

FaultNet: A Deep Convolutional Neural Network for bearing fault classification

Rishikesh Magar, Lalit Ghule, Junhan Li, Yang Zhao and Amir Barati Farimani

Abstract—The increased presence of advanced sensors on the production floors has led to the collection of datasets that can provide significant insights into machine health. An important and reliable indicator of machine health, vibration signal data can provide us a greater understanding of different faults occurring in mechanical systems. In this work, we analyze vibration signal data of mechanical systems with bearings by combining different signal processing and coupling them machine learning techniques to classify different types of bearing faults. We also highlight the importance of using different signal processing methods and analyze their effect on accuracy for bearing fault detection. Apart from the traditional machine learning algorithms we also propose a convolutional neural network FaultNet which can effectively determine the type of bearing fault with a high degree of accuracy. The distinguishing factor of this work is the idea of channels proposed to extract more information from the signal, we have stacked the ‘Mean’ and ‘Median’ channels to raw signal to extract more useful features to classify the signals with greater accuracy.

Index Terms— Convolutional Neural Network, FaultNet, Featurization, Machine Learning

I. INTRODUCTION

With the advent of the 4th industrial revolution, industries across the globe are using artificial intelligence (AI) to improve their processes and increase efficiency to meet the ever-rising customer demands. In this rapidly changing landscape of technology, organizations across the globe, have increased the presence of sensors on the production floor with a motivation of gathering data that can give them valuable insights about their processes[1]. This sensory data contains rich information about the machine and its effective analyses using AI can contribute significantly towards preventive maintenance, quality control and increased process efficiency[2]. Realizing these obvious benefits of cost-effective tools like AI, organizations across the world are turning towards smarter technologies.

Driven by the keenness from the industry to embrace advanced digital technologies, many researchers are using different signal processing methods and coupling them machine learning algorithms to address some of the complicated research problems. For example, Thomazella et al. used digital signal processing techniques such as short-time Fourier transform (STFT) and ratio of power (ROP) to extract features from vibrations signals captured to monitor chatter phenomenon during grinding process [3]. In another paper on signal processing on vibration data, Zoltan et al., have demonstrated

that signal processing techniques such as Discrete Wavelet and Wavelet Packet Transform are effective in extracting features from frequency domain for fault detection [4]. Their simulated results proved the techniques are even capable of predicting abnormalities exploring long-term tendencies of the detected signals. Signal processing techniques have been performed on acoustic signals as well. Adam et al. have proposed a signal processing technique named MSAF-RATIO-24-MULTIEXPANDED-FILTER-8. This technique is used on acoustic signals captured from electric motors and extracted features are used to classify motor faults [5].

In this work, we will be focusing on the vibration signal data and analyze the different methods of fault detection in bearings using vibration signals. Vibration data has many applications in the areas of structural weakness or looseness, rotating component looseness and validating the presence of resonance. The optimal monitoring of vibration signal can thus help the analyzing the machine performance more effectively, improve the efficiency and more importantly give us insights about machine health. With bearing failure being one of the major contributors to the downtime of industrial machines, it is very important to address this problem with high reliability and reduce the break-down of machines [6][7][8]. As determined by Zhang et al., for rotating machine health monitoring, vibration signal is very important as it contains rich information regarding machine health [9]. Therefore, the analysis of vibration data may help us in detection and prevention of faults in bearing. In their study, Samanta B et al. used time domain statistical features extracted from vibrations signals to classify faults using artificial neural network[11]. This study was one of the earliest attempts to utilize the capabilities of deep learning for bearing fault detection using vibration signal data. Apart from signal featurization, some researchers have also used wavelet decomposition to extract relevant information from the signal. In a study conducted on wavelet transformation of vibrations signals for fault diagnosis, Sun W et al., use a combination of discrete wavelet transforms and envelope analysis using which they extract the characteristic spectrum of rolling bearing vibration data. Subsequently, a spectrum cross-correlation coefficient is then applied to identify different operating conditions of rolling bearings[10]. Based on this coefficient, different vibration signals are then classified.

Building upon the previous works, we use different signal featurization methods to extract 14 features from the raw vibration signals to classify bearing faults using machine learning and deep learning approaches. In order to comprehensively analyze the signal data, we also implemented wavelet decomposition on the raw signal and couple it with machine learning approaches to evaluate its performance for

bearing fault classification. With the motivation of developing a generalized model, we evaluate our machine learning and deep learning approaches on two major publicly available datasets for bearing fault classification.

The first dataset that we analyze has been developed by Case Western Reserve University (CWRU) bearing center[12]. The dataset from the CWRU bearing center will be referred as CWRU dataset throughout the paper. The CWRU dataset is one of the important datasets in this research area and has been widely used by researchers to benchmark the performance of their models. In their study, Smith et al., have proposed a benchmark for the CWRU dataset using three different techniques. They have carefully analysed the different ball faults and compared the signal data amongst the faults. Thus, articulating the difference among signals data when different types of fault occur. However, they do not use the signal featurization techniques that we have employed and don't use any deep learning models[13]. Many researchers recently have used different deep learning models on the CWRU bearing dataset. In their review paper, Zhang et al., have compiled a comprehensive list of different methods used by researchers working in this area. Based on their review on different methods, it is evident that best performing deep learning models have accuracies in the range of (97%-99%)[14]. Thus, proving the compatibility and effectiveness of deep learning methods to bearing fault diagnostics problem.

Most of deep learning architecture used for bearing fault diagnosis is based on Convolutional Neural Network (CNN). Guo et al. propose hierarchical adaptive deep convolution network for bearing fault size prediction. In their paper they convert the signal data into 32x32 array and use CNNs to accomplish the task. However, their work does not use other information available from signal data like skewness, kurtosis, impulse factor, rms value[15]. In order to take advantage of the signal data available, another methodology used Pan et al., employs 1D CNN with LSTM, in their paper one-dimensional CNN and LSTM are combined into one unified structure by using the CNNs output as input to the LSTM to identify the bearing fault types[16]. They also compare the usage of nine different featurization techniques and using them with different traditional machine learning algorithms. However, Pan et al., do not use stacked median and mean channel in their work and use a more computationally heavy framework by combining the CNN and LSTM approaches. Most of the state-of-the-art works report an accuracy of more than 98% in bearing fault detection. Guo et al., in their paper used Stacked denoising Autoencoders have obtained an accuracy of 99.83%. However, they separate the data as per size of the fault and then make predictions and have only 4 classes in their predictions[17]. As the dataset created by CWRU bearing center is able to mimic the actual operating conditions the dataset contains some noisy signals as is expected in actual environment. Therefore, the use of SDAE has been made particularly by researchers to make their predictions more resilient to the noise in the dataset[18][19]. Another approach used by Li et al., combines the convolutional neural network and Dempster-Shafer theory-based evidence fusion. In their work, they demonstrate adaptability to different loads and report an accuracy 98.92% [20]. Another CNN based approach named LiftingNet by Pan et al., proposes split, predict

and update blocks that are accurately able predict the bearing faults and is adaptable to different motor speeds and loads. However, their approach is not able to gauge the size of the fault[21]. Our FaultNet can not only predict the type of fault but also the size of the fault based on the input signal.

The second dataset that we have considered in this work is the Paderborn University Data Center bearing dataset[22]. From here onwards, the dataset from Paderborn University will be referred as Paderborn dataset throughout the paper. The dataset has vibration as well as motor current signal captured on the test-rig. In the paper proposing the dataset, to extract the features, Fast Fourier Transform (FFT) and power spectral density (PSD) are performed on vibration and motor current signal. After feature extraction and feature selection, 18 features emerge for motor current signals and 15 features are extracted for the vibration signal data[23]. Using conventional machine learning approaches, Karatzinis et al, achieved the highest accuracy of 98%. However, they do not use advanced deep learning techniques which may possibly increase the accuracy. In another study Zhong et al., transform the signal using Short time Fourier Transform (STFT) and use CNN to classify the bearing fault [24]. On the transformed signal domain, they apply CNN. The average accuracy achieved is 97.4%. Compared to their work, our model yields better results with directly using raw signal. Bin Li et al., have implemented 1 dimensional CNN architecture and the best result achieved by them is 98.3 % accuracy in fault classification. However, they have not explored the 2D CNN method to improve the results. In another study Pandhare et al., have implemented 2D CNN for the bearing fault classification on the Paderborn dataset[25]. In their work, they have demonstrated 2D CNN on 3 different signal types – raw time domain signal, envelope spectrum, and spectrogram. The maximum accuracy achieved is with spectrogram. For raw signals, the achieved accuracy of 95% is slightly lower when compared to the other studies. Another group of researchers, Wang et al, have proposed a method to use 1D CNN as well as 2D CNN together to predict the fault class in the Paderborn dataset [26]. They have concatenated the 1D CNN output with 2D CNN output before passing it on to fully connected neural network for classification. Their resultant accuracy for classification task is 98.58%. However, their approach is computationally expensive and hence may not be very suitable for online deployment

In this paper, we propose FaultNet, a CNN based model to determine different types of bearing faults with a high accuracy. The aim of this paper is to set a benchmark for bearing fault detection using the conventional machine learning algorithms and deep learning techniques on CWRU and Paderborn datasets. It is important to note that base architecture for both the datasets is the same and the performance of FaultNet is not dataset specific, suggesting wide applicability and deployability of the model to detect different types of bearing faults. We achieve state of the art accuracies for both dataset while proposing a different methodology to extract features from the data. We also study different signal processing techniques and compare accuracies of the traditional machine learning algorithms when combining different types of signal features and our own 2D CNN model.

II. DATASET PREPROCESSING

A. Case Western Reserve University Bearing Dataset

The test rig to generate the dataset consists of 2 hp electric motor to the left, driving a shaft on which a torque transducer and encoder is mounted in the middle coupled to a dynamometer in the right. The torque is applied to the shaft via a dynamometer and electronic control system. The test rig also includes bearings at both drive end (DE) and fan end (FE) of the motor. The bearing at the DE and FE are 6205-2RS JEM and 6203-2RS JEM, respectively. The 6205-bearing used for data collection is a Single Row Deep Groove Radial Ball Bearing with inner diameter 25mm, outer diameter 52mm and 15mm width. To collect the vibration signal data single point faults were artificially induced using electro-discharge machining (EDM) with fault diameters from 7 to 28 mils (0.18 to 0.71mm). The motor loads varied from 0 to 3 hp (approximate motor speeds of 1720 to 1797 rpm). The vibration data was collected using accelerometers, which were attached to the housing with magnetic bases. The data was collected with two sampling frequencies, one with 12,000 samples per second, and 48,000 samples per second and was processed using MATLAB®. In their study, the DE & FE bearing data for the normal (N), inner race fault (IF), outer race fault (OF), and the rolling element(ball) fault (BF) conditions was acquired for fault pattern classification where the fault diameters were selected to be 7 mils, 14 mils, and 21 mils. Ten different conditions are investigated to verify the accuracy of the proposed method in consideration of multiple fault patterns. The vibration signals of ten health conditions are referred in the table 1. In this paper, we used the data from the drive end of the test rig. The sampling frequency chosen is 48 kHz with the load condition being 2 HP at 1750 rpm. To analyze and classify different bearing faults we do some preprocessing steps on the dataset. The rotating speed of the shaft is 1750 rpm and sampling frequency is 48 kHz implies that approximately 1670 data points will be collected for one revolution. Out of 1670 data points, the first 35 points and last 35 points are ignored to account for the noise in the data. Finally, 467600 number of data points of each fault class are chosen and divided into 280 samples, with 1670 data points for each class. Thus, we have 2800 samples with 10 different classes with 280 samples each. Further details, which introduce the test set-up and other data collected, can be found at the CWRU Bearing Data Center website.

TABLE I: CWRU BEARING HEALTH CONDITIONS AND CLASS LABELS

Health Condition	Fault size (mm)	Total dataset	class labels
Normal	-	280	0
ball fault	0.18	280	1
ball fault	0.36	280	2
ball fault	0.53	280	3
inner race fault	0.18	280	4
inner race fault	0.36	280	5
inner race fault	0.53	280	6
outer race fault	0.18	280	7
outer race fault	0.36	280	8
outer race fault	0.53	280	9

B. Paderborn University Dataset

This dataset is generated using 32 bearings. The bearing type used for this dataset generation is 6203, which stands for Deep Groove Ball Bearings with dimensions (inner diameter, outer diameter and width) – 17X40X12mm. Out of 32, 6 bearings are healthy, 12 bearings have artificially created defects and remaining 14 bearings are naturally damaged. The artificial defects have been created by using drilling, EDM, and electric engraving machine. The artificial defects are produced on both, inner and outer race. The natural damages are produced by accelerated lifetime tests. The detailed description can be found in the paper[22].

Further, the bearings' samples can be divided into 3 classes, healthy, inner race fault, and outer race fault. By this classification, there are 6 healthy bearings, 11 inner race fault bearings and 12 outer race fault bearings. This amounts to 29 bearings in total. Remaining 3 bearings are omitted due to their nature of fault. These 3 bearings have inner as well as outer race fault. In the study conducted by Paderborn University, the authors have classified these bearings on the grounds of the maximum contributing fault. If the inner race damage is more compared to outer race, the bearing is classified as inner race fault bearing. For the current model, we used 29 bearings data which can be classified distinctively. The data set is generated with multiple combinations of rpm, torque, and load. For the purpose of this study, we use the following combination. N=1500 rpm, load torque=0.7 Nm and Radial force=1000 N.

Each bearing is used 20 times to generate 20 signals with one fixed combination. The signal generated is a vibration signal for 4 sec with sampling frequency of 64kHz. That means, in a signal there are 2,56,000 data points. To avoid initial and ending noise and disturbance, the sample signal is clipped off for the first 1/16th part and the last 1/16th part. Eventually, signal used, has 2,24,000 data points which are used further for featurization. In total 2320 signals have been used for classification.

III. FEATURIZATION

In every machine learning process, feature engineering plays a very important role and can significantly affect the performance of an algorithm. Feature engineering can directly help the machine learning algorithm to identify the underlying patterns and effectively improve the accuracy of the model. For signal data, featurization includes deriving different features from raw signals such as time domain, time-frequency domain, etc. The vibration signals from machinery components are in general considered to be non-stationary. The non-stationary signals mean that, the frequencies present in a signal vary with time [27]. Therefore, it is important to extract features from time domain as well as time-frequency domain to capture the time varying nature of frequencies present in a signal. In this paper, the features extracted from raw signal data include multiple time and time-frequency domain features. Some of the statistical time domain features that we extract include mean, variance, standard deviation, root mean square (RMS). Moreover, features such as kurtosis and skewness are also extracted as these signals are not stationary. In their paper, Caesarendra et al., give us some physical insights into the features as they report the approximate values of kurtosis and skewness for a normal

bearing to be 3 and 1 respectively [28]. Hence, for bearings which are faulty, we expect to have kurtosis and skewness values shifted from 3 and 1. Another important observation we made was that for faulty bearings, the bearing signal amplitude undergoes abrupt changes when rolling elements pass over defective region of the bearing. These abrupt changes are responsible in disturbing the overall distribution of signal and therefore can act as an important clue in detecting faulty bearings. Generally, value of kurtosis increases, and skewness may change to negative or positive side for faulty bearings. Apart from these features, dimensionless features such as crest factor, shape factor, impulse factor are also extracted. Shape factor is affected by the shape but is independent of the dimension. Crest factor is a measure of an impact when a rolling element comes in contact with raceway. Table 2 summarizes all the 14 features extracted from the raw signal data along with their mathematical formulae used.

TABLE II: FEATURES AND THE MATHEMATICAL FORMULAE USED TO CALCULATE VALUES FOR EACH SIGNALS

No	Feature	Formula
1	Mean	$\text{Mean} = \frac{1}{n} \sum_{i=1}^n x_i$
2	Absolute mean	$\text{Abs Mean} = \frac{1}{n} \sum_{i=1}^n x_i $
3	Maximum	$\text{Maximum} = \max(x_i)$
4	Minimum	$\text{Minimum} = \min(x_i)$
5	Peak to Peak	$\text{Maximum} - \text{Minimum}$
6	Absolute max	$\text{Abs Max} = \max(x_i)$
7	Root Mean Square	$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
8	Variance	$\text{Var} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
9	Clearance factor	$\text{Clf} = \frac{\text{Absolute max}}{\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2}$
10	Kurtosis	$\text{Kurt} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n \times \text{var}^2} - 3$
11	Skewness	$= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}\right)^3}$
12	Impulse Factor	$\text{IF} = \frac{\text{Abs Max}}{\frac{1}{n} \sum_{i=1}^n x_i }$
13	Crest Factor	$\text{CF} = \frac{\text{Abs Max}}{\text{RMS}}$
14	Shape Factor	$\text{SF} = \frac{\text{RMS}}{\frac{1}{n} \sum_{i=1}^n x_i }$

Time-frequency domain representation methods such as short-time Fourier transform (STFT), wavelet transform and Wigner-Ville distribution (WVD) are commonly used for non-stationary or transient signal. These methods implement a mapping of one-dimensional time-domain signals to a two-dimensional function of time and frequency. The objective is to provide a true time-frequency representation of a signal. Similar, to the methods presented in the review conducted by Feng et al., on time-frequency analysis methods for machinery fault diagnosis, we

decomposed these raw signals by employing wavelet decomposition package (WPD) for the extraction of time-frequency domain features[29].

IV. RESULTS WITH SIGNAL FEATURIZATION

A. Case Western Reserve University

We used all the 14 features in Table 2 and evaluated the performance of different shallow learning algorithms. The train-test split used was 80-20% and average 5 fold cross validation accuracy has been reported. Amongst all the models we tried, random forest yielded the highest accuracy (figure 2(a)). An important functionality of the random forest algorithm is that it provides feature importance which gives the user important insights about the features. The feature importance score of the top 5 features calculated using Random Forest has been demonstrated in figure 1(a). It was observed that absolute mean, variance, RMS, shape factor and the mean are 5 most important signal features. Subsequently, to analyze the effects of these important features on the accuracy multiple combinations of these features were used in conjunction with machine learning models to classify the bearing faults. As expected, the accuracy improves when a model is trained with a greater number of features (figure 2(a)). We also implemented wavelet decomposition on the raw signal data for the CWRU dataset extracted the same 14 features from Table 2 on the decomposed signal. Similar to our earlier observations, Random Forest had the best performance with 90% fault detection accuracy on the wavelet decomposed signal model (Figure 3(a)). It was observed that wavelet decomposition level 2 and level 3 had a slightly higher performance. After testing multiple algorithms and different signal featurization techniques it was observed that the performance did not improve beyond 90% accuracy, to further improve the accuracy we then tried deep learning approaches.

B. Paderborn University

Like CWRU dataset, we used a train-test split of 80-20% and report the average five-fold cross validation accuracy. For the Paderborn data random forest algorithm, gave the highest accuracy. Utilizing the feature importance functionality of random forest, top five important features were calculated. It is important to note that, out of top five important features, four features are common in CWRU dataset (figure 1(b)). Similar to CWRU multiple combination of different features were tried and the accuracy of different shallow learning methods was evaluated (figure 2(b)). We also tested for three different decomposition levels to check the effect of wavelet decomposition on overall accuracy of the model. As demonstrated in figure 3(b), it is observed that, there is a slight increase in accuracy with the decomposition level. To further evaluate the results, we analyzed the results of the best performing random forest in greater depth and found that the algorithm was not able to classify outer race faults correctly. Most of the error occurs when classifying between 'Inner race fault' and 'Outer race fault' category while classifying outer race faults. However, inner race faults are being classified correctly.

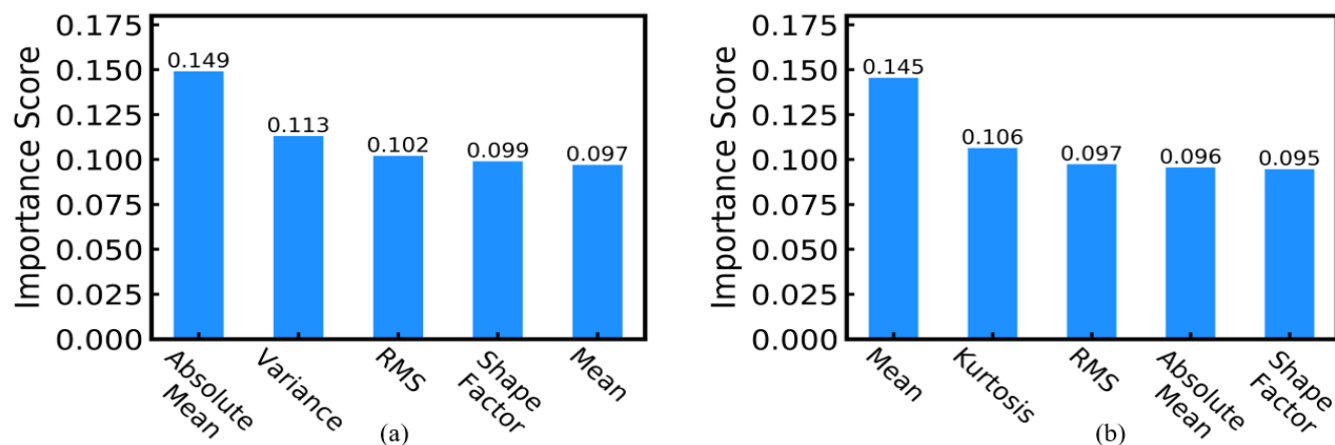


Fig. 1. Feature importance based on Random Forest results. (a) shows top 5 important features obtained on CWRU bearing dataset. Similarly (b) shows the results obtained for Paderborn University dataset. For both the datasets, out of 5 important features.

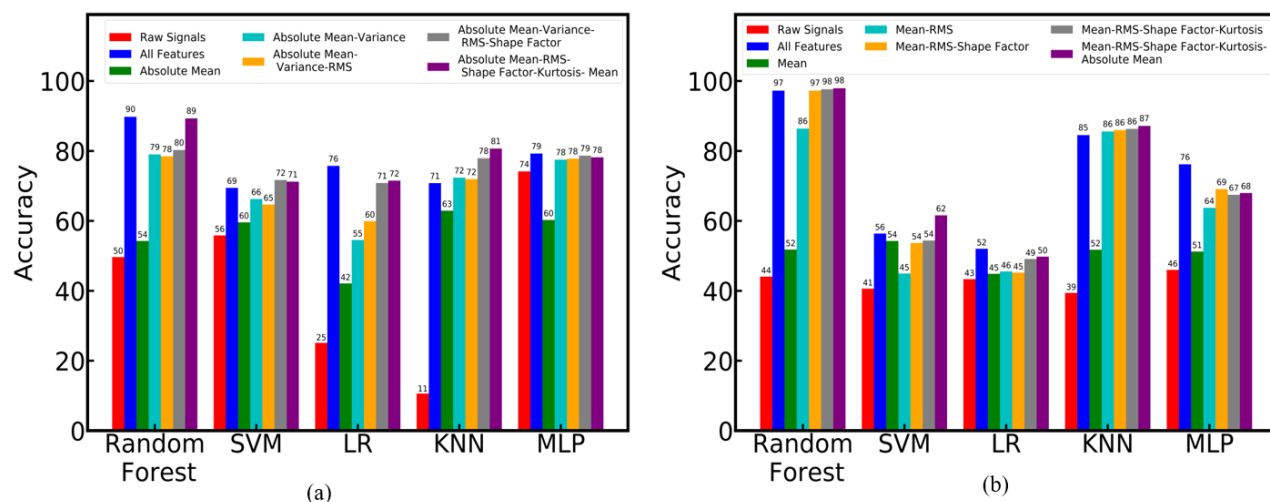


Fig. 2. Classification accuracy with raw data, all features, and top five important features (a) shows accuracy for CWRU dataset for five different classification algorithms. It is evident from the figure that accuracy improves as number of features increases. Only using raw data yields the lowest accuracy. Similarly (b) shows the results obtained for Paderborn University dataset.

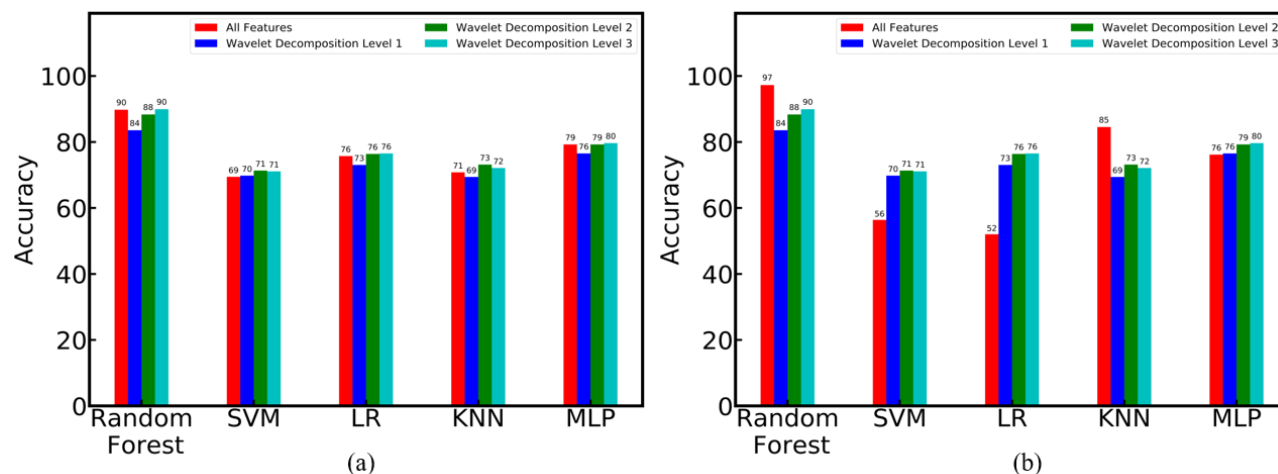


Fig. 3. Classification accuracy using all 14 features on different signal wavelet decomposition level (a) shows accuracy for CWRU dataset. Decomposing signals at different levels result in slightly higher accuracy. Also, for some ML algorithms decomposed signals gives better accuracy compared to raw signals. Similarly (b) shows the results obtained for Paderborn University dataset.

V. CONVOLUTIONAL NEURAL NETWORK

To further improve the classification performance, we developed FaultNet a CNN based architecture that takes raw signal data as input without any pre-processing. CNNs were used because they possess a special ability of extracting relevant features from the data, given a task of prediction. In the FaultNet architecture, we have 2 convolutional and 2 fully connected layers are employed, with the final fully connected layer being the classification layer. We also use max pooling between the first- and second-layer convolutional layer to ensure the selectivity of the most important features. The addition of the of pooling has also led to decreased computational times making the FaultNet relatively inexpensive architecture. The details of the CNN architecture are demonstrated in figure 5. Since the FaultNet is a 2D CNN architecture the amplitude vs time signal needs to be converted in to 2D array. We do this by folding the signal in two-dimensional data and treating the signal data as images. The FaultNet presented here, is applied on both the datasets, and gives state of the art accuracies for the CWRU and Paderborn datasets.

As we know, the raw signals CWRU dataset contains 2800 signals of 1600 data points. The signal data is converted into a 2D array of shape 40×40 . Therefore, we have 2800 signals in the form of 2D arrays of shape 40×40 . Similarly, for Paderborn University dataset, signals of 250,000 datapoints are split into 100 smaller signals of shape 50×50 2D arrays. The convolution operation is performed on the 2D data and subsequently passed it to a fully connected layer which does the classification. The convolution operation is defined by:

$$y(t) = x(t) * w(t) \quad (1)$$

where $y(t)$ is the output of convolution operation between the input $x(t)$ and weight of convolutional filter $w(t)$. The 'ReLU' non-linear activation is applied after each layer and soft-max activation is applied to the outputs of the neurons in the final layer[30] [31]. The learning rate is kept constant at 0.001 and the Adam optimizer is used. We use a batch size of 128 for training and also a drop-out of 0.25 is added in the final layer to prevent overfitting in the neural network [32]. The overall network architecture is shown in the figure 4.

VI. RESULTS WITH CONVOLUTIONAL NEURAL NETWORK

A. Case Western Reserve University Dataset

Initially we tried to predict the bearing fault class with only raw signal as the feature. However, the 5-fold accuracy maxed out at 95.27%. With the state-of-the-art model achieving accuracies in the range 97-100%. We devised a methodology to incorporate more signal information so that the model can learn the signal features better. Thus, we came up with an idea of using mean and median channels to augment the raw signal in a bid to improve the accuracy. In order to generate new channels, a sliding window with length 10 was used as a filter. For every given sample signal data, the filter scans through the whole sample data from the front to the end. To get the same

quantity of data points as the sample data, nine '0' were replenished at the end of the sample signal data and the filter has been set to shift by length 1 for each time. Consequently, a new channel, which had the same shape size with the original channel of sample signal data, is generated by all outputs while the filter completed the data traversal. For the first additional channel, a mean filter was applied to create mean channel. For the second channel, mean filter is substituted with median filter to generate the median channel. We combined new channels with original channel as the new input for 2D convolution model. With increasing number of channels, the accuracy improves simultaneously. The average accuracy has already improved to 98.50% as evident from the figure 5(a). Similar improvements in performance were seen on the Paderborn dataset (figure 5(b)). Apart from the improved accuracy, we observed that the deviation of accuracy was lowered over the five folds of test datasets. Leading us to conclude that channel addition not only improves accuracy but also generates more stable model in terms of accuracy.

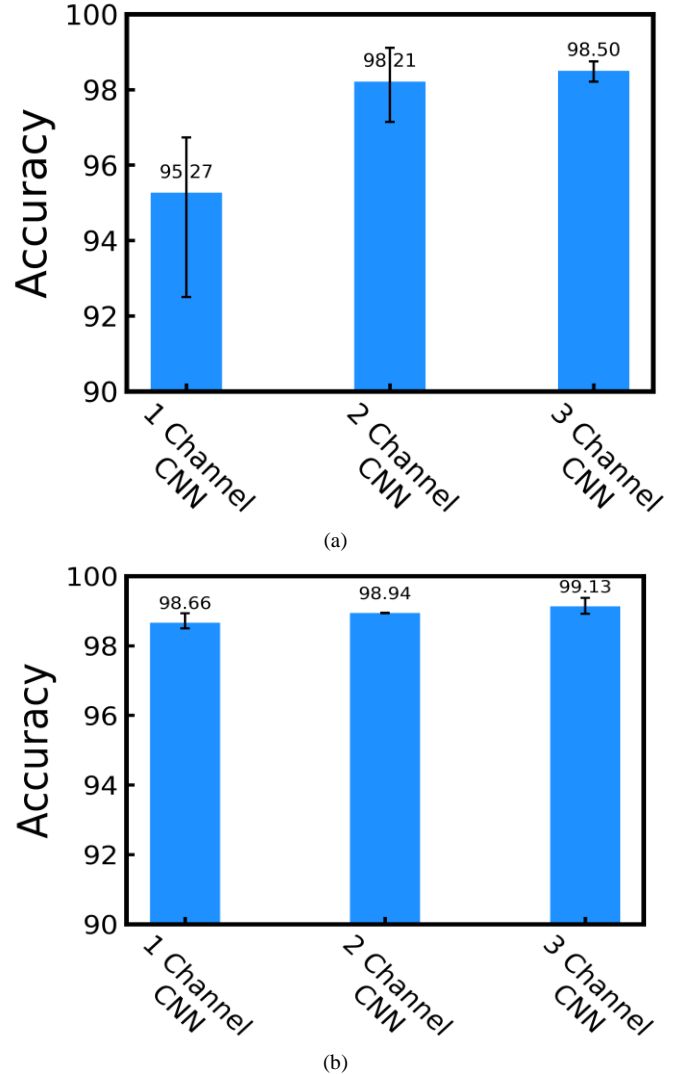


Fig. 5. Comparison of accuracies obtained for different CNN models. (a) CWRU dataset with one-channel, two channel and three channel approach. (b) Paderborn dataset with one-channel, two-channel and three channel approach

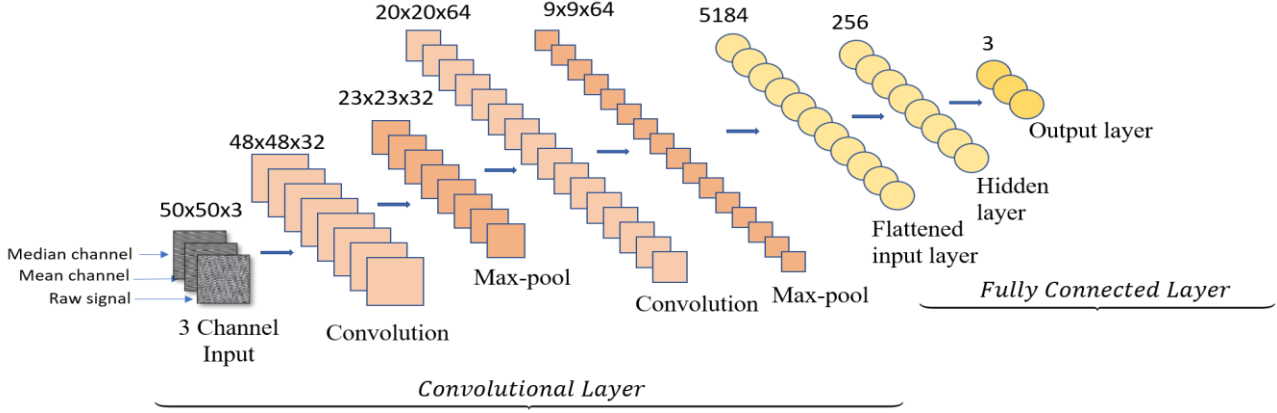


Fig. 4. Convolutional Neural Network Architecture. 2D CNN architecture used for Paderborn University dataset. There are 2 convolution and 2 max-pool layers along with fully connected neural network with 3 outputs. After each convolution layer, a ‘Batch-norm’ and ‘ReLU’ activation is applied.

VII. CONCLUSION

In this paper, a systematic approach towards data-driven vibration-based diagnosis of faults in rolling element bearings is demonstrated. We have benchmarked the performance of different machine learning algorithms by using the featurized signal data and deep learning approaches for the CWRU and Paderborn datasets. Five-fold accuracies of ~99% are obtained for both the datasets indicating state of the art performance. It is found out that, comparatively smaller CNN structure along with multidimensional input i.e., number of channels, work the best for better classification accuracy. For classification, it is important to have enough number of differentiating features between classes. As we stack a greater number of channels, algorithm is able to extract more features compared to single channel input. Each additional channel could be considered as a feature map of the input which provides more information about the input. This is analogous to grayscale and RGB images. If we convert an RGB image to grayscale image, often, there happens to be information loss leading to poor performance. Hence, adding more information improves the performance of our model.

The novelty of this work is in the concise CNN structure, also, in the concept of augmenting 2D raw signal with its mean and median value channels to extract more meaningful feature for CNN. We have demonstrated that the CNN structure devised here improves upon previous methods and has highly competitive performance compared with state-of-the-art methods.

We believe this work can pave the way for online fault detection in case of bearings which could be extremely beneficial for industries. Our approach can be extended to similar types of datasets. The github repository for FaultNet can be found at <https://github.com/BaratiLab/FaultNet>.

VIII. REFERENCES:

- [1] J. Lee, E. Lapira, B. Bagheri, and H. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manuf. Lett.*, vol. 1, no. 1, pp. 38–41, Oct. 2013, doi: 10.1016/j.mfglet.2013.09.005.
- [2] Kim, D., Kim, T.J.Y., Wang, X. et al., "Smart Machining Process Using Machine Learning: A Review and Perspective on Machining Industry," *Int. J. Precis. Eng. Manuf.-Green Technol.*, pp. 555–568, Aug. 2018, doi: <https://doi.org/10.1007/s40684-018-0057-y>.
- [3] R. Thomazella, W. N. Lopes, P. R. Aguiar, F. A. Alexandre, A. A. Fiocchi, and E. C. Bianchi, "Digital signal processing for self-vibration monitoring in grinding: A new approach based on the time-frequency analysis of vibration signals," *Measurement*, vol. 145, pp. 71–83, Oct. 2019, doi: 10.1016/j.measurement.2019.05.079.
- [4] Z. Germán-Salló and G. Strnad, "Signal processing methods in fault detection in manufacturing systems," *11th Int. Conf. Interdiscip. Eng. INTER-ENG 2017 5-6 Oct. 2017 Tirgu Mures Rom.*, vol. 22, pp. 613–620, Jan. 2018, doi: 10.1016/j.promfg.2018.03.089.
- [5] A. Glowacz, "Acoustic fault analysis of three commutator motors," *Mech. Syst. Signal Process.*, vol. 133, p. 106226, Nov. 2019, doi: 10.1016/j.ymssp.2019.07.007.
- [6] L. Jiang, T. Shi, and J. Xuan, "Fault diagnosis of rolling bearings based on Marginal Fisher analysis," *J. Vib. Control*, vol. 20, no. 3, pp. 470–480, Nov. 2012, doi: 10.1177/1077546312463747.
- [7] B. Van Hecke, Y. Qu, and D. He, "Bearing fault diagnosis based on a new acoustic emission sensor technique," *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.*, vol. 229, no. 2, pp. 105–118, Nov. 2014, doi: 10.1177/1748006X14558900.
- [8] J. Xu, S. Tong, F. Cong, and Y. Zhang, "The application of time–frequency reconstruction and correlation matching for rolling bearing fault diagnosis," *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 229, no. 17, pp. 3291–3295, Apr. 2015, doi: 10.1177/0954406215584397.
- [9] X. Zhang, N. Hu, L. Hu, L. Chen, and Z. Cheng, "A bearing fault diagnosis method based on the low-dimensional compressed vibration signal," *Adv. Mech. Eng.*, vol. 7, no. 7, p. 1687814015593442, Jul. 2015, doi: 10.1177/1687814015593442.
- [10] W. Sun, G. An Yang, Q. Chen, A. Palazoglu, and K. Feng, "Fault diagnosis of rolling bearing based on wavelet transform and envelope spectrum correlation," *J. Vib. Control*, vol. 19, no. 6, pp. 924–941, Mar. 2012, doi: 10.1177/1077546311435348.
- [11] B. SAMANTA and K. R. AL-BALUSHI, "ARTIFICIAL NEURAL NETWORK BASED FAULT DIAGNOSTICS OF ROLLING ELEMENT BEARINGS USING TIME-DOMAIN FEATURES," *Mech. Syst. Signal Process.*, vol. 17, no. 2, pp. 317–328, Mar. 2003, doi: 10.1006/mssp.2001.1462.
- [12] "<http://csegroups.case.edu/bearingdatacenter/home>."
- [13] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64–65, pp. 100–131, Dec. 2015, doi: 10.1016/j.ymssp.2015.04.021.
- [14] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020, doi: 10.1109/ACCESS.2020.2972859.
- [15] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016, doi: 10.1016/j.measurement.2016.07.054.
- [16] H. Pan, X. He, S. Tang, and F. Meng, "An Improved Bearing Fault Diagnosis Method using One-Dimensional CNN and LSTM," *Stroj. Vestn. - J. Mech. Eng.*, 2018, doi: 10.5545/sv-jme.2018.5249.
- [17] X. Guo, C. Shen, and L. Chen, "Deep Fault Recognizer: An Integrated Model to Denoise and Extract Features for Fault Diagnosis in Rotating Machinery," *Appl. Sci.*, vol. 7, no. 1, p. 41, Dec. 2016.
- [18] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017, doi: 10.1016/j.sigpro.2016.07.028.
- [19] F. Xu, W. tai P. Tse, and Y. L. Tse, "Roller bearing fault diagnosis using stacked denoising autoencoder in deep learning and Gath–Geva clustering algorithm without principal component analysis and data label," *Appl. Soft Comput.*, vol. 73, pp. 898–913, Dec. 2018, doi: 10.1016/j.asoc.2018.09.037.
- [20] S. Li, G. Liu, X. Tang, J. Lu, and J. Hu, "An Ensemble Deep Convolutional Neural Network Model with Improved D-S Evidence Fusion for Bearing Fault Diagnosis," *Sensors*, vol. 17, no. 8, p. 1729, Jul. 2017, doi: 10.3390/s17081729.
- [21] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, "LiftingNet: A Novel Deep Learning Network With Layerwise Feature Learning From Noisy Mechanical Data for Fault Classification," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4973–4982, Jun. 2018, doi: 10.1109/TIE.2017.2767540.
- [22] Christian Lessmeier, James Kuria Kimotho, Detmar Zimmer, Walter Sextro, "Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification," *Conf. Eur. Conf. Progn. Health Manag. Soc.*, Jul. 2016.
- [23] G. Karatzinis, Y. S. Boutalis, and Y. L. Karnavas, "Motor Fault Detection and Diagnosis Using Fuzzy Cognitive Networks with Functional Weights," in *2018 26th Mediterranean Conference on Control and Automation (MED)*, Jun. 2018, pp. 709–714, doi: 10.1109/MED.2018.8443043.
- [24] D. Zhong, W. Guo, and D. He, "An Intelligent Fault Diagnosis Method based on STFT and Convolutional Neural Network for Bearings Under Variable Working

- Conditions,” in *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, Oct. 2019, pp. 1–6, doi: 10.1109/PHM-Qingdao46334.2019.8943026.
- [25] V. Pandhare, J. Singh, and J. Lee, “Convolutional Neural Network Based Rolling-Element Bearing Fault Diagnosis for Naturally Occurring and Progressing Defects Using Time-Frequency Domain Features,” in *2019 Prognostics and System Health Management Conference (PHM-Paris)*, May 2019, pp. 320–326, doi: 10.1109/PHM-Paris.2019.00061.
 - [26] D. Wang, Q. Guo, Y. Song, S. Gao, and Y. Li, “Application of Multiscale Learning Neural Network Based on CNN in Bearing Fault Diagnosis,” *J. Signal Process. Syst.*, vol. 91, no. 10, pp. 1205–1217, Oct. 2019, doi: 10.1007/s11265-019-01461-w.
 - [27] James Kuria Kimotho, Walter Sextro, “An approach for feature extraction and selection from non-trending data for machinery prognosis,” *Second Eur. Conf. Progn. Health Manag. Soc. 2014*, 2014.
 - [28] W. Caesarendra and T. Tjahjowidodo, “A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing,” *Machines*, vol. 5, no. 4, p. 21, Sep. 2017.
 - [29] Z. Feng, M. Liang, and F. Chu, “Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples,” *Cond. Monit. Mach. Non-Station. Oper.*, vol. 38, no. 1, pp. 165–205, Jul. 2013, doi: 10.1016/j.ymssp.2013.01.017.
 - [30] Vinod Nair, Geoffrey E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” *ICML*, 2006.
 - [31] Bridle J.S., “Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition,” *Soulié FF Héault J Eds Neurocomputing NATO ASI Ser. Ser. F Comput. Syst. Sci. Vol 68 Springer Berl. Heidelb.*, 1990, doi: https://doi.org/10.1007/978-3-642-76153-9_28.
 - [32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.