Transformer-Based Signal Inference for Electrified Vehicle Powertrains

Stan Muñoz Gutiérrez ✉ 🏠
Institute of Software Technology, TU Graz, Austria

Adil Mukhtar ✉
Institute of Software Technology, TU Graz, Austria

Franz Wotawa[1] ✉ 🏠
Institute of Software Technology, TU Graz, Austria

Abstract

The scarcity of labeled data for intelligent diagnosis of non-linear technical systems is a common problem for developing robust and reliable real-world applications. Several deep learning approaches have been developed to address this challenge, including self-supervised learning, representation learning, and transfer learning. Due largely to their powerful attention mechanisms, transformers excel at capturing long-term dependencies across multichannel and multi-modal signals in sequential data, making them suitable candidates for time series modeling. Despite their potential, studies applying transformers for diagnostic functions, especially in signal reconstruction through representation learning, remain limited. This paper aims to narrow this gap by identifying the requirements and potential of transformer self-attention mechanisms for developing auto-associative inference engines that learn exclusively from healthy behavioral data. We apply a transformer backbone for signal reconstruction using simulated data from a simplified powertrain. Feedback from these experiments, and the reviewed evidence from the literature, allows us to conclude that autoencoder and autoregressive approaches are potentiated by transformers.

2012 ACM Subject Classification Hardware → Fault models and test metrics; Computing methodologies → Machine learning algorithms; Hardware → Failure recovery, maintenance and self-repair

Keywords and phrases Signal Inference, Deep Learning, Self-Supervised Learning, Multimodal Transformer Autoencoder, Electric Vehicle, Powertrain, Electric Motor

Digital Object Identifier 10.4230/OASIcs.DX.2024.29

Category Short Paper

Supplementary Material *Software (MATLAB Notebook)*: `https://zenodo.org/records/13879822`

Funding This study was conducted within the framework of the project ARCHIMEDES, receiving funding from the Key Digital Technologies Joint Undertaking (KDT JU) - the Public-Private Partnership for research, development, and innovation under Horizon Europe – and National Authorities under Grant Agreement No 101112295 funded by the European Union and the FFG under Grant Agreement No FO999899377.

1 Introduction

Data-driven diagnosis solutions rely on characterizations of the target system's healthy, and sometimes abnormal, observed behavior. Machine learning models are usually preferred when diagnosis tasks are needed for systems exhibiting partial observability, high dimensionality, non-linear dynamics, temporal multiscale dependencies, delays, low signal-to-noise ratio, and high epistemic uncertainty. Deep learning methods excel at addressing challenging diagnosis

[1] corresponding author

35th International Conference on Principles of Diagnosis and Resilient Systems (DX 2024).
Editors: Ingo Pill, Avraham Natan, and Franz Wotawa; Article No. 29; pp. 29:1–29:14
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

problems, provided sufficient data is available, and a proper methodology is followed for their training and validation. In 2017, the transformer architecture[36] marked a paradigm shift regarding the handling of sequential data in deep learning applications. Attention-based models have since become central to developing natural language processing (NLP) solutions, with transformers positioning themselves as state-of-the-art technologies. They outperform long short-term memory (LSTM) networks, and recurrent neural networks (RNNs) in various tasks, including language translation, text classification, and natural language reasoning. Their performance is primarily attributed to their attention mechanisms, which enable them to exploit intricate contextual temporal dependencies of sequential behavior.

The performance of transformer architectures extends beyond natural language processing (NLP) applications. According to the taxonomy identified by Islam et al. [14], there are 5 main transformer-based application domains: NLP, computer vision (CV), multi-modality (MM), audio/speech (A&S), and signal processing (SP)[2]. Research in the MM, A&S, and SP domains is important to diagnosis solutions because they handle digital signals as inputs rather than text. CV category is relevant too, as it is common practice to transform signals into images for deep learning applications in diagnosis. Notably, transformers are increasingly being considered for medical diagnosis based on biomedical images or physiological signals.

Our research objective is to design transformer models that enable fault detection and diagnosis using self-supervised learning. Our contributions can be summarized as:

- Identification of pre-processing steps necessary to input the transformer model.
- Identification of meta-requirements for a development framework for transformer-based intelligent diagnosis of electric drives.
- A proof of concept transformer autoencoder architecture for multimodal signal inference.
- A discussion on the suitability of these models for deployment in electrified powertrains.

2 Background and Literature Review

2.1 Learning Modalities

Machine learning (ML) techniques are increasingly being adopted in various scientific disciplines [17] and industrial applications [1]. The key characteristic of ML techniques is the ability to extract and learn *linear* and *non-linear* decision boundaries from within the observational data, i.e., feature space, of an underlying system. Hence, the objective is to approximate a learning function $F(\cdot)$ for mapping the decision, such as classification or regression, through predictor variables, called feature space $\{x_1, x_2, x_3, ..., x_n\}$. Different learning modalities exist, including *supervised* [23, 16], *unsupervised* [34], and *semi-supervised* [35, 26]. Supervised methods primarily require annotated labels for each single observation [16] in the form of input/output pairs, whereas, unsupervised methods heuristically infer the model of the unique representation(s) present in the data and do not require prior labels. Semi-supervised techniques can exploit both unlabeled and label data. In our case, we focus on *self-supervised* learning, a type of unsupervised learning that seeks to uncover the inherent structural dependencies within the data samples. Additionally, our model is based on the concept of *autoencoders*, which will be introduced next.

[2] The classification is based on a comprehensive literature review that selected 641 works. The relative number of papers shows that applications outside NLP are significant: NLP (40.0%, 257 papers), CV (30.7%, 197 papers), MM (14.7%, 94 papers), A&S (10.9%, 70 papers), and SP (3.6%, 23 papers). Some examples of applications outside the NLP domain include image detection, segmentation and classification, visual automatic captions, speech recognition, and medical signal processing.

2.2 Autoencoders

In 1993, Hinton and Zemel [13] understood that unsupervised learning algorithms can be seen as variants of PCA or vector quantization and that the *autoencoder* model (AE) encompasses both. In a recent literature review, Yang et al. [39] presents the foundations of *representational learning* from the perspective of the autoencoder family, covering their contribution to the field of intelligent fault diagnosis. The architecture of autoencoders entails dimensionality reduction, which aligns well with Hinton and Zemel's theoretical formulation that interprets them as solving a minimum-length description problem. When implemented using deep learning models, they are conceptually divided into two components: the *encoder* and the *decoder*. The encoder is typically depicted as a funnel-like structure that progressively projects a high-dimensional input into a lower-dimensional space, also referred to as the *latent space*; the decoder takes this representation and projects it back to the original input space, a process referred to as *reconstruction*. Learning algorithms for the parameters of such structures are driven primarily by a loss function designed to minimize the reconstruction error. A well-accepted taxonomy of AEs categorizes them into (1) denoising, (2) sparse, (3) contractive, and (4) variational. We will only introduce denoising autoencoders here, the reader can find more about the other categories in [27, 15, 39].

In [37], Vincent et al. pose the question of what constitutes a good representation. In answering this question, they come out with desiderata that we rephrase as follows: (1) preserve as much information as possible about the input, (2) embedding in a *pre-defined* representational space, (3) invariance to partial destruction of the input, and (4) recall from partial observable inputs. Therefore, denoising autoencoders (DAEs) are nuanced and explicit about requiring *robust representations* and capturing the internal structural associations needed to solve the *fill-in-the-blanks* problem. The learning in DAEs is based on a simple but powerful principle, *the autoencoder should repair partially corrupted inputs*.

2.3 Transformers for anomaly detection in time series

In [29], Shin et al. obtained state-of-the-art anomaly detection performance in time series using *AnoFormer*, a pure transformer-based generative adversarial network framework. Self-supervised training was performed using a two-stage masking strategy. In the first stage, masks were randomly obtained from a predefined pool of masks, then in the second stage, an entropy-based resampling strategy was applied. In their framework, Shin et al. work directly in the time domain, building a set of tokens from a quantized version of the normalized signal. A matrix-based linear embedding is automatically learned, followed by standard sinusoidal positional encoding. The training was adversarial, using only normal behavior data. The generator and the discriminator (critic) were implemented using transformer encoders. Once the generator is trained, the decision about normal or abnormal behavior is based on the generator's reconstruction error of the input signal. AnoFormer outperformed BeatGAN [43] (adversarial reconstruction-based CNN model); TadGAN, RAE-Ensemble, and RAMED (RNN-based models); as well as Anomaly Transformer (transformer-based) for the four datasets in the benchmark when evaluated using AUROC, AUPRC, and F1-score.

2.4 Attention-based diagnostic models for electric motors

In [18], Li et al. analyze the performance of a deep learning architecture enhanced with an attention mechanism when diagnosing bearing faults. They apply an attention-as-weighing approach, where a positive weight captures the relative importance of each input segment. High-level representations of the input are computed by a deep learning architecture that stacks convolutional layers followed by a long short-term memory (LSTM) model with a fully connected layer at the output end. Attention weights are computed by feedforward single-layer applied to the high-level representations. Finally, the high-level representations of the segments are linearly combined using the attention weights generating the inputs used for the classification stage that is implemented with a softmax regression layer. They analyzed three different input formats, (1) raw vibration signal, (2) envelope spectrum, and (3) frequency spectrum, and benchmarked their architecture against traditional deep neural network architectures with and without their attention mechanism in place. Attention mechanisms enhanced diagnosis performance across input formats and model variants.

Transformer-based diagnosis of mechanical faults has been demonstrated by Ding et al. [10]. Their study engineered a transformer architecture for rolling bearing diagnosis that selects as inputs the synchrosqueezed wavelet transforms (SWT) of vibration signals to capture time-frequency information that is then split into sequential fragments. Using each sequential fragment as input, a tokenization process was applied to obtain high-dimensional embeddings of the time-frequency fragments by applying learnable linear mappings and considering 1-D and 2-D learnable positional encodings. After the tokenization stage, a transformer encoder is trained to predict the next token, treating token characterization as an *autoregressive prediction problem.* Finally, the fault diagnosis is performed by a multilayer perceptron using one-hot encoding and a softmax layer at the output. Ding et al. benchmarked their model against a multilayer perceptron, a CNN, and a gated recurrent unit neural network (GRU). Their experiments show that their model outperforms the others across different operating conditions and signal-to-noise ratios.

3 Diagnosis of EVPs

3.1 Fault modes in motor systems

In [6], Chen et al. present a taxonomy of common fault types occurring in permanent magnet synchronous motors (PMSMs), which are the most prevalent type of motor used in powertrains of electrified road vehicles[41]. Faults are classified as electrical, mechanical, and magnetic. Faults obey "mutual catalytic" relationships among them, meaning that manifested faults are likely to trigger dependent faults. These dependent faults rapidly intensify each other, compromising the safety and integrity of the system and shortening its remaining useful life. Because of this, we are interested in the early detection of manifested faults also known as *incipient* faults, whose detection and diagnosis are difficult due to their low energy spread-out frequency components [18].

Ullah and Hur [30] review the detection and identification of inter-turn short (-circuit) faults (ITSFs) and irreversible demagnetization faults (IDFs) occurring in permanent magnet (PM) type machines. ITSFs are caused by insulation damage in the winding of PM motors causing a short circuit in the same phase of the machine [3]. Meanwhile, IDFs are due to the nonlinear high sensitivity of PMs to their operating temperature, overload conditions, mechanical stressors, reverse magnetic fields resulting from an ITSFs, and aging. Ullah and Hur identity fault indices based on: (1) stator current (SC), (2) stator voltage (SV), (3) parameter estimation (PE), (4) magnetic flux (MF), and (5) mechanical output (MO).

[3] ITSFs are one type of winding insulation faults, other types include coil-to-coil (CCFs), phase-to-phase (PPSFs), and phase-to-ground (PGSFs) [9].

3.2 Sensors

Prognosis and health monitoring (PHM) applications often require sensors that monitor modalities independent of the system's primary functional requirements. For instance, consider the use of thermal sensors in various powertrain components. These sensors are safety-critical and provide key parameters for models of their *physics of degradation*, thus contributing to an accurate estimation of the components' remaining useful life (RUL). In the case of Electric Vehicle Powertrains (EVPs), a combination of embedded and networked sensors should be selected, with some networked sensors capable of self-diagnosis.

Diagnosing sensor faults effectively is hard because they manifest in several ways, including bias sensor fault (BiasSF) [40, 42], cyclic sensor fault (CycSF) [40], erratic sensor fault (ErrSF) [40], gain fault on sensor (GFOS) [28], sensor abrupt fault (SAF) [40], sensor omission fault [3], stuck sensor fault (StckSF) [40], and spike sensor fault (SpkSF) [40].

3.3 Signal modalities

3.3.1 Electric

SV signals are used extensively in fault-tolerant control. It relies on measuring the voltage at the neutral point of the stator windings and the DC midpoint of the inverter. The zero-sequence-voltage components (ZSVCs) are informative of anomalies because they are unaffected by the drive on healthy motors, yet, they are highly coupled with the speed when an ITSF manifests. In particular, the first harmonic that can be efficiently computed by the Vold-Kalman filtering order tracking (VKF-OT) algorithm is suitable to be used as the basis for ITSFs indicators [33, 32]. SC signals are the inputs to many algorithms for fault detection of ITSFs and are the concern of the machine current signature analysis (MCSA) field. MCSA uses different spectral characterizations based on Hilbert-Huang, Wavelet, or Fourier transformations. Ullah and Hur [30], report on several AI techniques that take SC inputs and have shown to be successful in implementing fault detection, diagnosis, and severity assessment of ITSFs: [22], [24], [25]. In particular, [25] and [22] were designed to assess the severity of faults using a particle swarm optimization (PSO) algorithm, the former, and a multilayer perception (MLP), the latter. Both [25] and [24] can disambiguate ITSFs from load fluctuations across different operating conditions.

3.3.2 Vibration

In [44], Zou et al. diagnoses ball faults (BallFs), inner raceway faults (IRFs), and outer raceway faults (ORFs) from vibration signals of gearboxes. They treat vibration signals with Ensemble empirical mode decomposition (EEMD). EEMD conducts an ensemble of experimental trials, where independent Gaussian noise signals are added to the observed vibration signal. After decomposing each one of the signals of the ensemble using empirical mode decomposition (EMD) into the intrinsic mode functions IMFs, the method combines the signals within each one of the IMFs, the effect is that the Gaussian signals cancel each other out, but also increase the signal-to-noise ratio. EEMD. They apply a *kurtosis filter*, consisting of an IMF selection mechanism that selects the k=6 IMFs that are most informative for the faults. An LSTM uses the 6 IMFs and can achieve accuracies of 99.98%. In [20], Li uses images with spectral information of vibration signals to diagnose rotor unbalance (RU), shaft misalignment (SM), BallFs, IRFs, and ORFs. Variational mode decomposition (VMD) was applied to the vibration signal. Then the intrinsic mode functions (IMFs) components were transformed using the Hilbert transformation. From these transformed components, the

Hilbert marginal spectrum was computed. Finally, this was transformed into an image that was the input to three different deep neural network models: ResNet101, GoogLeNet, and AlexNet. ResNet101 obtained the highest diagnostic accuracy (94%).

3.3.3 Acoustic

In the work of Choudhary et al. [8], CNN models have been trained to diagnose ORFs, IRFs, and BallFs using exclusively acoustic signals. Using *transfer learning* target also diagnosis of BRBFs, rotor misalignments (RMAs), and BFs. The accuracy in the target domain is close to that of the source domain (95.8% and 94.98% no load, 94.2% and 97.68% with load). In [12], Glowacz study diagnosis and severity assessment of faulty ring of squirrel-cage (FRSCs), and BRBFs (1 or to broken bars) using acoustic signals. He compares the performance of a nearest neighbor classifier (NNC), a MLP, and a word-code-based classification (WCBC). MLP outperformed the other approaches, exhibiting perfect accuracy.

3.4 Multi-signal

Multi-signals, meaning signals from the same modality monitoring subsystems with coupled dynamics, are informative of sensor faults. In [28], a current sensor fault in an inverter was diagnosed by estimating the gain from the three-phase currents, reference torque, and motor speed. After a feature selection step, the signals from all current sensors were used to diagnose the GFOS for the current sensor of a single phase of the inverter. In [7], Chen and Li performed sensor fusion from vibration signals, simultaneously acquired from three accelerometers monitoring a motor system. The inputs to the model are 18 indicators computed in the time and frequency domains. The model comprises a bank of 18 sparse autoencoders (SAEs). These SAEs compute a *sparse overcomplete representation* that is then fused into a single channel. The fused feature signals are the inputs of a deep belief network DBN that diagnoses IRFs and ORFs and assesses their severities.

3.5 Multimodal

Liang et al. [19], used single-phase currents and vibration signals to diagnose ITSFs. By applying sparse linear decomposition (SLD) and the algorithm orthogonal matching pursuit (OMP), they computed a representation from which a feature vector was obtained. An SVM was then trained to detect ITSFs in PMSMs. The SVM achieves perfect accuracy on a small testing dataset of 6 feature vectors (3 healthy, and 3 abnormal). Ullah et al. [31] integrate vibration and current signals to diagnose IDFs and BFs of PMSMs using a CNN (VGG-16 architecture). They convert both the FFT spectrum of current signals and the temporal vibration signals to 64×64 2D images with 3 channels (one for each modality and the third zero padded). Using this setting, they obtain 96.56% accuracy with real data from an experimental platform. In [21], Ma et al. integrate vibration and acoustic information using a deep coupling autoencoder trained with a loss function that uses information-theoretic measures of signal correlations. Their model diagnoses gearboxes (GB) defects and BFs, outperforming single modality and a variant based on simple signal stacking.

3.6 Input format

The input format of signals refers to the domain of representation. Some of the most widely used include the fast Fourier transform (FFT), wavelet transform (WT), wavelet packet decomposition (WPD), short-time Fourier transform, empirical mode decomposition (EMD), Hilbert transform (HT) and Hilbert-Huang transform (EMD + HT). Transient vibrations produce signals that are not stationary, the choice of representation can enhance or hinter conveying this information to the diagnostic model. We refer the reader to [11] for a detailed discussion and analysis of several of the most important input formats.

4 The signal inference transformer

Our signal inference transformer is part of the research supported by the project ARCHIMEDES [2], for the acceleration of the energy transition, safety, and security for the future society 5.0.

4.1 Metarequirements

The high-level requirements for the signal inference transformer can be summarized in the following list:

1. The deep learning model should be trained exclusively on data representative of the healthy behavior of the system learned in self-supervised learning mode.
2. Multimodal support for signals relevant to automotive powertrains of electrified vehicles.
3. Inference of signals should support the prediction of future scenarios given projected control signals and operational state.
4. Inference of whole signal channels (virtual sensors).
5. Explanations (signal reconstruction) of arbitrary channels.
6. Knowledge injection by simulation models that enable parameter estimation (relevant for fault detection and prognosis).
7. Knowledge injection for intelligent masking by definition of causal constraints.
8. A flexible open-source processing pipeline should be defined and implemented to automate as much as possible the development of the model.
9. Objective metrics should be automatically computed and human-legible reports in the spirit and facilitating quality assurance.
10. The framework facilitates downstream adaptation to implement several diagnostic functions via semi-supervised learning.

4.2 Conceptual solution

A conceptual diagram of the transformer-based signal inference solution is shown in Figure 1. The transformer is fed with a sequence that integrates the sequential presentation of multimodal sequential tokens augmented with control tokens. A key aspect of future research is the appropriate tokenization for the signals relevant to the diagnostic functions of electrified vehicle powertrains. The controller is a module that parameterizes the sequencer handling query tokens and should not be confused with the vehicle's controller signals[4].

5 Case Study

A model of a simplified powertrain [38] was used to generate simulation data with meaningful signals that could be used to prototype ideas and develop our processing pipeline and solution. The signals shown in Figure 1 are part of the simulation dataset.

[4] The vehicle's control signals are encoded as multimodal channels.

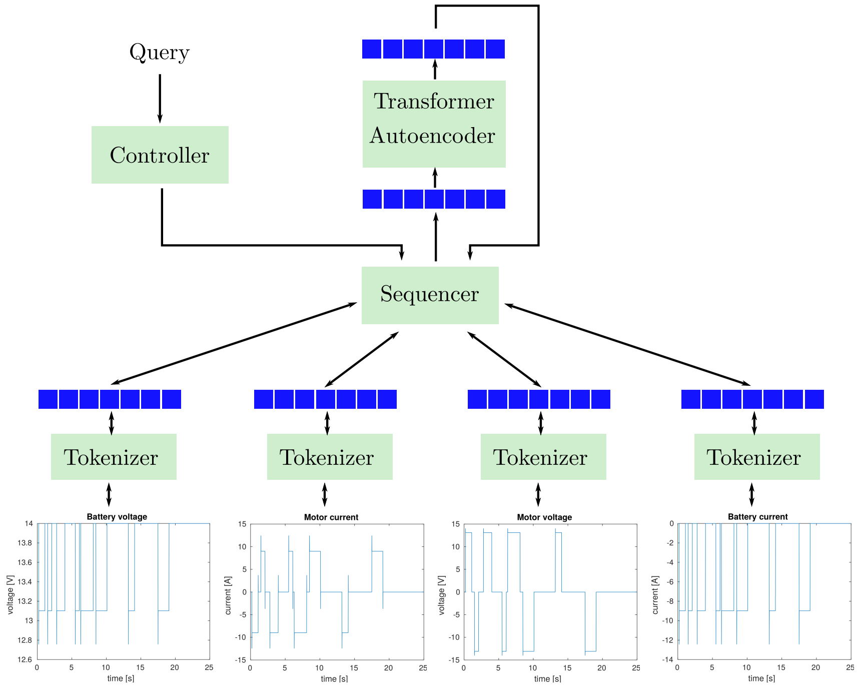


Figure 1 Transformer-based signal inference box.

5.1 Transformer architecture

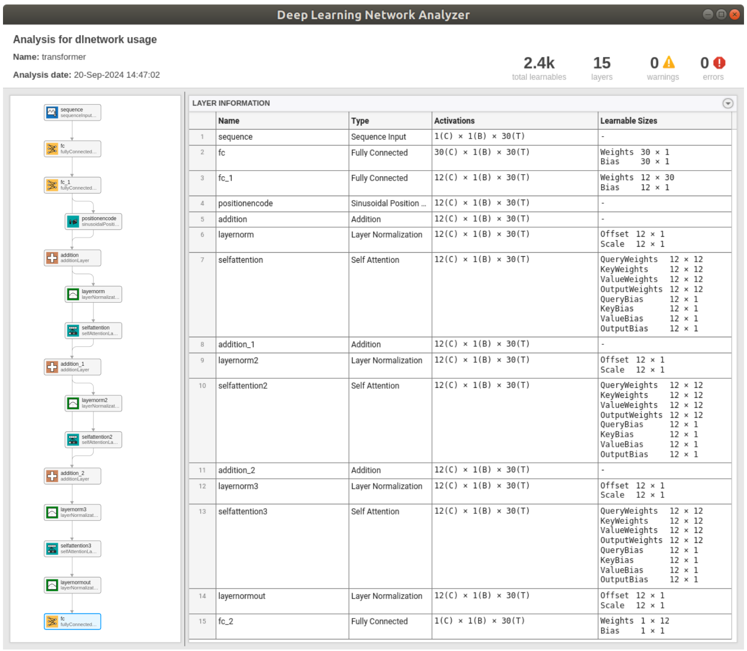
We evaluated the definition of an architecture in the platform MATLAB. Figure 2 shows one of the several examples of architectures we are exploring. These architectures comprise an autoencoder implemented with a transformer encoder backbone. For anomaly and fault-detection applications, it is often sufficient to have this type of architecture, therefore we are starting there. The number of modules in the transformer encoder is a customizable parameter. We have found encouraging results, even with naive representations for the tokens, such as representing masks with negative constants.

5.2 Self-supervised approach

The training of the autoencoder follows the DAE methodology, although we don't want to refer only to masking as the corruption of information for the transformer to denoise but would like to have a richer set of control tokens in place. We found encouraging applications in the any-to-any models that are capable of inference across several modalities[4]. We are considering pre-training schemes to assist with the transformer encoder before plugging it into the decoder (see Figure 1 in Bao et al. [5]).

5.2.1 The processing pipeline

In Figure 3, we illustrate our preprocessing pipeline with an example. The computer simulation model signals were acquired at 10 kHz. The first preprocessing step consists of scaling and downsampling the signals, in the future model, several temporal scales need to be accounted for. The next step splits the signal into windows (10 ms length in the example), that can potentially overlap (50% in the example). The input sequence is obtained from the window by stacking and interleaving samples so that synchronous samples remain



Analysis for dlnetwork usage

Name: transformer

Analysis date: 20-Sep-2024 14:47:02

2.4k total learnables · 15 layers · 0 ⚠ warnings · 0 ❗ errors

LAYER INFORMATION

	Name	Type	Activations	Learnable Sizes
1	sequence	Sequence Input	1(C) × 1(B) × 30(T)	-
2	fc	Fully Connected	30(C) × 1(B) × 30(T)	Weights 30 × 1 Bias 30 × 1
3	fc_1	Fully Connected	12(C) × 1(B) × 30(T)	Weights 12 × 30 Bias 12 × 1
4	positionencode	Sinusoidal Position ...	12(C) × 1(B) × 30(T)	-
5	addition	Addition	12(C) × 1(B) × 30(T)	-
6	layernorm	Layer Normalization	12(C) × 1(B) × 30(T)	Offset 12 × 1 Scale 12 × 1
7	selfattention	Self Attention	12(C) × 1(B) × 30(T)	QueryWeights 12 × 12 KeyWeights 12 × 12 ValueWeights 12 × 12 OutputWeights 12 × 12 QueryBias 12 × 1 KeyBias 12 × 1 ValueBias 12 × 1 OutputBias 12 × 1
8	addition_1	Addition	12(C) × 1(B) × 30(T)	-
9	layernorm2	Layer Normalization	12(C) × 1(B) × 30(T)	Offset 12 × 1 Scale 12 × 1
10	selfattention2	Self Attention	12(C) × 1(B) × 30(T)	QueryWeights 12 × 12 KeyWeights 12 × 12 ValueWeights 12 × 12 OutputWeights 12 × 12 QueryBias 12 × 1 KeyBias 12 × 1 ValueBias 12 × 1 OutputBias 12 × 1
11	addition_2	Addition	12(C) × 1(B) × 30(T)	-
12	layernorm3	Layer Normalization	12(C) × 1(B) × 30(T)	Offset 12 × 1 Scale 12 × 1
13	selfattention3	Self Attention	12(C) × 1(B) × 30(T)	QueryWeights 12 × 12 KeyWeights 12 × 12 ValueWeights 12 × 12 OutputWeights 12 × 12 QueryBias 12 × 1 KeyBias 12 × 1 ValueBias 12 × 1 OutputBias 12 × 1
14	layernormout	Layer Normalization	12(C) × 1(B) × 30(T)	Offset 12 × 1 Scale 12 × 1
15	fc_2	Fully Connected	1(C) × 1(B) × 30(T)	Weights 1 × 12 Bias 1 × 1

Figure 2 A potential transformer encoder architecture.

close to each other. Note that we do not use multiple channels for different signals because that would treat them as the same token, and we would lose resolution in the queries. For simulation data, the system stays in steady-states most of the time. Using raw window data would result in highly imbalanced datasets. To overcome this, we run a *uniqueness filter* (UF) with a tolerance factor (0.001 for the example). After this step, for the 25 s run shown in the plots, the 4999 windows shown in subfigure (4) became only 48 (subfigure (5)). It is interesting to note that by overlaying the windows in the original signals, the windows happen to capture the signal transitions only (subfigure (A)).

5.3 Datasets generation

To avoid *data leakage*, we don't partition the training, validation, and testing sets at the window level because each central window overlaps with its two neighbors. Knowing that the unique windows are scarce, we instead computed *segments of contiguous windows* (SCWs). In the example, there are 15 such segments (one segment can be seen in Figure 3) (B). We define training, validation, and test ratios as 0.6, 0.2, and 0.2 respectively. After random sampling without replacement, in the example, 9 segments were allocated for training, 3 for validation, and 3 for testing.

5.4 Self-supervised learning by masking

We trained the system by presenting random masked versions of the training samples at the input and the corresponding complete (unmasked) sequence at the output. The loss function is the mean squared (reconstruction) error (MSE).

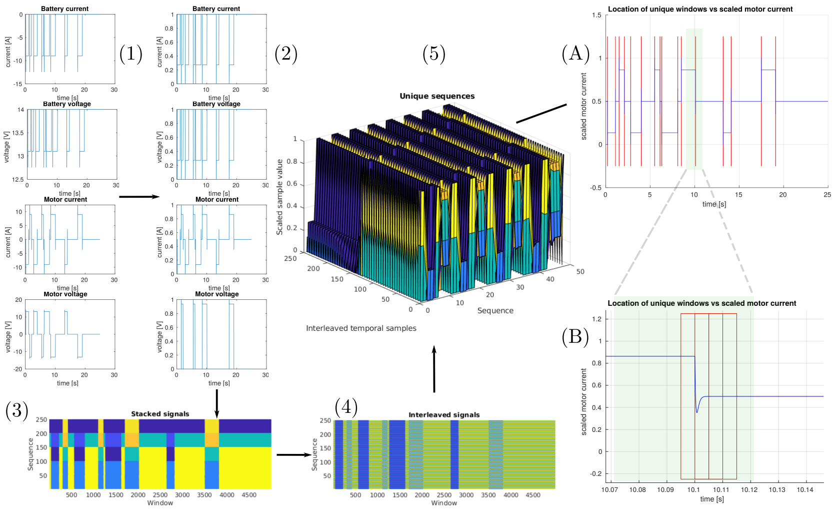


Figure 3 Preprocessing pipeline: (1) raw signals (h-bridge signal is omitted), (2) downsampling and scaling, (3) window extraction and stacking, (4) signal interleaving, (5) unique windows; (A) overlay of the unique windows on the motor current, and (B) zoomed region showing a segment of three overlapped consecutive windows.

5.5 Preliminary results

Figure 4 shows a histogram of reconstruction errors for the test dataset, after 196 epochs of training (MSE: training 0.0034, validation 0.0042, testing 0.0091). We provide an open-source interactive notebook in supplementary material with all parameters used in the example.

6 Discussion and future work

The progressive refinement of the AE concept from an information-theoretic perspective to that of a deep self-associative machine is opening new avenues for addressing challenging problems in PHM. Evidence from the literature supports the notion that integrating multimodal information enhances diagnostic performance. Attending to that conclusion, we have added multimodality to our desiderata. Although we are not yet in a position to present concrete results in this direction, the research community has shown transformers to be able to handle tens of modalities [4]. *Tokenization* is key to this aim, as provides a *unified representation* digestible by a transformer architecture. Predicting future scenarios is becoming relevant in the context of driving automation and eco-driving. The functional requirements in our desiderata are ambitious but follow naturally from achieving accurate learning of deep structural dependencies in the signals.

Although a simple idea, learning by filling in the blanks requires a deeper understanding of how to properly dose and structure those blanks. An advantage of diagnosing technical systems is that our machines have stringent specifications, and are designed and evaluated by mathematical and computational models. We added *Knowledge injection by simulation models* to our requirements because it introduces interesting possibilities. For the transformer architecture, it does not matter if the signal is learned from synthetic data or real-world data. Simulated data can be fed into the system during training and serve as a baseline for real-world data, while also addressing scenarios for rare events that might never be observed in actual data. Moreover, we understand *qualitatively* the dependencies among subcomponents and to some degree of the signals. This is an advantage as compared with

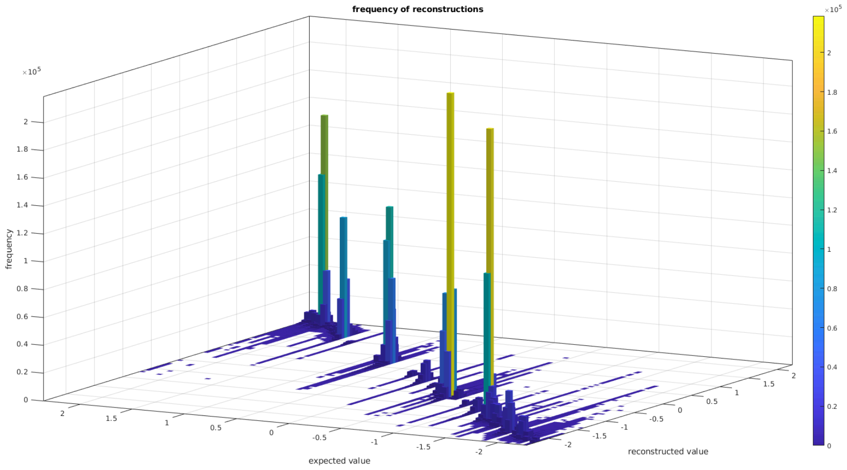


Figure 4 Preliminary results: (a) Histogram of reconstruction errors for the test dataset.

other fields because masking does not need to try so many combinations but a few *engineered* causal links. In other words, we don't need to start from *tabula rasa*, by incorporating engineering knowledge and constraints, the transformer autoencoder can be made more reliable by eliminating spurious dependencies not distinguishable from the data alone.

Having finite dictionaries aligns well with the problem of handling big data in modern vehicles, compacting the requirements and bandwidths. The challenge consists of optimizing the tokenization process while satisfying the desiderata of Vincent et al. [37]. Compressing is concomitant to reducing the energy footprint, another good aspect of these approaches. Digital twins in the context of Industry 4.0 and Society 5.0 would greatly benefit from standardized dictionaries, task-agnostic pre-trained and optimized embeddings, and open pre-trained foundational transformer models.

An important consideration for the application of transformer-based signal inference is the computational cost and energy demands associated with these technologies. It remains unclear whether current advancements are mature and efficient enough to be viable and environmentally friendly for onboard implementation in electric vehicles. Although numerous studies are exploring the feasibility of bringing transformers to edge computing for IoT, the timeline for these approaches to become mainstream remains uncertain. Nevertheless, we recognize other promising use cases where signal inference systems could offer significant value. For instance, simulations are integral to the qualification process in automotive development. The potential of these technologies to create mock digital twins could greatly enhance evaluation procedures by incorporating generative capabilities. This is particularly relevant given the proven effectiveness of variational autoencoders in other domains, which could be adapted to generate subsystem behaviors during development, thereby strengthening the overall quality assurance process.

7 Final remarks

For centuries, we have accumulated extensive knowledge in electric drive engineering. This knowledge, encapsulated in mathematical and computational models, enables us to make remarkably accurate predictions of the expected healthy behavior of electric drives across a wide range of conditions within their specified operational envelope.

While the physics of failure is well understood for most common faults, emerging technologies challenge our grasp of the highly coupled, nonlinear multiphysics resulting from the deep integration of electrical and mechanical components in electric vehicle powertrains.

A wide array of classical and data-driven techniques exists for detecting and diagnosing manifested faults in these systems. However, a significant gap remains in incorporating prognostics, partly due to the limited understanding and validation of theoretical models of hybrid high-power electronics degradation and the accumulated aging of motor components as a function of their operational history across varying environmental conditions.

Although transformer architectures are relatively new to the deep learning paradigm, they hold promise for unifying existing knowledge captured by computational models with real vehicle operational data to deliver the adaptability and learning required for the condition monitoring and prognosis of highly automated vehicles. Our conceptual multimodal transformer autoencoder seeks to introduce a foundational technology that can catalyze further advancements through its high adaptability while maintaining safety and quality standards in electrified vehicle powertrains.

References

1 Sheena Angra and Sachin Ahuja. Machine learning and its applications: A review. In *2017 international conference on big data analytics and computational intelligence (ICBDAC)*, pages 57–60. IEEE, 2017.

2 ARCHIMEDES Consortium. ARCHIMEDES ensures lifetime with digital means. `https://www.archimedesproject.eu/`, 2024. Accessed on September 9, 2024.

3 Ritik Argawal, Dattatraya Kalel, M Harshit, Arun D Domnic, and R Raja Singh. Sensor Fault Detection using Machine Learning Technique for Automobile Drive Applications. In *2021 National Power Electronics Conference (NPEC)*, pages 1–6, 2021. `doi:10.1109/NPEC52100.2021.9672546`.

4 Roman Bachmann, Oğuzhan Fatih Kar, David Mizrahi, Ali Garjani, Mingfei Gao, David Griffiths, Jiaming Hu, Afshin Dehghan, and Amir Zamir. 4M-21: An Any-to-Any Vision Model for Tens of Tasks and Modalities, 2024. `doi:10.48550/arXiv.2406.09406`.

5 Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. Beit: Bert pre-training of image transformers, 2022. `doi:10.48550/arXiv.2106.08254`.

6 Yong Chen, Siyuan Liang, Wanfu Li, Hong Liang, and Chengdong Wang. Faults and Diagnosis Methods of Permanent Magnet Synchronous Motors: A Review. *Applied Sciences*, 9(10):2116, May 2019. `doi:10.3390/app9102116`.

7 Zhuyun Chen and Weihua Li. Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network. *IEEE Transactions on Instrumentation and Measurement*, 66(7):1693–1702, 2017. `doi:10.1109/TIM.2017.2669947`.

8 Anurag Choudhary, Tauheed Mian, Shahab Fatima, and B K Panigrahi. Deep Transfer Learning Based Fault Diagnosis of Electric Vehicle Motor. In *2022 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*, pages 1–6. IEEE, December 2022. `doi:10.1109/PEDES56012.2022.10080274`.

9 Yao Da, Xiaodong Shi, and Mahesh Krishnamurthy. A New Approach to Fault Diagnostics for Permanent Magnet Synchronous Machines Using Electromagnetic Signature Analysis. *IEEE Transactions on Power Electronics*, 28(8):4104–4112, 2013. `doi:10.1109/TPEL.2012.2227808`.

10 Yifei Ding, Minping Jia, Qiuhua Miao, and Yudong Cao. A novel time–frequency Transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings. *Mechanical Systems and Signal Processing*, 168:108616, 2022. `doi:10.1016/j.ymssp.2021.108616`.

11 Robert X Gao and Ruqiang Yan. Non-stationary signal processing for bearing health monitoring. *International journal of manufacturing research*, 1(1):18–40, 2006. `doi:10.1504/IJMR.2006.010701`.

12 Adam Glowacz. Acoustic based fault diagnosis of three-phase induction motor. *Applied Acoustics*, 137:82–89, 2018. doi:10.1016/j.apacoust.2018.03.010.

13 Geoffrey E Hinton and Richard Zemel. Autoencoders, minimum description length and helmholtz free energy. *Advances in neural information processing systems*, 6, 1993.

14 Saidul Islam, Hanae Elmekki, Ahmed Elsebai, Jamal Bentahar, Nagat Drawel, Gaith Rjoub, and Witold Pedrycz. A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications*, 241:122666, 2024. doi:10.1016/j.eswa.2023.122666.

15 Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

16 Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007. URL: http://www.booksonline.iospress.nl/Content/View.aspx?piid=6950.

17 Yogesh Kumar, Komalpreet Kaur, and Gurpreet Singh. Machine learning aspects and its applications towards different research areas. In *2020 International conference on computation, automation and knowledge management (ICCAKM)*, pages 150–156. IEEE, 2020.

18 Xiang Li, Wei Zhang, and Qian Ding. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Processing*, 161:136–154, 2019. doi:10.1016/j.sigpro.2019.03.019.

19 Siyuan Liang, Yong Chen, Hong Liang, and Xu Li. Sparse Representation and SVM Diagnosis Method Inter-Turn Short-Circuit Fault in PMSM. *Applied Sciences*, 9(2):224, January 2019. doi:10.3390/app9020224.

20 Shih-Lin Lin. Application Combining VMD and ResNet101 in Intelligent Diagnosis of Motor Faults. *Sensors (Basel, Switzerland)*, 21(18):6065, 2021. doi:10.3390/S21186065.

21 Meng Ma, Chuang Sun, and Xuefeng Chen. Deep Coupling Autoencoder for Fault Diagnosis With Multimodal Sensory Data. *IEEE Transactions on Industrial Informatics*, 14(3):1137–1145, 2018. doi:10.1109/TII.2018.2793246.

22 S.S. Moosavi, A. Djerdir, Y. Ait-Amirat, and D.A. Khaburi. ANN based fault diagnosis of permanent magnet synchronous motor under stator winding shorted turn. *Electric Power Systems Research*, 125:67–82, 2015. doi:10.1016/j.epsr.2015.03.024.

23 Vladimir Nasteski. An overview of the supervised machine learning methods. *Horizons. b*, 4(51-62):56, 2017.

24 Yaw Nyanteh, Chris Edrington, Sanjeev Srivastava, and David Cartes. Application of artificial intelligence to real-time fault detection in permanent-magnet synchronous machines. *IEEE Transactions on Industry Applications*, 49(3):1205–1214, 2013. doi:10.1109/TIA.2013.2253081.

25 Yaw D. Nyanteh, Sanjeev K. Srivastava, Chris S. Edrington, and David A. Cartes. Application of artificial intelligence to stator winding fault diagnosis in permanent magnet synchronous machines. *Electric Power Systems Research*, 103:201–213, 2013. doi:10.1016/j.epsr.2013.05.018.

26 YCAP Reddy, P Viswanath, and B Eswara Reddy. Semi-supervised learning: A brief review. *Int. J. Eng. Technol*, 7(1.8):81, 2018.

27 Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. Contractive auto-encoders: explicit invariance during feature extraction. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 833–840, Madison, WI, USA, 2011. Omnipress. URL: https://icml.cc/2011/papers/455_icmlpaper.pdf.

28 Tunan Shen, Yuping Chen, Christian Thulfaut, and Hans-Christian Reuss. A Data Based Diagnostic Method for Current Sensor Fault in Permanent Magnet Synchronous Motors (PMSM). In *IECON Proceedings (Industrial Electronics Conference)*, volume 2019-Octob, pages 5979–5985, 2019. doi:10.1109/IECON.2019.8927667.

29 Ah-Hyung Shin, Seong Tae Kim, and Gyeong-Moon Park. Time Series Anomaly Detection Using Transformer-Based GAN With Two-Step Masking. *IEEE Access*, 11:74035–74047, 2023. doi:10.1109/ACCESS.2023.3289921.

30 Zia Ullah and Jin Hur. A Comprehensive Review of Winding Short Circuit Fault and Irreversible Demagnetization Fault Detection in PM Type Machines. *Energies*, 11(12):3309, November 2018. doi:10.3390/en11123309.

31 Zia Ullah, Bilal Ahmad Lodhi, and Jin Hur. Detection and Identification of Demagnetization and Bearing Faults in PMSM Using Transfer Learning-Based VGG. *Energies (Basel)*, 13(15):3834, 2020.

32 Julio-César Urresty, Jordi-Roger Riba, and Luis Romeral. Diagnosis of interturn faults in pmsms operating under nonstationary conditions by applying order tracking filtering. *IEEE Transactions on Power Electronics*, 28(1):507–515, 2013. doi:10.1109/TPEL.2012.2198077.

33 Julio-César Urresty, Jordi-Roger Riba, and Luís Romeral. Application of the zero-sequence voltage component to detect stator winding inter-turn faults in PMSMs. *Electric Power Systems Research*, 89:38–44, 2012. doi:10.1016/j.epsr.2012.02.012.

34 Muhammad Usama, Junaid Qadir, Aunn Raza, Hunain Arif, Kok-Lim Alvin Yau, Yehia Elkhatib, Amir Hussain, and Ala Al-Fuqaha. Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE access*, 7:65579–65615, 2019. doi:10.1109/ACCESS.2019.2916648.

35 Jesper E Van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine learning*, 109(2):373–440, 2020. doi:10.1007/S10994-019-05855-6.

36 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. arXiv:1706.03762, doi:10.48550/arXiv.1706.03762.

37 Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1390156.1390294.

38 Franz Wotawa, David Kaufmann, Adil Mukhtar, Iulia Nica, Florian Klück, Hermann Felbinger, Petr Blaha, Matus Kozovsky, Zdenek Havranek, and Martin Dosedel. Real-Time Predictive Maintenance - Model-Based, Simulation-Based and Machine Learning Based Diagnosis. In *Artificial Intelligence for Digitising Industry – Applications*. River Publishers, 2011.

39 Zheng Yang, Binbin Xu, Wei Luo, and Fei Chen. Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review. *Measurement*, 189:110460, 2022. doi:10.1016/j.measurement.2021.110460.

40 Hongkun Zhang and Wenjun Li. A New Method of Sensor Fault Diagnosis Based on a Wavelet Packet Neural Network for Hybrid Electric Vehicles. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1143–1147, 2016. doi:10.1109/CISP-BMEI.2016.7852886.

41 Jiyu Zhang, Giorgio Rizzoni, and Qadeer Ahmed. Fault Modelling for Hierarchical Fault Diagnosis and Prognosis. In The American Society of Mechanical Engineers, editor, *Dynamic Systems and Control Conference*, volume 2, October 2013. doi:10.1115/DSCC2013-3825.

42 Chunheng Zhao, Yi Li, Matthew Wessner, Chinmay Rathod, and Pierluigi Pisu. Support-Vector Machine Approach for Robust Fault Diagnosis of Electric Vehicle Permanent Magnet Synchronous Motor. In *Annual Conference of the PHM Society*, volume 12(1), page 10, November 2020. doi:10.36001/phmconf.2020.v12i1.1291.

43 Bin Zhou, Shenghua Liu, Bryan Hooi, Xueqi Cheng, and Jing Ye. BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series. In *IJCAI*, volume 2019, pages 4433–4439, 2019. doi:10.24963/IJCAI.2019/616.

44 Ping Zou, Baocun Hou, Lei Jiang, and Zhenji Zhang. Bearing Fault Diagnosis Method Based on EEMD and LSTM. *International Journal of Computers, Communications and Control*, 15(1), February 2020. doi:10.15837/IJCCC.2020.1.3780.