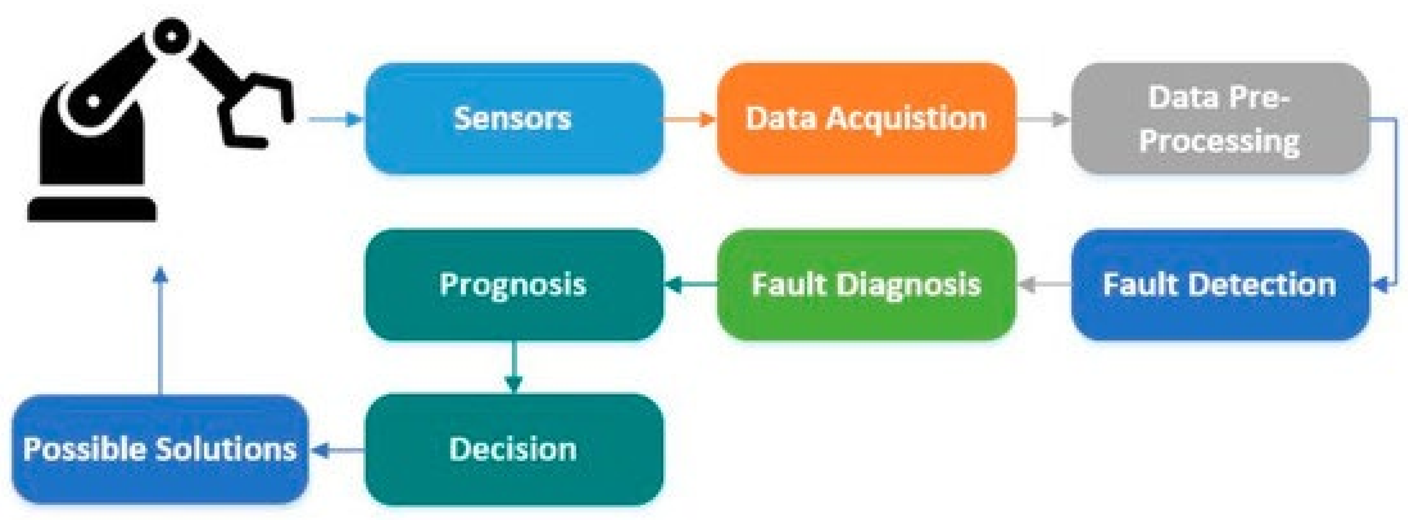
Transformers for Classification of Timeseries Data

Centre for Mechatronics and Hybrid Technologies (CMHT)
Department of Mechanical Engineering, McMaster University
Jonathan Wong, M.A.Sc Candidate; Dr. Saeid Habibi, Supervisor

EECOMOBILITY (ORF) & HEVPD&D CREATE

What is FDD?



FDD or **Fault Detection and Diagnosis** is a method in which data is analyzed in order to remedy possible failures in a mechanical system. Outcomes of applying this method can produce prognosis' at a system-level for important decision-making in the industry.

There are many types of methods that can be used for fault detection: data-driven methods which utilize artificial intelligence and machine learning algorithms, model-based methods which simulate the system at a component level, and signal processing methods which transforms and augment the output from sensor data directly. We will focus on a **data-driven method** for deep learning as well as **signal-based processing**.

Transformer Architecture

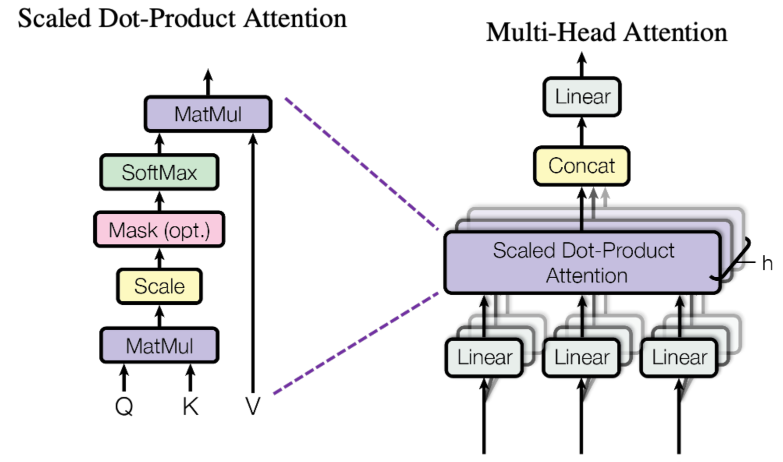


Figure 2: Multi-Head Attention Block

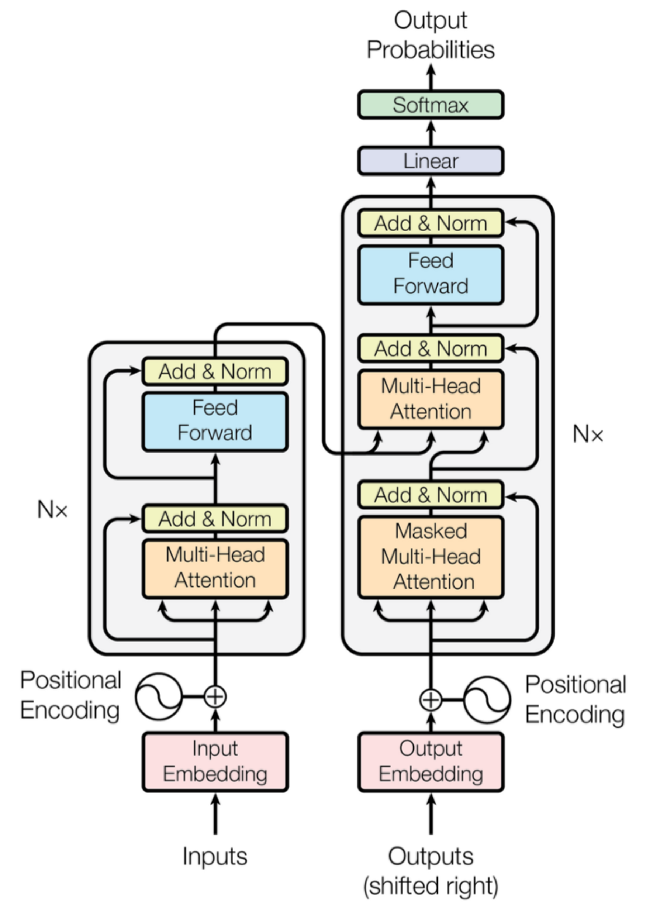


Figure 1: The Transformer - model architecture.

The Transformer Neural Network is a deep-learning model that utilizes "**attention**" in order to store memory and link sequences together. It was originally developed by Google in 2017 and published with a paper titled, "**Attention is All You Need**". With this new type of NN that rivals the performance of LSTMs and RNNs, a new **state-of-the-art** technology was born. Traditionally used for Natural Language Processing, it is now adapted to tackle different types of problems such as detecting faults in an electric motor.

Proposed Network Design

Parameter	Decision
Data Split (Train/Validate/Test)	80/10/10
Optimizer	Adam
Loss Function	Sparse Categorical Cross Entropy
Epochs	363 pre-tuning, 31 post-tuning = 394 total
Activation Function	ReLU
Input Layer Shape	[(None, 480, 1)]
Output Layer Shape	(None, 8)

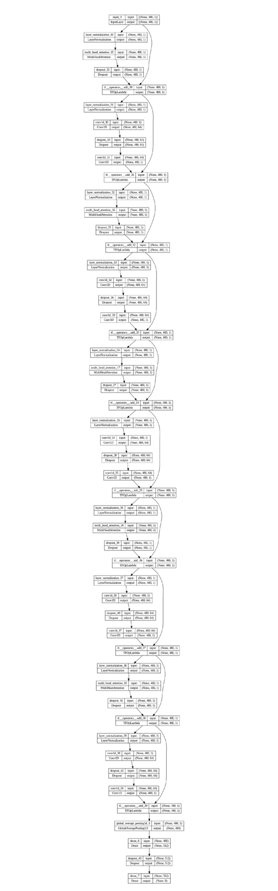
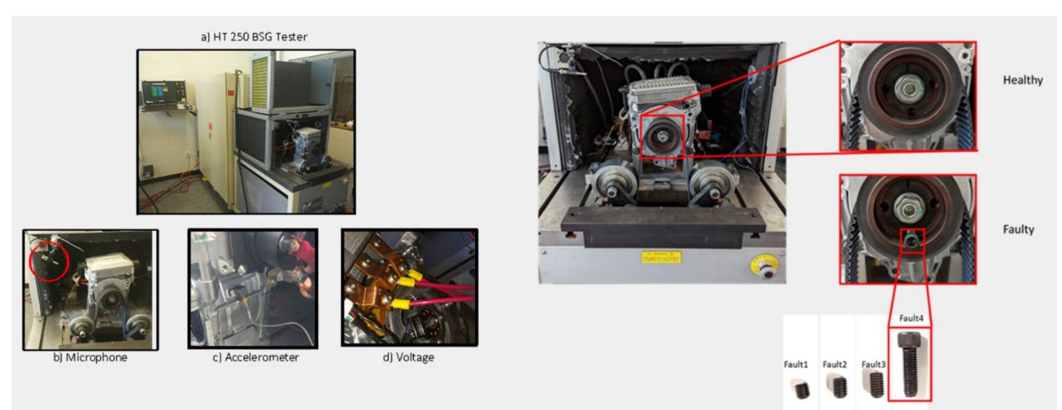


Figure 3: Proposed Transformer Architecture

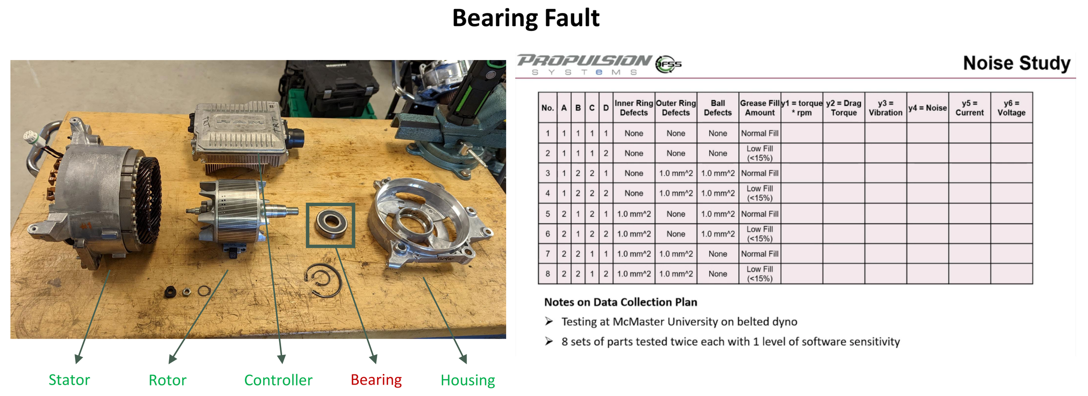
Thousands of parameters are optimized during training of the transformer model but there are some parameters we can adjust that will be used to define the model: Optimizer Type = Adam, Loss Function = sparse categorical cross entropy, Tuning Algorithm = Hyperband. These choices help the model become more **computationally efficient** and **increase accuracy**.

Mass Imbalance & Bearing Faults

Mass Imbalance Fault



Bearing Fault



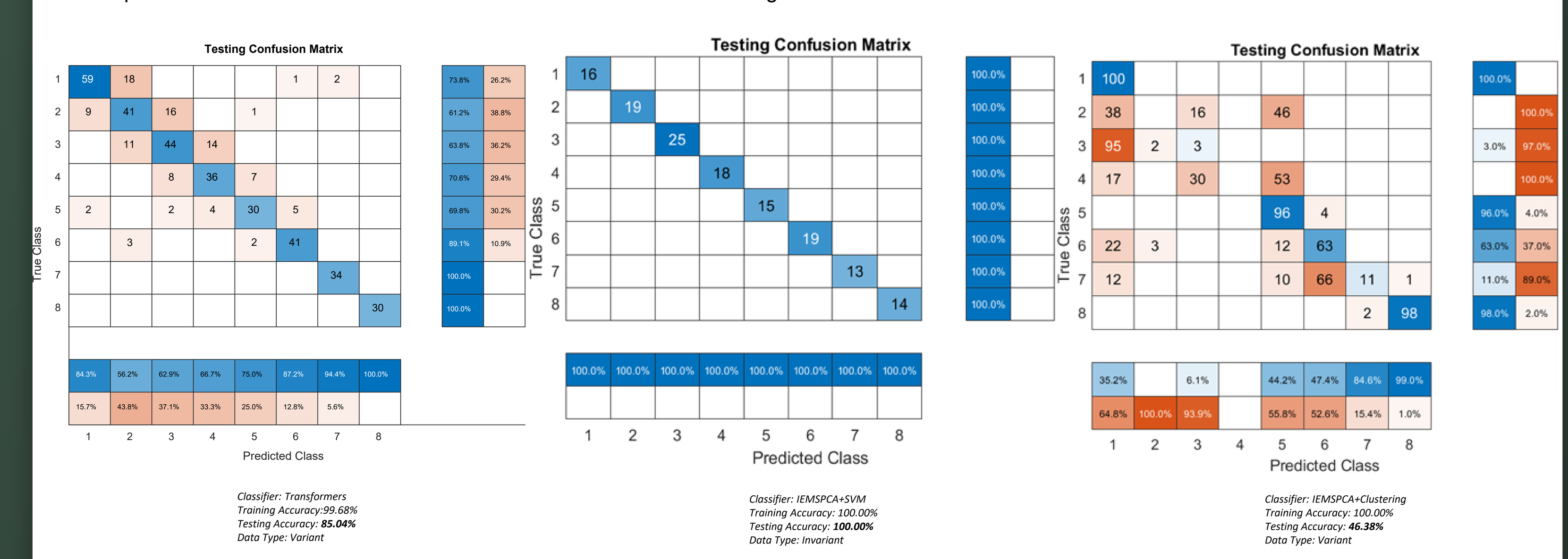
Experiments were conducted at CMHT to collect test data for a series of faults: The two main groups of tests are categorized as Mass Imbalance and Bearing faults. Both experiments were carried out with the same apparatus and setup: a **Stellantis eTorque belted starter generator** is mounted on a **D&V Electronics HT-250 Tester**. Power is supplied to the electric motor in which the output torque is measured by the dynamometer. An assortment of sensors are affixed to the test bench including: 2x microphones, 2x accelerometers, and 3x current/voltage sensors. Each test run iterates through **pre-determined setpoints** of RPM and Torque supplied by the motor. Data from the sensors are stored in the on-board computer.

In **Experiment #1**, a set screw or bolt of varying size Is inserted into the threaded hole of the pulley. It is asymmetrical in shape so that the mass becomes imbalanced during rotation. The size of the bolts range from **1.200g - 21.996g**.

In **Experiment #2**, we investigate different types of bearing faults such as **reduced lubrication, inner, outer, and ball bearing surface defects**. Each test begins with a teardown of the BSG and etching a mark onto the affected surface that is to be tested.

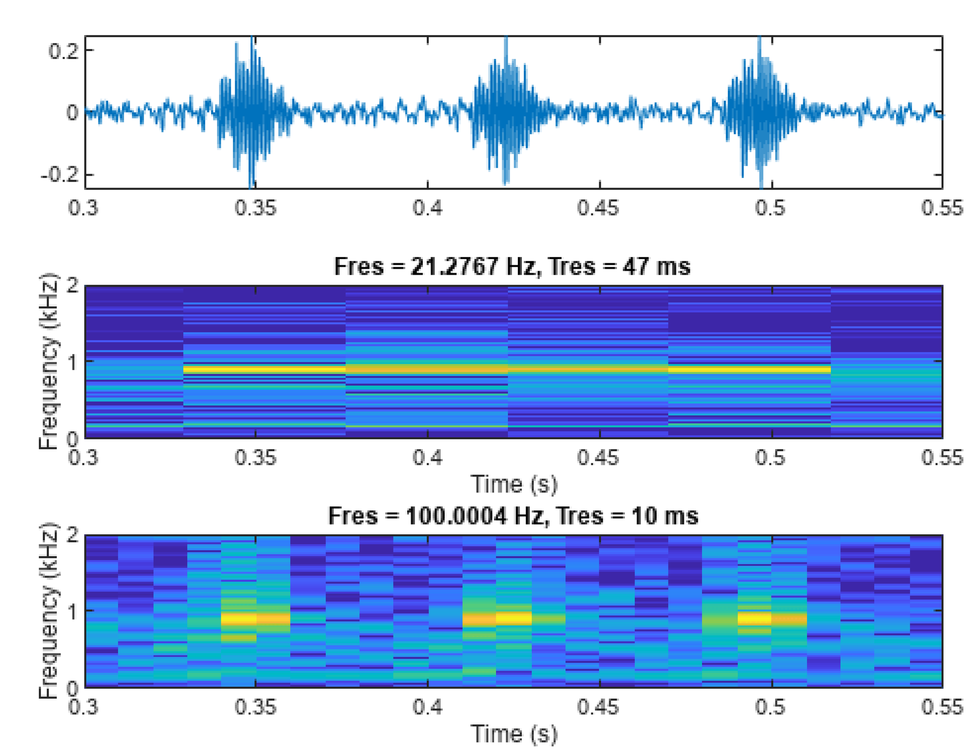
Initial Results

Comparison of Transformers vs IEMSPCA+SVM vs IEMSPCA+Clustering



The initial results from testing of Experiment #1 show a **high degree of accuracy** in the test data. In an AUC chart, the model has been measured to have an excellent measure of **separability of 0.97**. Although the IEMSPCA+SVM method has excellent training and testing accuracy, the datatype used for that model was invariant. Finally a clustering method with IEMSPCA feature selection performed well in training but not as strong in the testing stage. Work is currently being conducted to test the model on the dataset from Experiment #2.

Time-Frequency Analysis



Many ecological, epidemiological, and physical data records come in the form of time series. A time series is a **sequence of observations** recorded at a succession of time intervals. The value of the series at some time t is generally not independent of its value at, say, t−1 . We use **specialized statistics** to analyze time series and **specialized data structures** to represent them in Python. Time series data commonly show **periodic behavior**. This periodic behavior can be very complex. **Spectral analysis** is a technique that allows us to discover underlying periodicities. To perform spectral analysis, we first must transform data from **time domain to frequency domain**.

McMaster University · University of Windsor · Western · CMHT Centre for Mechatronics and Hybrid Technologies · Ontario Research Fund

We acknowledge the support of the Ontario Research Fund: Research Excellence Program.

NSERC CRSNG

We acknowledge the support of the Natural Sciences and Engineering Council of Canada (NSERC), which invests annually over $1 billion in people, discovery and innovation.