
A MULTI-TASK LEARNING FOR CAVITATION DETECTION AND CAVITATION INTENSITY RECOGNITION OF VALVE ACOUSTIC SIGNALS*

Yu Sha Xidian University FIAS [†] XF-IJRC [‡]	Johannes Faber FIAS	Shuiping Gou Xidian University	Bo Liu Xidian University	Wei Li FIAS	Stefan Schramm FIAS
---	-------------------------------	--	------------------------------------	-----------------------	-------------------------------

Horst Stoecker FIAS Goethe Universität GSI [§]	Thomas Steckenreiter SAMSOM AG	Domagoj Vnucec SAMSOM AG	Nadine Wetzstein SAMSOM AG	Andreas Widl SAMSOM AG
---	--	------------------------------------	--------------------------------------	----------------------------------

Kai Zhou[¶]
FIAS

ABSTRACT

With the rapid development of smart manufacturing, data-driven machinery health management has received a growing attention. As one of the most popular methods in machinery health management, deep learning (DL) has achieved remarkable successes. However, due to the issues of limited samples and poor separability of different cavitation states of acoustic signals, which greatly hinder the eventual performance of DL modes for cavitation intensity recognition and cavitation detection. Also different tasks were performed separately conventionally. In this work, a novel multi-task learning framework for simultaneous cavitation detection and cavitation intensity recognition framework using 1-D double hierarchical residual networks (1-D DHRN) is proposed for analyzing valves acoustic signals. Firstly, a data augmentation method based on sliding window with fast Fourier transform (Swin-FFT) is developed to alleviate the small-sample issue confronted in this study. Secondly, a 1-D double hierarchical residual block (1-D DHRB) is constructed to capture sensitive features from the frequency domain acoustic signals of valve. Then, a new structure of 1-D DHRN is proposed. Finally, the devised 1-D DHRN is evaluated on two datasets of valve acoustic signals without noise (*Dataset 1* and *Dataset 2*) and one dataset of valve acoustic signals with realistic surrounding noise (*Dataset 3*) provided by SAMSON AG (Frankfurt). Our method has achieved state-of-the-art results. The prediction accuracies of 1-D DHRN for cavitation intensity recognition are as high as **93.75%**, **94.31%** and **100%**, which indicates that 1-D DHRN outperforms other DL models and conventional methods. At the same time, the testing accuracies of 1-D DHRN for cavitation detection are as high as **97.02%**, **97.64%** and **100%**. In addition, 1-D DHRN has also been tested for different frequencies of samples and shows excellent results for frequency of samples that mobile phones can accommodate.

* Accepted by the *Engineering Applications of Artificial Intelligence* on 19 April 2022.

[†] FIAS: Frankfurt Institute for Advanced Studies

[‡] XF-IJRC: Xidian-FIAS international Joint Research Center

[§] GSI: GSI Helmholtzzentrum für Schwerionenforschung GmbH

[¶] Kai Zhou is the corresponding author. Email: zhou@fias.uni-frankfurt.de

Keywords Valves acoustics signal; Cavitation intensity recognition; Cavitation detection; Multi-task learning; 1-D convolutional neural network and 1-D double hierarchical residual block

1 Introduction

Cavitation is a dynamical phenomenon occurs in fluid dynamics, during which bubbles, also called cavities, form (and later collapse) inside the liquid when the local pressure (e.g. at the contact point of a liquid and some solid surface) is lower than the liquid's vapor pressure [1, 2]. When the bubble flows to places where the liquid pressure exceeds the vapor pressure, the bubble collapses and the implosion instantaneously produces a great impact (shock waves) and high temperature [3, 4]. Furthermore, severe pitting and wear can be induced on the solid surface [5].

Generally, cavitation can potentially bring dangers to process plants, especially for valves, pumps, pipes or propellers. These potential dangers might quickly cause damage to components of the plants and loss of efficiency [6]. On one hand, corrosion and destruction to valves, pumps or pipes can be induced by the cavitation occurrence [7], vibration of the internal structure and noise loudness levels might also increase [8] due to the bubble collapse and rupture [9]. On the other hand, the fluid bulk properties and flow rate inside the plants can be altered by the cavitation [10, 11] which further can lead to standstill of the process. Consequently, plants such as valves, pipes, pumps and related components are at the largest risk confronting cavitation. Many industries are fighting against cavitation, such as SAMSON AG for valves and other manufacturing plants [12]. In the worst case, cavitation can lead to the closure of the factory, e.g. for situations with test rack system including control valve. Therefore, it is vital to detect the cavitation of process plants (e.g. valves, pumps or pipes) at its early stages for intervention so as to ensure security and reduce any possible economic loss. Conventionally cavitation is detected by comparing the fault and healthy conditions of specific devices under the monitored signals. According to the type of monitoring sensors, it can be divided into vibration signals based cavitation detection and acoustic signals based detection. Recently, some researchers have explored machine learning application for cavitation detection, based on features handcrafted from the vibration or acoustic signals [13].

Sakthivel et al. [14] extracted 11 statistical features from the vibration time domain signals. These features are then fed into the C4.5 decision tree [15] to classify the bearing fault, seal fault, impeller fault, bearing and impeller fault together with cavitation. Muralidharan et al. [16] used the Continuous Wavelet Transform (CWT) [17] to replace the statistical feature extraction from vibration signals. Then the CWT achieved features are taken as input of the decision tree algorithm for similar classification task as in [14]. In addition, Muralidharan et al. [18] also studied the influence of different families and levels of the CWT on fault diagnosis of single-piece centrifugal oils using Support Vector Machine (SVM). Samanta et al. [19] extracted features from the original and pre-processed signals as the input of two different classifiers, the SVM and the artificial neural network (ANN), to identify normal and defective bearings. The parameters of the SVM and ANN are optimized by genetic algorithms [20], and the results explained the importance of feature selection to the classifier. Yang et al. [21] extracted 4 statistical features from the vibration time domain signals as the input of the SVM to detect cavitation and no cavitation of the butterfly valve. Bordoloi et al. [22] proposed a SVM method using directly the vibration signal data of bearing block and pump casing to diagnose blockage level and cavitation intensity. Panda et al. [23] extracted statistical features from the vibration time domain signals of the pump as the input of SVM to distinguish cavitation and flow block. Rapur et al. [24] proposed an intelligent detection method based on SVM to identify mechanical fault and flow rate using combination of motor line current and pump vibration signal as input. Shervani-abar. [25] proposed a multi-class cavitation detection method based on the vibration signal of the axial flow pump using SVM.

Zouari et al. [26] proposed a vibration signal fault detection method for centrifugal pumps using neural network and neuro-fuzzy technology. Rajakarunakaran et al. [27] proposed a centrifugal pump fault detection using a feedforward neural network and a binary adaptive resonance network (ART1). Siano et al. [28] proposed a method combining ANN and nonlinear regression to diagnose cavitation of time domain vibration signals. Nasiri et al. [29] extracted features from the vibration signal of the centrifugal pump as the input of the neural network to identify cavitation. Jia et al. [30] proposed a deep neural network (DNN) to directly extract features from the original rolling element bearings and planetary gearboxes data set for fault diagnosis. Zhao et al. [31] proposed a deep learning method to extract features from non-stationary vibration signals and diagnose centrifugal pump faults. Tiwari et al. [32] extracted 6 statistical features from the time domain pressure data, and then fed these features as input into the neural network to classify blockage and cavitation. Potocnik et al. [33] extracted spectral and psychoacoustic features from the valve acoustic data and then took these features to be input of a variety of machine learning algorithms to classify the cavitation, flow noise, whistling and rattling. Muralidharan et al. [34] presented the use of Naïve Bayes algorithm and Bayes net algorithm for fault diagnosis through discrete wavelet features extracted from vibration signals of good and faulty conditions of the components of centrifugal pump. Liu et al. [35] proposed a fault diagnosis method for gear pumps based on the ensemble empirical mode decomposition (EEMD) and Bayesian networks. It decomposes the vibration signal by the EEMD method and calculates the energy of the intrinsic mode functions (IMFs) as fault features and adds these

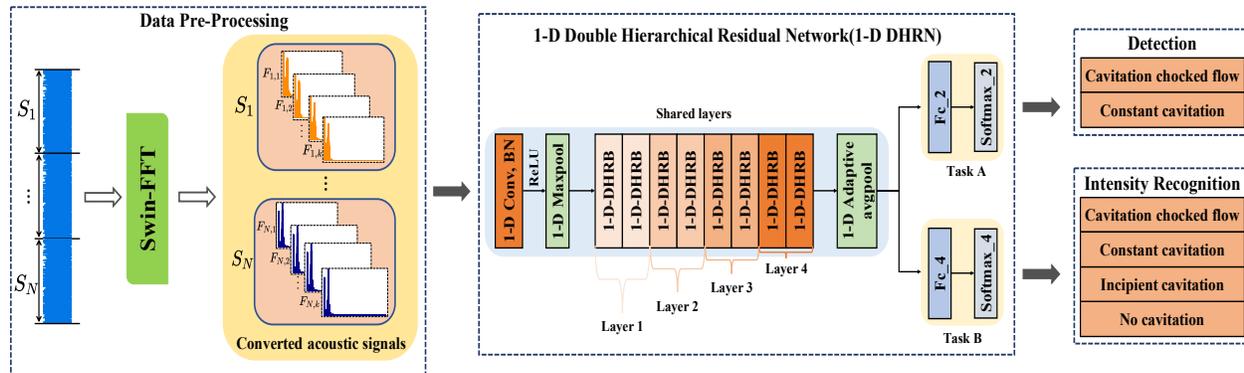


Figure 1: The framework of multi-task learning for cavitation detection and cavitation intensity recognition of valves acoustic signals using 1-D double hierarchical residual network. Task A is cavitation detection ("cavitation" and "non cavitation"). Task B is cavitation intensity recognition ("cavitation choked flow", "constant cavitation", "incipient cavitation" and "non cavitation").

features to the Bayesian network. Tang et al. [36] studied intelligent fault diagnosis of hydraulic piston pumps based on deep learning and Bayesian optimization, where the Bayesian Optimization (BO) algorithm helps the neural network to automatically select hyperparameters.

It can be found from the literature review that many traditional machine learning methods (like, decision tree, SVM or ANN with shallow architectures) have been explored for cavitation or fault detection. However, the above methods suffer from two main weaknesses. First, the shallow network architecture can not capture sensitive features, especially when the acoustic signals are contaminated by background noise. Second, manual feature extraction and selection are time-consuming and require domain specific knowledge [37] which also might introduce bias. Fortunately, state-of-the-art deep learning methods have strong ability in automatically extracting relevant features from original data. This advantage enable the application of deep learning to be transferred and inspired from speech recognition and computer vision to cavitation detection and diagnosis.

Deep convolutional neural networks (DCNN) have been successfully applied in image classification and also natural language processing (NLP) [38]. DCNN integrate the extraction of low, medium and high level features and classifiers in an end-to-end hierarchical fashion. The "level" of features can be achieved by the number of layers (depth) in the stack. The depth of convolutional neural networks (CNN) has gradually increased with architecture evolved from AlexNet [39], VGG [40], Inception [41] to Residual networks (ResNet) [42]. Recently, CNNs with 1-D convolutional layers have been widely applied on time series data, such as, electrocardiogram detection [43], electroencephalogram diagnosis [44], bearing fault diagnosis [45, 46, 47] and so on. The used 1-D convolution can acquire sensitive features from local segments of the entire time series data which are uncorrelated with each other [45, 46, 47, 48, 43, 44]. Therefore, time series data analysis can take 1-D CNN for efficient feature representation.

In the literature, 1-D convolution has been successfully applied to do fault diagnosis. However, it's not yet clear how it works in cavitation detection and cavitation intensity recognition of valves, and under a combined task-demanding, which in practice is more required. Also, the modern residual structure was not discussed in before within this area. In this study, targeted at a simultaneous multi-task learning, we introduce a special 1-D CNN based model that exploits 1-D double hierarchical residual blocks (1-D DHRB) as feature extractors. This novel method is termed as 1-D double hierarchical residual network (1-D DHRN). The architecture of our multi-task learning for cavitation detection and cavitation intensity recognition using 1-D DHRN is shown in Figure 1. Our 1-D DHRN is inspired by the ideas that, respective filed [49] and residual structure [42] with large enough perspective areas enable acquisition on more detailed information, and, the skip connection can increase the network non-linearity capability. Compared to our previous research [50], 1-D DHRN is a multi-task and end-to-end approach for cavitation detection and cavitation intensity recognition based upon the modern deep learning. In contrast to manual feature extraction that has to be performed in our previous research, the proposed 1-D DHRN here is able to capture sensitive features by itself and does not require specific knowledge, it thus allows the simultaneous cavitation detection and cavitation intensity recognition. Moreover, our proposed approach is more compatible with the requirements of the real-world industry. The main contributions of this work are summarized in follows:

- In order to tackle the small-sample problem, the sliding window (Swin-FFT) data augmentation method is introduced in the study.

- The 1-D Double Hierarchical Residual Blocks (1-D DHRB) with large kernel are proposed as an automatic feature extractor to capture sensitive features of valve acoustic signals.
- The 1-D Double Hierarchical Residual Network (1-D DHRN) is developed for valve cavitation detection and cavitation intensity recognition using directly the emitted acoustic signal.
- Multi-task learning for simultaneous cavitation detection and cavitation intensity recognition is presented.
- The proposed method is tested on three different datasets of acoustic signals measured by SAMSON AG and compared with the state-of-the-art methods.
- The impact of different sampling rate of the acoustic signals on our proposed method for cavitation detection and cavitation intensity recognition is investigated.

The remainder of this paper is organized as follows. Section 2 introduces the acquisition of datasets for valve cavitation acoustic signals. Section 3 describes the details of the proposed 1-D DHRN model. Section 4 introduces different case studies for each specific task. Section 5 discusses the effect of different experimental conditions on 1-D DHRN. The conclusion and future research outlook are given in Section 6.

2 Experimental Setup

SAMSON AG devised a test rack with control valve (SAMSON AG type 3241, DN 80, PN40, Kvs 25 with positioner type 3730-6) and running water as the flowing medium inside to generate different flow status by gauging operation conditions accordingly: upstream pressure, downstream pressure and the valve stroke. The test bench is equipped with a set of sensors to measure the temperature of the test medium T , the pressures upstream p_1 and downstream p_2 of the test valve and the volumetric flow rate Q . Additionally, for the test valve mounted inside the bench, the absolute valve stroke h and the sound intensity L_p with a special sensor directly mounted on the valve body are measured. Furthermore, the test bench includes two additional control valves upstream and downstream of the test valve. A control system controls the pumps as well as these two valves to modify the total volumetric flow through the test bench. The two additional valves are used to vary the upstream and downstream pressure around the test valve. All sensors and the test valve are mounted between these two external control valves (more details see Table 1). The pipes between the external valves and the test valve are long enough to ensure an undisturbed flow for proper measurements. Water, which can also be heated, was used as the test medium for all tests. The water temperature was maintained at 25 – 40°C in order to hold the vapor pressure nearly constant and to eliminate the influence of temperature on the cavitation.

Two piezo elements were introduced to record directly the structure-borne noise of the valve: one placed on the body of the control valve and the other to the NAMUR at the bonnet. Furthermore, two microphones were placed in a distance of 1m to the control valve to record the airborne noise in a frequency range between 40 Hz to 20 kHz: one high-end microphone as well as a low cost microphone which typically is used for smartphones. The test section in figure 2 of is in a separate room and all other noise sources from the plant like the pump and the upstream and downstream throttling valves are in another room (below the best section room, in the cellar). Therefore, there is a spatial separation of the test section to all possible noise sources. Also bellow expansion joints are integrated in the pipe system in order to the transport of vibrations into the test section. All these measures ensure that a surrounding noise of maximum 55 dB were present during the tests. The preparation of the different states is very simple. The principle procedure is described in the standard IEC 60534-8-2. The considered states depend on the differential pressure ratio $((p_1 - p_2)/(p_1 - p_v))$ of the valve. Therefore, the valve opening will be held constant during the test. The temperature of the medium is during the test constant, which means the vapor pressure will be held constant. Furthermore, the upstream pressure of the valve will also be held constant (e.g. 10 bar (a)) with the process control system of the plant. Therefore, in the consequence the different states will be adjusted by varying the downstream pressure in steps of 0.1, . . . , 0.4 bar starting at 1bar (a). For each step the downstream pressure will be adjusted with the process control system of the plant and based on the steady state of the flow the noise measurement concerning the adjusted step will be started. As a result, there is for a constant valve opening and a constant valve upstream pressure a noise emission characteristic (noise emission versus differential pressure ratio). And based on the noise emission characteristic the different states will be determined. This extensive procedure for determination of cavitation states cannot be realized in customers plants.

The cavitation we are identifying here takes place at control valves and the fluid inside is water. Cavitation takes place local at the narrowest throttling area, where the pressure falls to its minimum in the system, below the vapor pressure. Then, vapor bubbles are formed up in the liquid. These vapor bubbles implode when subjected to higher pressures with intensive sound pressure level. Depending on cavitation intensity erosion effects in the valve itself can occur additionally. Within SAMSON AG experimental setup in the laboratory, differential pressure at various constant upstream pressures of the control valve is varied, also different operation conditions (e.g. valve opening ratio, temperature) were adjusted, by which, five flow status are induced with corresponding acoustic signal measured for

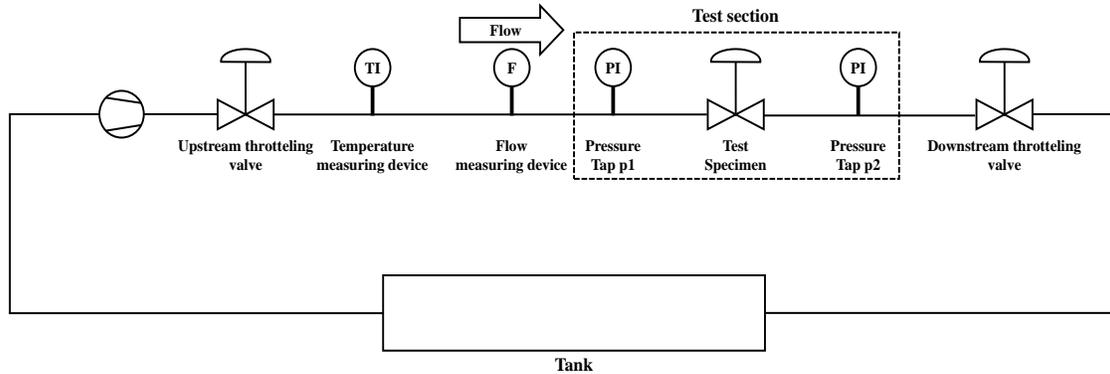


Figure 2: Schematic view of the test rack at SAMSON AG (Figure provided by SAMSON AG).

Table 1: The detailed information of various measurement sensors

Physical parameters	Sensor types	Measuring ranges	Tolerances
Upstream pressure	SAMSON Type 6054	0-16 bar	± 0.09 bar
Downstream pressure	SAMSON Type 6054	0-16 bar	± 0.09 bar
Flow rate	Krohne Type M950	0-180 m ³ /h	± 0.9 m ³ /h
Valve stroke	Sylvac Type s229	0-50 mm	± 0.001 mm
Medium temperature	SAMSON Type 5204	-20-150 °C	± 0.45 °C
Structure-borne noise (at valve body)	Vallen Type VS45-H	20-450 kHz	Sensitivity -63 dB re 1V/ μ bar (Accuracy according to individual calibration)
Structure-borne noise (at valve body)	PCB Type M353 B17	0-30000 Hz	Sensitivity 10 mV/g, Accuracy ± 3 dB
Structure-borne noise (at NAMUR at bonnet)	PCB Type M353 B17	0-30000 Hz	Sensitivity 10 mV/g, Accuracy ± 3 dB

many events: cavitation choked flow, constant cavitation, incipient cavitation, turbulent flow and background no-flow. The turbulent flow is a flow through the control valve without any cavitation noise. Starting at a certain differential pressure ratio within a certain range only a few vapor bubbles will be generated and the implosion of these bubbles is causing an increase of the noise emission, which is defined as incipient cavitation. By increasing the concentration of vapor bubbles, the noise emission is increasing up to a maximum noise which is defined as constant cavitation. Achieving the noise maximum also the concentration of the vapor bubbles is maximum which leads to a choked flow condition with cavitation and a decreasing noise behavior.

Figure 3 shows an exemplary comparison of time waveforms of the acoustic signals for cavitation choked flow, constant cavitation incipient and non cavitation. It is difficult to distinguish the sensitive features directly in Figure 3. Therefore, in this study an intelligent 1-D DHRN method is used to extract the sensitive features from the raw acoustic signals and to recognize the levels of cavitation intensity.

3 Methods

This section presents the multi-task learning framework of cavitation detection and cavitation intensity recognition using 1-D DHRN, including data augmentation method Swin-FFT, 1-D DHRB construction, and the final 1-D DHRN structures.

3.1 Architecture Overview

The overview architecture of our multi-task learning framework of cavitation detection and cavitation intensity recognition using 1-D double hierarchical residual networks (1-D DHRN) is illustrated in Figure 1. First, the Swin-FFT (introduced in subsection 3.2) method is used for data augmentation to handle the small-sample issue in the data pre-processing module. Before data augmentation, the total data is split into a training set and testing set (with a ratio of 80% : 20%). Performing the train/test splitting in advance can ensure that a piece of signal data after any data augmentation would only exit in the training set or the testing set. Then, the pieces of the frequency domain acoustic

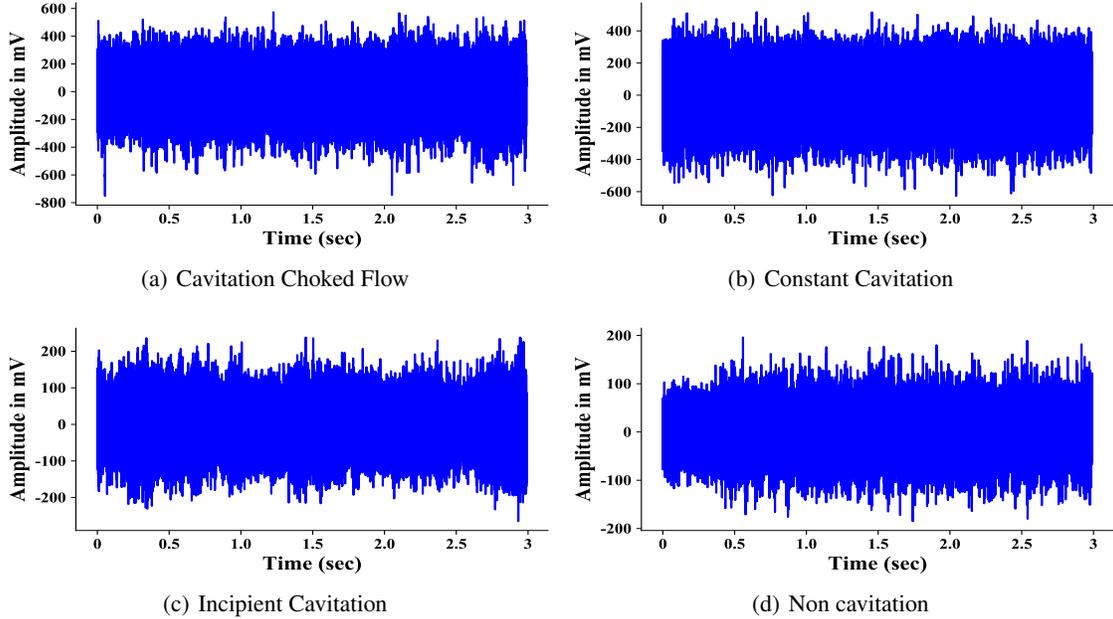


Figure 3: Waveforms of acoustic signals for cavitation choked flow (a), constant cavitation (b), incipient cavitation (c) and non-cavitation (d), respectively.

signals after Swin-FFT are fed as the input of the 1-D DHRN (introduced in subsection 3.3). Finally, the results of cavitation detection and cavitation intensity recognition on the valve are achieved simultaneously. And the Swin-FFT, 1-D pooling layers (max pooling and adaptive average pooling) do not introduce additional parameters. Only our convolutional and fully concatenated layers introduce additional parameters.

In our method, cavitation detection is taken as Task A and cavitation intensity recognition as Task B. Cavitation intensity recognition and cavitation detection correspond to different practical industrial applications. The cavitation detection and cavitation intensity recognition apply the mechanism of hard parameter sharing in multi-task learning [51]. The two tasks share features extracted from the hidden layer. Then, the two tasks can be performed through different output layers. The essence of our proposed method is to minimize the cavitation detection error and the cavitation intensity recognition error, respectively. In other words, we transform data-driven multi-task supervised learning into an optimization problem, to achieve the minima of the two objective functions of cavitation detection \mathcal{L}' and cavitation intensity recognition \mathcal{L}'' by choosing appropriate hyperparameters, as follows.

$$\mathcal{L}' = CE(f(X_{train}; \theta'), Y'_{train}) \quad (1)$$

$$\mathcal{L}'' = CE(f(X_{train}; \theta''), Y''_{train}) \quad (2)$$

where, $CE(\cdot)$ is the cross-entropy function. The basic process of different optimization methods is similar, and we choose the Adam optimization method. First, the output of $f' = f(X_{train}; \theta')$, $f'' = f(X_{train}; \theta'')$ and the optimization objectives \mathcal{L}' , \mathcal{L}'' of the model are computed using the initial parameters. The network parameters θ' and θ'' are then tuned to decrease the objective functions from the final layer to the first layer. This process is repeated until the proper model and a small fit error, i.e. loss function value. And the proposed 1-D DHRN is summarized in Algorithm 1.

3.2 Data Augmentation

In general, machine learning is driven by big data[52]. However, our *Dataset 1*, *Dataset 2* and *Dataset 3* only have 356, 806 and 160 measured acoustic signals, respectively. So data augmentation is essential in handling this kind of problem with small-sample issue. Data augmentation can improve the performance of the model and prevent over-fitting, it can also bring into the model the desired invariance for the task and robustness for the extracted features[53, 54].

Considering the purposed maintaining for steady flow status (i.e. for each individual measurement it is the same fluid status class within 3 s or 25 s record duration) in each recorded data sample and the fine resolution for the sensor, one can

Algorithm 1 Pseudocode of 1-D DHRN

Require: Sub-sequences $(X_{train}, (Y'_{train}, Y''_{train}))$ of the training set, Sub-sequences $(X_{val}, (Y'_{val}, Y''_{val}))$ of the validation set, learning rate: 0.0001, batch size: 4

Ensure: results of cavitation detection and cavitation intensity recognition

repeat

Random sorting of sub-sequences in training set $(X_{train}, Y'_{train}, Y''_{train})$.

Random initialization of 1D DHRN weights θ .

for epoch = 0, 1, ..., N **do**

Select sub-sequences of specified batch size from training set $(X_{train}, (Y'_{train}, Y''_{train}))$.

The forward propagation algorithm calculates the net input $z(l)$ and activation value $\sigma(l)$ for each layer of the 1D DHRN.

Calculating the loss function:

$\mathcal{L}' = CE(f(X_{train}; \theta'), Y'_{train}) \triangleright$ cavitation detection loss

$\mathcal{L}'' = CE(f(X_{train}; \theta''), Y''_{train}) \triangleright$ cavitation intensity recognition loss

Update parameters of cavitation detection and cavitation intensity recognition:

$\min CE(f(X_{train}; \theta), Y'_{train}) \Rightarrow \partial \mathcal{L}' / \partial \theta'$

$\min CE(f(X_{train}; \theta), Y''_{train}) \Rightarrow \partial \mathcal{L}'' / \partial \theta''$

Save parameters of cavitation detection and cavitation intensity recognition in current epoch.

end for

until

1D DHRN model on validation set for $\mathcal{L}'(X_{train}, Y'_{train})$ and $\mathcal{L}''(X_{train}, Y''_{train})$ not falling.

for $i = 0, 1, \dots, len((X_{test}, (Y'_{test}, Y''_{test})))$ **do**

The output of cavitation detection and cavitation intensity recognition:

$f'(X_{test}) \triangleright$ cavitation detection

$f''(X_{test}) \triangleright$ cavitation intensity recognition

Calculate the accuracy of cavitation detection and cavitation intensity recognition:

$Acc(f'(X_{test}), Y'_{test})$ and $Acc(f''(X_{test}), Y''_{test})$

end for

actually split every sample into several pieces with each still holding enough essential information to decipher the flow status, but also with independent characteristics per piece due to the intrinsic randomness of the noise emission-given the piece is not so short. Therefore, we propose here a data augmentation method based on sliding window with fast Fourier transform (Swin-FFT), see Figure 4. The method is divided into two steps:

Step 1: The signal data is split by a sliding window.

Step 2: The time-domain data is transformed into frequency-domain by fast Fourier transform (FFT).

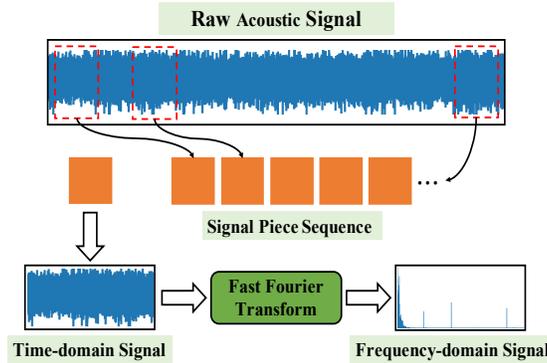


Figure 4: Data augmentation based on sliding window with fast fourier transform (Swin-FFT).

The Swin-FFT contains only one parameter, which is the size of window W_{size} . In our following study, the value of W_{size} will be analyzed and determined by experiments, which can reduce the effect of experts' bias on cavitation intensity recognition and cavitation detection (see section 4) as much as possible.

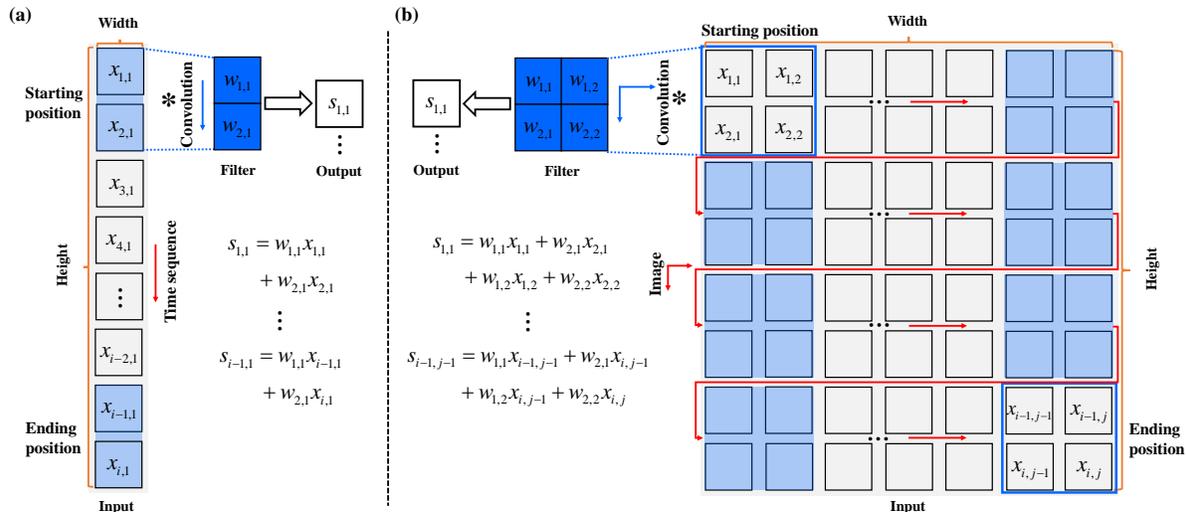


Figure 5: Comparison of working principle between 1-D and 2-D convolution. Figure (a) and Figure (b) show the working principle of 1-D and 2-D convolution, respectively.

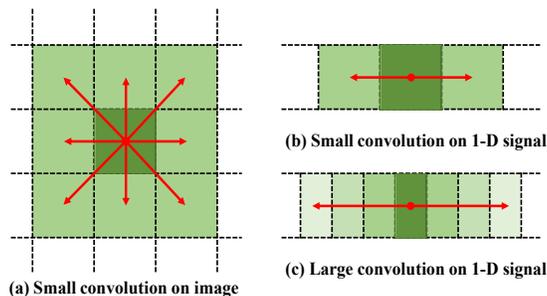


Figure 6: Comparison of small convolution kernel and large convolution kernel in catching information on signals and images. Figure (a) shows that a small convolution kernel can obtain neighbourhood information from eight directions for one pixel on images. Figures (b) and (c) show the ability of the small and large convolution kernels to capture neighbourhood information for one data point on the signal, respectively.

3.3 1-D Double Hierarchical Residual Networks

3.3.1 Comparison Between 1-D and 2-D Convolution

1-D convolution and 2-D convolution are very similar in structure, except that the filters are slid in a different way. For 1-D convolution operation, the convolution kernel slides towards only one direction, i.e. weighted summation in the width or height direction. In 2-D convolution, the convolution kernel slides into both horizontal and vertical directions, i.e. weighting and summing in both the width and height directions. Figure 5 compares the different sliding manners of the filters for 1-D and 2-D convolution operations.

In the case of 1-D time series data as shown in Figure 5(a), each element represents the original acoustic signal point measured by the sensor at each moment in time, and the column represents the original signal from one sensor or one channel with time directed downwards in the sketch. In general, the 1-D convolution kernel only slides vertically along the time axis (height) to extract features. The size of the 1-D convolution kernel determines the size of the receptive field for the filter, i.e. the number of samples (moments) to be calculated by the convolution operation. The stride of the 1-D convolution kernel is defined as the distance moved at each step on the time series signal, and it presents the accuracy of the extraction. In the case of 2-D array image-like data, the pixels of an image consist of a number of small blocks, the positions of which are presented by x and y direction (Figure 5(b)). The width and height of the image is the sum of all x and y , respectively. The size of the 2-D convolution kernel, i.e. its width and height, determines the scope of the convolution operation at each step. Similarly, the stride of the 2-D convolution kernel is defined as the distance

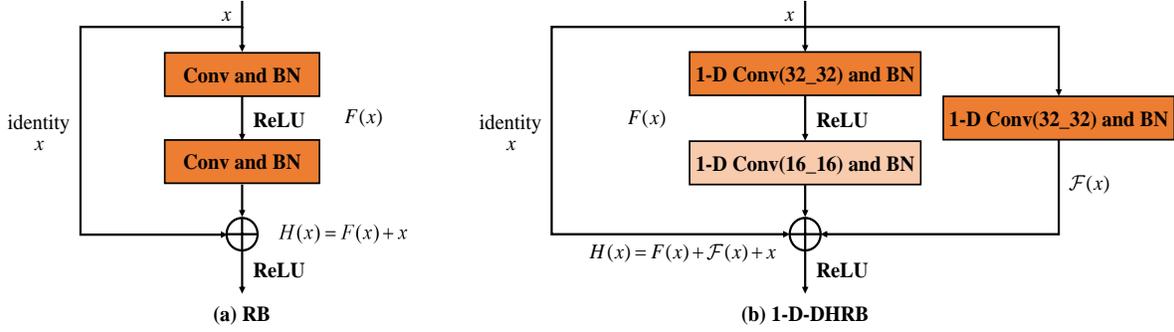


Figure 7: Comparison of structure the RB and 1-D DHRB. Figure (a) shows traditional residual block. Figure (b) is the proposed the 1-D double hierarchical residual block (1-D DHRB).

moved in each step over the width or height of the 2-D array (gray scale image). In general, the 2-D convolution is applied to the 2-D array or gray scale image in the order of left to right and top to bottom.

Regardless of whether it is a 1-D or 2-D convolution kernel, they all have the same and significant hyperparameters, such as the convolution kernel size and stride size, etc. They share weights when sliding to extract local features. However, the convolution with different kernel sizes have different roles on time series signals and images (see Figure 6). In general, a small convolution kernel (3×3) can obtain information on eight neighbourhood directions for one pixel on an image, but it can obtain information on only two neighbourhood directions for one data point on a signal. To solve this problem, we use 1-D DHRB (introduced in 3.3.2) with a large kernel on one-dimensional signals to obtain more neighbourhood information, although the information still comes from only the two neighbourhood directions.

3.3.2 1-D Double Hierarchical Residual Block

Residual block (RB) is an important module of the residual networks (ResNet). RB is based on the idea of connecting blocks of convolutional layers by using skip connection, as shown in Figure 7(a). The output of the RB is $H(x) = F(x) + x$. RB changes the network learning objective from $H(x)$ to $H(x) - x$. This structure can help optimizing the trainable parameters in error backpropagation to avoid the problem of vanishing and exploding gradients. Practically it can also enable the building of deeper network structures.

Motivated by above, we developed a one dimensional convolution-based double hierarchical residual block (1-D DHRB) to improve valve cavitation detection and intensity recognition. The 1-D DHRB consists of several convolutional layers of different filter size (Conv), batch normalization layers (BN), rectified linear unit (ReLU) activation function and two shortcuts, as shown in Figure 7(b). Both RB and 1-D DHRB have two Conv and BN layers. However, the size of two Conv layers of 1-D-DHRB are different and hierarchical compared to the traditional RB. The filter size of the first Conv layer is 32×32 , and the second Conv layer has a filter size of 16×16 (The filter size has been determined through several experiments). 1-D DHRB uses large convolution kernels compared to RB, which can capture sensitive features of the acoustic signal with fewer layers of the network. The size of the first Conv layer is twice the size of the second Conv layer, which helps the network to focus more on the sensitive features of the feature map from the first Conv layer.

The 1-D DHRB adds one more shortcut compared to the traditional RB. One shortcut for 1-D DHRB is the identity x , the other shortcut for 1-D DHRB is a non-linear function $\mathcal{F}(x)$ of identity x through one Conv and BN layer. The output of 1-D DHRB is given by concatenation as

$$H(x) = F(x) + \mathcal{F}(x) + x \quad (3)$$

Several 1-D DHRBs of different kernel number are stacked after the first convolutional layer in 1-D DHRN. The 1-D DHRN has a depth of 18 and it consists of one 1-D convolutional layer, one maximum pooling layer, several 1-D DHRBs with different numbers of kernels, one global average pooling layer and fully connected layer. Table 2 shows the relevant parameters of the 1-D DHRN architecture. These layers enable 1-D DHRN to automatically extract features and classify cavitation and the levels of cavitation intensity.

- 1-D convolutional layer. The essence of the convolutional layer is a set of convolutional kernels. In this research, each convolutional layer has 64, 128, 256 and 512 convolutional kernels of sizes 32 and 16, respectively. In the one-dimensional forward propagation process, each filter convolves with the input vector to generate a corresponding output featured vector. For the k -th filter, the j -th of its output feature vector at layer l is defined

Table 2: Parameters of the proposed 1-D DHRN architecture

No.	Layer type	Kernel size/Stride	Kernel number
1	1-D Convolution	32/1	64
2	Layer 1	[32, 16]/1	64
3	Layer 2	[32, 16]/2	128
4	Layer 3	[32, 16]/2	256
5	Layer 4	[32, 16]/2	512
6	Fully-connected	512	1
7	Softmax_4	4	1
8	Softmax_2	2	1

as follows.

$$y_{j,k}^l = \sigma(\mathbf{W}_k^l * \mathbf{X}_i^{l-1} + b_{j,k}^l) \quad (4)$$

where, σ is the activation function and the Rectified Linear Units (ReLU) is applied in this research, \mathbf{X}_i^{l-1} denotes the i -th local region at layer $l-1$, \mathbf{W}_k^l and $b_{j,k}^l$ denote the weight vector and bias of the k -th filter kernel at layer l .

- 1-D pooling layer. The pooling function uses the overall statistical features of the neighbouring outputs at a location to replace the network's output at the location. The pooling layer is usually added after the convolutional layer to perform the down-sampling operation. This operation can compress the feature vector and removes redundant information, and thus reducing the complexity of the network, decreasing computation and reducing memory consumption, etc. In this study, max pooling and adaptive average pooling are applied to different positions of the 1-D DHRN. The adaptive average pooling is a special pooling of average pooling and its output size is always H . The max pooling and average pooling are shown in Equation (5) and Equation (6), respectively.

$$y_{i,k}^l = \max_{(i-1)H+1 \leq j \leq iH} \{y_{j,k}^{l-1}\} \quad (5)$$

$$y_{i,k}^l = \frac{1}{S} \sum_{m=0}^{S-1} \{y_{j,k}^{l-1}\} \quad (6)$$

where, $y_{j,k}^{l-1}$ is the j -th element of the output feature vector of the k -th filter in layer $l-1$ and H is its height, $y_{i,k}^l$ is the i -th element of the new feature vector after the down-sampling operation in layer l , S is the filter size.

- Fully connected layer. Each node of the fully connected layer is connected to all nodes of the previous layer and it can combine class-distinctive local information on convolutional or pooling layers. These information can be flattened into a vector and fed to the fully connected layer. The output \mathbf{y}^l of the l -th fully connected layer is defined as follows.

$$\mathbf{y}^l = (\mathbf{W}^l)^T \mathbf{y}^{l-1} + \mathbf{b}^l \quad (7)$$

where, \mathbf{y}^{l-1} , \mathbf{W}^l and \mathbf{b}^l are the input vector, weight matrix and bias vector of the fully connected layer at layer l , respectively.

- Softmax layer. The softmax layer is also called the output layer. It is usually used to predict the multi-class classification. The softmax function is defined as follows.

$$p_c^{(n)} = \frac{\exp(\boldsymbol{\theta}_c^T \mathbf{y}^{(n),L-1})}{\sum_{c=1}^C \exp(\boldsymbol{\theta}_c^T \mathbf{y}^{(n),L-1})} \quad (8)$$

where L is the maximum of layer number, the superscript n denotes that n -th sample, the subscript c denotes the c -th class of total classification number C , $\mathbf{y}^{(n),L-1}$ is the output vector of the previous layer, and $\boldsymbol{\theta}_c$ is the weight vector for the c -th class.

For deep neural networks, the loss function is generally used to evaluate the degree of inconsistency between the predicted value of the model and the true value. It is a non-negative real-valued function, and the smaller the loss

		Predicted Class	
		Predicted Value : Positive (+)	Predicted Value : Negative (-)
True Class	True Value : Positive (+)	TP True Positive	FN False Negative
	True Value : Negative (-)	FP False Positive	TN True Negative

Figure 8: The Confusion Matrix.

function, the better the robustness of the model. The cross-entropy loss function is the most popular loss function and is defined as follows.

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C \text{sign}(\text{pred}^{(n)} = c) \log(p_c^{(n)}) \quad (9)$$

where, N is the total number of samples, $\text{sign}(\cdot)$ denotes the sign function returning 1 for real and 0 for fake. And $\text{pred}^{(n)}$ represents the prediction result for the n -th sample.

Neural networks are trained to determine the weights and biases of the network. In this research, Adam optimisation algorithm is used to update the weights and biases of the network during back propagation. The Adam optimisation algorithm [55] is an extension of the stochastic gradient descent (SGD) [56, 57] method and has been introduced in the literature [58, 59, 60].

4 Experiments

4.1 Overview

In our experiments, first evaluation metrics are introduced. Then, *Dataset 1*, *Dataset 2* and *Dataset 3* of the cavitation detection and cavitation intensity recognition are described. And we analyze the effect of the window size W_{size} of Swin-FFT on the performance of 1-D DHRN. Finally, we compare the performance of 1-D DHRN and other state-of-the-art deep learning and machine learning models on *Dataset 1*, *Dataset 2* and *Dataset 3*.

In all experiments, we use adaptive moment estimation (Adam) [55] as our optimizer with learning rate set to be 1×10^{-4} . Both our model and the model of the comparison state-of-the-art method are trained on NVIDIA GPU and the number of training epochs is set to 100 on all datasets, upon which by monitoring the loss of validation in learning curve we observe saturation which gives convergence indication (see Appendix). These setting ensures that all trained models are more fairly analyzed for performance comparisons. In addition, both our method and the baseline methods are trained with the same data augmentation (Swin-FFT) in our implementation. Our comparison methods include five different approaches from both conventional machine learning and deep learning. Where, the SVM [21], Decision Tree [14] and 1-D CNN [25] are methods by other researchers for cavitation. The XGBoost + ASFE [50] and 1-D ResNet-18 are our previous research and our benchmark, respectively. Our source code and data are released at <https://github.com/CavitationDetection/1-D-DHRN>.

4.2 Evaluation Metric

In order to evaluate the model after training, four metrics are chosen to comprehensively assess the model performance: Accuracy, Precision, Recall and F1-score. We first calculate the confusion matrix [61] to more conveniently define the evaluation metrics and visualize model performance. In the confusion matrix shown in Figure 8, each column represents the predicted class, and each row represents the actual class. TP (True Positive) is the fraction of positive samples those got correctly predicted by the model, and TN (True Negative) is for the correctly predicted negative samples. FP (False Positive) means the incorrectly classified positive samples those should be negative actually, and FN (False Negative) is for the incorrectly predicted negative samples.

Accuracy, Precision, Recall and F1-score are common evaluation metrics for classification problems, defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

4.3 Cavitation Intensity Recognition: Dataset 1

4.3.1 Data Description

In this subsection, the proposed 1-D DHRN is evaluated on the condition monitoring of cavitation dataset (*Dataset 1*) provided by SAMSON AG in Frankfurt. The hardware of this experiment is shown in Figure 2, and five flow status are induced in the acoustic signal data by varying the differential pressure at various constant upstream pressures of the control valve different operation conditions: cavitation choked flow, constant cavitation, incipient cavitation, turbulent flow and background no-flow, as shown in Table 4. The turbulent flow and background no-flow are non-cavitation conditions. The experiments have been conducted using seven different valve strokes at four different upstream pressures and the operating parameters are shown in Table 5. *Dataset 1* has a total of 356 acoustic signal samples and the frequency of samples is 1562.5 kHz within time duration of 3 s. It should be noted that the *Dataset 1* has been measured by SAMSON AG in a professional environment. The details of the training set, validation set and test set for *Dataset 1* without Swin-FFT is shown in Table 3.

Table 3: The details of the training set, validation set and test set for *Dataset 1*, *Dataset 2* and *Dataset 3* without Swin-FFT.

	Training set				Validation set				Test set			
	choked flow	constant	incipient	non	choked flow	constant	incipient	non	choked flow	constant	incipient	non
<i>Dataset 1</i>	53	68	29	109	5	7	3	12	14	18	8	30
<i>Dataset 2</i>	107	286	47	143	11	31	5	15	30	79	12	40
<i>Dataset 3</i>	29	29	29	29	3	3	3	3	8	8	8	8

4.3.2 Results

The effect of W_{size} on the accuracy of cavitation intensity recognition is studied (other metrics see Appendix). The W_{size} is set to be 2234720, 1167360, 778240, 583680, 466944, 389120, 333531, 291840, 259413 and 233472. The results of cavitation intensity recognition of *Dataset 1* are shown in Table 6 and Figure 9.

From Table 6 and Figure 9, it can be seen that the accuracy of 1-D DHRN gradually increases along with reducing window size when $2334720 \geq W_{size} \geq 466944$. And 1-D DHRN achieves the best cavitation intensity recognition accuracy of 93.75% when W_{size} is 466944. When $466944 \geq W_{size} \geq 233472$, the accuracy of 1-D DHRN progressively decreases but remains above 90%. Compared to other methods in the literature, the accuracy of 1-D DHRN is

Table 4: Details of the flow status condition of *Dataset 1*.

	Flow status	Number of samples
Cavitation	Cavitation choked flow	72
	Constant cavitation	93
	Incipient cavitation	40
Non cavitation	Turbulent flow	118
	No flow	33

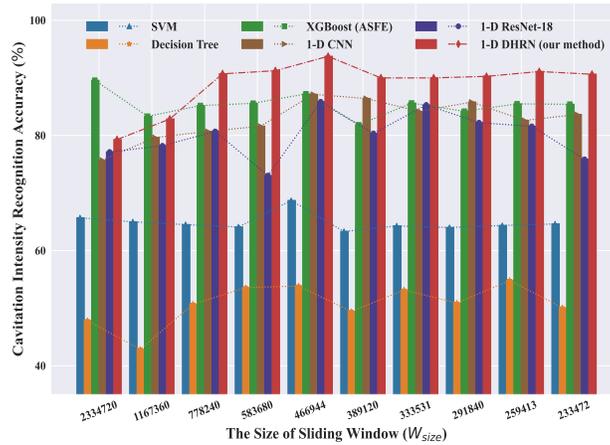
Table 5: Operation parameters of *Dataset 1*.

No.	Valve stroke (<i>mm</i>)	Upstream pressure (<i>bar(a)</i>)	Temperature ($^{\circ}\text{C}$)
1	15	[10,9,6,4]	25-50
2	13.5	[10,9,6,4]	25-50
3	11.25	[10,9,6,4]	25-50
4	7.50	[10,9,6,4]	25-50
5	3.75	[10,9,6,4]	25-50
6	1.50	[10,9,6,4]	25-50
7	0.75	[10,9,6,4]	25-50

Table 6: Accuracy results of different W_{size} of cavitation intensity recognition in *Dataset 1* (%).

Method	Window Size (W_{size})									
	2334720	1167360	778240	583680	466944	389120	333531	291840	259413	233472
SVM [21]	65.71	65.00	65.42	64.11	68.71	63.33	64.29	64.02	64.37	64.64
Decision Tree [14]	47.86	42.86	50.71	53.57	53.86	49.44	53.16	50.89	54.84	50.07
1-D CNN [25]	75.71	79.64	80.71	81.61	87.14	86.43	84.18	85.80	82.62	83.64
XGBoost + ASFE [50]	89.58	83.33	85.19	85.59	87.22	81.83	85.62	84.20	85.49	85.42
1-D Resnet-18*	77.14	78.21	80.71	73.04	85.86	80.36	85.31	82.23	81.59	75.86
1-D DHRN (our method)	79.29	82.86	90.71	91.25	93.75	90.00	90.00	90.27	91.11	90.64

* our baseline: 1-D Resnet-18

Figure 9: The effect of W_{size} on cavitation intensity recognition accuracy for 1-D DHRN and comparison methods in *Dataset 1*.

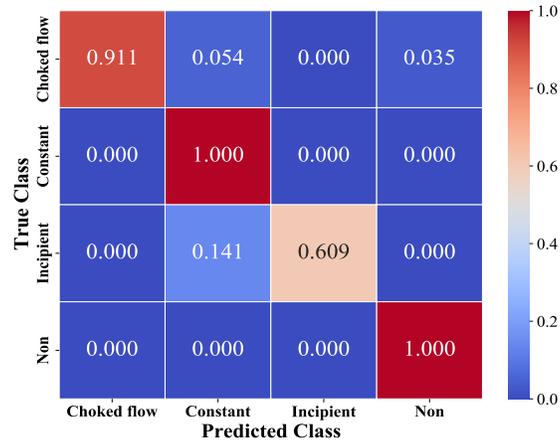


Figure 10: The confusion matrix for the best cavitation intensity recognition accuracy of the 1-D DHRN in *Dataset 1*.

improved by 25.04%, 39.89%, 6.53%, 6.61% and 7.89% compared to SVM, Decision Tree, XGBoost (ASFE), 1-D CNN and 1-D ResNet-18, respectively.

In general, the performances of 1-D DHRN are obviously better than SVM, Decision Tree, XGBoost (ASFE), 1-D CNN and also 1-D ResNet-18 at each particular value of W_{size} . The best cavitation intensity recognition accuracies for the SVM, Decision Tree, XGBoost (ASFE), 1-D CNN and 1-D ResNet-18 methods are 68.71%, 54.84%, 89.58%, 87.14% and 85.86%, respectively.

When W_{size} is large such as 2334720 and 1167360, the 1-D DHRN has poorer performance compared to XGBoost (ASFE), which is mainly because of the small amount of data obtained from data augmentation. However, even for this setup, the advantage of our approach compared to conventional XGBoost (ASFE) is the end-to-end implementation of cavitation intensity identification using only the valve acoustic signal, instead of feature extraction separately.

To demonstrate the cavitation intensity recognition results in more detail, the confusion matrix of the testing accuracy is evaluated in Figure 10. It can be seen that the constant cavitation and the non cavitation are most easily to be distinguished with the highest accuracy compared to other levels of cavitation intensity. This can be explained by the fact that the constant cavitation and the non cavitation have special features that distinguish them from other states of cavitation. It can also be indicated that our 1-D DHRB structure can capture more sensitive cavitation features. However, the incipient cavitation is challenging to be recognized, which has the lowest identification accuracy compared to choked flow and constant cavitation. This is understandable because in physics it is the critical state between cavitation and non cavitation. Technically another reason is that the amount of incipient cavitation data is very small. In the study of Lehmann and Young [62] it also shows that the end stages of cavitation can be more easily detected than incipient cavitation. From a practical perspective, it's acceptable that the incipient cavitation is recognized as the constant cavitation to still give rise alarm to the process/plant. With this consideration, the accuracy of the incipient cavitation in our method reaches 75% for *Dataset 1*.

4.4 Cavitation Intensity Recognition: Dataset 2

4.4.1 Data Description

In this case study, the proposed 1-D DHRN is tested on condition monitoring of cavitation dataset (*Dataset 2*) provided by SAMSON AG in Frankfurt. The principle of the experimental setup is shown in Figure 2. *Dataset 2* uses different valves and sensors compared to *Dataset 1*. The five flow states are shown in Table 7. The experiments have been conducted using seven different valve strokes at three different upstream pressures and the operating parameters are shown in Table 8. The amount of *Dataset 2* is 806 and the frequency of samples is 1562.5 kHz within time duration of 25 s. It should be noticed that the *Dataset 2* has been collected by SAMSON AG in a professional environment. The details of the training set, validation set and test set for *Dataset 2* without Swin-FFT is shown in Table 3.

Table 7: Details of the flow status condition of *Dataset 2*.

Flow status		Number of samples
Cavitation	Cavitation choked flow	148
	Constant cavitation	396
	Incipient cavitation	64
Non cavitation	Turbulent flow	183
	No flow	15

Table 8: Operation parameters of *Dataset 2*.

No.	Valve stroke (mm)	Upstream pressure (bar(a))	Temperature (°C)
1	60	[10,6,4]	23-52
2	55	[10,6,4]	23-52
3	45	[10,6,4]	23-52
4	30	[10,6,4]	23-52
5	25	[10,6,4]	23-52
6	15	[10,6,4]	23-52
7	6	[10,6,4]	23-52

4.4.2 Results

The effect of W_{size} on the accuracy of cavitation intensity recognition is studied (see Appendix of other metrics). The W_{size} is set to be 2234720, 1167360, 778240, 583680 and 466944. The comparison results of cavitation intensity recognition of *Dataset 2* are shown in Table 9 and Figure 11.

From Table 9, it can be found that 1-D DHRN gets the best cavitation intensity recognition accuracy of 94.31% when W_{size} is 778240. The accuracy of 1-D DHRN gradually increases when $2334720 \geq W_{size} \geq 778240$. When $778240 > W_{size} \geq 466944$, the accuracy of 1-D DHRN progressively decreases and remains above 92%. The best cavitation intensity recognition of results from the SVM, Decision Tree, XGBoost (ASFE), 1-D CNN and 1-D Resnet-18 are 50.49%, 65.07%, 89.22%, 86.13% and 91.71%, respectively, which were surpassed by 40.82%, 29.24%, 5.09%, 8.18% and 2.6% compared to 1-D DHRN. As can be seen from Figure 11, 1-D DHRN has achieved with higher accuracy than all other approaches at each value of W_{size} .

The confusion matrix for testing accuracy is evaluated and shown in Figure 12. It can be noted that choked flow, constant cavitation and non-cavitation are most easily identified with the highest accuracy compared to the incipient cavitation. And the recognition accuracy of the constant cavitation and the non cavitation reaches 100%. Take a practical point of view (i.e., can cause alarm) as mentioned earlier, the recognition accuracy of initial cavitation can reach 80%. It can also be seen that our method achieves excellent performance on datasets with different types of valves and sensors, which demonstrates the generalisation capability of the methodology.

Table 9: Accuracy results of different W_{size} of cavitation intensity recognition in *Dataset 2* (%).

Method	Window Size (W_{size})				
	2334720	1167360	778240	583680	466944
SVM [21]	50.49	49.73	47.88	48.56	50.27
Decision Tree [14]	65.07	62.86	62.27	62.39	60.29
1-D CNN [25]	85.05	85.09	86.13	85.12	85.11
XGBoost + ASFE [50]	84.93	88.00	89.22	88.71	87.01
1-D Resnet-18*	87.34	89.58	91.71	90.57	90.00
1-D DHRN (our method)	89.39	91.25	94.31	92.71	92.29

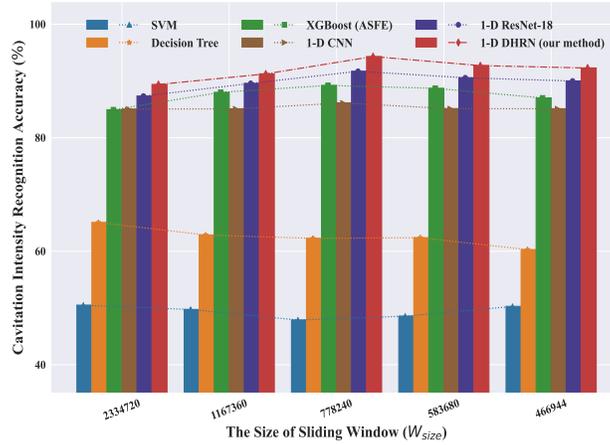


Figure 11: The effect of W_{size} on cavitation intensity recognition accuracy for 1-D DHRN and comparison methods in **Dataset 2**.

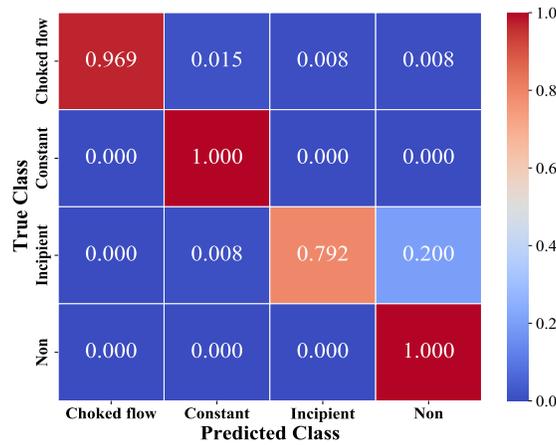


Figure 12: The confusion matrix for the best cavitation intensity recognition accuracy of the 1-D DHRN in **Dataset 2**.

4.5 Cavitation Intensity Recognition: Dataset 3

4.5.1 Data Description

As the third case study, the proposed 1-D DHRN is evaluated on the condition monitoring of cavitation dataset (**Dataset 3**) provided by SAMSON AG in Frankfurt. **Dataset 3** is different from **Dataset 1** and **Dataset 2** by carrying the noise of the real working environment. The five flows states about **Dataset 3** are shown in Table 10. The experimental operational parameters about **Dataset 3** are shown in Table 11. The quantity of **Dataset 3** is 160 for each flow states and the frequency of samples is 1562.5 kHz within time duration of 25 s. It should be noted out that the **Dataset 3** has been collected by SAMSON AG inside a professional environment. The details of the training set, validation set and test set for **Dataset 3** without Swin-FFT is shown in Table 3.

4.5.2 Results

The influence of W_{size} on the accuracy of cavitation intensity recognition is investigated (see Appendix of other metrics). The W_{size} is the same as the W_{size} of **Dataset 2**. The accuracy of cavitation intensity recognition of **Dataset 3** is shown in Table 12 and Figure 13.

From Table 12 and Figure 13, it can be observed that the best accuracy of our method is 100%. And the accuracy of our method is above 99% for every value of W_{size} . Our method has achieved 100% accuracy of recognition of all cavitation states (see Figure 14).

Table 10: Details of the flow status condition of *Dataset 3*.

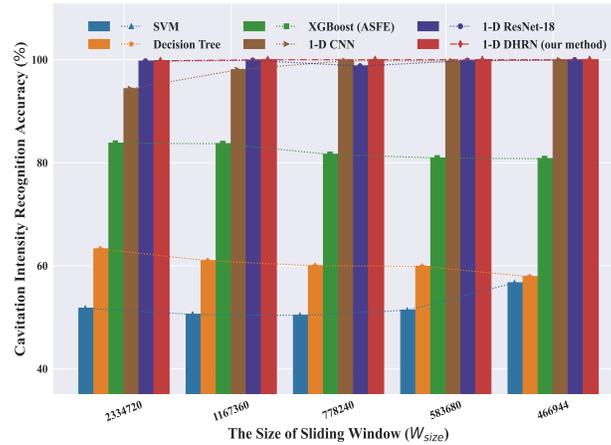
Flow status		Number of samples
Cavitation	Cavitation choked flow	40
	Constant cavitation	40
	Incipient cavitation	40
Non cavitation	Turbulent flow	40

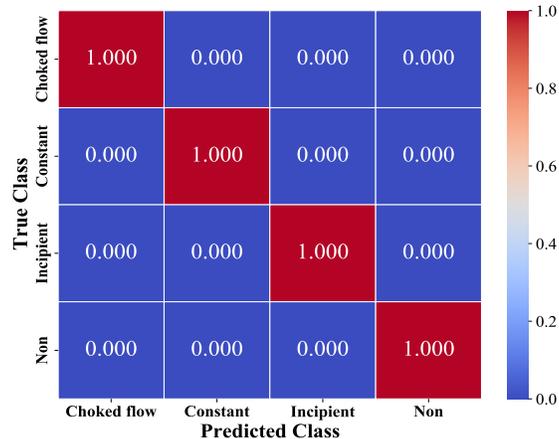
Table 11: Operation parameters of *Dataset 3*.

No.	Valve stroke (<i>mm</i>)	Upstream pressure (<i>bar(a)</i>)	Temperature ($^{\circ}\text{C}$)
1	15	10	32-39

Table 12: Accuracy results of different W_{size} of cavitation intensity recognition in *Dataset 3* (%).

Method	Window Size (W_{size})				
	2334720	1167360	778240	583680	466944
SVM [21]	51.76	50.57	50.38	51.37	56.70
Decision Tree [14]	63.28	60.98	60.00	59.90	57.91
1-D CNN [25]	94.37	98.06	99.81	99.75	99.94
XGBoost + ASFE [50]	83.79	83.71	81.63	80.92	80.80
1-D Resnet-18*	99.72	99.88	98.77	99.85	99.96
1-D DHRN (our method)	99.81	100.00	100.00	100.00	100.00

Figure 13: The effect of W_{size} on cavitation intensity recognition accuracy for 1-D DHRN and comparison methods in *Dataset 3*.

Figure 14: The confusion matrix for the best cavitation intensity recognition accuracy of the 1-D DHRN in *Dataset 3*.Table 13: Accuracy results of different W_{size} of cavitation detection in *Dataset 1* (%).

Method	Window Size (W_{size})									
	2334720	1167360	778240	583680	466944	389120	333531	291840	259413	233472
SVM [21]	81.43	82.86	80.24	79.82	83.29	79.29	79.29	80.00	80.56	79.14
Decision Tree [14]	73.57	76.07	74.29	76.96	77.00	76.55	77.86	76.07	79.44	81.86
1-D CNN [25]	94.29	94.29	94.05	94.46	94.14	95.24	94.90	95.09	95.71	95.21
XGBoost + ASFE [50]	93.56	92.36	92.36	91.32	91.25	91.08	88.99	89.93	89.51	90.00
1-D Resnet-18*	95.00	95.71	95.71	95.54	95.43	96.19	96.12	95.89	95.48	95.64
1-D DHRN (our method)	96.43	96.71	96.67	96.43	96.43	97.02	96.82	96.07	96.03	96.23

The reasons why our method and compared methods have achieved excellent results on *Dataset 3* compared to *Dataset 1* and *Dataset 2* are as follows. First, although *Dataset 3* are with real background noise compared to *Dataset 1* and *Dataset 2*, our data augmentation of the Swin-FFT operation can filter most of the noise (see subsection 5.1) in frequency domain. Second, *Dataset 3* are obtained with only one setup for the valve stroke and upstream pressure (see Table 11 and subsection 5.2). Third, *Dataset 3* is balanced for each flow state especially for cavitation state (see Table 10 and subsection 5.2). From the above, we can infer that different valve openings and upstream pressures can affect the accuracy of the cavitation intensity recognition.

4.6 Cavitation Detection

In this subsection, the proposed 1-D DHRN is evaluated on condition monitoring of cavitation dataset (*Dataset 1*, *Dataset 2* and *Dataset 3*) for the binary classification task, cavitation detection. The comparison of accuracies on *Dataset 1*, *Dataset 2* and *Dataset 3* are shown in Tables 13, 14 and 15 (see Appendix of other metrics).

Table 14: Accuracy results of different W_{size} of cavitation detection in *Dataset 2* (%).

Method	Window Size (W_{size})				
	2334720	1167360	778240	583680	466944
SVM [21]	89.83	89.36	89.06	89.92	89.30
Decision Tree [14]	89.58	88.59	89.96	89.84	88.97
1-D CNN [25]	94.29	95.24	94.05	94.46	94.14
XGBoost + ASFE [50]	90.56	90.73	91.33	91.02	90.43
1-D Resnet-18*	95.48	96.19	95.71	95.54	95.43
1-D DHRN (our method)	96.67	97.02	97.64	96.43	96.43

Table 15: Accuracy results of different W_{size} of cavitation detection in **Dataset 3** (%).

Method	Window Size (W_{size})				
	2334720	1167360	778240	583680	466944
SVM [21]	100.00	100.00	100.00	100.00	100.00
Decision Tree [14]	100.00	100.00	100.00	100.00	100.00
1-D CNN [25]	100.00	100.00	100.00	100.00	100.00
XGBoost + ASFE [50]	100.00	100.00	100.00	100.00	100.00
1-D Resnet-18*	100.00	100.00	100.00	100.00	100.00
1-D DHRN (our method)	100.00	100.00	100.00	100.00	100.00

Dataset 1 From Table 13, it can be found that 1-D DHRN achieves the best accuracy, 97.02%, when W_{size} is 389120. 1-D DHRN is always better than other methods under the same value of W_{size} . And the accuracy of the 1-D DHRN is above 96% at each value of W_{size} .

Dataset 2 From Table 14, it can be noted that 1-D DHRN achieves the best accuracy of 97.64% when W_{size} is 778240. 1-D DHRN has better accuracy compared to other methods under the same value of W_{size} . And the accuracy of the 1-D DHRN is above 96% for each value of W_{size} .

Dataset 3 As can be seen from Table 15, both the 1-D DHRN and comparative methods have achieved 100% accuracy for each W_{size} values. It can be concluded that the cavitation and non-cavitation features of **Dataset 3** are more easily distinguished compared to **Dataset 1** and **Dataset 2**.

5 Discussions

5.1 Analysis of Swin-FFT

In order to verify that the Swin-FFT can filter out most of the noise, we select the data from **Dataset 2** which was obtained under the same operation as **Dataset 3**. Then, the data are processed using the Swin-FFT. Finally, we compare the shape of each cavitation state from **Dataset 2** without noise and **Dataset 3** with noise, as shown in Figure 15. From Figure 15, it can be seen that the spectral structures of **Dataset 2** without noise and **Dataset 3** with noise are basically similar when transformed into the frequency domain by Swin-FFT. This indicates that Swin-FFT can filter most of the noise.

5.2 Analysis of Data Acquisition Operations

In order to determine whether the data obtained by the mixed operation and the single operation have an impact on the results, a total of 46 data are selected from **Dataset 2** for valve stroke equals to 15 mm and upstream pressure equals to 10 bar(a), where the numbers of cavitation choked flow, constant cavitation, incipient cavitation and non cavitation are 15, 24, 3 and 4, respectively. Then, the selected data are processed by Swin-FFT and fed into the 1-D DHRN. Finally, the results of cavitation detection and cavitation intensity recognition are obtained and shown in Figure 16.

As can be seen from Figure 16, the accuracies of cavitation intensity recognition and cavitation detection are higher with single operation data compared to mixed operation data. Therefore, we conclude that for data obtained from a single valve stroke and upstream pressure, the performance of cavitation intensity recognition and cavitation detection is better. We noted that the accuracy of cavitation intensity recognition is not 100% in this test, it can be thus inferred that the balance of data for each cavitation state is important.

5.3 Analysis of Downsampling

In practical applications, the resolution of sensors represent the quality of the data obtained. The ability to recognize different levels of cavitation intensity with the data obtained from low level sensors becomes very significant and challenging. In this research, 1-D DHRN is evaluated for cavitation intensity recognition and cavitation detection under the original frequency of samples ($F_s = 1562500$ Hz), one-half, one-quarter, one-sixth, one-eighth of the original frequency (sampling rate) of samples (781250 Hz, 390625 Hz, 260416 Hz and 195312 Hz) and that of the mobile phone can mostly accommodate (48000 Hz $\approx F_s/32$), respectively. The cavitation intensity recognition and cavitation detection accuracy of the 1-D DHRN model has been compared again in **Dataset 1**, **Dataset 2** and **Dataset 3** with different frequencies of samples, as shown in Figure 17.

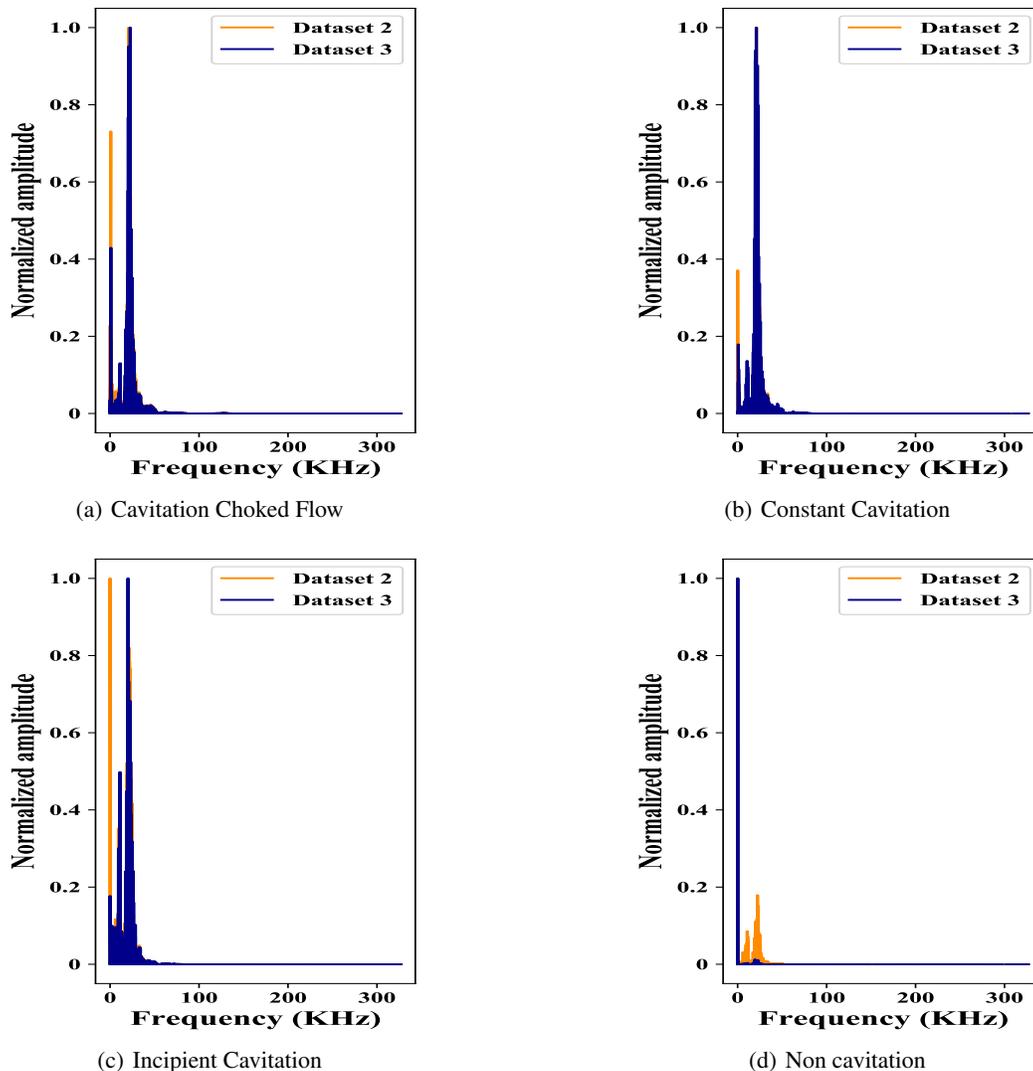


Figure 15: Examples of cavitation choked flow (a), constant cavitation (b), incipient cavitation (c) and non-cavitation (d) for *Dataset 2* and *Dataset 3* converted by Swin-FFT.

• **Dataset 1 and Dataset 2.** For cavitation intensity recognition, although the 1-D DHRN model begins to suffer from gradual reduction in accuracy as the frequency of samples decreases, the accuracy always remains above 84% when $1562500 \text{ Hz} \geq F_s \geq 195312 \text{ Hz}$. And even for mobile phone setup with the frequency of samples equals to 48000 Hz (one-32nd of the original signal frequency of samples), our 1-D DHRN has achieved 72.83% and 77.65% accuracy on the two datasets. For cavitation detection, the accuracy of the 1-D DHRN model progressively decreases as the frequency of samples decreases, but the accuracy of cavitation detection remains above 93% at each all values of sampling frequency. When the frequency of samples equals to 48000 Hz, our method gives 93.29% and 93.57% accuracies, respectively. And when the frequency of samples is reduced to one-32nd of the original signal frequency of samples, the accuracy of 1-D DHRN is only decreased by 3.73% and 4.07%.

• **Dataset 3.** For cavitation intensity recognition, the accuracy of our method begins to reduce when the frequency of sample is 260416 Hz ($260416 \text{ Hz} \approx F_s/6$). The accuracy of our method always remains above 98% when $260416 \text{ Hz} \geq F_s \geq 195312 \text{ Hz}$. And our method has achieved 90.17% accuracy when the frequency of sample is equal to 48000 Hz. For cavitation detection, the accuracy of our method always remains above 99% at each particular value of F_s . When the frequency of samples is equal to 48000 Hz, our method obtains 99.28%, which is reduced by only 0.72% of the accuracy when compared to the original frequency of samples.

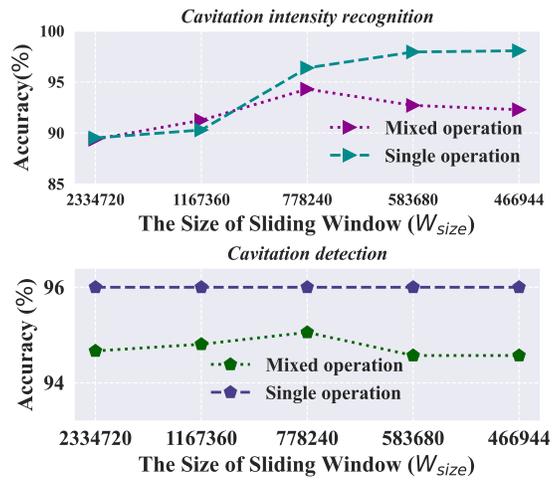


Figure 16: The effect of data diversity, with data obtained from single valve stroke and upstream pressure and mixed valve strokes and upstream pressures, on the results.

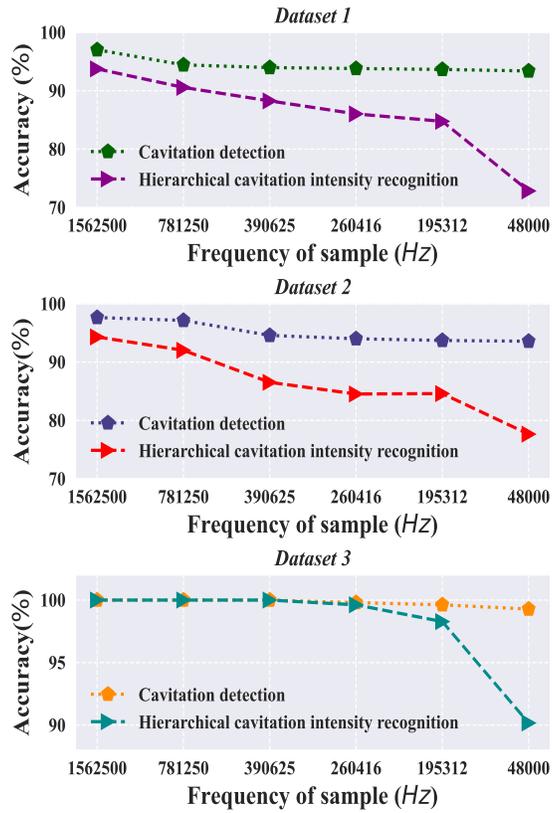


Figure 17: The results of cavitation intensity recognition and cavitation detection on different frequencies of samples for *Dataset 1*, *Dataset 2* and *Dataset 3*.

5.4 Learning and Optimization

In our paper, cavitation detection and cavitation intensity recognition are multi-class classification problem in data-driven supervised learning. The essence of supervised learning is to minimize the cumulative classification error, i.e. $\arg \min_{\theta} \sum_i e(f(x_i; \theta), y_i^*)$, where y^* denotes reference or "ground truth" and $f(x; \theta)$ is the output of the neural network. We typically optimize, i.e. train, with a stochastic gradient descent (SGD) optimizer. And we rely on auto-diff to compute the gradient w.r.t. weights θ , $\partial f / \partial \theta$. In recent years, there has been a lot of research on optimizers [59, 60, 63, 64], and the Adam optimizer is used in our research to optimize errors. In addition, the choice of error function is important for learning and optimization of supervised learning. For classification problems, the cross-entropy function is usually employed as the error function, since it can avoid the problem of learning rate decreasing due to gradient dispersion.

5.5 Extending Applications

Our proposed 1-D DHRB is motivated by the residual block, and the residual block is the core idea of the ResNet [42]. As ResNet has been proven to be a very effective convolutional neural network structure, which is widely employed in fields such as image classification [65], medical image recognition [66] and object recognition. In addition, 1D CNN have been successfully applied to vibration signals [67]. In addition, 1D CNN have been successfully applied to vibration signals [46], ECG [68, 69] and EEG [70, 71]. Therefore, our proposed 1D DHRN can also be employed for vibration signals, ECG, EEG and other signals.

6 Conclusion and Outlook

In this work we present a novel multi-task learning framework for cavitation intensity recognition ("cavitation choked flow", "constant cavitation", "incipient cavitation" and "non cavitation") and cavitation detection ("cavitation" and "non cavitation") using 1-D double hierarchical residual networks (1-D DHRN). To the best of our knowledge, besides to be the first application of 1-D convolutional neural networks based on 1-D residual blocks to perform cavitation intensity recognition and cavitation detection, other main contributions to this paper are summarized as the following four points. Firstly, a sliding window with fast fourier transform (Swin-FFT) data augmentation method is introduced to mitigate the few-short learning problem. Secondly, the 1-D double hierarchical residual blocks (1-D DHRB) with large kernels is proposed as a feature extractor to capture sensitive features of valves acoustic signals. Thirdly, a new structure of 1-D DHRN has been constructed through stacking 1-D DHRB for cavitation intensity recognition, which has in total 18 layers for the whole structure. With deeper network and better feature extraction layers, the proposed 1-D DHRN could improve the final prediction performance on cavitation intensity recognition and cavitation detection. Finally, 1-D DHRN has been tested on three dataset of valve acoustic signals and achieved significant results for the two tasks simultaneously. Moreover, the structure of 1-D DHRN has also been evaluated on different frequencies of samples and showed excellent results for even the frequency of samples which the mobile phones can hold. These results have validated the good performance of our proposed 1-D DHRN both in the field of cavitation intensity recognition and cavitation detection.

Although the proposed method has improved recognition of the incipient cavitation recognition, which is still relatively low. In future work, convolutional neural networks (CNN) have achieved remarkable results in mechanical device health management due to their stronger capability of representation learning. Therefore, we directly extract deeper and more expressive valve cavitation features using the 1D CNN to enhance the performance of valve cavitation detection and cavitation intensity recognition.

Acknowledgements

This research is supported by Xidian - FIAS International Joint Research Center (Y. S.), by the AI grant at FIAS through SAMSON AG (J. F., K. Z.), by the BMBF funding through the ErUM - Data project (K. Z.), by SAMSON AG (D. V., T. S., A. W.), by the Walter GreinerGesellschaft zur Förderung der physikalischen Grundla - genforschung e.V. through the Judah M. Eisenberg Laureatus Chair at Goethe Universität Frankfurt am Main (H. S.), by the NVIDIA GPU grant through NVIDIA Corporation (K. Z.).

References

- [1] A Masjedian Jazi and H Rahimzadeh. Detecting cavitation in globe valves by two methods: Characteristic diagrams and acoustic analysis. *Applied Acoustics*, 70(11-12):1440–1445, 2009.

- [2] Rui Zhao, Rong-qing Xu, Zhong-hua Shen, Jian Lu, and Xiao-wu Ni. Experimental investigation of the collapse of laser-generated cavitation bubbles near a solid boundary. *Optics & Laser Technology*, 39(5):968–972, 2007.
- [3] EA Brujan, GS Keen, A Vogel, and JR Blake. The final stage of the collapse of a cavitation bubble close to a rigid boundary. *Physics of fluids*, 14(1):85–92, 2002.
- [4] Marion Bonnier and Olivier Eiff. Experimental investigation of the collapse of a turbulent wake in a stably stratified fluid. *Physics of fluids*, 14(2):791–801, 2002.
- [5] QN Song, YG Zheng, DR Ni, and ZY Ma. Corrosion and cavitation erosion behaviors of friction stir processed ni-al bronze: effect of processing parameters and position in the stirred zone. *Corrosion*, 70(3):261–270, 2014.
- [6] Kristoffer McKee, Gareth Forbes, Muhammad Ilyas Mazhar, Rodney Entwistle, and Ian Howard. A review of major centrifugal pump failure modes with application to the water supply and sewerage industries. In *ICOMS Asset Management Conference Proceedings*. Asset Management Council, 2011.
- [7] C Schleih, E Viennet, M Deeken, H Ding, Y Xia, S Lowry, and H Murrenhoff. 3d-cfd simulation of an axial piston displacement unit. In *Ninth International Fluid Power Conference, Aachen, Germany, Mar*, pages 24–26, 2014.
- [8] Fang-long Yin, Song-lin Nie, Shu-han Xiao, and Wei Hou. Numerical and experimental study of cavitation performance in sea water hydraulic axial piston pump. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 230(8):716–735, 2016.
- [9] Christopher E Brennen. *Cavitation and bubble dynamics*. Cambridge University Press, 2014.
- [10] Hossein Gholizadeh, Richard Burton, and Greg Schoenau. Fluid bulk modulus: comparison of low pressure models. *International journal of fluid power*, 13(1):7–16, 2012.
- [11] Y Liu, J Lin, TA Dean, and DC J Farrugia. A numerical and experimental study of cavitation in a hot tensile axisymmetric testpiece. *The Journal of strain analysis for engineering design*, 40(6):571–586, 2005.
- [12] Brad Clarke and Kari Oksanen. Evaluating cavitation solutions—past and present. *Opflow*, 37(7):18–23, 2011.
- [13] Radu-Emil Precup, Plamen Angelov, Bruno Sielly Jales Costa, and Moamar Sayed-Mouchaweh. An overview on fault diagnosis and nature-inspired optimal control of industrial process applications. *Computers in Industry*, 74:75–94, 2015.
- [14] NR Sakthivel, V Sugumaran, and SJESwA Babudevasenapati. Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. *Expert Systems with Applications*, 37(6):4040–4049, 2010.
- [15] J Ross Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [16] V Muralidharan and V Sugumaran. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement*, 46(1):353–359, 2013.
- [17] Raghuvver Rao. Wavelet transforms. *Encyclopedia of imaging science and technology*, 2002.
- [18] V Muralidharan, V Sugumaran, and V Indira. Fault diagnosis of monoblock centrifugal pump using svm. *Engineering Science and Technology, an International Journal*, 17(3):152–157, 2014.
- [19] BISWAJIT Samanta, KR Al-Balushi, and SA Al-Araimi. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering applications of artificial intelligence*, 16(7-8):657–665, 2003.
- [20] John H Holland. Genetic algorithms. *Scientific american*, 267(1):66–73, 1992.
- [21] Bo-Suk Yang, Won-Woo Hwang, Myung-Han Ko, and Soo-Jong Lee. Cavitation detection of butterfly valve using support vector machines. *Journal of sound and vibration*, 287(1-2):25–43, 2005.
- [22] DJ Bordoloi and Rajiv Tiwari. Identification of suction flow blockages and casing cavitations in centrifugal pumps by optimal support vector machine techniques. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 39(8):2957–2968, 2017.
- [23] Asish Kumar Panda, Janani Shruti Rapur, and Rajiv Tiwari. Prediction of flow blockages and impending cavitation in centrifugal pumps using support vector machine (svm) algorithms based on vibration measurements. *Measurement*, 130:44–56, 2018.
- [24] Janani Shruti Rapur and Rajiv Tiwari. Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 40(6):1–21, 2018.

- [25] Mohammad Taghi Shervani-Tabar, Mir Mohammad Etefagh, Saeed Lotfan, and Hamed Safarzadeh. Cavitation intensity monitoring in an axial flow pump based on vibration signals using multi-class support vector machine. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 232(17):3013–3026, 2018.
- [26] Rafik Zouari, Sophie Sieg-Zieba, and Menad Sidahmed. Fault detection system for centrifugal pumps using neural networks and neuro-fuzzy techniques. *Surveillance*, 5:11–13, 2004.
- [27] S Rajakarunakaran, P Venkumar, D Devaraj, and K Surya Prakasa Rao. Artificial neural network approach for fault detection in rotary system. *Applied Soft Computing*, 8(1):740–748, 2008.
- [28] D Siano and MA Panza. Diagnostic method by using vibration analysis for pump fault detection. *Energy Procedia*, 148:10–17, 2018.
- [29] MR Nasiri, MJ Mahjoob, and H Vahid-Alizadeh. Vibration signature analysis for detecting cavitation in centrifugal pumps using neural networks. In *2011 IEEE International Conference on Mechatronics*, pages 632–635. IEEE, 2011.
- [30] Feng Jia, Yaguo Lei, Jing Lin, Xin Zhou, and Na Lu. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72:303–315, 2016.
- [31] Wanlin Zhao, Zili Wang, Chen Lu, Jian Ma, and Lianfeng Li. Fault diagnosis for centrifugal pumps using deep learning and softmax regression. In *2016 12th world congress on intelligent control and automation (WCICA)*, pages 165–169. IEEE, 2016.
- [32] Rajiv Tiwari, DJ Bordoloi, and Aakash Dewangan. Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm. *Measurement*, 173:108676, 2021.
- [33] Primož Potočnik, Borja Olmos, Lučka Vodopivec, Egon Susič, and Edvard Govekar. Condition classification of heating systems valves based on acoustic features and machine learning. *Applied Acoustics*, 174:107736, 2021.
- [34] V Muralidharan and V Sugumaran. A comparative study of naïve bayes classifier and bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Applied Soft Computing*, 12(8):2023–2029, 2012.
- [35] Zengkai Liu, Yonghong Liu, Hongkai Shan, Baoping Cai, and Qing Huang. A fault diagnosis methodology for gear pump based on eemd and bayesian network. *PloS one*, 10(5):e0125703, 2015.
- [36] Shengnan Tang, Yong Zhu, and Shouqi Yuan. Intelligent fault diagnosis of hydraulic piston pump based on deep learning and bayesian optimization. *ISA Transactions*, 2022.
- [37] Chengjin Qin, Jianfeng Tao, Haotian Shi, Dengyu Xiao, Bingchu Li, and Chengliang Liu. A novel chebyshev-wavelet-based approach for accurate and fast prediction of milling stability. *Precision Engineering*, 62:244–255, 2020.
- [38] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377, 2018.
- [39] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [40] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [41] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [42] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [43] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. Real-time patient-specific eeg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2015.
- [44] Roneel V Sharan and Shlomo Berkovsky. Epileptic seizure detection using multi-channel eeg wavelet power spectra and 1-d convolutional neural networks. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 545–548. IEEE, 2020.
- [45] Turker Ince, Serkan Kiranyaz, Levent Eren, Murat Askar, and Moncef Gabbouj. Real-time motor fault detection by 1-d convolutional neural networks. *IEEE Transactions on Industrial Electronics*, 63(11):7067–7075, 2016.

- [46] Wei Zhang, Chuanhao Li, Gaoliang Peng, Yuanhang Chen, and Zhujun Zhang. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, 100:439–453, 2018.
- [47] Feng Jia, Yaguo Lei, Na Lu, and Saibo Xing. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mechanical Systems and Signal Processing*, 110:349–367, 2018.
- [48] Charissa Ann Ronao and Sung-Bae Cho. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications*, 59:235–244, 2016.
- [49] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 4905–4913, 2016.
- [50] Yu Sha, Johannes Faber, Shuiping Gou, Bo Liu, Wei Li, Stefan Schramm, Horst Stoecker, Thomas Steckenreiter, Domagoj Vnucce, Nadine Wetzstein, Andreas Widl, and Kai Zhou. An acoustic signal cavitation detection framework based on xgboost with adaptive selection feature engineering. *Measurement*, 192:110897, 2022.
- [51] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.
- [52] Alexandra L'heureux, Katarina Grolinger, Hany F Elyamany, and Miriam AM Capretz. Machine learning with big data: Challenges and approaches. *Ieee Access*, 5:7776–7797, 2017.
- [53] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019.
- [54] Long-Gang Pang, Kai Zhou, Nan Su, Hannah Petersen, Horst Stöcker, and Xin-Nian Wang. An equation-of-state-meter of quantum chromodynamics transition from deep learning. *Nature Commun.*, 9(1):210, 2018.
- [55] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [56] Léon Bottou. Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pages 421–436. Springer, 2012.
- [57] Léon Bottou. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9):142, 1998.
- [58] Diederik P Kingma and Jimmy Ba. Adam: a method for stochastic optimization (2014). *arXiv preprint arXiv:1412.6980*, 180, 2014.
- [59] Sebastian Bock, Josef Goppold, and Martin Weiß. An improvement of the convergence proof of the adam-optimizer. *arXiv preprint arXiv:1804.10587*, 2018.
- [60] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.
- [61] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.
- [62] AF Lehman and JO Young. Experimental investigations of incipient and desinent cavitation. *Journal of Basic Engineering*, pages 275—284, 1964.
- [63] Iuliu Alexandru Zamfirache, Radu-Emil Precup, Raul-Cristian Roman, and Emil M Petriu. Policy iteration reinforcement learning-based control using a grey wolf optimizer algorithm. *Information Sciences*, 585:162–175, 2022.
- [64] Yi-Fei Pu, Ji-Liu Zhou, Yi Zhang, Ni Zhang, Guo Huang, and Patrick Siarry. Fractional extreme value adaptive training method: fractional steepest descent approach. *IEEE transactions on neural networks and learning systems*, 26(4):653–662, 2013.
- [65] Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.
- [66] Swarnambiga Ayyachamy, Varghese Alex, Mahendra Khened, and Ganapathy Krishnamurthi. Medical image retrieval using resnet-18. In *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, volume 10954, page 1095410. International Society for Optics and Photonics, 2019.
- [67] Muhammad Rashid, Muhammad Attique Khan, Majed Alhaisoni, Shui-Hua Wang, Syed Rameez Naqvi, Amjad Rehman, and Tanzila Saba. A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection. *Sustainability*, 12(12):5037, 2020.

- [68] Fatiha Bouaziz, Hamouche Oulhadj, Daoud Boutana, and Patrick Siarry. Automatic eeg arrhythmias classification scheme based on the conjoint use of the multi-layer perceptron neural network and a new improved metaheuristic approach. *IET Signal Processing*, 13(8):726–735, 2019.
- [69] Adriana Albu, Radu-Emil Precup, and Teodor-Adrian Teban. Results and challenges of artificial neural networks used for decision-making and control in medical applications. *Facta Universitatis, Series: Mechanical Engineering*, 17(3):285–308, 2019.
- [70] Prabhat Kumar Upadhyay and Chetna Nagpal. Wavelet based performance analysis of svm and rbf kernel for classifying stress conditions of sleep eeg. *SCIENCE AND TECHNOLOGY*, 23(3):292–310, 2020.
- [71] Ioan-Daniel Borlea, Radu-Emil Precup, Alexandra-Bianca Borlea, and Daniel Iercan. A unified form of fuzzy c-means and k-means algorithms and its partitional implementation. *Knowledge-Based Systems*, 214:106731, 2021.

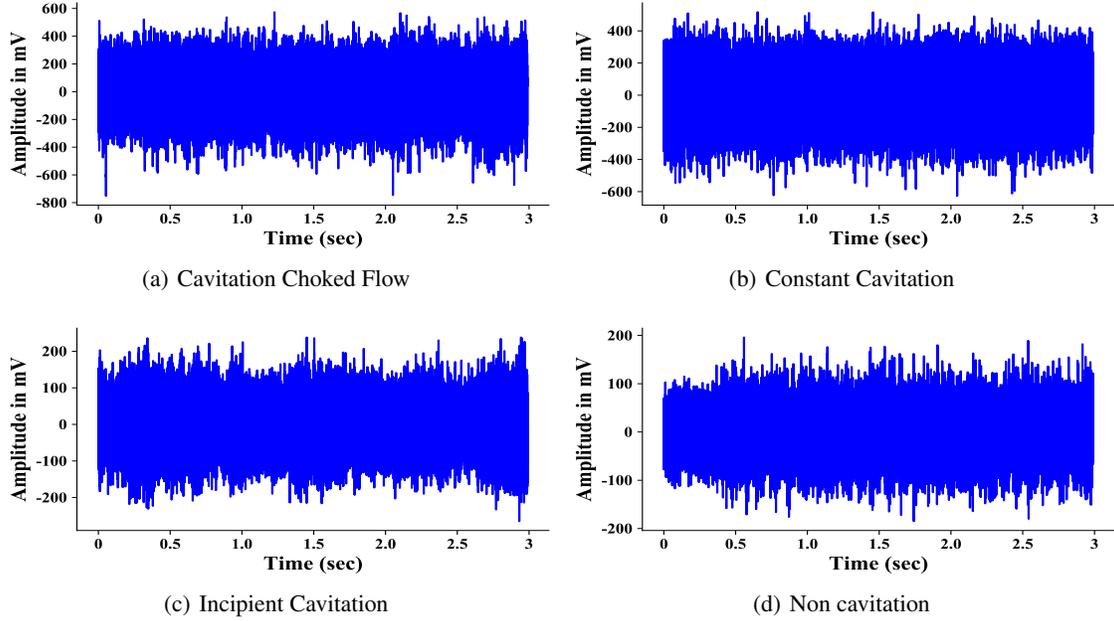


Figure A1: Examples of (a)-(d) for cavitation choked flow, constant cavitation, incipient cavitation and non-cavitation in *Dataset 1*.

Examples of cavitation states

Figures A1, A2 and A3 show examples of different cavitation states for *Dataset 1*, *Dataset 2* and *Dataset 3*, respectively.

Training and validation sets loss curves

In order to quantitatively show that our good results are not due to overfitting of our model. We show the loss curves of our method on three different real-world cavitation datasets (*Dataset 1*, *Dataset 2* and *Dataset 3*), as shown in Figures A4 and A5. As can be seen in Figures A4 and A5, our method does not overfitting on three different real-world cavitation datasets (*Dataset 1*, *Dataset 2* and *Dataset 3*). By monitoring the loss of validation in learning curve, we observe saturation which indicates that our model has converged.

Cavitation Intensity Recognition

The precision, recall and F1-score of 1-D DHRN (our method) and the compared methods in *Dataset 1*, *Dataset 2* and *Dataset 3* are shown in Table A1, Table A2 and A3.

Cavitation Detection

The precision, recall and F1-score of 1-D DHRN (our method) and the compared methods in *Dataset 1*, *Dataset 2* and *Dataset 3* are shown in Table A4, Table A5 and A6.

The accuracy of the 1-D DHRN (our method) and comparison methods during different W_{size} in *Dataset 1*, *Dataset 2* and *Dataset 3* are shown in Figures A6, A7 and A8. The confusion probability matrix of our method for best accuracy in *Dataset 1*, *Dataset 2* and *Dataset 3* are shown in Figures A9, A10 and A11.

Downsampling analysis

The accuracy, precision, recall and F1-scores of 1-D DHRN for cavitation detection during different frequencies of samples in *Dataset 1*, *Dataset 2* and *Dataset 3* are shown in Table A7.

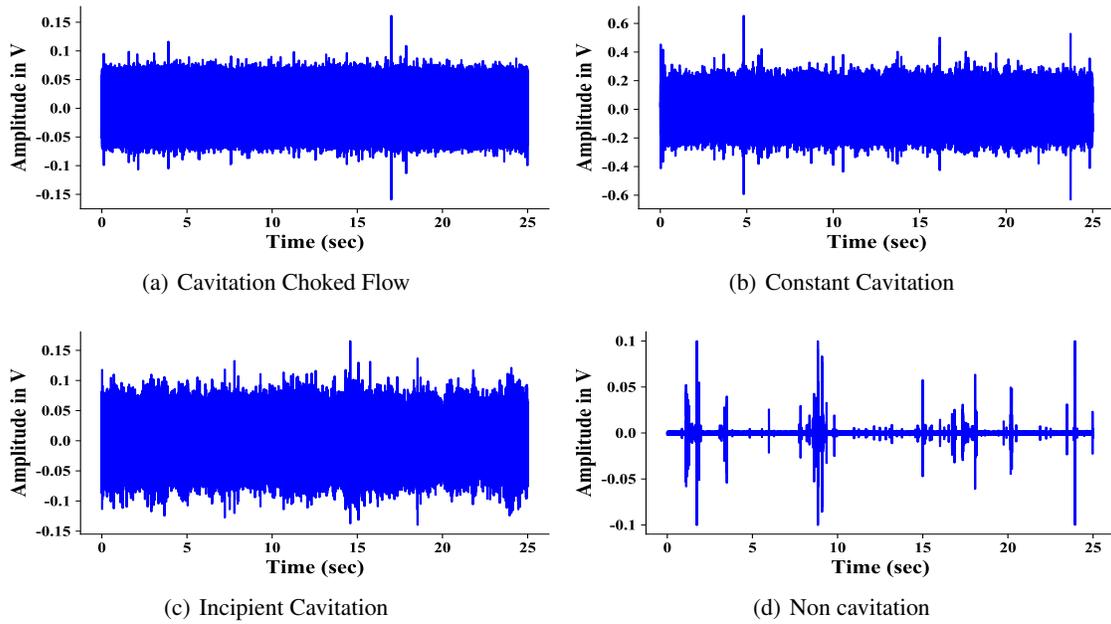


Figure A2: Examples of (a)-(d) for cavitation choked flow, constant cavitation, incipient cavitation and non-cavitation in *Dataset 2*.

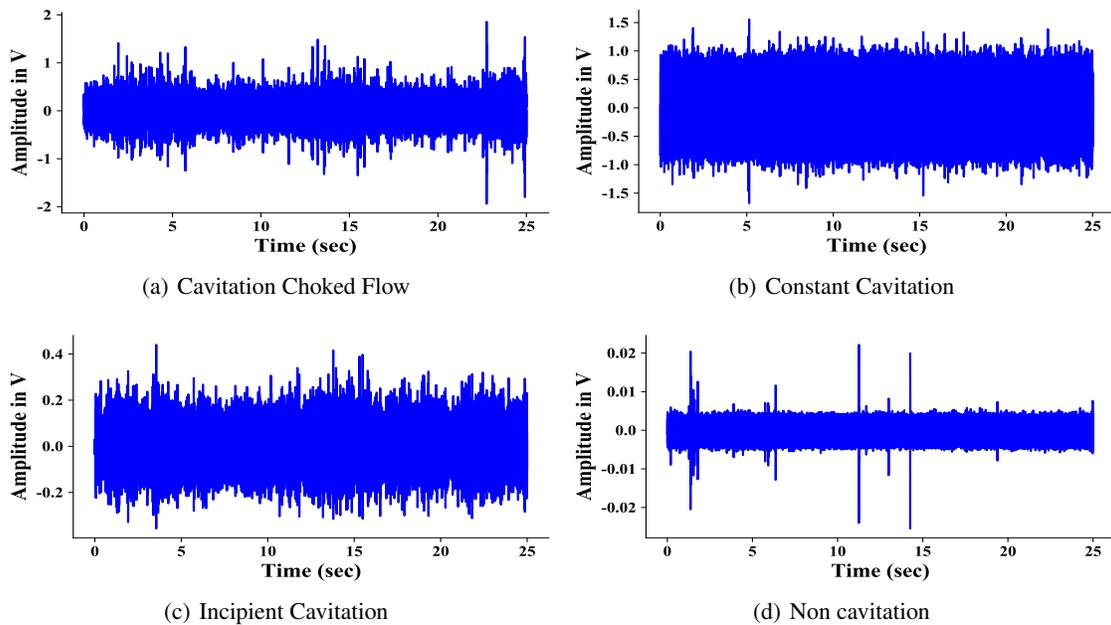


Figure A3: Examples of (a)-(d) for cavitation choked flow, constant cavitation, incipient cavitation and non-cavitation in *Dataset 3*.

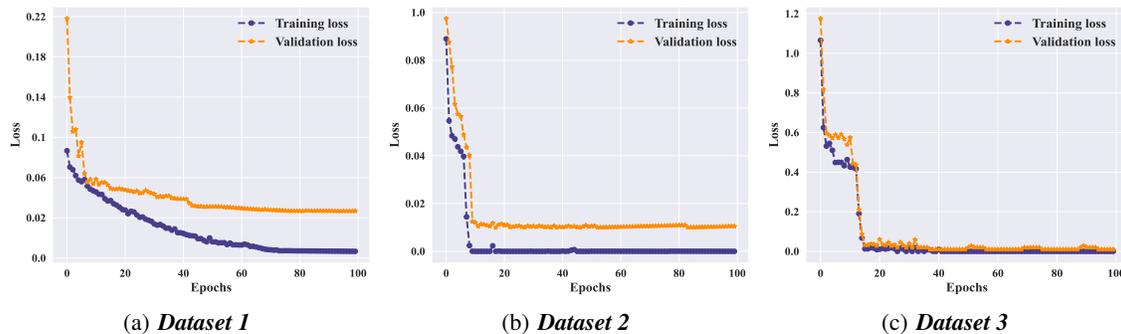


Figure A4: The loss curves of our method for cavitation intensity recognition in three different real-world cavitation datasets (*Dataset 1*, *Dataset 2* and *Dataset 3*).

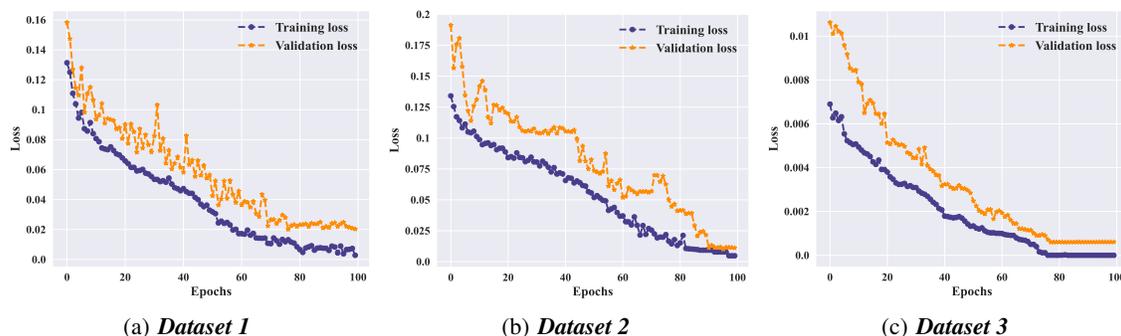


Figure A5: The loss curves of our method for cavitation detection in three different real-world cavitation datasets (*Dataset 1*, *Dataset 2* and *Dataset 3*).

Table A1: Precision, Recall and F1-score results of different W_{size} of the 1-D DHRN and comparison methods for cavitation intensity recognition in *Dataset 1* (%).

Metric	Method	Window Size (W_{size})									
		2334720	1167360	778240	583680	466944	389120	333531	291840	259413	233472
Precision	SVM [21]	40.93	31.90	31.77	31.59	49.15	31.09	31.78	31.71	31.71	31.84
	Decision Tree [14]	46.51	43.84	58.12	44.47	48.81	48.33	48.25	44.79	52.32	45.78
	1-D CNN [25]	64.90	61.34	73.86	84.33	90.42	88.13	81.41	86.22	84.03	82.55
	XGBoost + ASFE [50]	89.21	83.28	85.07	85.59	86.99	80.29	84.03	78.46	83.33	83.38
	1-D Resnet-18*	59.02	56.41	78.24	56.05	74.43	58.93	87.82	77.60	88.27	69.89
	1-D DHRN	59.20	80.96	91.28	91.61	95.72	90.93	92.60	91.09	91.18	93.08
Recall	SVM [21]	48.65	46.81	46.53	45.99	50.38	45.28	46.15	45.92	46.16	46.71
	Decision Tree [14]	43.86	38.78	46.32	47.10	51.38	45.53	47.66	45.66	49.26	44.31
	1-D CNN [25]	65.98	67.83	72.71	72.77	74.20	75.72	77.60	75.55	73.81	73.94
	XGBoost + ASFE [50]	89.58	83.33	85.19	85.59	87.22	81.83	85.62	84.20	85.49	85.42
	1-D Resnet-18*	62.70	65.49	70.16	59.77	73.84	68.09	80.43	75.33	72.85	67.45
	1-D DHRN	65.38	72.40	83.27	84.24	88.00	84.12	79.74	82.30	84.56	83.09
F1-score	SVM [21]	41.01	37.93	37.75	37.39	42.35	36.87	37.64	37.51	37.59	37.28
	Decision Tree [14]	40.96	37.19	43.66	42.03	43.25	41.60	42.86	41.87	47.34	41.47
	1-D CNN [25]	63.52	64.23	71.69	72.81	71.61	74.64	77.37	75.18	73.82	73.28
	XGBoost + ASFE [50]	89.46	82.57	85.06	85.39	87.09	80.52	84.32	81.20	83.72	83.41
	1-D Resnet-18*	60.37	60.06	72.16	56.58	70.92	62.98	83.14	75.70	77.36	66.81
	1-D DHRN	62.05	74.05	85.40	86.32	90.53	85.70	80.25	83.87	86.24	85.46

Table A2: Precision, Recall and F1-score results of different W_{size} of the 1-D DHRN and comparison methods for cavitation intensity recognition in **Dataset 2** (%).

Metric	Method	Window Size (W_{size})				
		2334720	1167360	778240	583680	466944
Precision	SVM [21]	52.37	51.78	49.50	49.76	51.62
	Decision Tree [14]	66.52	63.48	63.20	63.20	60.94
	1-D CNN [25]	85.54	86.51	87.25	86.76	86.56
	XGBoost + ASFE [50]	85.44	88.82	89.65	89.21	87.66
	1-D Resnet-18*	88.26	89.93	91.93	93.53	89.43
	1-D DHRN	89.41	91.19	95.14	92.17	92.03
Recall	SVM [21]	50.66	49.89	48.07	49.06	50.40
	Decision Tree [14]	64.84	62.63	62.06	62.17	60.04
	1-D CNN [25]	85.10	85.13	86.31	85.42	85.31
	XGBoost + ASFE [50]	84.89	88.04	89.18	88.67	87.01
	1-D Resnet-18*	87.38	89.47	86.87	84.06	84.57
	1-D DHRN	89.38	82.83	94.02	89.20	87.43
F1-score	SVM [21]	51.02	50.28	48.24	49.41	50.63
	Decision Tree [14]	64.96	62.47	62.31	62.40	60.28
	1-D CNN [25]	84.84	85.14	86.78	86.08	85.93
	XGBoost + ASFE [50]	84.75	87.82	88.87	88.38	86.82
	1-D Resnet-18*	87.22	89.70	88.85	86.94	86.11
	1-D DHRN	89.39	86.81	94.58	90.18	89.24

Table A3: Precision, Recall and F1-score results of different W_{size} of the 1-D DHRN and comparison methods for cavitation intensity recognition in **Dataset 3** (%).

Metric	Method	Window Size (W_{size})				
		2334720	1167360	778240	583680	466944
Precision	SVM [21]	52.37	51.78	49.50	49.76	51.62
	Decision Tree [14]	66.52	63.48	63.20	63.20	60.94
	1-D CNN [25]	85.54	86.51	87.25	86.76	86.56
	XGBoost + ASFE [50]	85.44	88.82	89.65	89.21	87.66
	1-D Resnet-18*	88.26	89.93	91.93	93.53	89.43
	1-D DHRN	89.41	91.19	95.14	92.17	92.03
Recall	SVM [21]	50.66	49.89	48.07	49.06	50.40
	Decision Tree [14]	64.84	62.63	62.06	62.17	60.04
	1-D CNN [25]	85.10	85.13	86.31	85.42	85.31
	XGBoost + ASFE [50]	84.89	88.04	89.18	88.67	87.01
	1-D Resnet-18*	87.38	89.47	86.87	84.06	84.57
	1-D DHRN	89.38	82.83	94.02	89.20	87.43
F1-score	SVM [21]	51.02	50.28	48.24	49.41	50.63
	Decision Tree [14]	64.96	62.47	62.31	62.40	60.28
	1-D CNN [25]	84.84	85.14	86.78	86.08	85.93
	XGBoost + ASFE [50]	84.75	87.82	88.87	88.38	86.82
	1-D Resnet-18*	87.22	89.70	88.85	86.94	86.11
	1-D DHRN	89.39	86.81	94.58	90.18	89.24

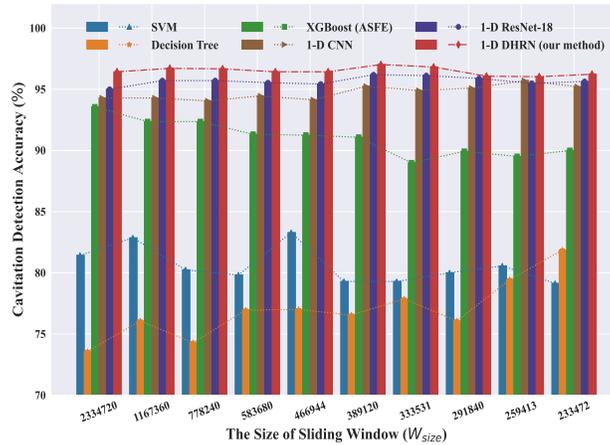
Figure A6: The effect of W_{size} on cavitation detection accuracy under different methods in **Dataset 1**.

Table A4: Precision, Recall and F1-score results of different W_{size} of the 1-D DHRN and comparison methods for cavitation detection in **Dataset 1** (%).

Metric	Method	Window Size (W_{size})									
		2334720	1167360	778240	583680	466944	389120	333531	291840	259413	233472
Precision	SVM [21]	73.61	73.37	71.95	70.82	73.40	69.87	69.62	70.85	72.38	79.15
	Decision Tree [14]	92.59	88.41	88.30	90.51	99.29	89.00	98.56	84.42	97.31	92.61
	1-D CNN [25]	94.64	94.17	93.95	95.14	94.66	95.35	94.96	95.63	95.53	95.84
	XGBoost + ASFE [50]	93.14	92.61	92.48	91.63	91.57	91.83	89.26	90.28	89.83	90.29
	1-D Resnet-18*	94.72	95.63	95.45	95.26	95.59	95.91	95.85	95.63	95.29	95.37
	1-D DHRN	96.15	96.55	96.45	96.16	96.19	96.75	96.58	96.02	96.05	96.09
Recall	SVM [21]	88.33	94.17	88.33	90.00	95.67	90.83	91.67	90.63	88.33	85.83
	Decision Tree [14]	41.67	50.83	46.11	51.67	46.67	51.67	49.05	54.17	53.52	62.67
	1-D CNN [25]	93.75	94.17	93.89	93.75	93.50	94.93	94.61	94.48	95.76	62.67
	XGBoost + ASFE [50]	93.56	92.36	92.36	91.32	91.25	91.09	88.99	89.93	89.51	90.00
	1-D Resnet-18*	95.42	95.63	96.25	95.99	95.08	96.63	96.58	96.12	95.51	95.90
	1-D DHRN	96.88	96.25	97.08	96.77	96.46	97.40	96.07	96.30	96.25	95.90
F1-score	SVM [21]	80.30	82.48	79.30	79.27	83.06	78.99	79.14	79.52	79.57	77.91
	Decision Tree [14]	57.47	64.55	60.58	65.78	63.49	65.38	65.51	65.99	69.06	74.75
	1-D CNN [25]	94.11	94.17	93.92	94.27	93.95	95.12	94.77	94.93	95.64	95.05
	XGBoost + ASFE [50]	93.08	92.39	92.39	91.38	91.31	91.19	89.06	90.00	89.58	90.07
	1-D Resnet-18*	94.94	95.63	95.67	95.49	95.31	96.15	96.08	95.83	95.39	95.56
	1-D DHRN	96.51	96.40	96.76	96.46	96.32	97.07	96.32	96.16	96.15	95.99

Table A5: Precision, Recall and F1-score results of different W_{size} of the 1-D DHRN and comparison methods for cavitation detection in **Dataset 2** (%).

Metric	Method	Window Size (W_{size})				
		2334720	1167360	778240	583680	466944
Precision	SVM [21]	87.88	86.98	85.60	86.19	84.70
	Decision Tree [14]	88.20	82.64	86.21	86.14	82.42
	1-D CNN [25]	94.64	95.35	93.95	95.14	94.66
	XGBoost + ASFE [50]	90.54	90.75	91.53	91.24	90.37
	1-D Resnet-18*	95.85	95.91	95.45	95.26	95.59
	1-D DHRN	96.45	98.32	98.47	96.16	96.19
Recall	SVM [21]	69.71	76.66	68.62	72.00	70.81
	Decision Tree [14]	68.27	69.93	72.15	71.69	72.10
	1-D CNN [25]	93.75	94.93	93.89	93.75	93.50
	XGBoost + ASFE [50]	90.56	90.73	91.33	91.02	90.43
	1-D Resnet-18*	95.48	96.63	96.25	95.99	95.08
	1-D DHRN	94.08	95.40	95.38	96.77	96.46
F1-score	SVM [21]	77.75	89.04	76.17	78.46	77.13
	Decision Tree [14]	76.96	75.76	78.56	78.25	76.92
	1-D CNN [25]	94.11	95.12	93.92	94.27	93.95
	XGBoost + ASFE [50]	90.16	90.32	90.90	90.54	90.04
	1-D Resnet-18*	95.47	96.15	95.67	95.49	95.31
	1-D DHRN	96.76	96.83	96.90	96.46	96.32

Table A6: Precision, Recall and F1-score results of different W_{size} of the 1-D DHRN and comparison methods for cavitation detection in **Dataset 3** (%).

Metric	Method	Window Size (W_{size})				
		2334720	1167360	778240	583680	466944
Precision	SVM [21]	100.00	100.00	100.00	100.00	100.00
	Decision Tree [14]	100.00	100.00	100.00	100.00	100.00
	1-D CNN [25]	100.00	100.00	100.00	100.00	100.00
	XGBoost + ASFE [50]	100.00	100.00	100.00	100.00	100.00
	1-D Resnet-18*	100.00	100.00	100.00	100.00	100.00
	1-D DHRN	100.00	100.00	100.00	100.00	100.00
Recall	SVM [21]	100.00	100.00	100.00	100.00	100.00
	Decision Tree [14]	100.00	100.00	100.00	100.00	100.00
	1-D CNN [25]	100.00	100.00	100.00	100.00	100.00
	XGBoost + ASFE [50]	100.00	100.00	100.00	100.00	100.00
	1-D Resnet-18*	100.00	100.00	100.00	100.00	100.00
	1-D DHRN	100.00	100.00	100.00	100.00	100.00
F1-score	SVM [21]	100.00	100.00	100.00	100.00	100.00
	Decision Tree [14]	100.00	100.00	100.00	100.00	100.00
	1-D CNN [25]	100.00	100.00	100.00	100.00	100.00
	XGBoost + ASFE [50]	100.00	100.00	100.00	100.00	100.00
	1-D Resnet-18*	100.00	100.00	100.00	100.00	100.00
	1-D DHRN	100.00	100.00	100.00	100.00	100.00

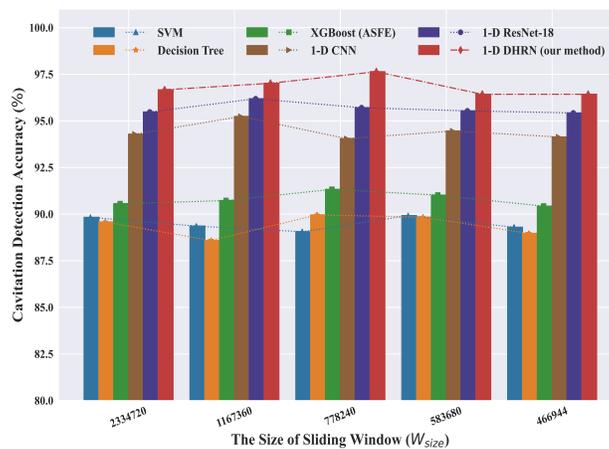


Figure A7: The effect of W_{size} on cavitation detection accuracy under different methods in **Dataset 2**.

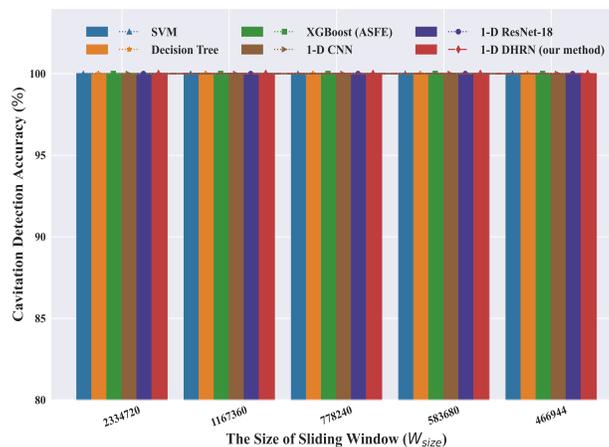


Figure A8: The effect of W_{size} on cavitation detection accuracy under different methods in **Dataset 3**.

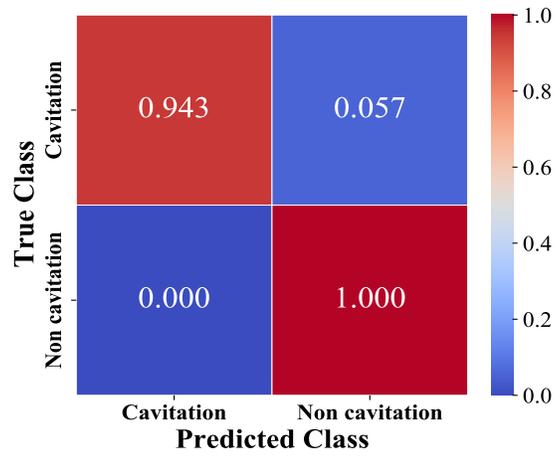


Figure A9: The confusion matrix for the best cavitation detection accuracy of the 1-D DHRN in *Dataset 1*.

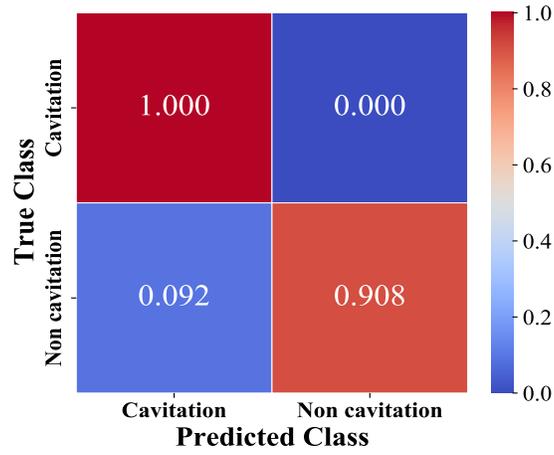


Figure A10: The confusion matrix for the best cavitation detection accuracy of the 1-D DHRN in *Dataset 2*.

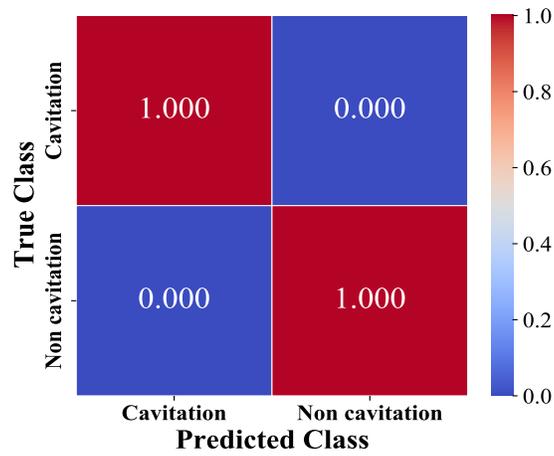


Figure A11: The confusion matrix for the best cavitation detection accuracy of the 1-D DHRN in *Dataset 3*.

Table A7: Accuracy, Precision, Recall and F1-score results of different frequencies of samples FS of the 1-D DHRN for cavitation detection in **Dataset 1**, **Dataset 2** and **Dataset 3** (%).

Data	FS	Metrics			
		Accuracy	Precision	Recall	F1-score
Dataset1	1562500	97.02	96.75	97.40	97.07
	781250	94.43	94.25	95.12	94.39
	390625	93.97	93.71	94.49	93.91
	260416	93.81	93.52	94.12	93.73
	195312	93.65	93.55	94.44	93.61
	48000	93.39	93.29	94.18	93.35
	Dataset2	1562500	97.64	98.47	95.38
781250		97.14	98.36	95.29	96.80
390625		94.57	94.40	94.54	94.47
260416		94.00	93.78	94.63	93.95
195312		93.71	93.45	94.21	93.65
48000		93.57	93.45	94.33	93.53
Dataset3		1562500	100.00	100.00	100.00
	781250	100.00	100.00	100.00	100.00
	390625	100.00	100.00	100.00	100.00
	260416	99.80	99.61	99.87	99.74
	195312	99.62	99.26	99.75	99.50
	48000	99.28	98.61	99.52	99.06

The accuracy, precision, recall and F1-scores of 1-D DHRN for cavitation intensity recognition during different frequencies of samples in **Dataset 1**, **Dataset 2** and **Dataset 3** are shown in Table A8.

Table A8: Accuracy, Precision, Recall and F1-score results of different frequencies of samples FS of the 1-D DHRN for cavitation intensity recognition in *Dataset 1*, *Dataset 2* and *Dataset 3* (%).

Data	FS	Metrics			
		Accuracy	Precision	Recall	F1-score
Dataset1	1562500	93.75	95.72	88.00	90.53
	781250	90.57	91.49	84.10	86.18
	390625	88.26	89.47	81.63	83.03
	260416	86.03	89.38	78.25	81.19
	195312	84.76	88.04	77.43	80.76
	48000	72.83	73.26	69.92	65.59
Dataset2	1562500	94.31	95.14	94.02	94.58
	781250	92.02	92.17	91.03	91.60
	390625	86.53	87.25	86.56	86.39
	260416	84.53	85.57	84.52	84.50
	195312	84.60	85.35	84.52	84.43
	48000	77.65	80.37	77.66	78.09
Dataset3	1562500	100.00	100.00	100.00	100.00
	781250	100.00	100.00	100.00	100.00
	390625	100.00	100.00	100.00	100.00
	260416	99.62	99.26	99.75	99.50
	195312	98.28	98.39	98.28	98.28
	48000	90.17	92.95	90.17	89.78