

A Deep Learning Approach based on MLP-mixer Models for Bearing Fault Diagnosis

I. Abstract

In order to achieve high accuracy, deep learning model required large a substantial amount of training data.
⇒ Propose a new deep learning model based on [MLP-Mixer](#) model and CNN that can dramatically improve training performance even with a small amount of dataset.

II. Model

Three main contributions of this paper:

- Propose a novel model called the Hybrid CNN-MLP model, which combines the strength of both the convolutional neural network (CNN) and the [MLP-Mixer](#) model.
- During the training and testing phase, the raw vibration signals in the time domain are transferred to the frequency domain to improve the extraction of information.
- 60 Training sample ⇒ Achieve 98.07% accuracy

II.1 Methodology

- Model takes a spectrogram representation of signal as its input. ⇒ Short time Fourier Transform technique is utilized.

- The Short time Fourier transform could be express as follow:

$$X(t, f) = \int_{-\infty}^{\infty} x(t)w(t - \tau).e^{-i2\pi f\tau}d\tau$$

$X(t, f)$ is the Short Fourier transformation results of signal $x(t)$ and f is frequency, $w(t)$ is window function (Hann or Gaussian). The spectrogram is generated by taking square of $X(t, f)$: $S = |X(t, f)|^2$ (**Input of model**).

- **Horizontal display times, vertical axis represent frequency**

II.2 Model architecture

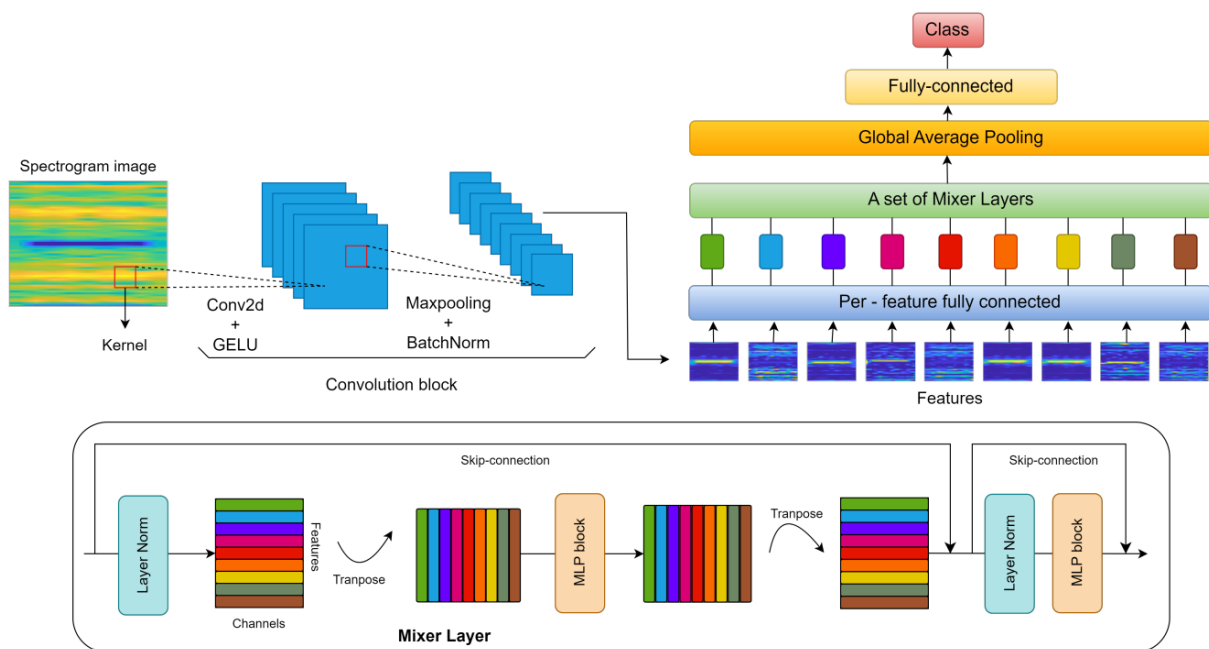
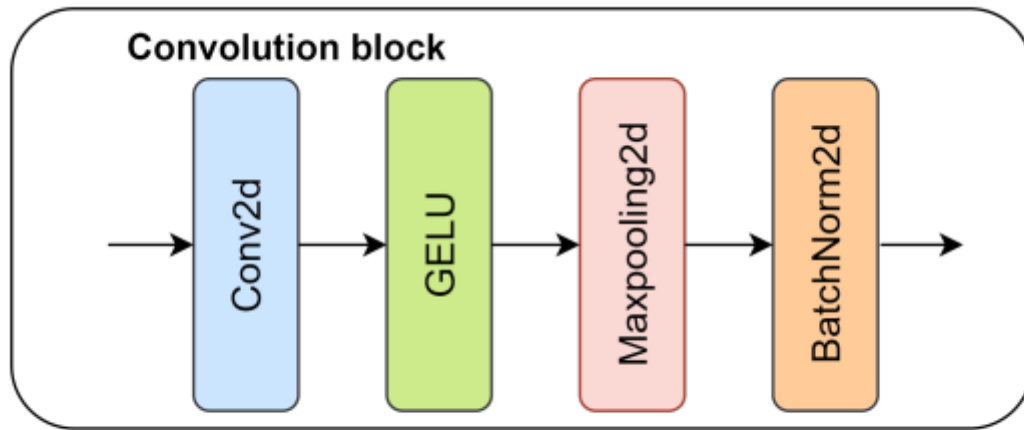


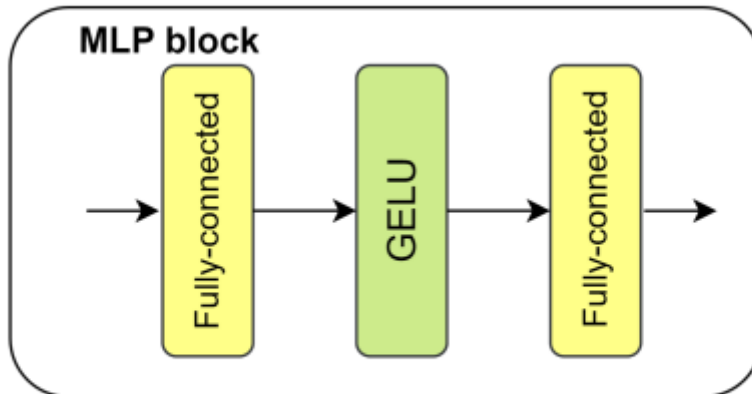
Fig. 1: General structure of the hybrid model CNN-MLP architecture. The network includes a group of Mixer layers and convolution layers that are combined together. A Mixer Layer comprises both feature-mixing and channel-mixing. Each of these mixings includes a MLP block and a LayerNorm.

- CNN filters for filtering the local feature input.
+ Structure of convolution block:



(b) Convolution block

- MLP-mixer model to capture global information.



(a) MLP block

II.3 Config model

In this model:

- Input spectrogram image size: 64×64 .
- Apply two convolutional layers, each utilizing a 3×3 kernel filter size
- Maxpooling with a 2×2 to reduce dims off input image.
- MLP-mixer with 8 depths and 256 hidden dimensions.

III. Experiment and results

Dataset

Dataset produced by Case Western Reserve University (CWRU). In this dataset:

- Data was collected while the machine was operating under three different loads of 1, 2, and 3 horsepower.
- Labels as A, B and C respectively.
- There are three categories of faults, each fault is divided into three variations, measuring 0.007 inches, 0.014 inches and 0.021 inches respectively. \Rightarrow We have 10 classes ($3 \times 3 + 1$ - normal states).

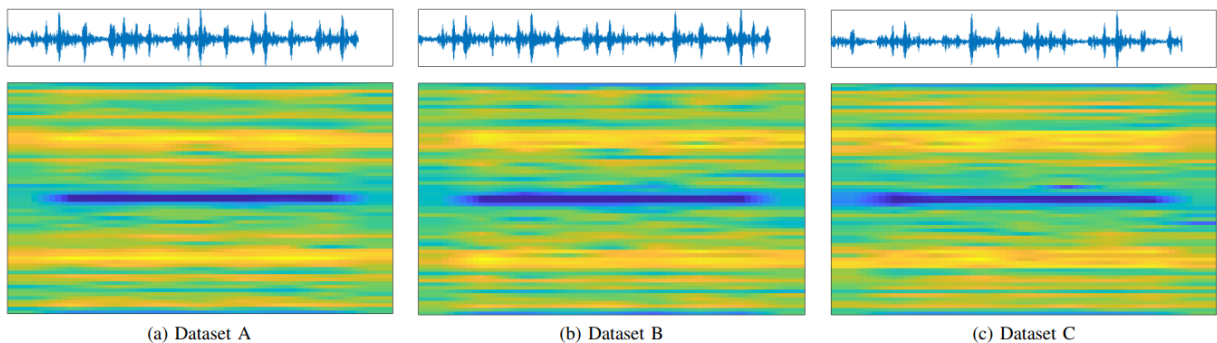


Fig. 3: The vibration signal (top) of Inner Race 0.014 fault and the corresponding spectrum image (bottom) in (a) Dataset A, (b) Dataset B, (c) Dataset C

- Each datasets contain 6600 training images and 250 test images. D is composed of A,B and C wit 19800 training images and 750 test images.

Results

98.07% despite trained on only 60 samples

the model still faces difficulties in accurately diagnosing bearing faults when faced with certain real-world challenges. The effectiveness of this technique is

demonstrated through experimental results and comparisons.

TABLE I: The accuracy (in %) of the model when it is trained from one work condition and tested on a different work condition.

Dataset		Number of training samples			
Train set	Test set	90	300	3000	6600
A	A	98.12	98.47	99.76	100
	B	93.20	91.84	97.68	99.04
	C	77.28	82.00	94.90	95.88
B	A	82.60	85.14	77.64	69.10
	B	97.89	98.69	99.85	100
	C	87.92	88.90	94.24	96.2
C	A	56.96	50.12	45.44	42.48
	B	89.56	81.88	69.20	66.55
	C	97.03	98.45	99.36	99.91

TABLE II: The accuracy (in %) on the same test dataset of comparative models when using a different number of training samples

Model	Number of training samples				
	60	90	300	6000	19800
One-shot	79.33	88.41	95.45	99.65	99.79
Five-shot	82.80	91.37	95.65	99.66	99.77
WDCNN	64.93	70.13	88.27	99.33	99.45
MLP-mixer	93.33	96.47	98.26	99.47	99.92
Ensemble learning (WDCNN + MLP-mixer)	93.73	96.77	98.46	99.60	99.95
Hybrid model (ours)	98.07	98.86	99.37	99.91	100