

# WaveNet based Autoencoder Model: Vibration Analysis on Centrifugal Pump for Degradation Estimation.

Fan-yun Yen<sup>1</sup> and Ziad Katrib<sup>2</sup>

<sup>1,2</sup>*BKO Services, Houston, Texas, 77002, USA*

*fred@bkoservices.com*

*ziad@bkoservices.com*

## ABSTRACT

Centrifugal pumps are essential equipment in a wide range of industries. Pump maintenance costs make up a considerable amount of their operational life cycle costs. Therefore, estimating pumps health condition is rather critical. Traditional vibration analysis usually done by extracting features through frequency domain and the vibration analysis require critical domain knowledge. This paper presents a novel perspective of utilizing vibrational signal by combining autoencoder and Wavenet, providing a set of embeddings that contain essential characteristics for these high-frequency vibration signal and degradation status without significant insight into the domain.

## 1. INTRODUCTION

Centrifugal pumps are versatile and have been used in a wide range of applications such as agricultural services, wastewater services, and other industrial services. The mechanism behind the pump is converting rotational kinetic energy to induce flow or raise pressure of liquid. Boiler feedwater pump (BFP) is an important piece of equipment in a thermal power generation plant. Generally, the cost of the pump itself only accounts for less than 20% of its life cost and about 30% - 35% of the life cost is allocated to pump operation and maintenance. Understanding the degradation status of the pumping system provides important insight into maintenance schedule and can help reduce maintenance costs. Traditionally, engineers evaluate the performance and/or find faults by observing the vibrational signal, specifically, looking at the power spectrum density of the vibrational signal measured at different locations. However, such vibration analysis requires substantial domain knowledge and experience to accommodate all the variables

caused by various conditions like different pump OEM models, sizes in different plants, units, and facilities. Often Vibration Analyst must bin the vibration signal according to a predetermined frequency bins, thereby potentially removing useful markers about vibration health.

Long short-term memory (LSTM) and autoencoders have been used to process sequential data. However, most of these applications were focused on relatively lower sampling rate, like time series data and making prediction from it (Gensler et al., 2017) or require prior feature extracting steps such as Short Time Fourier Transformation (STFT) (Marchi et al., 2015). Those methods suffer from not leaning the high frequency characteristics and can be rather time consuming, whereas our proposed method is able to ingest longer sequential data while largely reducing computational time without applying any preprocessing steps to the raw data.

This paper presents a novel way of conducting vibration analysis on pump bearings to determine the degradation trend, without requiring expert domain knowledge, by extracting useful information using a WaveNet (Aaron van den Oord et al., 2016) based autoencoder (Hinton & Zemel, 1994) on the historical vibration data. WaveNet is known for processing raw audio data and building generative models. Unlike recurrent neural network (RNN), WaveNet is capable of handling much longer sequential data, which is very suitable for high frequency signals like sound and vibration signals. The autoencoder model extract essential information for reconstructing the input data. The embeddings from the autoencoders can represent the characteristics of the input data. By combining the two techniques, we were able to compress the vibration data 12x and extract the embeddings from raw vibration data and use them to estimate the degradation status of pumps. We pre-selected a collection of vibration data from pumps under “normal” condition. The degradation trend is estimated by computing the distance of the embeddings from “normal” data to new inputs. Such model provides additional information on pump condition vis-a-vis vibration data with no prior domain knowledge. This technique can assist decision making and reduce costs from improper operation and maintenance.

---

Fan-yun Yen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 2. METHODS AND EXPERIMENT

### 2.1. Vibration Data Preprocessing

The model was trained on 216 measurements from a group of six pumps and each measurement was taken at 1600 Hz sampling rate with length of 4096. Figure 1 shows an example of one measured vibration data in three directions (channels). The vibration data was taken by a handheld device on the pump surface in three directions (X,Y,Z). Therefore, for a single vibration measurement, the size is 4096 x 3.

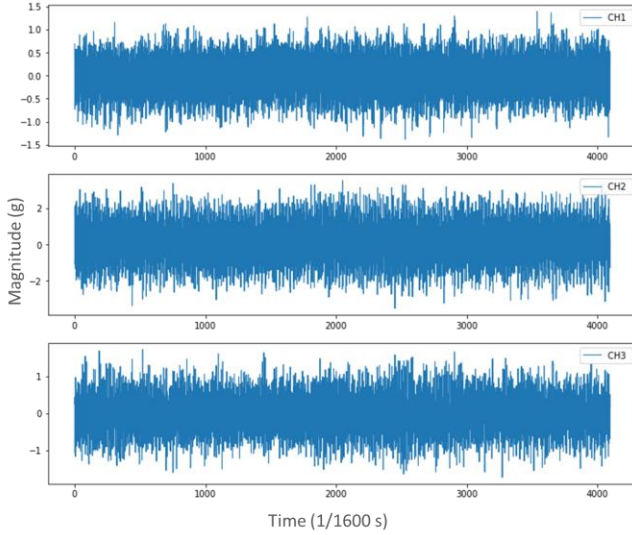


Figure 1. Raw vibration data measured from the surface of a pump.

A standard scaler was applied to each feature before further analysis. A standard score,  $z$ , for a given value  $x$  was calculated as:

$$z = (x - u) / s, \quad (1)$$

where  $u$  is the mean of training measurement value for a feature and  $s$  is the standard deviation of the training measurement value for that feature.

Training samples are a list of sequential data consisting of vibration measurement. Figure 2 shows an illustration of how we took a sliding window size of 512 and stride of 128 for each measurement. Each box shows one window taken as a training sample. Thus, size of one training sample is 512 x 3. The sliding window is applied on each measurement individually to make sure we have a continuous sequence within each one of the training samples (no training sample contains signal from two different measurements).

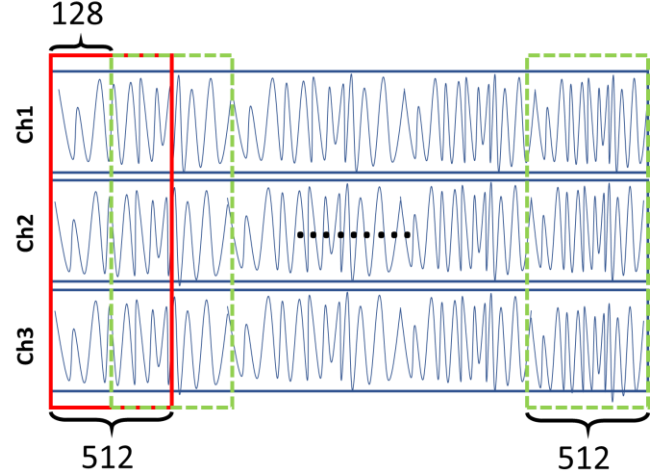


Figure 2. Illustration of sliding window of size 512 applying on one measurement of vibration signal.

### 2.2. WaveNet

WaveNet was first introduced by Google's DeepMind in 2016. In contrast to RNN and other attention models, WaveNet was designed to accommodate long sequential data sets like sound waves and vibration data.

The core idea of WaveNet is taking conditional probability of raw audio signal at each sample point and join it with the previous sample points. The joint probability of a waveform  $x = \{x_1, x_2 \dots x_n\}$  is formulated as:

$$p(x) = \prod_{t=1}^T p(x_t | x_1, x_2, \dots, x_{t-1}) \quad (2)$$

Therefore, each data point in a sequential sample carries information from the previous time steps within the sequence. Figure 3 visualizes one block of 5 layers WaveNet with three hidden layers, an input layer and an output layer. The dilation rate increases in power of two for each addition layer.

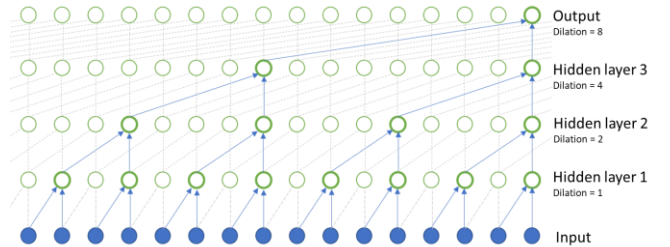


Figure 3. Visualization of dilated casual CNN.

The activation function used in WaveNet is a gated activation unit that was introduced in the Pixel recurrent neural networks (Aaron Van Den Oord et al., 2016). The gated unit is formulated as

$$z = \tanh(W_{f,k} * x) \odot \sigma(W_{g,k} * x), \quad (3)$$

where  $*$  and  $\odot$  are convolution operator and element wise multiplication operator, respectively,  $\tanh(\cdot)$  and  $\sigma(\cdot)$  are the tanh activation function and sigmoid activation function,  $f$ ,  $g$ , and  $k$  denote the filter, gate, and layer index, respectively.

Here we implemented the skip connection and residual that were used in the original WaveNet to accelerate the training process. The idea of skipping connection is to avoid the vanishing gradient problem, so the gradient would not be too small and the weights could be updated relatively fast. For further detail, please refer to the original WaveNet publication (Aaron van den Oord et al., 2016).

### 2.3. Autoencoder

Autoencoders are sometimes considered unsupervised or semi supervised learning. The focus of an autoencoder is force the model to compress the input data and learn a representation (embedding) out of the input data while eliminating the noise and unwanted signals to accomplish dimensionality reduction. Figure 4 shows the basic architecture of an autoencoder. Autoencoders generally have series of fully connected layers as encoder followed by a bottleneck (embedding) layer then reconstruct the input data through a decoder, which mirrors the encoder structure.

The idea of this structure is to force the model to compress input data and to only retain the essential information (embedding layer) for reconstructing the input data. The target is to minimize the difference between the raw input data and the reconstructed signal. Here we used mean squared error (MSE) between labels and predictions as loss function, which can be formulated as

$$L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4)$$

where  $L(y, \hat{y})$  is the loss function given labels  $y$  and prediction values  $\hat{y}$ ,  $n$  is the number of labels.

Unlike principle component analysis (PCA) which finds the hyperplane underlying the input data to reduce dimensionality, autoencoder introduces nonlinearity to the model and learns the representation of the input data.

To add to the robustness of the proposed model, random Gaussian noise was added to the input training data so the model will predict the original signal from corrupted input (Vincent et al., 2008).

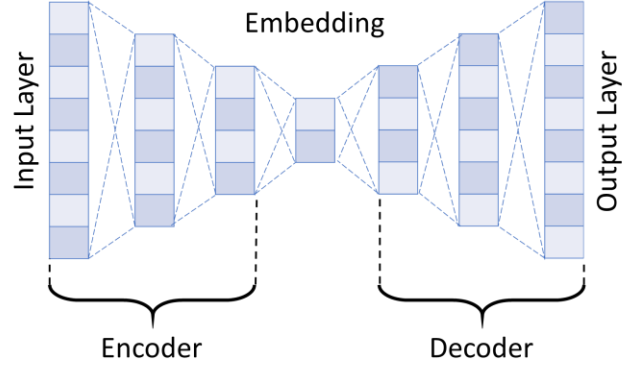


Figure 4. Autoencoder architecture.

### 2.4. Model Architecture

This paper proposes a fusion neural network that combines autoencoder and WaveNet to extract the essential information out of a set of sequential data. Figure 5 shows the architecture of the proposed model. The model is divided into three blocks: encoder segment, embedding layer, and the decoder.

The encoder is to extract the information from input data by compressing the data to a single vector of embeddings. Input sequential data in size of  $512 \times 3$  first go through a  $1 \times 1$  1D CNN layer with 64 filters followed by three blocks of 8 layers WaveNet to process the data while maintaining the sequential order of the data points. After the WaveNet, two sets of compression steps composed by one  $1 \times 1$  1D CNN and a max pooling layer with a pooling rate = 2. The input data is compressed 12 folds from  $512 \times 3$  into the embeddings, a single vector with 128 elements. The decoder is a reverse process to the encoder except instead of max pooling layers, the decoder combines the 1D CNN layer with up-sampling layer to reconstruct the data from embeddings back to the same size as the input data.

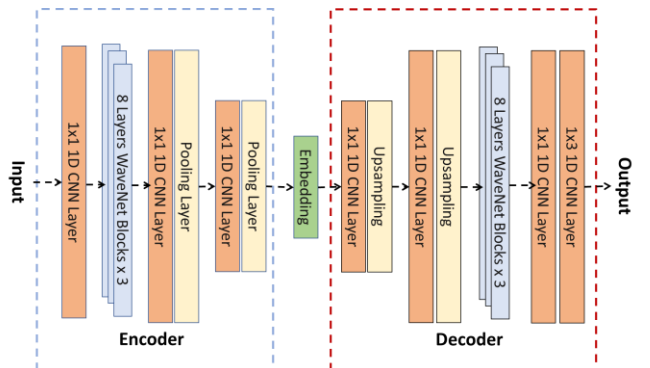


Figure 5. Schema of proposed model.

### 3. RESULTS

#### 3.1. Training

The model was trained on a of 216 vibration measurements from six pumps, with A total number of 6264 training samples. Adam optimizer was used to reduce the loss function defined in section 2.3. Training samples were shuffled, and 20 percent of training samples were set as validation samples. After 100 epochs of training with batch size of 32, the MSE reduced to  $\sim 0.16$  and validation loss was about  $\sim 0.2$ . The proposed model was developed using TensorFlow 2.x (Abadi et al., 2016) and WaveNet implementation in this proposed model was built upon the WaveNet example in (Géron, 2017).

#### 3.2. Signal Reconstruction

Input vibration signal is compressed into embedding representation via encoder then the decoder reconstructs the vibration signal from the embeddings. Figure 6 shows the input vibration data in three directions of a pump and the signal reconstructed by the decoder. The proposed autoencoder model reconstructed the vibration signal from all three channels well, as shown in Figure 6.

Pump vibration data contain critical information regarding pump mechanical and flow conditions. The vibration on pump surface is influenced by the mechanical components and flow conditions of a pump. A main way to investigate a pump status through vibration signal is analyzing the pump vibration at different multiple of the pump rotational running speed.

Figure 7 presents the power spectral density (PSD) of the input vibration signal as well as the reconstructed vibration signal. In the plot, we can clearly see peaks all over the spectrum. These peaks are the multiples of pump rotational speed and these multiples are usually referred as 1x, 2x, 3x...etc. As shown in the Figure 7, the reconstructed signal in all three channels are able to reconstruct the multiples in the frequency domain.

We also observed that the reconstructed signals behaved like projection operation. As seen in the PSD, the reconstructed signal retained the peak amplitudes and “projected” them to a flatter baseline. This frequency domain normalization suppressed the frequency bands that have higher amplitude and brought up the bands that are relatively low on power. This action enables a more general comparison between the multiples of running speed. That said the embeddings contains the essential information of the multiples of rotational running speed which can represent the pump operation condition at the time of measurement.

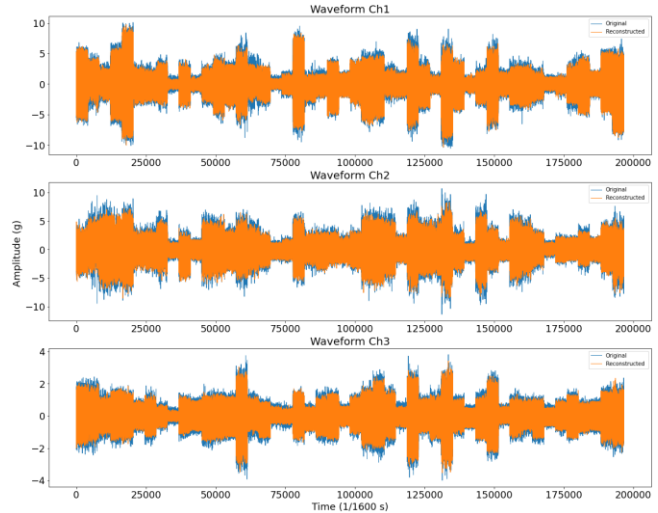


Figure 6. Input vibration signal measured at three directions and the reconstructed signal.

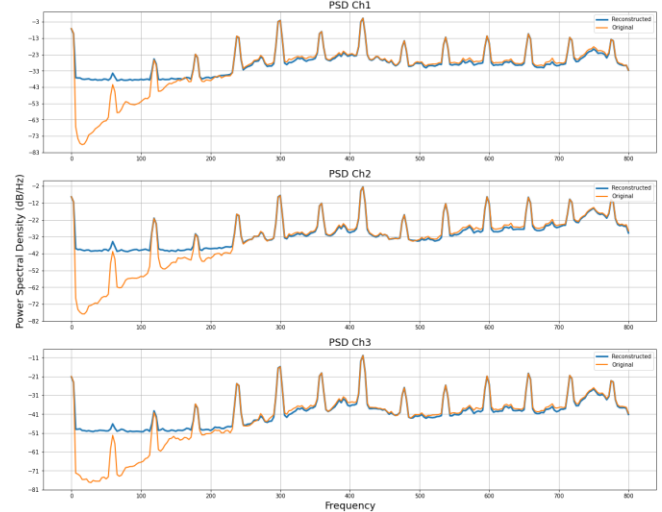


Figure 7. Power Spectral Density of input signals and reconstructed signals.

#### 3.3. Degradation Estimation

The proposed model reconstructed the vibration signal and normalized the signal in frequency domain which made the vibration signal comparable across different measurements. The frequency normalized signal was reconstructed from the embeddings that contained the essential characteristic of the input vibration signal and can represent the operational condition of the pump.

This paper demonstrates a simple comparison of pump vibration signal by calculating the Euclidean distance between the embedding of input vibration signal and the reference embedding. The reference embedding was selected manually for each pump based on maintenance records and vibration analysis report. The goal is to select the



measurement that is considered normal operating condition of the pump. Therefore, the distance between the embeddings of input signals and the reference signals show how far the measurement is from the normal operating condition. This deviation is considered a degradation indicator.

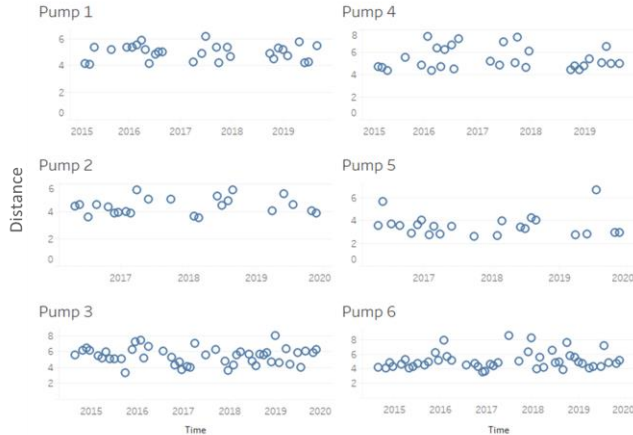


Figure 8. Euclidean distance of vibration signal embeddings to the reference measurement embedding.

Figure 8 shows the embedding distances of measurements at different time points with respect to the reference measurements for each pump. By calculating the distance between new embeddings and normal embedding, we can estimate the degradation condition of a pump given new vibration signal inputs. Here we are looking for the pump degradation, meaning when a pump gradually deteriorates to a certain point where it would cause serious impact to the pump operation. For example, as we can see in pump 6, the change in embedding distance has a peak around mid-2016. This could indicate the pump was not running under normal operating condition at that point in time and the embedding distances of its previous measurements showed that they gradually deviated from normal operating condition as the distance increased.

Please note that the distance values here are not an absolute measurement, meaning the range and/or amplitude can be different from pump to pump. Thus, the abnormal example shown in this paper was from a manual inspection. We were able to align several maintenance events with the degradation patterns from the distance measurement, large distance values corresponded to maintenance records. However, due to the lack of historical maintenance records this could not be confirmed for all degradation points. It is possible to quantify the “degree of degradation” with properly labeled data, however, we do not have enough maintenance record for every data point, and this could be accomplished in the future.

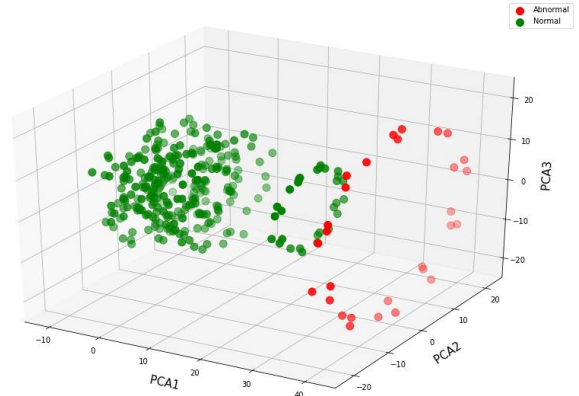


Figure 9. First three principle components from the embeddings of sample input data.

To get a better intuitive view and understanding on what happens in the embedding level, we visualized the first three principle components of embeddings from input samples. Figure 9 shows the first three principle components of the large distance peak at mid-2016 (considered as abnormal in red) from pump 6 and the data points (consider as normal in green) before that peak. We can see clear separation between the cluster of normal data and abnormal data.

#### 4. DISCUSSION

Centrifugal pumps serve an important part in various industries. Optimization of pump maintenance schedule has been a popular topic in wide range of fields due to the high cost of pump maintenance. Estimation of pump health condition and its degradation play an important role in reducing maintenance cost. This paper presented a novel way of conducting pump condition evaluation by extracting the representation from vibration signals and making the measurements comparable across different time points.

The model integrated autoencoder with WaveNet, which was designed for processing raw audio signals, to extract the representation of vibration signal via compressing input signal into a much smaller vector referred to as embedding. One advantage of using WaveNet is that the model is capable of ingesting much longer sequential data compared to other attention based neural networks like (RNNs). This is critical because processing a longer sequential data allows the embedding vector to encapsulate richer information of each sample point. Another benefit, WaveNet is based on casual convolution and no recurrent connection; Thus, it is much faster than RNNs. One important contribution of this model is that it bypasses the prior domain knowledge for pump vibration analysis and allows analysis of operating condition directly. Lastly, the embedding extracted from vibration signal makes the vibration signal comparable across measurements.

This evaluation process we presented here is a preliminary estimation that can be improved in different ways. For example, the process can be made more robust by using different matrices to estimate the distances between measurements. Furthermore, with sufficient properly labeled data through maintenance records, the level of degradation could be quantified for better estimations.

The proposed model can be valuable for extracting features from long sequential data that have higher sampling rate like audio or vibration data. The extracted representation (embedding) makes the input signal comparable across different measurements. The model also has great potential to be utilized in larger machine learning models. For instance, it can serve as feature extractor in a classification model. The degradation estimation could be extended to evaluate the remaining useful life for an industrial equipment.

## REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016*.
- Gensler, A., Henze, J., Sick, B., & Raabe, N. (2017). Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks. *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, 2858–2865.  
<https://doi.org/10.1109/SMC.2016.7844673>
- Géron, A. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems. In *O'Reilly Media*.
- Hinton, G. E., & Zemel, R. S. (1994). Autoencoders, Minimum Description Length and Helmholtz free Energy. *Advances in Neural Information Processing Systems*. <https://doi.org/10.1021/jp906511z>
- Marchi, E., Vesperini, F., Eyben, F., Squartini, S., & Schuller, B. (2015). A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2015-Augus(289021)*, 1996–2000.  
<https://doi.org/10.1109/ICASSP.2015.7178320>
- Oord, Aaron van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio*. September. <http://arxiv.org/abs/1609.03499>
- Van Den Oord, Aäron, Kalchbrenner, N., & Kavukcuoglu, K. (2016). Pixel recurrent neural networks. *33rd International Conference on Machine Learning, ICML 2016*, 4, 2611–2620.
- Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*. <https://doi.org/10.1145/1390156.1390294>

## BIOGRAPHIES

**Fan-yun Yen** obtained his BS in Biomedical Engineering from the Ming Chuan University, Taipei, Taiwan in 2011 and holds a PhD in Biomedical Engineering from University of Houston, Houston, Texas, United States, 2019. His researches were focused on 3D modeling of chromosome territories and analysis of their organization within nuclei. During his years in the University of Houston, he has experiences in processing, analyzing, and modeling various types of data such as medical images, electrical biosignals, and genetic sequences. He joined BKO Services as a data scientist in 2020 and has been working on applying machine learning algorithms, statistic principle, and data analytics methods to solve industrial problems.

**Ziad Katrib** holds a BE in Mechanical Engineering from the American University of Beirut, an MS in Data Science from Texas A&M Statistic Department, and an MBA from University of Texas at Tyler. He has extensive experience in using technology to drive value within the energy and manufacturing industries. During the last 14 years, he focused on solving operational and technical problems through building simulation products and Machine Learning based platforms. As a co-founder and Data Science practice lead at BKO Services, he leads a team of engineers, data scientists, and software developers. The team has deployed physics and Statistics based Models (Digital Twins) for equipment diagnostics, forecasting and optimizing industrial equipment operations. His work focused not only on industrial machine learning research and application but also on deploying those methods into software stacks and mobile applications