THREE-CHANNEL DEEP TRANSFER LEARNING MODEL FOR HEALTH DIAGNOSIS OF ROTATING MACHINERY USING DATA REGULARIZATION AND DIFFERENT OVERLAPPING DATA RATIOS

WEI-LUN TSAO [1], MING-JONG TSAI*[1], KANG-MENG LIU[1]

[1] Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, Taiwan
E-MAIL: waylon217@gmail.com, mjtsai@mail.ntust.edu.tw, angus98524@gmail.com

Abstract:

This paper presents a deep transfer learning model for assessing the health diagnosis of rotating machinery, employing a unique three-channel input. The input is created by fusing a time-domain grayscale image with two time-frequency images generated from original vibration signals. The approach was evaluated on two publicly available datasets, CWRU (A) and Mendeley (B), both of which include nine fault types—covering normal operation as well as defects in the inner race, outer race, and ball components. The VGG19-based model with three-channel input was constructed, trained, and assessed using four key metrics: accuracy, recall, confusion matrix, and precision-recall (P-R) curve. In experiments involving transfer learning between the CWRU (A) and Mendeley (B) datasets, the model achieved accuracies of 88.94%, 94.61%, and 96.23% for the corresponding configurations. These results indicate that the three-channel input configuration, especially when paired with an 87.5% overlapping data ratios and regularization, significantly outperforms both alternative setups and the traditional VGG19 transfer learning model.

Keywords:

rotating machine, health diagnosis, fault diagnosis, deep learning, transfer learning, VGG19

1. Introduction

Bearing health is a critical concern in rotating machinery analysis, as bearings are particularly susceptible to failures under harsh conditions. This fact emphasizes the importance of accurate bearing fault diagnosis [1–3]. Diagnostic techniques for bearing faults can be grouped into three main categories: machine learning, deep learning, and transfer learning. For example, W. Zhang et al. [4] proposed a Self-Supervised Joint Learning (SSJL) method that integrates self-supervised and supervised learning strategies to effectively extract fault features from unlabeled data. Their approach converts vibration signals into three-channel images, consisting of a time-domain image, a short-wavelet time-frequency image, and a continuous wavelet transform time-frequency image, which together enhance feature recognition. When tested on two representative motor bearing datasets, the SSJL method achieved higher diagnostic accuracy with minimal reliance on labeled data, outperforming existing alternatives.

C. H. Wang [5] utilized a deep convolutional neural network with single-channel grayscale images as input. Transfer learning between CWRU and Mendeley datasets showed high accuracy, with VGG19 performing exceptionally well. The results demonstrate the model's strong adaptability to unfamiliar data.

M. Ahsan et al. [6] introduced a fault diagnosis technique for industrial rotating machinery that utilizes the Short-Time Fourier Transform (STFT). Their method involves extracting STFT features from both test and reference signals and comparing these features using metrics like Euclidean distance and cosine similarity, enabling precise fault detection. Additionally, the incorporation of variable speed vibration data further improves the method's effectiveness.

R.M. Ahsan et al. [7] introduced a fault diagnosis technique for industrial rotating machinery that utilizes the Short-Time Fourier Transform (STFT). Their method involves extracting STFT features from both test and reference signals and comparing these features using metrics like Euclidean distance and cosine similarity, enabling precise fault detection. Additionally, the incorporation of variable speed vibration data further improves the method's effectiveness.

Zewen Li et al. [8] presented a comprehensive review of CNNs, covering 1-D, 2-D, and multidimensional convolutions. The paper explored CNN history, key models, and state-of-the-art advancements, offering practical insights on hyper-parameter selection and applications across dimensions. It concluded with open challenges and future directions to guide further research in the field.

Junbo Zhou et al. [9] proposed a Whale Gray Wolf Optimization Algorithm (WGWOA) to enhance rolling bearing fault diagnosis by optimizing Variational Mode

Decomposition (VMD) and Support Vector Machine (SVM). Tested on the Case Western Reserve University dataset and laboratory data, the method achieved up to 100% accuracy, outperforming six comparative approaches. The results highlight WGWOA's superior optimization and diagnostic performance. In 2018, bearing vibration systems were set up and data were collected under time-varying rotational speed conditions in which the data set are available for further study [10,11].

Other Health diagnosis of rotating machine using deep transfer learning model with double-channel [12,13] and triple-channel inputs [12,14] are presented under fixed sliding step of 1200 (71% of overlapping data ratios) for data extension without data regularization.

Yi Zhang et al. [15] proposed an innovative approach that employs the Time-Reassigned Multi-Synchro Squeezing Transform (TMSST) to convert vibration signals into time-frequency feature maps for CNN-based fault diagnosis. When evaluated on the Case Western Reserve University dataset, their method achieved a global accuracy of 95.67%, outperforming other techniques and demonstrating robust fault detection performance under a variety of conditions.

This paper presents three-Channel Deep Transfer Learning Model for Health Diagnosis of Rotating Machinery using data regularization and different overlapping data ratios of 87.5% (sliding step of 512) and 75% (sliding step of 1024). A comparison is also given.

2. Technical background

2.1. Short-time Fourier Transform

The Short-Time Fourier Transform (STFT) is a modification of the traditional Fourier Transform. Rather than simply yielding a frequency–magnitude plot, STFT divides the original signal into smaller segments and applies the Fast Fourier Transform (FFT) to each segment individually. This process preserves both the temporal characteristics of the signal and its frequency information, resulting in a more comprehensive depiction. Before carrying out the FFT, a suitable window function is chosen—often the Hamming window [16]. The mathematical window function expression for the Hamming window $\omega[n]$ is given in eq. (1)

$$\omega[n] = \begin{cases} 0.54 + 0.46\,cos\left({}^{2\pi n}/_{(N-1)}\right), n = 0,1,2,\dots N-1 \\ 0, \qquad\qquad otherwise \end{cases} \quad (1)$$

where n is the number of samples in a given frame, and N represents the total number of samples in the frame (window size).

After selecting Hamming window as the window function, the Discrete Short-Time Fourier Transform of x[n] can be obtained by Eq. (2).

$$STFT\{x[n]\}(m,\omega) = X(m,\omega) = \sum_{n=0}^{N-1} x[n]\omega[n-m]e^{-i\omega n} \quad (2)$$

where $\omega[n]$ represents window function, $x[n]$ represents n^{th} discrete signal, m is a discrete number , N is an integer number of data size.

2.2. VGG19 Structure

VGG19 is a Convolutional Neural Network (CNN) developed by the Visual Geometry Group [8][9]. Renowned for its straightforward structure, ease of interpretation, and simplicity in implementation, it has seen widespread use in image classification tasks. Trained on the ImageNet dataset which encompasses over 1,000 categories and around one million images covering a wide range of real-world objects and scenes VGG19 demonstrates strong generalization capabilities when used as a pre-trained model.

The VGG19 architecture consists of 19 layers in total, comprising 16 convolutional layers and three fully connected layers. A distinctive feature of VGG19 is its approach of stacking multiple 3×3 convolutional layers, rather than employing larger convolution kernels. This design choice increases the network's depth and enhances its capacity to learn fine-grained features. The final three fully connected layers handle classification, with the output layer providing the resulting predictions. Given its well-structured architecture and pre-trained weights, which exhibit robust generalization, VGG19 is frequently employed in transfer learning scenarios. Practitioners often customize the model by modifying its input, removing the original fully connected layers, and introducing additional hidden layers or classifiers to meet specific needs.

3. Methodology

In this study, transfer learning is employed as the primary method for examining vibration signals under varying rotational speeds and across different datasets. The aim is to build a model capable of classifying fault types, ultimately achieving fault identification and accelerating the training process through transfer learning. The model is fed preprocessed data from two publicly available internet datasets, and the trained model's performance in distinguishing among fault types is then evaluated. This process is divided into two key stages: data preprocessing and transfer learning.

3.1. Data preprocessing

Data preprocessing involves handling and transforming raw data before applying machine learning or deep learning methods, and it is crucial for achieving successful outcomes. This process includes steps such as noise filtering, feature scaling, dimensionality reduction, and data balancing to enhance the original data's utility. In our study, the proposed data preprocessing strategy is designed to reduce reliance on expert intervention by enhancing the features of vibration signals. To accomplish this, one-dimensional vibration signals from the time domain was converted into two-dimensional grayscale and time-frequency images. These images are then merged into multi-channel images that serve as inputs for both training and testing. Finally, data augmentation techniques are applied to increase the dataset size too.

3.2. Sliding sampling and 2D image conversion

Due to the relatively limited size of the dataset used in this study and the desire to extract more features, the first step in the two-dimensional data conversion process is to employ sliding sampling. The principle of sliding sampling is illustrated in Fig.1, where the original signal (left) undergoes sliding sampling to produce the two sub-data sequences on the right. This method involves extracting subsequences from a signal, whether it is continuous or ordered sequentially in time. In this process, new data samples are generated by sliding a fixed-size window along the original signal and extracting the signal from each window. In this experiment, there are 12,000 data points per second, resulting in a total of 120,000 data points for each category. The window size N set in this study is 4096 points, meaning that each extracted segment consists of 4096 data points. The sliding step is set at 512 or 1024 and overlapped data ratio is 87.5% (512/4096) or 75% (1024/4096).

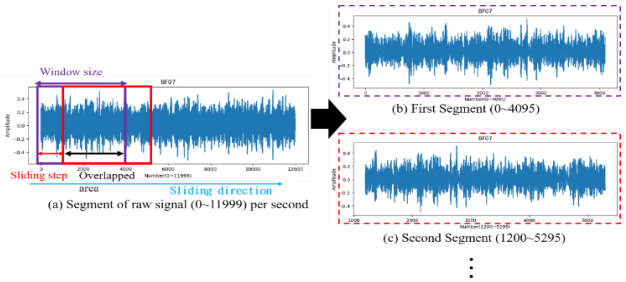

FIGURE 1. Sliding step and overlapped ratio diagram

After sliding steps with different overlapped ratio, more data set can be obtained. The quantity of preprocessed data with window size of 4096 and sliding step of s can be calculated by eq. (3).

$$n_a(i) \approx \frac{120000 - 4096}{s} + 1 \quad ; \quad i = 0,1,2,\dots 9 \qquad (3)$$

where s is the sliding step, $n_a(i)$ is the sub dataset for each faulty category.

The time-domain and time-frequency domain signals are transformed into two-dimensional images using different methods. As shown in Fig.2, for the time-domain data, the 4096 data points are stacked into a 64x64 two-dimensional image, where each column represents 64 data points. In the case of the time-frequency domain data, each small window is transformed into 64x64 two-dimensional image after STFT.

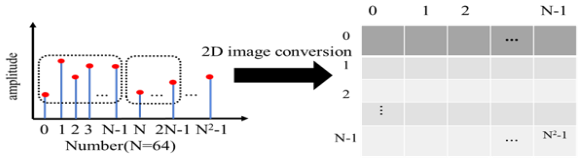

FIGURE 2. 2D time-domain image conversion diagram

Subsequently, the 2D time-domain images and time-frequency images are normalized by eq. (4).

$$X = \text{round}\left(255 \times \frac{x - \min(x)}{max(x) - \min(x)}\right) \qquad (4)$$

In the final step, one two-dimensional time-domain image and two time-frequency images are combined to create a three-channel 64x64 image as shown in Fig.3.

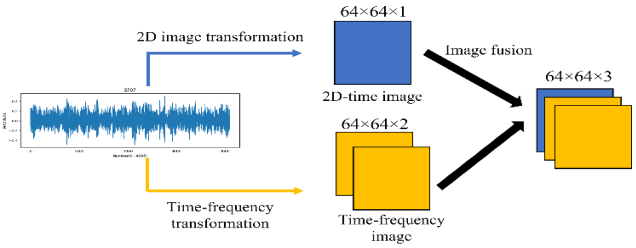

FIGURE 3. Three-channel image composition diagram

3.3. Data augmentation

Data augmentation is a widely used approach in deep learning that involves applying a series of transformations to the original data to create an expanded dataset. This augmented data can help improve training and enhance the robustness of the resulting model. In this study, which centers on vibration signal analysis, Gaussian noise was selected for augmentation to maintain the signal's integrity while boosting the model's noise resistance. Gaussian noise refers to noise that follows a Gaussian distribution in the frequency domain. The equation for Gaussian noise, *f(x)*, is provided in eq. (5).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

where x is the input signal, μ is the mean, and σ is the standard deviation.

3.4. Model establishment and transfer learning

In this study, a model based on VGG19 with three additional hidden layers and one output layer was built [10, 11]. The model establishment process can be divided into two stages: pre-training and transfer learning. In the pre-training stage, the first dataset is split into training, validation, and test sets in proportions of 49%, 21%, and 30%, respectively. After the model is trained, various performance indicators are observed to determine if they meet the criteria.

4. Experimental result and discussion

The study utilizes two publicly available open-source datasets to validate the proposed methodology. The first dataset is Bearing Fault Dataset A from Case Western Reserve University (CWRU) [10], and the second is Bearing Fault Dataset B from Mendeley [11].

In the CWRU dataset, the bearing faults are categorized into Inner Race Faults (IRF), Outer Race Faults (ORF), and Ball Faults (BF). Each of these fault types is further subdivided into three groups based on varying fault depths (0.007, 0.014, and 0.021 inches), alongside a Normal (N) class, culminating in a total of 10 classes. The experimental settings for the rotating machinery in this study are detailed in Table 1.

TABLE 1. Dataset information

Item	Description
Fault type	Normal(N), Inner Race (IR x3), Outer Race (OR x3), Ball (B x3)
Sampling frequency	12 kHz (12,000 samples/sec)
Sampling time	10 sec
Total samples	120,000 points (each fault)
Shaft rotation speed	1797 rpm, 1772 rpm
Fault category	0.007inches (BF07, IRF07, ORF07) 0.014inches (BF14, IRF14, ORF14) 0.021inches (BF21, IRF21, ORF21)
Load	0 hp(A), 1 hp(A1)

To match the same number of categories as CWRU for transfer learning, the Mendeley dataset includes two classes of acceleration inner race faults (I-A), deceleration inner race faults (I-B); two classes of acceleration outer race faults (O-A), deceleration outer race faults (O-B); two classes of acceleration ball faults (B-A), deceleration ball faults (B-B), and one class of acceleration normal operation (H-C-1). There are 10 fault categories in total. Additionally, due to the significant difference in sampling frequency between Mendeley and CWRU, the sampling frequency of the Mendeley dataset was down-sampled from 200,000 Hz to 12,000 Hz.

4.1. Preprocessing result

After preprocessing, the original data is transformed into 64×64 two-dimensional time-domain and time-frequency images, which are mapped to grayscale images ranging from 0 to 255. In this study, a three-channel image method is used, where one 64×64 time-domain image is merged with two 64×64 time-frequency images to create a three-channel image.

In this experiment, both CWRU and Mendeley used a window size of 4096 and a sliding step size of 512 or 1024 for sliding sampling. As a result, the original data is divided into 2270 or 1140 parts through sliding sampling, yielding 2270 or 1140 2D time-domain images. During the conversion to time-frequency images for these 2270 or 1140 parts, each part has a window size of 128 and a moving step of 64, resulting in 2270 or 1140 time-frequency images.

4.2. Transfer learning result

This section presents the transfer learning outcomes using different input configurations across the two datasets. The single-channel input utilizes one time-frequency image, the two-channel input combines one 2D time-domain image with one time-frequency image, and the three-channel input integrates one 2D time-domain image with two time-frequency images. The performance of transfer learning with these various inputs is then illustrated using a confusion matrix and a precision-recall (P-R) curve.

4.3. Transfer learning from A to A1

The transfer learning outcomes under varying load conditions within the same dataset. The model is initially pre-trained using the CWRU dataset recorded at a 0 hp load (denoted as CWRU(A)). After preprocessing, the original dataset is divided into either 2270 or 1140 segments. Gaussian noise augmentation—with a mean of 0 and a standard deviation of 0.5—is then applied, expanding the dataset to

4540 or 2280 segments, covering a total of ten classes. These augmented segments are split into training, validation, and test sets in the ratios of 49%, 21%, and 30%, respectively.

The model's performance is evaluated under various sliding step sizes and with or without the use of regularization. In the training configuration, a batch size of 256 is used, the process runs for 200 epochs, and early stopping is set with a patience of 10 epochs. For the transfer learning phase, the dataset is also taken from CWRU but with a load of 1 hp (referred to as CWRU(A1)). Fig.4. shows the corresponding precision-recall (P-R) curves.

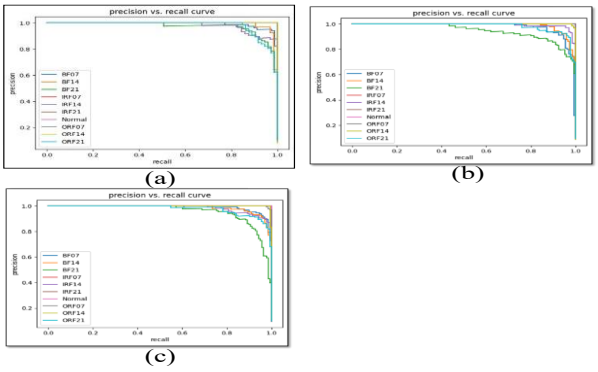

FIGURE 4. A to A1 P-R curve of different sliding step sliding step (a) 1024 (b) 1024 + Regularization (c) 512+ Regularization

4.4. Transfer learning from A to B

This section examines the transfer learning outcomes across different datasets. The pretrained model is built on the CWRU dataset captured at a 0 hp load (CWRU(A)), which consists of 1940 data points that are partitioned into training, validation, and testing sets at roughly 49%, 21%, and 30%, respectively. The training configuration uses a batch size of 256 and runs for 200 epochs with early stopping set to a patience of 10 epochs. For the transfer learning phase using the Mendeley dataset (B), Gaussian noise—with a mean of 0 and a standard deviation of 0.1—is applied for data augmentation. The resulting dataset, containing either 4540 or 2280 data points, is split into training, validation, and testing sets in the same proportions as CWRU(A). The model's performance is evaluated under varying sliding step sizes and both with and without regularization. In this phase, the training setup also employs a batch size of 256 and 200 epochs, but with early stopping configured to a patience of 15 epochs. Fig.5 displays the corresponding precision-recall (P-R) curves.

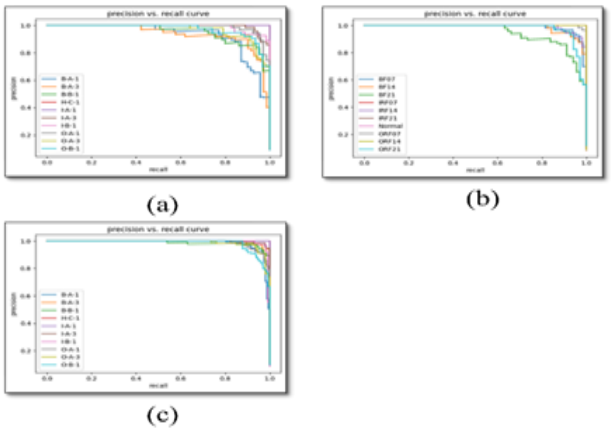

FIGURE 5. A to B P-R curve of different sliding step (a) 1024 (b) 1024 + Regularization (c) 512+ Regularization

TABLE 2. Comparison of Transfer Learning Results with Different Sliding Steps and Regularization Settings

sliding step		512 + Regularization	1024	1024 + Regularization
Transfer learning (A to A_1)	accuracy	**95.81%**	93.74%	94.36%
	training time	95.11s	20.26s	19.69s
Transfer learning (A to B)	accuracy	**96.23%**	88.94%	94.61%
	training time	129.87s	19.92s	60.55s

The comparison of accuracy and training time in this study is shown in Table 2. In transfer learning, the accuracy improvement from CWRU (A) to CWRU (A1) with regularization and learning rate adjustments is minimal, likely because both datasets are from the same source but have different rotational speeds. Even without these adjustments, the accuracy results are satisfactory.

However, from the Training and Validation Loss Curve, it is evident that overfitting occurs when the sliding step size is 1024 without regularization. Therefore, the model performs best when the sliding step size is 512 with regularization. For CWRU(A) to Mendeley(B), regularization and learning rate adjustments slightly increase training time but significantly improve accuracy. Comparing sliding step sizes of 512 and 1024, the smaller step size enhances accuracy but extends training time due to increased data samples.

5. Conclusions

This paper proposes a deep transfer learning model with three-channel input for health diagnosis of rotating machinery. The input consists of a three-channel image, which is composed of a time-domain grayscale image and two time-frequency images derived from the original time-domain signals. Testing was conducted on two similar open-source datasets, CWRU (A) and Mendeley (B), which include nine fault types based on normal conditions, inner race, outer race, and ball defects. Finally, a three-channel VGG19 model was constructed, trained, and validated, with a comparison of model performance using different sliding step sizes and the inclusion or exclusion of regularization.

In the two scenarios of transfer learning, using a three-channel input with a sliding step size of 512 (overlapped data ratios of 87.5%) and incorporating data regularization showed overall better performance in transfers from A to A1 and A to B. This improvement is reflected in increased overall accuracy and longer training time.

Acknowledgements

The authors would like to thank National Science and Technology Council, Taiwan for funding this research grant (NSTC113-2221-E011-082).

References

[1] Y. Ju, X. Tian, H. Liu, and L. Ma, "Fault detection of networked dynamical systems: A survey of trends and techniques," International Journal of Systems Science, vol. 52, no. 16, pp. 3390-3409, 2021.

[2] C. Li, J. V. De Oliveira, M. Cerrada, D. Cabrera, R. V. Sánchez, and G. Zurita, "A systematic review of fuzzy formalisms for bearing fault diagnosis," IEEE Transactions on Fuzzy Systems, vol. 27, no. 7, pp. 1362-1382, 2018.

[3] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-21, 2023.

[4] W. Zhang, D. Chen, and Y. Kong, "Self-supervised joint learning fault diagnosis method based on three-channel vibration images," Sensors, vol. 21, no. 14, p. 4774, 2021.

[5] C. H. Wang, "Deep transfer learning-based convolutional neural network for fault diagnosis of rotating machinery," Master's thesis, Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, 2023.

[6] M. Ahsan and M. M. Salah, "Similarity index of the STFT-based health diagnosis of variable speed rotating machines," Intelligent Systems with Applications, vol. 20, p. 200270, 2023.

[7] R. M. Souza, E. G. S. Nascimento, U. A. Miranda, W. J. D. Silva, and H. A. Lepikson, "Deep learning for diagnosis and classification of faults in industrial rotating machinery," Computers & Industrial Engineering, vol. 153, p. 107060, 2021.

[8] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE transactions on neural networks and learning systems, vol. 33, no. 12, pp. 6999-7019, 2021.

[9] J. Zhou, M. Xiao, Y. Niu, and G. Ji, "Rolling bearing fault diagnosis based on WGWOA-VMD-SVM," Sensors, vol. 22, no. 16, p. 6281, 2022.

[10] C. W. R. U. B. D. Center, "Case western reserve university bearing data center website," 2018.

[11] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," Data in brief, vol. 21, pp. 1745-1749, 2018.

[12] C. M. Liu, "Study of fault diagnosis of rotating machinery based on deep transfer learning model with multi-channel inputs," Master thesis, Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei, 2023.

[13] Kang-Meng Liu, and Ming-Jong Tsai*, H.-L. Wang. "Fault diagnosis of rotating machinery based on deep transfer learning model with two-channel input," 2024 ICNSE (#036), Kyoto, Japan, March 27-29, 2024.

[14] Kang-Meng Liu, Ming-Jong Tsai*, Hsiao-Ling Wang, and Wei-Lun Tsao, (2024, Aug), "Healthy diagnosis of rotating machine using deep transfer learning model with triple-channel inputs," 7th International Conference on Knowledge Innovation and Invention 2024 (2024 ICKII), Aug. 16-18, 2024, Nagoya, Japan. (#K240109)

[15] Y. Zhang, B. Li, and Z. Han, "Enhanced diagnostic method for rolling bearings using time-reassigned multi-synchro squeezing transform," Journal of Measurements in Engineering, vol. 12, No. 4, pp.671-685, 2024. https://doi.org/10.21595/jme.2024.2416

[16] F. J. HARRIS, "Chapter 3 - Multirate FIR Filters for Interpolating and Desampling", Handbook of Digital Signal Processing, Engineering Applications, pp.173-287, 1987. https://doi.org/10.1016/B978-0-08-050780-4.50008-4 , https://www.sciencedirect.com/topics/engineering/hamming-window