noise for generating data for measure the anomaly score. They use the encoded latent vector from the input data and that vector as the input of the generator. However, generating high resolution or sharp data via BiGAN is difficult because the encoder and generator of BiGAN share the parameter for each other. The GANomaly [21] simply use the separated parameter for encoder, generator, and discriminator respectively and it overcomes the limitation of BiGAN.

One other research tried to generate data consistently with avoiding blurred output [24]. For achieving their purpose, they apply the multiple hypothesis after the generator of the GAN, and select best case among hypotheses as a output. However, they still use the variational bound as same as Variational Auto-Encoder (VAE) [25] for training. Thus it can be regarded that has still limitation for generating sharped data.

The preprocessing is the one of the consideration to construct anomaly detection system with reducing the computational time. Preprocessing technique can be used for reducing dimension of the information [26–30]. However, in our case preprocessing does not needed and does not considered because PM sensor collects the value every hour that means dimension of the information is not high to process.

3. Proposed Architecture

In this section, we present the proposed architecture for anomaly detection in PM sensor. We firstly compare the property of two kinds of convolution. Then, we present the structure of neural network. Finally, we describe the anomaly detection procedure.

3.1. Reason for Using 1D Convolution

When use the image data as the input, the 2 dimensional (2D) convolutional layer can be used for constructing the neural network. Howerver, we use the 1 dimensional (1D) convolutional layer because our data is 1D signal data.

The signal data can have the multiple channel. For example, $PM_{2.5}$ and PM_{10} can represent first and second channel respectively. For processing the multi-channel signal data, 2D convolution can be used, but it is not efficient as shown in Figure 1.

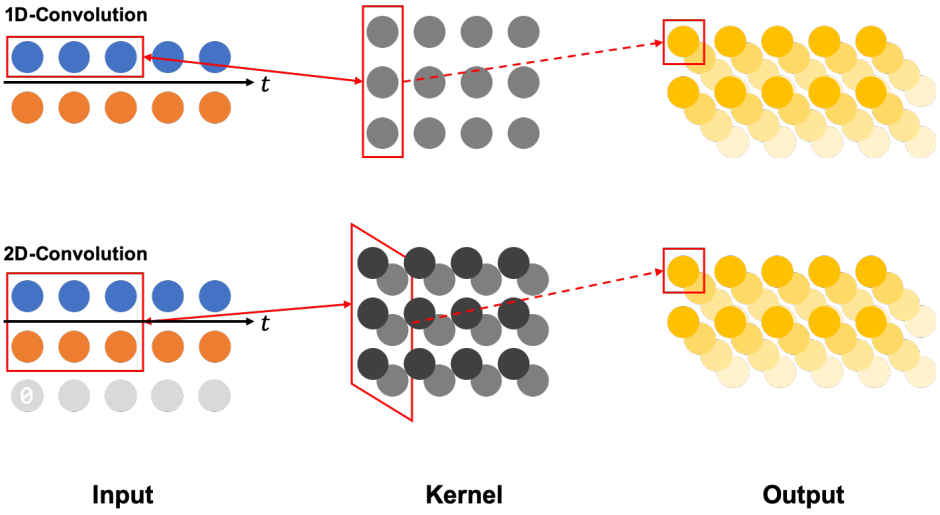

Figure 1. The examples of generating the output via 1D and 2D convolution. In 2D convolution case, the zero-padding method is needed for aggregating time information with maintaining feature dimension.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

In order to minimize the information loss for each channel, our purpose is only aggregating the time information via convolutional layer like recurrent neural network [31]. In the case of 1D convolution, time information can be reduced while maintaining channel information in natural. However, channel information of the input data will be aggregated to lower dimension than input in 2D convolution case. For making the output channel as same as channel of the input, the zero-padding can be used as shown in Figure 2 but the final channel may have less information than the front channels. Thus, we adopt the 1D convolutional layer for constructing the neural network.

3.2. *Multiple Hypothesis-based Architecture*

We have already summarized previous related works in Section 2. Using the variational bound with distribution assumption may generate blurry output. Also, we do not need to find the information that one to one matching between latent vector and generated data because we do not generate data from latent vector such as VAE case.

Theoretically, using multiple hypothesis can help produce more relevant results and can work more robustly [32–34]. The multiple hypothesis method also can be used for constructing neural network [24].

We construct the GANomaly-like neural network [21], because it can ease the several limitations of previous anomaly detection models such as blurred output and lower throughput. Also, we adopt the multiple hypothesis method for generating output consistently. The property of the multiple hypothesis can generate the output consistently because only the best hypothesis is selected as a output and the others are discarded. We call this procedure as hypothesis pruning.

We do not use the generated data from the random noise, but we only use them as a regularization term for avoiding overfitting and improving generalization performance. Also, the generated output from random noise, is regarded one of the hypotheses. The whole architecture of the neural network what we proposed in this paper is shown in Figure 2.

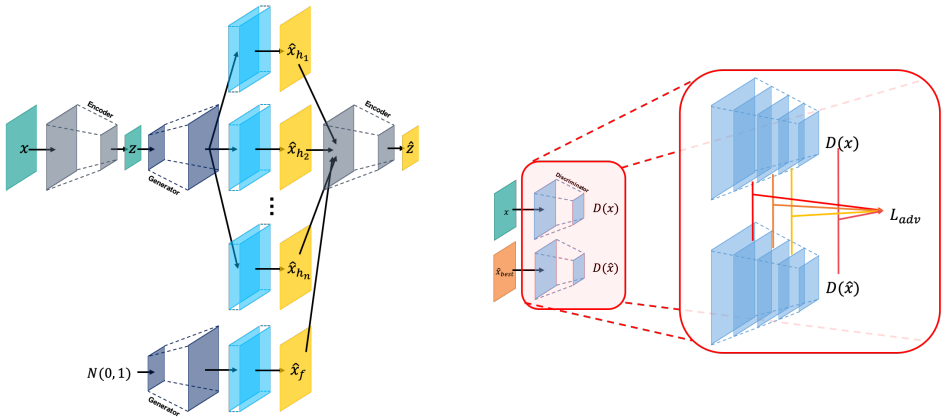

Figure 2. The architecture of proposed model Hypothesis Pruning GAN (HP-GAN). Pruning is conducted after generating the branches by multiple hypothesis. We use the single encoder, generator, and discriminator. We also use additional multiple hypothesis networks that colored with cyan.

The loss functions for training above neural network are shown as below equations. The symbol E, G, and D are representing encoder, generator, and discriminator respectively. Also, x, z, and f_l mean input data, encoded latent vector, and feature map of l-th layer. Three loss functions for encoder, generator, and discriminator are discripted as three loss function $\mathcal{L}_{enc}$, $\mathcal{L}_{gen}$, and $\mathcal{L}_{adv}$ respectively; shown in Equation 1, 2, and 7. Equation 1 and 2, encoding loss and generating loss, are based on Winner-Take-All (WTA) theory [35]. The Equation 7 is aggregated value from Equation 3 to 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

$$\mathcal{L}_{enc} = ||E(x) - E(G(E(x))_{best}||_2 = ||z - \hat{z}_{best}||_2 \tag{1}$$

$$\mathcal{L}_{gen} = |x - G(E(x))_{best}|_1 = |x - \hat{x}_{best}|_1 \tag{2}$$

$$\mathcal{L}_{adv_{noise}} = ||D(x) - D(\hat{x}_{noise})||_2 \tag{3}$$

$$\mathcal{L}_{adv_{best}} = ||D(x) - D(\hat{x}_{best})||_2 \tag{4}$$

$$\mathcal{L}_{adv_{others}} = ||D(x) - D(\hat{x}_{others})||_2 / number\ of\ others \tag{5}$$

$$\mathcal{L}_{adv_{feature}} = \sum_{l=0}^{L} ||f_l(x) - f_l(\hat{x}_{others})||_2 \tag{6}$$

$$\mathcal{L}_{adv} = \mathcal{L}_{adv_{noise}} + \mathcal{L}_{adv_{best}} + \mathcal{L}_{adv_{others}} + \mathcal{L}_{adv_{feature}} \tag{7}$$

For optimizing the parameters, we summarize above three loss function $\mathcal{L}_{enc}$, $\mathcal{L}_{gen}$, and $\mathcal{L}_{adv}$ with weighting parameters w_{enc}, w_{gen}, and w_{adv}. The weighting parameters work as hyperparameter, so it can be initialized with any value. In this paper, they are initialized with value 1, 50, and 1 respectively as same as GANomaly [21].

$$\mathcal{L} = w_{enc}\mathcal{L}_{enc} + w_{gen}\mathcal{L}_{gen} + w_{adv}\mathcal{L}_{adv} \tag{8}$$

Then, we train the above neural network via Algorithm 1. We use the Xavier initializer for initializing the neural network, and use Adam optimizer for optimizing the parameters.

Algorithm 1 Training algorithm.

Input: Measured PM values x, and random noise z_{noise}

Output: Restored PM values $\hat{X}$ with multiple hypothesis

Initialize parameters of neural network by Xavier initializer [36]
while the loss has not converged **do**
  Get $\hat{X}$ by forward propagation, and prune $\hat{X}$ as $\hat{x}_{best}$ or others
  Compute losses $\mathcal{L}_{enc}$, $\mathcal{L}_{gen}$, $\mathcal{L}_{adv}$, and total loss $\mathcal{L}$
  Update parameters by Adam optimizer [37]
end while

3.3. Anomaly Detection Method

We describe the anomaly detection procedure in this section. The abnormality decision method is simple as shown in Algorithm 2. For using above algorithm, input data must be measured via PM sensor firstly with containing two channel; $PM_{2.5}$ and PM_{10}. Then, restoration (generation) procedure is conducted via enter the input data to the neural network. The decision boundary θ is set by training data after training procedure as shown in Equation 9. The μ and σ represent the mean and standard deviation of the restoration error with best hypothesis. Of course, if the user want to change the sensitivity or specificity of the anomaly detection, θ can be adjusted.

$$\theta = \mu + (1.5 * \sigma) \tag{9}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

Algorithm 2 Abnormality decision algorithm.

Input: Measured PM values x

Output: Selected best restoration among the hypothesis $\hat{x}_{best}$

$\theta \leftarrow$ set threshold based on training data
if $||\hat{x} - \hat{x}_{best}||_2 \geqslant \theta$ **then**
  x is abnormal
else
  x is normal
end if

4. Experiments

In this section, we present the dataset for experiment that we collected and experimental results for various neural network architectures. For assessing each architecture, we use Area Under the Receiver Operating characteristics Curve (AUROC) [38], Area Under the Precision-Recall Curve (AUPRC) [39], and Mean Square Error (MSE) as the performance indicators.

4.1. Dataset

We have collected the PM dataset from 12 point location in Daegu Metropolitan Jungang Library, Korea. For collecting data, we use the TEOM-based sensor and the sensors are located at each collecting point. The collected dataset for experiment is shown in Table 2.

Table 2. The collected $PM_{2.5}$ and PM_{10} values via TEOM-based sensors. The dimension of each sample is 24 (hour) by 2 ($PM_{2.5}$ and PM_{10}).

	Normal	**Abnormal**	**Total**
Number of Sample	249	73	322

Each collected sample contains the information of one day with 1-hour unit. Before experiment, only the clear normal samples are classified by meteorologists. We use the portion of the normal set for a training procedure, the others of the normal set and the abnormal set are used for assesing the performance.

4.2. Experiment with Published Architecture

First of all, we conduct experiments for confirming which architecture among the previous studies is effective to anomaly detection. We adopt five models and reconstruct them using 1D convolutional layer [21–25] for experiment. Each network such as encoder or generator uses three convolutional block, and each convolutional block consists two convolutional layer with elu activation [40] and one max pooling layer at the last of the block. Also, two fully connected layers are used for the rear of the encoder (or discriminator) and the front of the generator. Each performance indicator is shown in Table 3 quantitatively. Also, the generated outputs are presented in Figure 3 for qualitively analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

Table 3. Measured performance of simple experiment with published architecture. The purpose of this experiment is confirming which style of the architecture can detect anomaly better.

Model	$PM_{2.5}$	PM_{10}	AUROC	AUPRC	MSE	Tr-Time	Te-Time
VAE [25]	O	O	0.92151	0.95629	0.01170	00:16:43	0.41122
	O	-	0.91699	0.95617	0.01424	**00:14:26**	0.30716
	-	O	0.92630	0.95643	0.01772	00:17:14	0.36180
GANomaly [21]	O	O	**0.93397**	**0.96485**	0.01080	00:20:10	0.34627
	O	-	0.92753	0.96209	0.00914	00:21:43	0.31621
	-	O	0.92301	0.95882	**0.01020**	00:24:21	**0.30096**
adVAE [22]	O	O	0.92849	0.95704	0.01716	00:23:42	0.40858
	O	-	0.92096	0.94832	0.02646	00:19:32	0.31841
	-	O	0.94233	0.95760	0.02264	00:19:28	0.32030
LVEAD [23]	O	O	0.88384	0.91724	0.01944	11:16:55	1.03436
	O	-	0.85370	0.87660	0.03764	10:37:55	1.03876
	-	O	0.91575	0.94738	0.01838	11:57:14	1.01579
ConAD [24]	O	O	0.92452	0.95638	0.02667	00:20:22	0.34142
	O	-	0.92342	0.95606	0.01524	00:30:05	0.31843
	-	O	0.91616	0.94154	0.01799	00:30:02	0.30392

Time consumption for training and test procedure of four architectures other than LVEAD are similar to each other but LVEAD needs much longer time. Thus, any architecture except for LVEAD can be adopted for anomaly detection in time consumption viewpoint. The GANomaly shows the highest AUROC and AUPRC. Also, it shows the lowest MSE among the above five architectures, so if who want to use the known neural network architecture without developing novel architecture GANomaly can be recommanded.

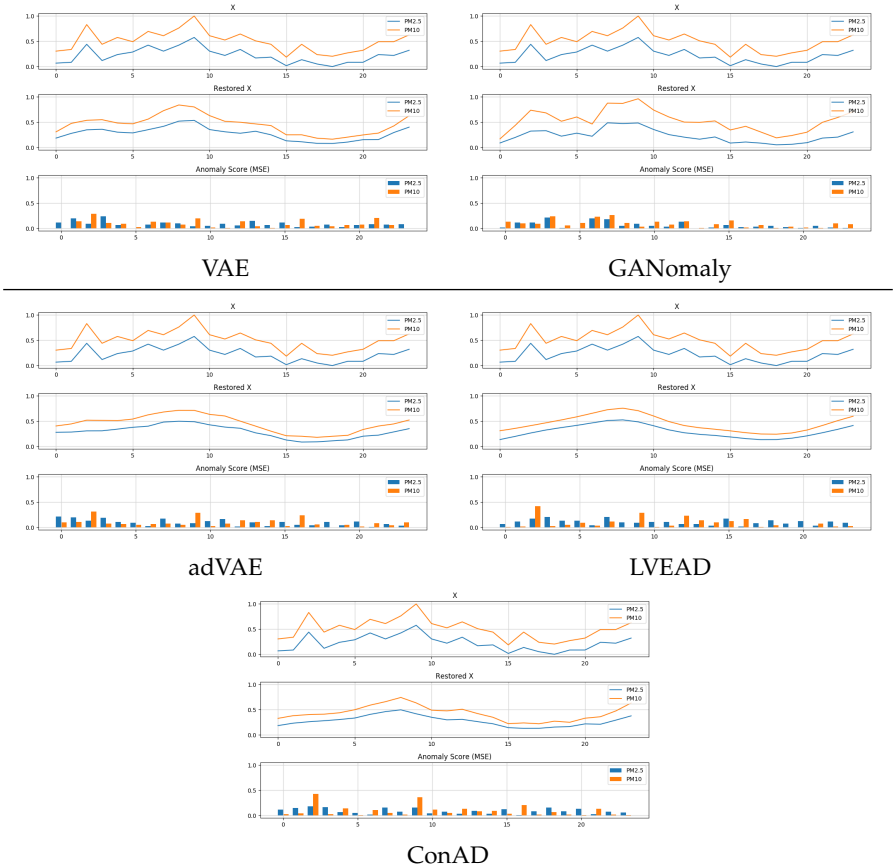

Figure 3. The generated output from the five architectures for qualitative analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

For qualitative results, GANomaly produces the output closest to the input, and ConAD and VAE follow after GANomaly. The common method of VAE and ConAD is variational bound that can be regarded as a reason for generating smoothed output. Thus, we conduct ablation study to find the cause of performance degradation in the next section.

4.3. Ablation Experiment

We construct the MP-GAN via referring previous studies. However, MP-GAN needs to confirm that can work better than other architecture via experiment, because it is novel and unprooved architecture. We compose the experiment to verify the ability of MP-GAN with our dataset. Also, several kinds of atchitecures are constructed for ablation experiment. Ablation experiment can help to find the cause of performance impediment [41,42].

The six ablation architecrtures are constructed for this experiment. Those are containing GANomaly and our MP-GAN. We use three conditions Latent vector Matching (LM), Variational Bound (VB), and Multiple Hypothesis (MH) for construct those neural networks. The purpose of LM is minimizing the Euclidean distance between latent vector z and $\hat{z}$; from x and $\hat{x}$ respectively. The VB pursues minimizing the Kullback–Leibler divergence between latent vector z and normal distribution. The last condition, MH, is used to generate the output consistently based on (WTA) theory.

The mini-batch size, training epoch, dimension of the latent vector z, and learning rate are used equal for all architecture and the value of those are 32, 1000, and 0.0001 respectively. Finally, the experimental result is shown in Table 4.

Table 4. The performance of the six ablation (combination) architectures. We use three condition Latent vector Matching (LM), Variational Bound (VB), and Multiple Hypothesis (MH) for ablation study.

Model	LM	VB	MH	$PM_{2.5}$	PM_{10}	AUROC	AUPRC	MSE
Ablation-1 (LM-GAN; GANomaly)	O	-	-	O	O	0.93397	0.96485	0.01080
				O	-	0.92753	0.96209	0.00914
				-	O	0.92301	0.95882	**0.01020**
Ablation-2 (VB-GAN)	-	O	-	O	O	0.90548	0.93130	0.02148
				O	-	0.94192	0.96536	0.01924
				-	O	0.86767	0.91166	0.04300
Ablation-3 (LMVB-GAN)	O	O	-	O	O	0.89521	0.93441	0.02510
				O	-	0.86096	0.90572	0.04022
				-	O	0.89205	0.93045	0.02974
Ablation-4 (LMMH-GAN; Ours)	O	-	O	O	O	0.91699	0.95303	0.02550
				O	-	0.93219	0.96312	0.01534
				-	O	0.92616	0.95424	0.01230
Ablation-5 (VBMH-GAN)	-	O	O	O	O	**0.94644**	**0.96462**	0.03578
				O	-	0.93027	0.96153	0.01486
				-	O	0.90068	0.94807	0.01404
Ablation-6 (LMVBMH-GAN)	O	O	O	O	O	0.92603	0.94032	0.03019
				O	-	0.93123	0.96179	0.01531
				-	O	0.92712	0.95730	0.01922

The Table 4 shows that VBMH-GAN has better AUROC, AUPRC than other ablated architectures. However, when analyzing the qulivatie reult, as shown in Figure 4, the generated data via LM-GAN (GANomaly) and LMMH-GAN (Ours; HP-GAN) show the better result than VBMH-GAN. Also, other VB based architectures commonly show the blurred result.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

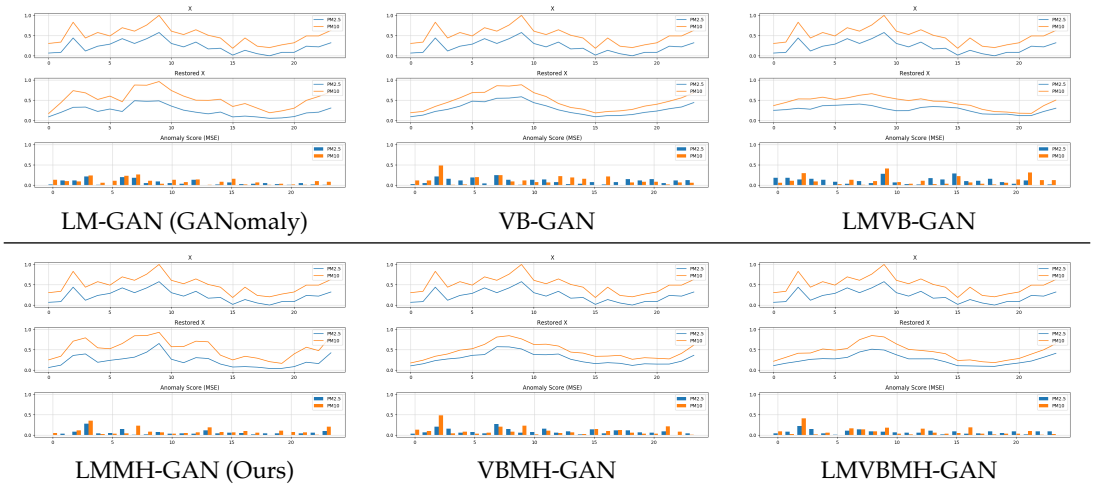

Figure 4. The generated output from the six ablation architectures for qualitative analysis.

In this experiment, we confirm that the quantitative result represents that the VBMH-GAN is the best architecture, but the qualitative result of them is not good. Thus, we need to conduct more experiment for verifying the above architectures using varied hyperparameters, because each model may need the specific hyperparameter for performing much better.

4.4. Experiment with Various Hyperparameter

In this section, we assess four of neural networks with various hyperparameter set. For experiment, we use the VAE as a baseline architecture that using VB. We also use GANomaly as one other baseline architecture that based on LM. The other two architectures are our HP-GAN (LMMH-GAN) and VBMH-GAN.

We use the kernel size, number of convolutional block, and learning rate as the hyperparemeter. Each convolutional block consists two convolutional layer with elu activation and one max pooling layer same as commented in Section 4.2. We compose the 108 kinds of the hyperparameter set for experiment via combining three hyperparameters as shown in Table 5.

Table 5. The hyperparameters for experiment. We combine these to 108 set.

Hyperparameter	Values
Kernel size	3, 5, 7, 9, 11, and 13
Number of convolutional block	2, 3, and 4
Learning rate	5e-3, 1e-3, 5e-4, 1e-4, 5e-5, and 1e-5

When closer to the last layer, the case of larger kernel size than the time axis dimension of the feature vector can be existed. However, it does not changing the amount of the feature information and balance between feature vectors, so we also use the large kernel size such as 13.

We present the measured performance as a surface form in Figure 5 and 6. The height of the surface shows the anomaly detection ability; higher surface means high performance. The flatten surface indicates that the neural network stably responds to hyperparameter changes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

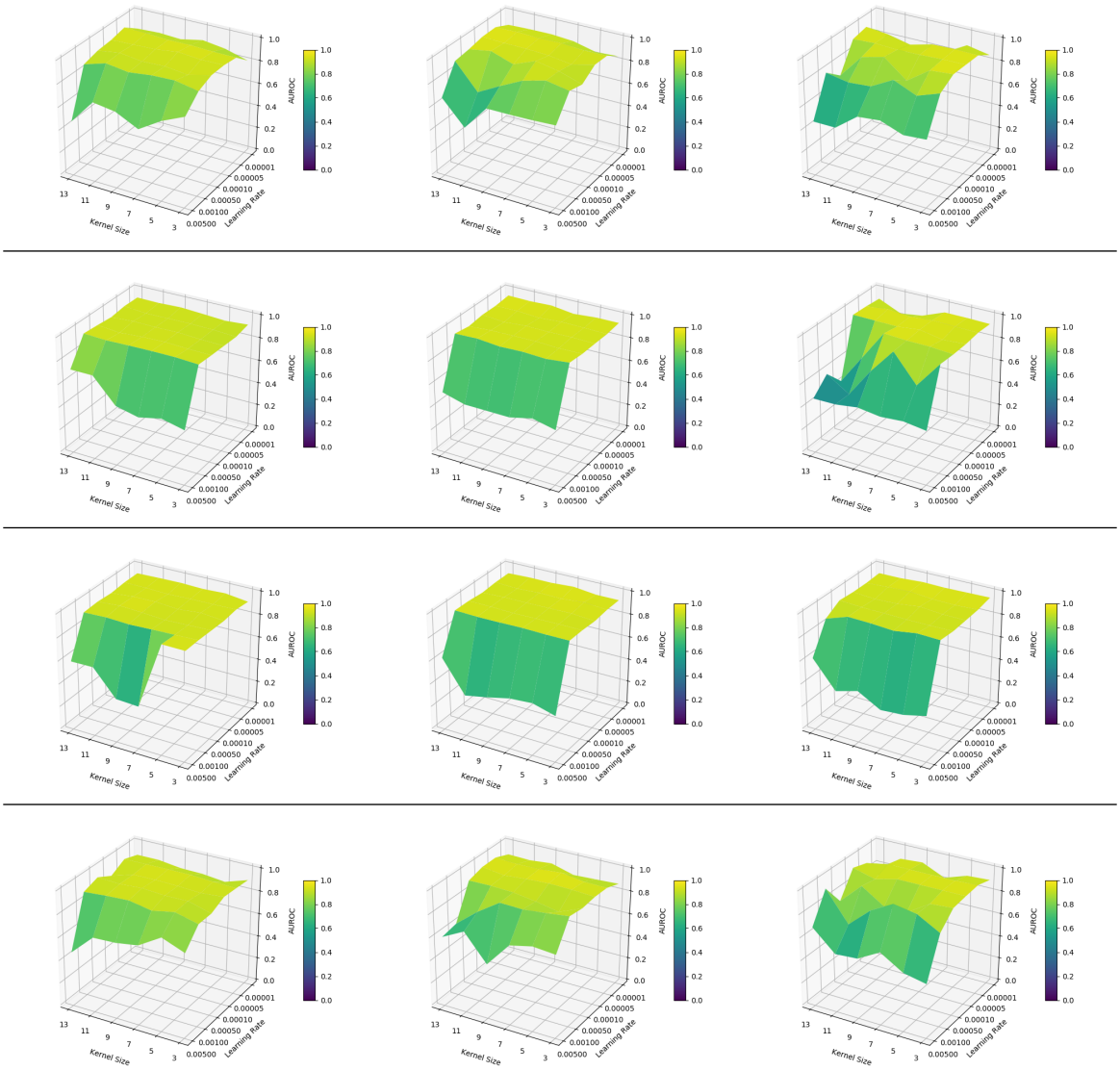

Figure 5. The surface of AUROC with various hyperparameter. The AUROC of VAE, GANomaly, and HP-GAN (LMMH-GAN), and VBMH-GAN are shown in each row sequentially. Each column shows the result of two, three, and four convolutional blocks from the left to right.

When the number of the convolutional block is two or three, the surfaces for each model look like similar and relatively flatten then four block cases. However, HP-GAN shows the most flatten surface in the four convolutional block case.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

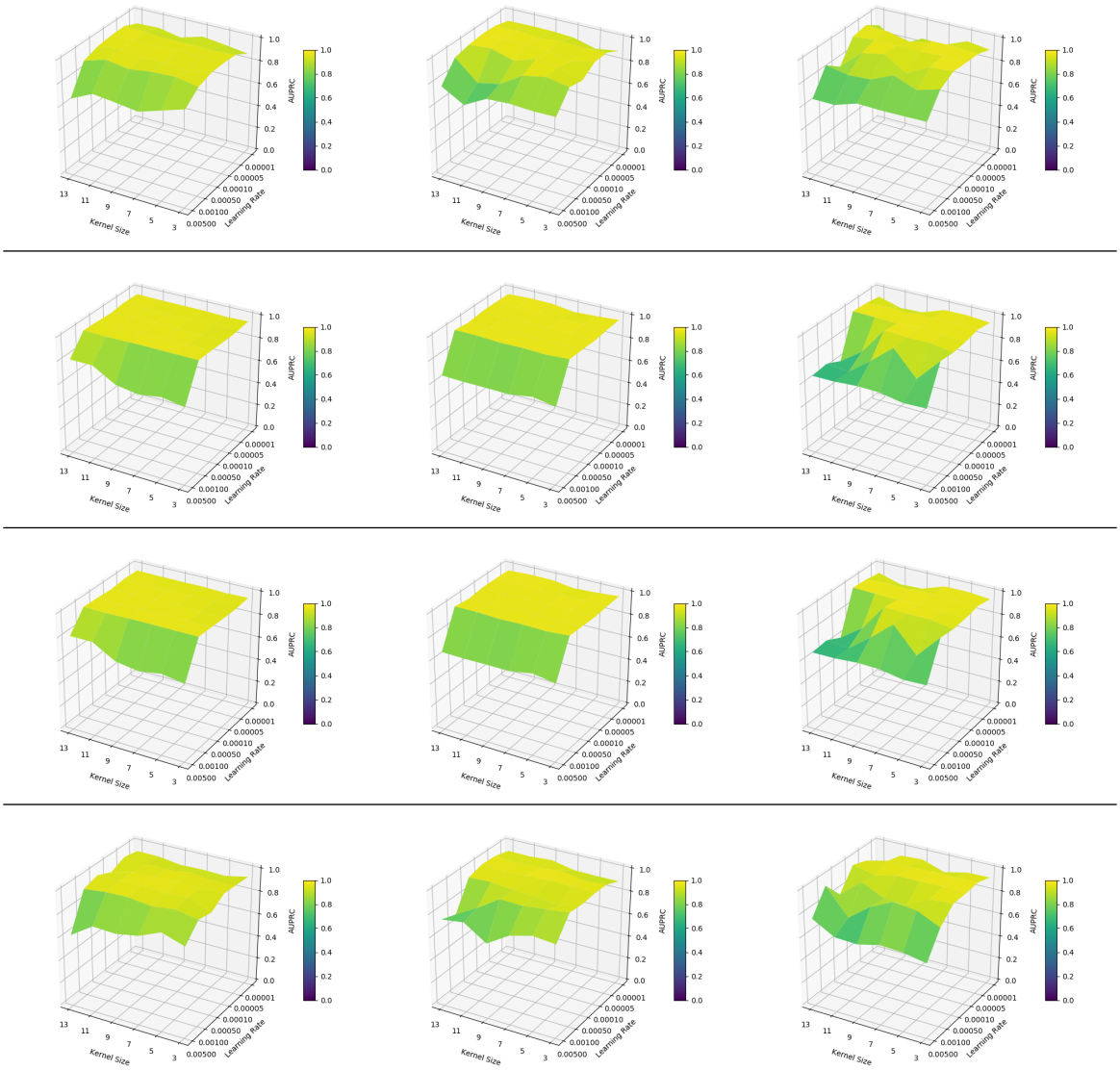

Figure 6. The surface of AUPRC with various hyperparameter. The order of the shown contents are same as Figure 5.

We also confirm that HP-GAN shows generally flatten surface in the AUPRC surfaces. We can conclude that HP-GAN may perform stably for anomaly detection than other architectures. Finally, we conduct the last experiment for confirming the performance with best hyperparameter for each model. The best hyperparameter set summarized in Table 6.

Table 6. The best hyperparameter set selected from AUROC and AUPRC surface.

	Kernel size	**Number of convolutional block**	**Learning rate**
VAE	7	3 (6 convolution)	5e-4
GANomaly	9	4 (8 convolution)	5e-4
HP-GAN (LMMH-GAN)	7	3 (6 convolution)	1e-5
VBMH-GAN	7	3 (6 convolution)	5e-5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 January 2020 doi:10.20944/preprints202001.0028.v1

We repeat the experiment for 30 times with randomly shuffled dataset with best hyperparameter set for monte calro estimation [43]. We present each performance indicator with mean $\pm$ standard deviation as shown in Table 7.

Table 7. The measured AUROC, AUPRC, and MSE are shown with mean $\pm$ standard deviation form. The experiment is conducted with best hyperparameter set for each architecture.

	AUROC	AUPRC	MSE
VAE	0.91753 $\pm$ 0.03284	0.94158 $\pm$ 0.02843	0.02544 $\pm$ 0.00864
GANomaly	0.91136 $\pm$ 0.09059	0.93682 $\pm$ 0.06223	0.02749 $\pm$ 0.04862
HP-GAN (LMMH-GAN)	**0.94781** $\pm$ **0.00976**	**0.96712** $\pm$ **0.00758**	**0.02301** $\pm$ **0.01600**
VBMH-GAN	0.94408 $\pm$ 0.02088	0.95909 $\pm$ 0.01957	0.03346 $\pm$ 0.05681

In Table 7, the higher the mean represents the better performance. On the other hand, the lower standard deviation means higher stability. The HP-GAN what we proposed in this paper shows the higher mean performance at every indicator. Also, HP-GAN has the higher stability because it has lowest standard deviation values at AUROC and AUPRC.

5. Conclusions

We experimentally show the cutting-edge performance at anomaly detection of our HP-GAN in TEOM-based PM sensor. The HP-GAN is trained by latent vector matching with multiple hypothesis based on WTA theory. Our neural network, HP-GAN, can generate the output more clearly with less blurring effect than other variational bound-based model when the input data is in normal category. The mean of AUROC and AUPRC of HP-GAN are 0.373% 0.0803% higer than second performance model VBMH-GAN. Also, mean of MSE is best (lowest) among the whole architectures. Thus, we finally conclude our HP-GAN, constructed based on latent vector matching and multiple hypothesis, as a cutting-edge architecture for anomaly detection.

Author Contributions: Conceptualization, Y.Park. and S.W.Chang.; methodology, software, validation, and formal analysis, Y.Park.; investigation, Y.Park. W.S.Park, Y.B.Kim and S.W.Chang.; resources, S.W.Chang.; data curation, W.S.Park and Y.B.Kim.; writing–original draft preparation, Y.Park.; writing–review and editing, Y.Park and S.W.Chang.; visualization, Y.Park.; supervision, S.W.Chang.

Funding: This research received no external funding.

Acknowledgments: Thank you to all the team and corporation members. They have supported this research via not only data collection but also equipment for the experiment. Furthermore, we also thank to all the director including Dong Chan Lim for supporting this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Miller, F.J.; Gardner, D.E.; Graham, J.A.; Jr., R.E.L.; Wilson, W.E.; Bachmann, J.D. Size Considerations for Establishing a Standard for Inhalable Particles. *Journal of the Air Pollution Control Association* **1979**, *29*, 610–615.
2. Faustini, A.; Stafoggia, M.; Berti, G.; Bisanti, L.; Chiusolo, M.; Cernigliaro, A.; Mallone, S.; Primerano, R.; Scarnato, C.; Simonato, L.; Vigotti, M.; Forastiere, F. The relationship between ambient particulate matter and respiratory mortality: a multi-city study in Italy. *European Respiratory Journal* **2011**, *38*, 538–547.
3. Du, Y.; Xu, X.; Chu, M.; Guo, Y.; Wang, J. Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. *Journal of thoracic disease* **2016**, *8*, E8—E19.
4. Hamra, G.B.; Guha, N.; Cohen, A.; Laden, F.; Raaschou-Nielsen, O.; Samet, J.M.; Vineis, P.; Forastiere, F.; Saldiva, P.; Yorifuji, T.; Loomis, D. Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives* **2014**, *122*, 906–911.

5. Occupational, W.H.O.; Team, E.H. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide : global update 2005 : summary of risk assessment, 2006.
6. Badura, M.; Batog, P.; Drzeniecka-Osiadacz, A.; Modzel, P. Evaluation of Low-Cost Sensors for Ambient PM2. 5 Monitoring. *Journal of Sensors* **2018**, *2018*.
7. Bulot, F.M.; Johnston, S.J.; Basford, P.J.; Easton, N.H.; Apetroaie-Cristea, M.; Foster, G.L.; Morris, A.K.; Cox, S.J.; Loxham, M. Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Scientific reports* **2019**, *9*, 7497.
8. Himmelblau, D.; Barker, R.; Suewatanakul, W. Fault Classification with The Aid of Artificial Neural Networks. In *Fault Detection, Supervision and Safety for Technical Processes 1991*; ISERMANN, A.; FREYERMUTH, B., Eds.; IFAC Symposia Series, Pergamon: Oxford, 1992; pp. 541 – 545.
9. Tellenbach, B.; Burkhart, M.; Schatzmann, D.; Gugelmann, D.; Sornette, D. Accurate network anomaly classification with generalized entropy metrics. *Computer Networks* **2011**, *55*, 3485 – 3502.
10. de la Hoz, E.; Ortiz, A.; Ortega, J.; de la Hoz, E. Network Anomaly Classification by Support Vector Classifiers Ensemble and Non-linear Projection Techniques. Hybrid Artificial Intelligent Systems; Pan, J.S.; Polycarpou, M.M.; Woźniak, M.; de Carvalho, A.C.P.L.F.; Quintián, H.; Corchado, E., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp. 103–111.
11. Jan, S.U.; Lee, Y.; Shin, J.; Koo, I. Sensor Fault Classification Based on Support Vector Machine and Statistical Time-Domain Features. *IEEE Access* **2017**, *5*, 8682–8690.
12. Lee, K.B.; Cheon, S.; Kim, C.O. A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes. *IEEE Transactions on Semiconductor Manufacturing* **2017**, *30*, 135–142.
13. Steinwart, I.; Hush, D.; Scovel, C. A classification framework for anomaly detection. *Journal of Machine Learning Research* **2005**, *6*, 211–232.
14. Sakurada, M.; Yairi, T. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. Proceedings of the MLSDA 2014 2Nd Workshop on Machine Learning for Sensory Data Analysis; ACM: New York, NY, USA, 2014; MLSDA'14, pp. 4:4–4:11.
15. Zhou, C.; Paffenroth, R.C. Anomaly Detection with Robust Deep Autoencoders. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: New York, NY, USA, 2017; KDD '17, pp. 665–674.
16. Malhotra, P.; Vig, L.; Shroff, G.; Agarwal, P. Long short term memory networks for anomaly detection in time series. Proceedings. Presses universitaires de Louvain, 2015, p. 89.
17. Chauhan, S.; Vig, L. Anomaly detection in ECG time signals via deep long short-term memory networks. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2015, pp. 1–7. doi:10.1109/DSAA.2015.7344872.
18. Xu, H.; Chen, W.; Zhao, N.; Li, Z.; Bu, J.; Li, Z.; Liu, Y.; Zhao, Y.; Pei, D.; Feng, Y.; Chen, J.; Wang, Z.; Qiao, H. Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. Proceedings of the 2018 World Wide Web Conference; International World Wide Web Conferences Steering Committee: Republic and Canton of Geneva, Switzerland, 2018; WWW '18, pp. 187–196.
19. Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. Information Processing in Medical Imaging; Niethammer, M.; Styner, M.; Aylward, S.; Zhu, H.; Oguz, I.; Yap, P.T.; Shen, D., Eds.; Springer International Publishing: Cham, 2017; pp. 146–157.
20. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial Feature Learning. *CoRR* **2016**, *abs/1605.09782*, [1605.09782].
21. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. GANomaly: Semi-supervised Anomaly Detection via Adversarial Training. Computer Vision – ACCV 2018; Jawahar, C.V.; Li, H.; Mori, G.; Schindler, K., Eds.; Springer International Publishing: Cham, 2019.
22. Wang, X.; Du, Y.; Lin, S.; Cui, P.; Yang, Y. Self-adversarial Variational Autoencoder with Gaussian Anomaly Prior Distribution for Anomaly Detection. *CoRR* **2019**, *abs/1903.00904*, [1903.00904].
23. Zhang, C.; Chen, Y. Time Series Anomaly Detection with Variational Autoencoders. *CoRR* **2019**, *abs/1907.01702*, [1907.01702].
24. Nguyen, D.T.; Lou, Z.; Klar, M.; Brox, T. Anomaly Detection With Multiple-Hypotheses Predictions. Proceedings of the 36th International Conference on Machine Learning; Chaudhuri, K.; Salakhutdinov,

R., Eds.; PMLR: Long Beach, California, USA, 2019; Vol. 97, *Proceedings of Machine Learning Research*, pp. 4800–4809.

25. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014.
26. Serpico, S.B.; D'Inca, M.; Melgani, F.; Moser, G. Comparison of feature reduction techniques for classification of hyperspectral remote-sensing data. Image and Signal Processing for Remote Sensing VIII; Serpico, S.B., Ed. International Society for Optics and Photonics, SPIE, 2003, Vol. 4885, pp. 347 – 358.
27. Jun Yan.; Benyu Zhang.; Ning Liu.; Shuicheng Yan.; Qiansheng Cheng.; Fan, W.; Qiang Yang.; Xi, W.; Zheng Chen. Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing. *IEEE Transactions on Knowledge and Data Engineering* **2006**, *18*, 320–333.
28. Thangavel, K.; Pethalakshmi, A. Dimensionality reduction based on rough set theory: A review. *Applied Soft Computing* **2009**, *9*, 1 – 12.
29. Balzanella, A.; Irpino, A.; Verde, R. Dimensionality Reduction Techniques for Streaming Time Series: A New Symbolic Approach. Classification as a Tool for Research; Locarek-Junge, H.; Weihs, C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2010; pp. 381–389.
30. Park, Y.; Yun, I.D. Fast Adaptive RNN Encoder–Decoder for Anomaly Detection in SMD Assembly Machine. *Sensors* **2018**, *18*.
31. Mikolov, T.; Karafiát, M.; Burget, L.; Černockỳ, J.; Khudanpur, S. Recurrent neural network based language model. Eleventh annual conference of the international speech communication association, 2010.
32. Blackman, S.S. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine* **2004**, *19*, 5–18.
33. Streller, D.; Dietmayer, K. Object tracking and classification using a multiple hypothesis approach. IEEE Intelligent Vehicles Symposium, 2004, 2004, pp. 808–812.
34. Rupprecht, C.; Laina, I.; DiPietro, R.; Baust, M.; Tombari, F.; Navab, N.; Hager, G.D. Learning in an Uncertain World: Representing Ambiguity Through Multiple Hypotheses. The IEEE International Conference on Computer Vision (ICCV), 2017.
35. Maass, W. On the Computational Power of Winner-Take-All. *Neural Computation* **2000**, *12*, 2519–2535.
36. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.
37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* **2014**.
38. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **2006**, *27*, 861 – 874.
39. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE* **2015**, *10*, 1–21.
40. Clevert, D.A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289* **2015**.
41. Anwar, S.; Hwang, K.; Sung, W. Structured Pruning of Deep Convolutional Neural Networks. *CoRR* **2015**, *abs/1512.08571*, [1512.08571].
42. Meyes, R.; Lu, M.; de Puiseau, C.W.; Meisen, T. Ablation Studies in Artificial Neural Networks. *CoRR* **2019**, *abs/1901.08644*, [1901.08644].
43. Kroese, D.P.; Brereton, T.; Taimre, T.; Botev, Z.I. Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics* **2014**, *6*, 386–392.