Application of Deep Learning Models for Aircraft Maintenance

Humberto Hayashi Sano[1], **Lilian Berton**[1]

[1]Institute of Science and Technology – Federal University of São Paulo (UNIFESP)
São José dos Campos, SP, Brazil

humberto.sano@embraer.com.br, lberton@unifesp.br

Abstract. *Neural networks provide useful approaches for determining solutions to complex nonlinear problems. The use of these models offers a feasible approach to help aircraft maintenance, especially health monitoring and fault detection. The technical complexity of aircraft systems poses many challenges for maintenance lines that need to optimize time, efficiency, and consistency. In this work, we first employ Convolutional Neural Networks (CNN), and Multi-Layer Perceptron (MLP) for the classification of aircraft Pressure Regulated Shutoff Valves (PRSOV). We classify a wide range of defects such as Friction, Charge and Discharge faults considering single and multi-failures. As a result of this work, we observed a significant improvement in the classification accuracy in the case of applying neural networks such as MLP (0.9962) and CNN (0.9937) when compared to a baseline KNN (0.8788).*

1. Introduction

Covid-19 and world inflation impacted the aviation industry increasing the need to optimize their costs [Iata 2021, Maneenop and Kotcharin 2020]. Maintenance should be effective ensuring the asset utilization is optimum and reducing the aircraft's 'hangar-time'. The complexity of aircraft maintenance, repair, and overhaul (MRO) generates demand for automatic systems for Prognostics and Health Monitoring (PHM), which need to ensure quality, security, and accountability.

The application of Artificial Neural Networks (ANN) in aircraft maintenance helps to analyze large amounts of data collected during the operation time and estimate a unit fault diagnosis efficiently and cost-effectively. A previous survey [Rengasamy et al. 2018] showed some deep learning approaches used for aircraft MRO such as Deep Auto-encoders (DAE), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Deep Belief Networks (DBN). They were employed in parts of the aircraft, such as aero engines, fuel systems, and actuators. The authors also point out the need for a benchmark data set, where researchers could test and compare their machine learning approaches for aerospace.

The aircraft is composed of many systems which are responsible for its functions. The *Environmental Control System* (ECS) is responsible for providing environmental conditioning to the cabin and cockpit. The ECS is composed of some valves which regulate the hot air extracted from the engines to other subsystems such as air conditioning, anti-ice, etc. These valves are called *Pressure Regulated Shutoff Valves* (PRSOV). Their internal components are susceptible to failures due to their operations at high pressures and temperature environments. PRSOV internal sub-components are shown in Figure 1.

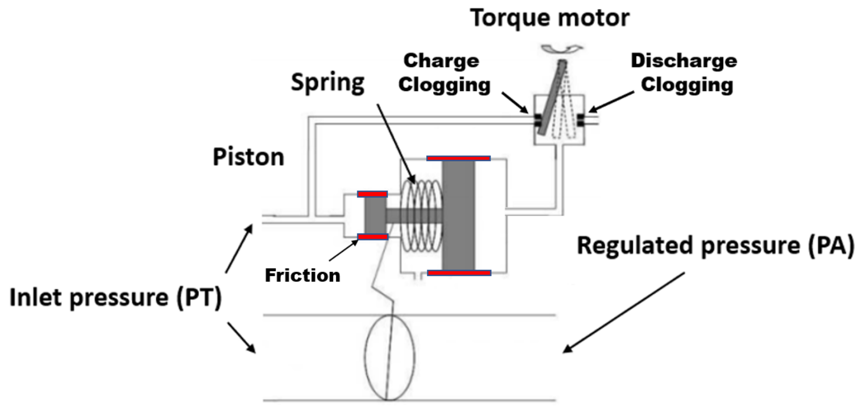

Figure 1. Internal scheme of Pressure Regulated Shutoff Valves (PRSOV) [Turcio et al. 2013].

There are few studies, [Sano et al. 2019], which considered the PRSOV multi-faults conditions. In [Castilho et al. 2018] only single-fault has been analyzed. According to [Sano et al. 2019], PRSOV multi-fault classification is a challenging task due to the overlap of PRSOV behavior under these scenarios.

Besides, there is also a lack of studies that analyzed the PRSOV diagnosis based on regulated pressure which is the information that is naturally available in an ECS system in the aircraft. For instance, in [Castilho et al. 2018] and [Sano et al. 2019], they were based on PRSOV opening and closing times which are information more appropriate for the test bench environment. Moreover, the evaluation of other more sophisticated machine learning algorithms (e.g. neural networks) to predict the PRSOV's health is demanded as well as an analysis of which features influence the PRSOV prediction.

The objective of this paper is to evaluate the effectiveness of Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) to perform the classification of the PRSOV. We considered the output responses due to their most relevant parameters variation (spring, charge and discharge chamber clogging levels). The contributions of this work are:

- Present the effectiveness of pressure regulated information applied in the MLP and CNN to predict the PRSOV health;
- Analyze the effect of hyper-parameters variation in both neural networks in the PRSOV multi-fault classification tasks;
- Provide a PRSOV standardized dataset (8,000 valve samples) that can help other researchers validate their approaches.

The results show that regulated pressure information applied on the MLP and CNN was able to classify PRSOV with accurate diagnostics higher than 99%. These results are higher than those obtained from the baseline KNN method (0.8788) and previous work [Sano et al. 2019]. This previous study found that other machine learning algorithms were able to diagnose multi-fault with a 94.3% of accuracy only. Besides, the usage of the regulated pressure in this present work, unlike the opening and closing times, is already monitored in the real system and the possibility to use it for the classification purpose is an improvement when compared to the previous study [Sano et al. 2019].

The remaining of this paper is organized as follows: Section 2 presents other works that have studied the application of machine learning to PHM with a focus to determine PRSOV health. Section 3 describes the obtaining dataset process, its preparation to be used as input of neural networks, the process of neural networks evaluations and their parameter tuning as well as the criteria to compare them. Section 4 describes the results obtained employing the concepts presented in section 3. Finally, Section 5 summarizes the main conclusions of this paper.

2. Related works

Some works already explore the PHM of PRSOV valves using machine learning techniques. In [Castilho et al. 2018], the PRSOV has been analyzed using output data from a Simulink model with a specific valve maneuver to obtain the PRSOV times as attributes for the machine learning algorithms. The author used Support Vector Machines (SVM) and Classification and Regression Trees (CART) to estimate the health state of the PRSOV. This work uses these techniques to classify some PRSOV individual failures independently.

In [Sano et al. 2019], the author showed that data-driven models based on different data sources (sensors data, fault messages, and reliability data) can provide better prognostics than traditional prognostics based on historical time-to-failure data. The following algorithms have been used in this study, k-Nearest Neighbors (k-NN), Artificial Neural Networks (ANN), CART, SVM, Bayesian generalized linear models (Bayes), Gradient Boosting with Regression Trees (Boosted Trees), Linear Regression (LR) and Random Forests (RF).

In [de Assis Silva et al. 2022], the authors studied the capability of some machine learning techniques (kNN, Decision tree, Random Forest, Ordinary Least Square, XGBoost) to perform the regression of PRSOV internal parameters. A significant effort to investigate the influence of feature selection based on operational and hysteresis parameters in the regression task has been performed in this study. However, the feature selection was a manual task.

Even though there are some studies using machine learning which tackled the PRSOV's diagnostic, this component has not been evaluated with deep learning, however, other works evaluate this approach in aircraft components. CNNs were used by [Fuan et al. 2017] that combined it with Particle Swarm Optimisation (PSO) to classify fault in rolling bearing in vibration signals collected from test rig with 8 different health. [Li et al. 2018] uses deep CNN to estimate Remaining Useful Life (RUL) and fault diagnosis of aircraft turbofan engines.

Auto-encoders were used by [Sarkar et al. 2016] for detection of cracks in thick multi-layer composites in aircraft, they analyzed videos of the composite coupons slowly bent until full fracture. [Reddy et al. 2016] employs AE on aircraft data for anomaly detection and fault disambiguation analyzing time-series data from multiple sensors. [Gao et al. 2017] use stacked Denoising Auto-encoders (SDAE) and Support Vector Machine (SVM) to predict the RUL of integrated modular avionics (IMA).

3. Methodology

This section presents the materials and methods used in the work. Subsection 3.1 describes the neural network models used in this work. Subsection 3.2 presents the dataset generation process based on the Simulink model simulation. Subsection 3.3 presents the setup, the training, and the test process used in this work. Subsection 3.4 presents the neural networks models configurations to be studied in this work as well as the evaluation process as validation and parameter tuning.

3.1. Neural network algorithms

An Artificial Neural Network (ANN) is a mathematical model inspired by biological neural networks. It consists of a connected group of artificial neurons that process information using non-linear approaches, this way, it can be used to find patterns in complex data [Goodfellow et al. 2016b].

A Multi-layer Perceptron (MLP) is an ANN with multiple layers, at least three: an input and an output layer with one or more hidden layers. It uses backpropagation, which is an efficient method for calculating the weight updates by computing the gradient of the loss function. For the training process, the network is created with random values in all of its weights and biases. Initially, the loss function will be high, and the aim of training the network is to reduce the loss function as low as possible. This way, the network can classify the training set with higher accuracy.

A Convolutional Neural Network (CNN) is an ANN with convolution layers with different filters and a fully connected layer at the end. The convolutions preserve the spatial relationship in the data points [Goodfellow et al. 2016a]. After passing the data through a convolutional layer, the output is normally passed through an activation function, such as sigmoid, or ReLu. The activation function adds non-linearity to the CNN. A basic CNN is composed of some convolutional layers, followed by an activation function, followed by a pooling layer. A pooling layer helps to reduce the spatial size of the representation, decreasing the required amount of computation. These layers can be repeated many times.

3.2. Dataset generation

In this study, the input data for the neural network are the regulated pressure which was provided by a validated Simulink model of PRSOV. The simulations have been performed for one cycle of PRSOV command in the torque motor as shown in Figure 1.

Each PRSOV sample was generated by varying its intrinsic parameters such as friction coefficient, charge, and discharge clogging levels. The values for each parameter were defined based on two lists of values picked up uniformly random distribution. The first list contains the normal values and another one with abnormal. These intervals have been defined based on valve specialist information.

In each simulation, we collected the regulated pressure in each timestamp. Each regulated pressure timestamp value (eg. Figure 2) was considered as a neural network input. In total there are 201 timestamp points corresponding to the features from the dataset.

In total, we obtained a number of 8,000 valve samples, each one in one of the following states as Healthy, isolated failures (Friction Fault, Charge Fault and Discharge

Fault) and simultaneous failures (Charge and Friction Simultaneous Faults, Charge and Discharge Simultaneous Faults, Discharge and Friction Simultaneous Faults and All Faults occurring simultaneously).

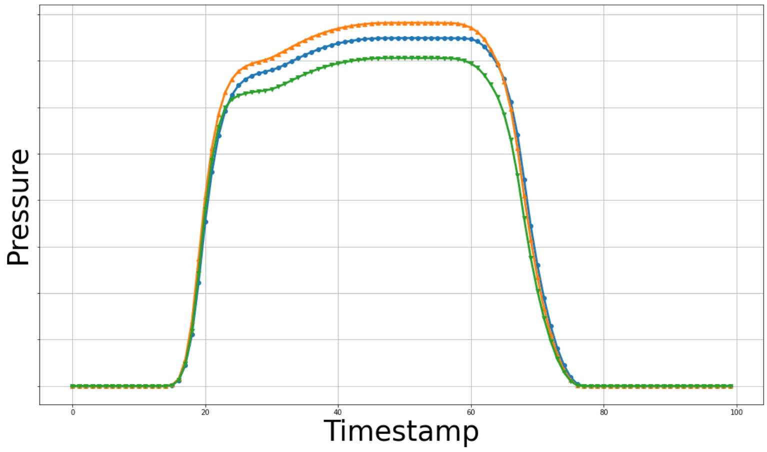

Figure 2. Pressure Regulated Shutoff Valves (PRSOV) pressure curves for three different valve samples.

Considering all the pressure values of each time step for all samples, we performed a standardization as described in Equation 1.

$$P_{Ni} = \frac{P_i - \bar{P_i}}{sd(P_i)} \quad (1)$$

Where, P_i is the set of pressure values of the valves in the timestamp at the column i, $\bar{P_i}$ and $sd(P_i)$ are the mean and the standard deviation of P_i, respectively.

We represented the PRSOV classes as the hot encoded as shown in Table 1. In this codification, each digit represents one neuron output in the output layer of the neural network.

Table 1. PRSOV classes represented as One Hot encoded.

Class Name	One Hot Encoded representation
Normal	10000000
Charge	01000000
Discharge	00100000
Friction	00010000
Charge and Discharge	00001000
Charge and Friction	00000100
Discharge and Friction	00000010
Charge and Discharge and Friction	00000001

The standardized dataset as defined in Equation 1 is available in [Sano and Berton 2022].

3.3. Experiments setup

We used the Keras library from Python to model the networks. The experiments have been executed in a Google Colab environment.

A portion of 10% of the total samples in the dataset was reserved as a test partition. This part represents new data that has not been used in the training process of the neural networks. This portion is made by preserving the percentage of samples for each class.

In the process of network topology evaluation, we used the K fold cross-validation method in the training process with $k = 10$. This represents a percentage of 10% of the training data as a validation set at each training process. During each training run, we captured the evolution of the loss values along the epochs and also the mean and deviation of the last 10 epochs.

After this topology comparison analysis, the best topology was trained with the train partition data. Finally, in order to verify the generalization capability of the trained model, the test partition was applied to this model and the loss metrics were evaluated.

3.4. PRSOV's healthy state classification

In this work, we studied the effectiveness of some MLP (Multi-Layer Perceptron) and CNN (Convolutional Neural Networks) models in the task of PRSOV healthy state classification based on the output pressure information resulting from the PRSOV Simulink model simulations.

We explored the variations of some hyper-parameters of the networks such as the number of layers/neurons as defined in Table 2 and evaluate their influence in the training and validation process. For CNN specifically, there are variations in the number of filters and kernel size in the convolution 1D layers and pooling size for the pooling layer.

Table 2. Neural network models configurations for classification.

MLP		CNN	
Network	**Layers**	**Network**	**Layers**
4N	DENSE (4N/Relu) DENSE (8N/Softmax)	M1	CONV1D (Filter = 1, Kernel = 8) AV POOLING (Size = 4) FLATTEN DENSE (16N/Relu) DENSE (8N/Softmax)
8N	DENSE (8N/Relu) DENSE (8N/Softmax)	M2	CONV1D (Filter = 2, Kernel = 8) AV POOLING (Size = 4) FLATTEN DENSE (16N/Relu) DENSE (8N/Softmax)
16N	DENSE (16N/Relu) DENSE (8N/Softmax)	M3	CONV1D (Filter = 1, Kernel = 16) AV POOLING (Size = 4) FLATTEN DENSE (16N/Relu) DENSE (8N/Softmax)
32N	DENSE (32N/Relu) DENSE (8N/Softmax)	M4	CONV1D (Filter = 1, Kernel = 8) AV POOLING (Size = 8) FLATTEN DENSE (16N/Relu) DENSE (8N/Softmax)
16-8N	DENSE (16N/Relu) DENSE (8N/Relu) DENSE (8N/Softmax)	M5	CONV1D (Filter = 1, Kernel = 8) AV POOLING (Size = 4) CONV1D (Filter = 1, Kernel = 8) AV POOLING (Size = 2) FLATTEN DENSE (16N/Relu) DENSE (8N/Softmax)

We performed the training process with the following configurations:

- Optimizer: Adam, Epochs: 50
- Loss criteria: Categorical cross entropy ($L(\hat{y}, y) = \sum_{k=1}^{K} y_i log(\hat{y_i})$), according to [Zafar et al. 2018]. Where K, y_s, and $\hat{y_s}$ are the number of classes, the actual and estimated value, respectively.

The training process and the test of the best configuration were performed following the steps defined in Section 3.3. In order to compare the performance of neural networks with another simpler machine learning method, we also evaluated the performance of some KNN models varying the value of K neighbors (1, 2, 5, 10, 20).

The metrics used to evaluate the classification errors were the accuracy and the confusion matrix. They were obtained from the application of the test partition data to the best MLP, CNN neural networks, and KNN configurations. In addition, we applied the Principal Component Analysis (PCA) decomposition of the input data and the information generated by the hidden layers for the best neural network configuration. The objective is to observe the separation of the labels after computing the data through the hidden layers.

4. Results

This section presents the results of applying the MLP and CNN network's topologies as described in Section 3.4 to perform the PRSOV classification.

Analyzing Figure 3, we can notice that for MLP with a number of neurons higher or equal to 8, the training and validation loss value tends to stabilize in a similar value (~ 0.01). However, the decay rate of the loss curves at the beginning of the learning process (low values of epochs) is directly proportional to the number of neurons. Another important point to be addressed is an increase in the loss (shifting up of loss curve) with the addition of a second hidden layer ($16 - 8N$).

Regarding the CNN models, taking M1 as a reference, increasing the filter numbers in the convolutional 1D layer from 1 to 2 filters (M2) or kernel size from 8 to 16 (M3) we observed a decrease in the loss value. However, with the increment in the pooling size from 4 to 8 (M4) or the addition of more than one convolutional layer (M5), there is an increase in loss value. We can observe that this last modification impacted the network results significantly (loss almost 10 times higher).

Analyzing the accuracy values for the best configurations, there is similar behavior of the loss training and validation, the addition of the second hidden layer in the MLP and the second convolutional 1D layer in the CNN causes a decrease in the accuracy. Based on the accuracy obtained using the validation portion as shown in Tables 3, the configurations MLP with 16 neurons and CNN model M2 had better performance compared with other configurations.

These two networks have close accuracy values however the CNN model has a significantly lower number of parameters when compared with the MLP as shown in Table 3.

Applying the test portion data to the MLP with 16 neurons and CNN (M2) resulted in a value of accuracy of 0.9962 and 0.9937. These values were significantly higher than the best KNN configuration (0.8788). These high accuracy values of both networks can

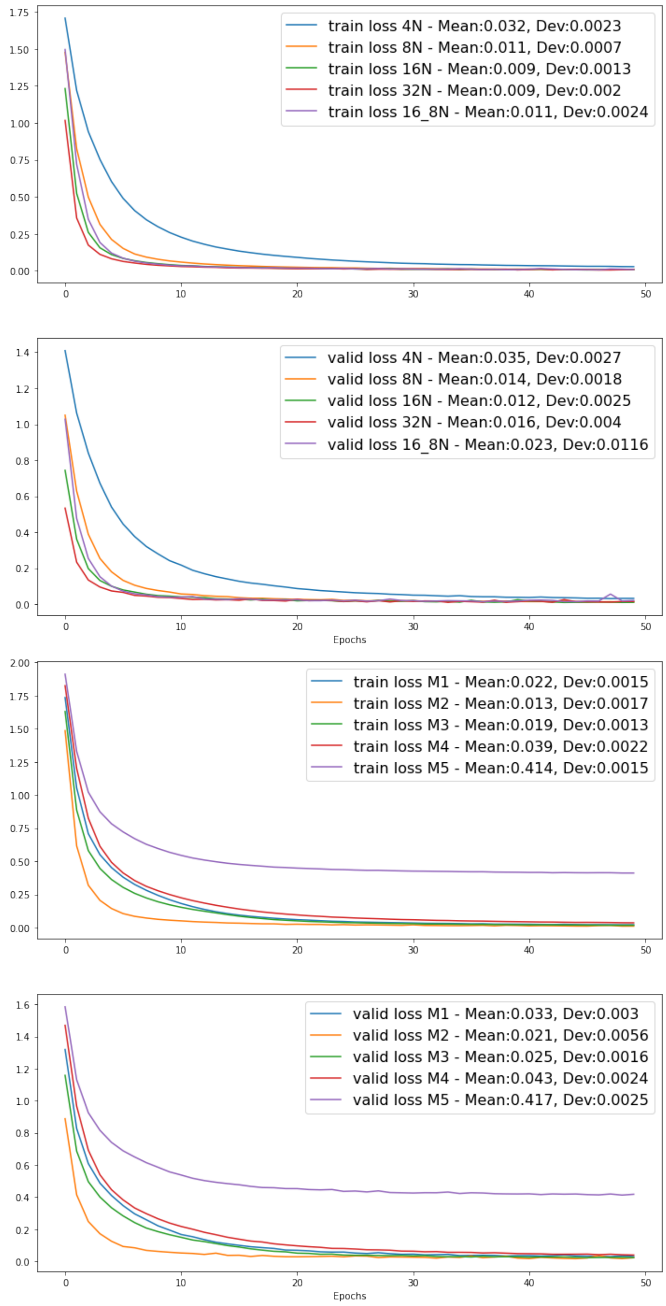

Figure 3. MLP (upper) and CNN (bottom) classification loss curves.

Table 3. Accuracy (Acc) and standard deviation (Dev) of test data portion applied to MLP, CNN, and KNN models.

MLP			CNN			KNN		
Config	Mean Acc	Dev	Config	Mean Acc	Dev	Config	Mean Acc	Dev
4 N	0.9898	0.00128	M1	0.9936	0.00048	K=1	0.883	0.0125
8 N	0.9972	0.00024	M2	0.9953	0.00051	K=2	0.855	0.0111
16 N	0.9975	0.00049	M3	0.9943	0.00041	K=5	0.882	0.0077
32 N	0.9974	0.00062	M4	0.9881	0.00092	K=10	0.880	0.0111
16/8 N	0.9965	0.00078	M5	0.8286	0.00063	K=20	0.862	0.0126

be observed in the confusion matrix, shown in Table 4. The number of right classification between them are close (MLP = 796, CNN = 795 and KNN=703). Most parts of the wrong classification in both neural network types are related to multiple faults.

Table 4. Confusion Matrix of MLP/CNN/KNN for single and multi-fault classification (Normal (N), Charge (C), Discharge (D), Friction (F), Charge and Discharge (CD), Charge and Friction (CF), Discharge and Friction (DF) and All Faults occurring simultaneously (CDF)).

	N	C	D	F	CD	CF	DF	CDF
Pred N	100/99/90	0/1/3	0/0/1	0/0/1	0/0/5	0/0/0	0/0/0	0/0/0
Pred C	0/0/0	100/100/97	0/0/0	0/0/1	0/0/0	0/0/2	0/0/0	0/0/0
Pred D	0/0/0	0/0/0	100/100/83	0/0/0	0/0/2	0/0/0	0/0/15	0/0/0
Pred F	1/1/3	0/0/0	0/0/0	99/99/88	0/0/1	0/0/2	0/0/1	0/0/5
Pred CD	0/0/10	0/0/0	0/0/2	0/0/4	100/100/74	0/0/0	0/0/2	0/0/8
Pred CF	0/0/0	1/2/2	0/0/0	0/0/2	0/0/0	99/98/96	0/0/0	0/0/0
Pred DF	0/0/0	0/0/0	0/0/6	0/0/2	0/0/3	0/0/0	100/100/89	0/0/0
Pred CDF	0/0/0	0/0/0	0/0/0	2/0/5	0/1/6	0/0/1	0/0/2	98/99/86

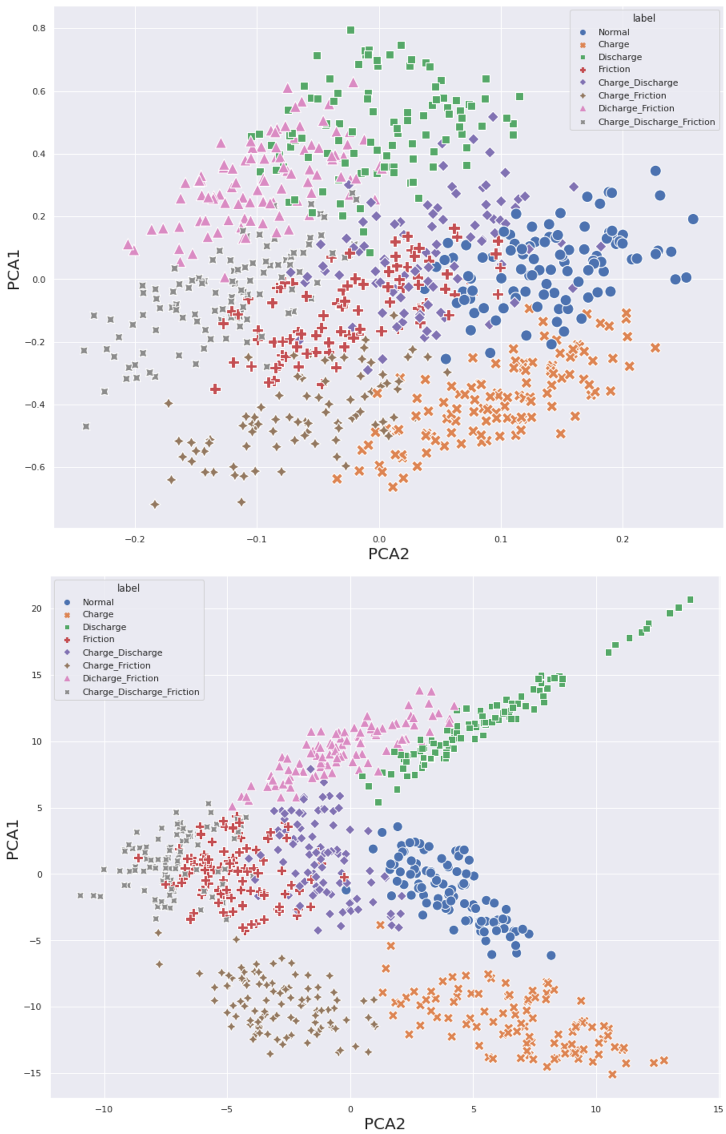

Figure 4. PCA decomposition of inputs (upper) and after MLP's hidden layers (bottom).

The difference in the classes overlap between the test partition data before and after passing through the MLP can be observed in the PCA analysis illustrated in the Figures 4 (left and right), respectively. They are the two main components of these data. We can notice a clear separation among the classes related to single faults. This behavior does not occur in multi-fault classes. The Friction class samples overlap with other classes such as CDF, and CD. These behaviors can be observed in the confusion matrix.

5. Conclusion

This work presented experiments to evaluate the effectiveness of MLP and CNN models in diagnosing the state of PRSOV healthy. The main results of this work are:

- The effectiveness of regulated pressure information to be used as an attribute to perform the PRSOV diagnostic prediction based on classification.
- The effectiveness of regulated pressure information to be used as an attribute to the neural network classification of PRSOV healthy;
- The classification improvements with the increase of neurons in MLP networks;
- The classification improvements with the increase of filter numbers and kernel size in the convolutional layer;
- Diminguish of classification effectiveness with the increase of pooling size or addition of more convolutional layers with its associated pooling layer in the network.

The results confirm the usage of ANN allows diagnostic failures in PRSOV. It can help to detect the problem in an aircraft accurately and efficiently minimizing costs.

References

Castilho, H. M., Nascimento, C. L., and Vianna, W. O. L. (2018). Aircraft bleed valve fault classification using support vector machines and classification trees. In *2018 Annual IEEE International Systems Conference (SysCon)*, pages 1–7.

de Assis Silva, L., Sano, H. H., and Júnior, C. L. N. (2022). Degradation estimation analysis of an aeronautical pneumatic valve using machine learning. In *2022 IEEE International Systems Conference (SysCon)*, pages 1–6.

Fuan, W., Hongkai, J., Haidong, S., Wenjing, D., and Shuaipeng, W. (2017). An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Measurement Science and Technology*, 28(9):095005.

Gao, Z., Ma, C., and Luo, Y. (2017). Rul prediction for ima based on deep regression method. In *2017 IEEE 10th International Workshop on Computational Intelligence and Applications (IWCIA)*, pages 25–31. IEEE.

Goodfellow, I., Bengio, Y., and Courville, A. (2016a). Convolutional networks. In *Deep learning*, volume 2016, pages 330–372. MIT Press Cambridge, MA, USA.

Goodfellow, I., Bengio, Y., and Courville, A. (2016b). *Deep learning*. MIT press.

Iata, A. (2021). The impact of covid-19 on aviation. `https://www.airlines.iata.org/news/the-impact-of-covid-19-on-aviation`.

Li, X., Ding, Q., and Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 172:1–11.

Maneenop, S. and Kotcharin, S. (2020). The impacts of covid-19 on the global airline industry: An event study approach. *Journal of air transport management*, 89:101920.

Reddy, K. K., Sarkar, S., Venugopalan, V., and Giering, M. (2016). Anomaly detection and fault disambiguation in large flight data: A multi-modal deep auto-encoder approach. In *Annual Conference of the PHM Society*, volume 8.

Rengasamy, D., Morvan, H. P., and Figueredo, G. P. (2018). Deep learning approaches to aircraft maintenance, repair and overhaul: A review. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 150–156. IEEE.

Sano, H., Malere, J., and Berton, L. (2019). Single and multiple failures diagnostics of pneumatic valves using machine learning. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 202–213, Porto Alegre, RS, Brasil. SBC.

Sano, H. H. and Berton, L. (2022). Dataset. `https://doi.org/10.5281/zenodo.7191171`.

Sarkar, S., Reddy, K. K., and Giering, M. (2016). Deep learning for structural health monitoring: A damage characterization application. In *Annual Conference of the PHM Society*, volume 8.

Turcio, W., Yoneyama, T., and Moreira, F. (2013). Quasi-lpv gain-scheduling control of a nonlinear aircraft pneumatic system. In *21st Mediterranean Conference on Control and Automation*, pages 341–350.

Zafar, I., Tzanidou, G., Burton, R., Patel, N., and Araujo, L. (2018). *Hands-On Convolutional Neural Networks with TensorFlow: Solve Computer Vision Problems with Modeling in TensorFlow and Python*. Packt Publishing.