# Deep Learning Pipeline for State-of-Health Classification of Electromagnetic Relays

Lucas Kirschbaum, Darius Roman, Valentin Robu, David Flynn
*Smart Systems Group (SSG)*
*Engineering and Physical Science, Heriot-Watt University*
Edinburgh EH14 4AS, U.K
lpk1@hw.ac.uk, dvr1@hw.ac.uk, v.robu@hw.ac.uk, d.flynn@hw.ac.uk

2021 IEEE 30th International Symposium on Industrial Electronics (ISIE) | 978-1-7281-9023-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIE45552.2021.9576278

***Abstract*—Industrial-scale component maintenance is shifting towards novel Predictive Maintenance (PdM) strategies supported by Big Data Analytics (BDA). This has resulted in an increased effort to implement Artificial Intelligence (AI) decision making into new maintenance paradigms. The transition of AI into industry faces significant challenges due to the inherent complexities of industrial operations, such as variability in components due to manufacturing, integration, dynamic operating environments and variable loading conditions. Therefore, AI in critical industrial systems requires more advanced capabilities such as robustness, scalability and verifiability. This paper presents the first Deep Learning (DL) based strategy for the classification of the State-Of-Health (SOH) of Electromagnetic Relays (EMR). The DL strategy scales with high-volumes of multivariate time-series data whilst automating labour intensive feature extraction requirements. The method proposed in our paper, combines a Convolutional-Auto-Encoder (CAE) with a Temporal Convolutional Neural Network (TCN), referred to as EMR-SOH CAE-TCN pipeline. Model uncertainty and SOH confidence bounds are approximated by Monte-Carlo dropout. Our pipeline is trained and evaluated on data generated from EMR life-cycle tests. We report a high classification accuracy and discriminatory power of the EMR-SOH classifier. The findings from our paper demonstrate the potential of AI pipelines for maintenance decision making of components in critical applications, providing a transferable AI based PdM solution that scales with large data quantities.**

***Index Terms*—Electromagnetic Relay, Predictive Maintenance, State-of-Health, Artificial Intelligence, Temporal Convolutional Neural Networks, Big Data Analytics**

## I. Introduction

Global demand for higher performance products and services from industrial systems, has led to a paradigm shift in maintenance strategies [1, 2, 3]. This trend has also been influenced by increasing safety compliance standards for existing and emergent systems [4, 5, 6], and a growth in industrial systems with coupled industrial networks e.g. transport and energy networks [7, 8]. Hence, companies are reverting to Predictive Maintenance (PdM) strategies to extract key decision making indicators, such as the State-Of-Health (SOH) or the Remaining Useful Life (RUL), from Big Data Analysis (BDA). Clearly, BDA represents an enabler for performance and economic enhancement from industrial systems, which will remain a predominant trend within the era of Industry 4.0. The omnipresence of BDA will be supported by a continuous integration of enabling technologies e.g. sensors, communications, processors [9].



BDA encompasses processing, pruning and pattern recognition in high volumes of multi-faceted data through statistics and Artificial Intelligence (AI). Besides prominent challenges such as reliable data management, streamlining and storage, a pressing aspect of BDA is the derivation of meaningful insights from multivariate time-series data in a PdM context. This entails: 1) dealing with high dimensional inputs in an efficient manner; 2) extracting relevant features, typically requiring expert knowledge; 3) improving utilisation of information from inter-sensor correlations; 4) long term dependencies and slow trends, often masked by noise - rendering many traditional time-series classification and forecasting methods unsuitable [10, 11]. Deep Learning (DL) is a promising candidate, as it supplies a scalable architectures for automatic feature extraction from raw data [12], and the ability to learn complex time dependent degradation processes [13]. Inter alia, solutions combining a set of DL techniques, capable of operating directly on large inputs while preserving sequential aspects have become a popular choice for classification and regression tasks alike presenting a variety of promising DL architectures. Predominantly, Convolutional Neural Networks (CNN), typically used for image recognition tasks [14, 15], have been adopted for time-series regression problems. CNN offer a scale-invariant method to extract local and global, spatial features, considering inter-sensor correlation, operating directly on raw multivariate input data [16]. However, to learn temporal dependencies, CNN can be combined with an auto-regressive model, such as a Long-Short-Term Memory (LSTM) network [10, 17, 18]. To improve performance of CNN-LSTM architectures, Shao et al. [19] adopt a multi-input architecture, which considers a statistical feature vector derived from windowed time-series data. Although accurate for auto-regressive tasks, Recurrent Neural Networks based architectures require considerable computational resources particularly when dealing with high volume inputs and evince deficits in retaining extended long-term dependencies [11]. In contrast, Temporal Convolutional Neural Networks (TCN) extend traditional CNN's applicability to sequence classification or forecasting problems, exhibiting improvements in training time as well as performance [20]. TCN are multi-layer neural networks utilising causal 1D-convolution in combination with

dilation to exponentially increase the size of the receptive field [21]. Applications implementing TCN as a substitute for LSTM architectures in a PHM context have only recently been presented by e.g. Liu et. al. [22] and Chen et. al. [23].

As an exemplar of the aforementioned challenges, our study focuses on the use case of the Electromagnetic-Relay (EMR), an electro-mechanical component widely used for switching, control and signal shaping [24]. EMRs play an essential role in safety critical applications, e.g. the nuclear, aviation, or railway industry as they isolate the switching circuitry, suppressing any potential of leakage current [25, 26]. The mechanical life of an EMR can be magnitudes higher than the electrical. A contact failure due to contact contamination or electrical wear is the predominant failure mode and essentially determines the EMR life; concomitant contact resistance has been identified as a key degradation indicator [27]. However, accurately measuring the contact resistance in-situ has proven to be difficult as change is typically in the mΩ range. To date, the majority of research focuses on physical models to derive the degradation process and predict contact resistance for PdM [28]. However, due to the dynamic nature of the degradation process, which is highly dependent on the specific asset (EMR characteristics) and specifics of the application (environment and direct loading conditions), deriving accurate physical models remains challenging. To address this, BDA has been explored to provide the means of diagnosis and prediction of EMR failure. Guo et al. [29] deploys a DL framework to derive EMR life estimates based on a set of engineered features, inter alia the contact resistance. In [30] an alternative prognostics approach estimating RUL for safety critical EMR, based on contact resistance measurements is presented. A physical model and a data-driven approach show equally promising predictive performance, though the authors conclude that the latter one is computational more efficient and adaptable to changing operational and environmental parameters.

To date, there has been no research considering EMR-SOH using BDA for in-situ analysis. Prior, BDA findings heavily rely on reference contact resistance measurements as degradation indicators. The methodology of our paper, utilises a DL strategy processing high-volume multivariate time-series data, i.e. continuous, passive voltage and current measurements. A SOH class is estimated from the incoming raw sensory data using a CNN-Autoencoder (CAE) for feature extraction in combination with a subsequent Temporal Convolutional Neural Network (TCN) for class prediction. To quantify model uncertainty, the proposed approach employs a computational efficient approximation of the Gaussian Process, namely the Monte-Carlo dropout. The presented DL strategy is referred to as EMR-SOH CAE-TCN pipeline and is evaluated using run-to-failure EMR data obtained via life-cycle experiments. A dedicated test platform was designed and developed to allow various loading conditions to accelerate degradation, such as elevated ambient temperature, whilst collecting high volume multi-sensory data to support the training and validation of the proposed methodology.

The remainder of the paper is organised as follows: Section II presents the working principle of the EMR and introduces the EMR life-cycle test platform and data collection. The EMR-SOH CAE-TCN pipeline is detailed in Section III and evaluated in Section IV. The primary conclusions and further research is discussed in Section V.

## II. Electromagnetic Relay Design of Experiment - Life-Cycle Test Platform

In this research, an accelerated life-cycle test platform and design of experiment (DoE) supports evaluation of EMR degradation and failure modes. The platform derives an extensive and complete (run-to-failure) set of experimental data. The varying degradation patterns support the design of a robust data pipeline for SOH estimation. The test platform allows variable loading conditions, e.g. elevated ambient temperature, enabling a wide range of possible test patterns and replicating the variability in real-world operating conditions. To allow for timely data acquisition, the test platform can operate four different EMR life-cycle tests in parallel.

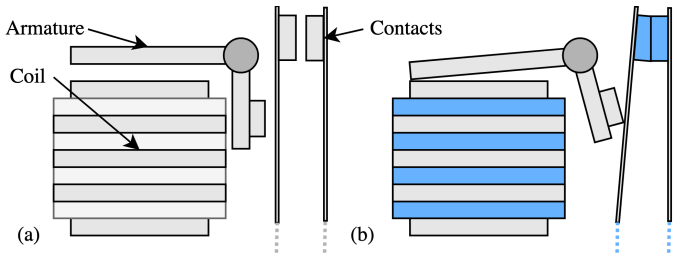


Fig. 1. Schematic of a SPST-NO EMR: (a) open contacts; (b) closed contacts.

### A. Electromagnetic Relay

EMR's are electro-mechanical components, consisting of electrical parts - such as the isolated input/ control circuit at the coil side - low voltage, low current - and mechanical contacts at the output circuit - high voltage, high current - refer to Fig. 1. The coil, an armature, a type of insulation, a pull-back spring, the contact strap and the contacts are the main parts of every EMR. The contacts are required to have a high electrical conductivity, while at the same time being resistant to mechanical wear and electrical erosion [24]. Contacts are typically either welded on the contact strap or designed as contact rivets. The contact material is typically copper with either a direct plating or sometimes an alloy layer in between. Typical plating materials are e.g. silver or gold [31].

For the purpose of this experiment a Single-Pole-Single-Throw Normally-Open (SPST-NO) EMR has been chosen. This type of EMR is commonly used in applications such as electrical protection circuits. The coil operates at 70 % of its specified 5 V rating at 106 mA. The contacts are rated at 10 A, 30 VDC. The contact resistance is 30 mΩ and the electrical durability is specified at min. 100,000 operations at a maximum of 1800 operations/hour. The EMR is not sealed but housed in a flux-protection casing. The rated maximum

TABLE I

EXPERIMENT DURATION LISTED AS NUMBER OF ACTUATIONS AND TIME TO FAILURE. DATA IS COMPRESSED BY LOWERING IT'S RESOLUTION.

| Experiment | Run#1 | | | | Run#2 | | | Run#3 | | | | | Run#4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Device | Dev#0 | Dev#1 | Dev#2 | Dev#3 | Dev#4 | Dev#5 | Dev#6 | Dev#7 | Dev#8 | Dev#9 | Dev#10 | Dev#11 | Dev#12 | Dev#13 | Dev#14 | Dev#15 |
| Actuations | 94190 | 120739 | 100868 | 109970 | 98968 | 174264 | 133251 | 162220 | 79826 | 77438 | 91543 | 102729 | 80014 | 198747 | 78968 | 113246 |
| Time to Failure [h] | 104.7 | 134.5 | 112.1 | 122.8 | 110.0 | 193.6 | 148.1 | 180.3 | 88.7 | 86.0 | 101.8 | 114.1 | 88.9 | 220.83 | 87.7 | 125.8 |
| Raw Data [GB] | 15.9 | 20.7 | 17.3 | 18.5 | 16.9 | 29.9 | 22.2 | 27.2 | 13.3 | 13.2 | 16.0 | 17.4 | 13.7 | 34.1 | 13.5 | 19.4 |
| Compressed [GB] | 5.1 | 5.6 | 5.5 | 5.9 | 5.4 | 9.5 | 7.2 | 8.7 | 4.3 | 4.2 | 5.1 | 5.5 | 4.3 | 10.7 | 4.3 | 6.1 |

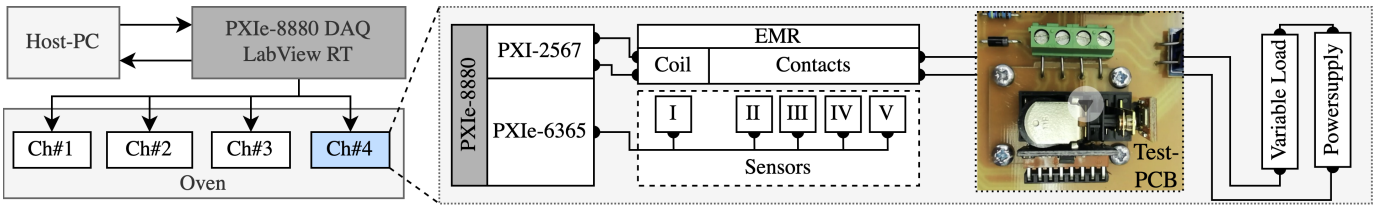


Fig. 2. An overview of the four channel EMR life-cycle experiment. Sensor measurements: (I) Coil Voltage, (II) Contact Voltage, (III) Contact Current, (IV) Infrared Contact Temperature, (V) Acoustic Emission (MEMS). Each channel is realized as an individual test-PCB, which allows a quick setup of the experiment and exchange of test components.

operation temperature is 70 °C. The contacts are copper-base, silver-plated rivets.

### B. Experimental Setup and Results

The experiment setup is depicted in Fig. 2. The test-stand is controlled using LabVIEW Real-Time and a National-Instruments PXIe8880 DAQ. The timely switching of the EMR is controlled via a PXI-2567 module. Data is sampled with a PXIe-6365 at 25 kHz for each actuation during the turn-on and turn-off phase for 5 ms. For each of the four test channels five measurements are taken, namely the *trigger-signal voltage*, *contact-side voltage*, *contact-side current*, *contact temperature* using an infrared temperature sensor, and *acoustic emission* during switching using a Micro-Electro-Mechanical Systems (MEMS) microphone. The test devices are placed in an oven to precisely control the ambient temperature. Two parallel variable resistors provide the load. In order to reduce the effects of load inductance, the resistors are set up with opposing winding.

Overall, 16 experiments were conducted at 60 °C, 30 VDC and 6 A, switching at 0.25 Hz/ 50% duty cycle which amounts to half of the rated maximum switching operations per hour. The results are displayed in Table I. All tested EMR failed due to the same failure-mode - *contacts stuck closed*, which was preceded by a sharp increase in contact surface temperature, as the contacts welded together. An extensive discussion relating to EMR failure modes can be found in [25].

## III. PROPOSED DEEP LEARNING PIPELINE

The EMR-SOH classification method is contained in a pipeline, a terminology referring to a sequence of hierarchical steps performed on each sample of the input data, i.e. the raw measured EMR multivariate time-series data. The pipeline's goal is to derive a prediction, which associates the input with a SOH class. As depicted in Fig. 3-(a) and Fig. 3-(b) the proposed pipeline differentiates between an offline and an online phase.

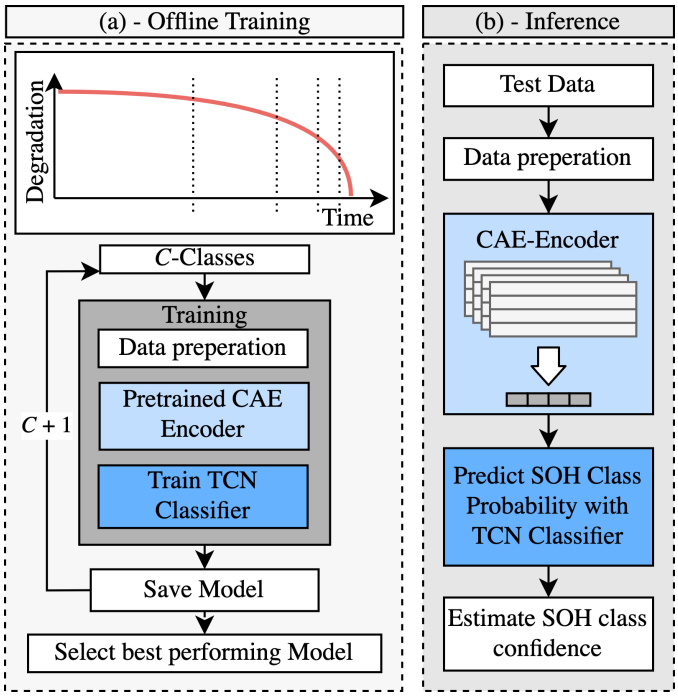


Fig. 3. Schematic of the proposed EMR Health-State Classifier.

During the *offline phase*, the model is trained, based on a set of historical life-cycle run-to-failure EMR data, which is segmented and pre-processed to act as input-data as further discussed in Section III-A. The employed DL architectures consists of a CAE for feature extraction, detailed in Section III-B, and TCN for sequence classification, as discussed in Section III-C. First, the CAE is pre-trained in an unsupervised manner to reduce the overall training time. Subsequently, the trained encoder-stage is refined in unison with the TCN. We consider the EMR-SOH estimation as a multi-class problem, by implementing a classification layer that transfers the model's prediction into class-probabilities. Therefore, a fully-

connected layer using *softmax* activation is employed, where the number of units is equal to the number of health state classes. Randomised-search is applied to determine the best model hyper-parameters and optimal number of hidden layers [32]. To reduce overfitting whilst improving generalisation dropout and regularisation, as well as rectified-linear-unit (ReLU) activation function for non-linearity are employed. Furthermore, different SOH class splits $C$ are reviewed. The *online phase* - inference - evaluates the performance of the trained model on the withheld test data. Each SOH class is estimated using the full DL pipeline. Now, utilising the Monte-Carlo dropout, we can determine the confidence for each class prediction, estimating the mean and the variance for each prediction, as explained in Section III-D. The final prediction supports the maintenance decision making during the entire EMR life-cycle. We further detail on each hierarchical step of the pipeline as follows.

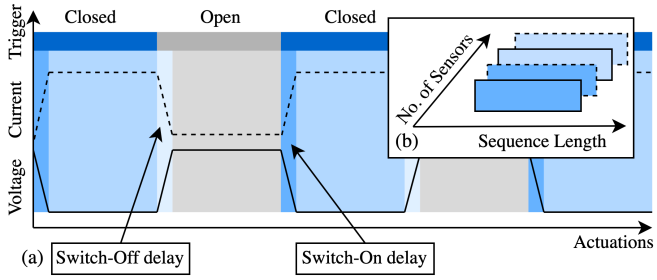


Fig. 4. (a) Schematic display of current and voltage on the EMR contact side during switching, closed - blue; open - grey; (b) the stacked channels of one actuation.

### A. Signal Segmentation and Data Preparation

Input data is pre-processed prior to training and inference, which entails windowing, scaling, and stacking of the measured sensor data. As shown in Fig. 4-(a) closing and opening the contacts is a mechanical process, which is not instantaneous. Data is recorded a short period prior to and during the transition - namely the *switch-on* and *switch-off* delay, i.e. the pick-up-time including the bounce-time and the release-time respectively [24, 29]. Therefore, we consider each actuation to consist out of four measurements as shown in Fig. 5: the current measurement during *switch-on* and *switch-off* and the voltage measurement during *switch-on* and *switch-off*. These four readings are stacked as a multi-channel 1D-vector, comparable to a four-channel image with a certain pixel-width and one-pixel height, refer to Fig. 4-(b).

### B. Feature Extraction

Automated feature extraction from the stacked raw multi-channel input data is accomplished using CAE, which is based on CNN. This DL architecture performs a set of subsequent operations on an input: (1) a linear operation, the convolution, slides a set of learnable kernels, i.e. its parameters are adjusted during training, over each position of the input. These convolutions are performed in parallel returning a set of output

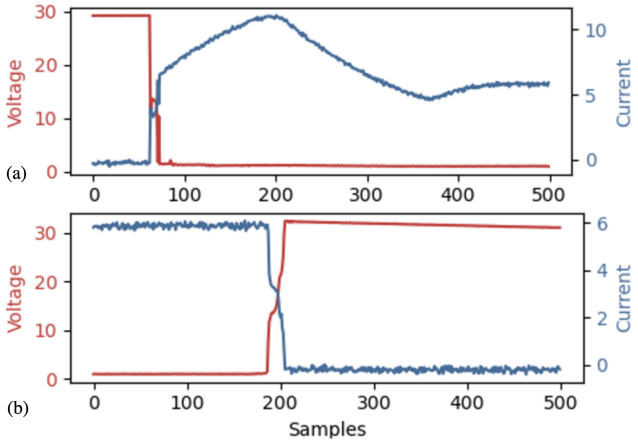


Fig. 5. Contact measurements: (a) Closing-Actuation during *switch-on* and (b) Opening-Actuation during *switch-off* recorded at 25 kHz.

features; (2) to increase non-linearity each feature map is activated using the ReLU activation function; (3) to improve invariance against small spatial changes in the input, i.e. translational invariance, pooling is applied, e.g. max-pooling [33].

Autoencoders (AE) are Neural Network (NN) structures, consisting of two building blocks: the *encoder* and the *decoder*. The (general) aim of the AE is to replicate it's input at the output. An under-complete AE constrains this copying task, i.e. learning the identity functions, as it encodes the input to a lower dimensional representation and respectively reconstruct this input from the latent space, trying to minimise a loss function [34]. The lower dimensional representation acts as a bottleneck, enforcing the encoder to learn the prominent features of the input. If non-linear activation functions are employed during encoding, the AE is capable of learning an expressive non-linear mapping of the input [33]. For a comprehensive study of AE's refer to [35].

To learn complex spatial non-linear feature-representation from large inputs, e.g. images or multivariate time-series data as in this instance, AE and CNN can be combined [36]. The so called CAE presents a computational efficient implementation, as weights are shared across the entire input. Multiple CAE layers can be stacked, which increases the complexity and robustness of the learned features [37].

### C. State-of-Health Classification

In order to predict the EMR-SOH class, we deploy TCN, which takes a sequence of the extracted features from CAE as input to predict the associate SOH. As the SOH, relates to an interval in the EMR life-cycle, one needs to consider the EMR degradation behaviour. Just like many other mechanical and electrical components, EMR exhibit slow long-term degradation patterns. An even segmentation of the entire degradation process could not account for an initial, long period depicting negligible wear. Hence, as proposed in [13], we consider a degradation model, which splits the SOH classes unevenly,

shortening towards the EOL. The threshold for the $c^{th}$ health state class $H_c$ of $C$ no. of classes is defined in Eq. 1.

$$H_c = 1 - 2^{-c} \;\; \{c \in \mathbb{Z}^+ | 0 \leq c \leq C\} \tag{1}$$

TCN is not a new DL architecture, but rather a combination of existing ones. It adapts CNN and incorporates mechanisms of RNN to retain sequence information [21]. The use of 1D-convolution layers, where each hidden layer has the same length as the input sequence, enables TCN to map each input to an output with the same length. TCN incorporates two, central structural DL elements, namely *Causal Dilated Convolution* and *Residual Blocks*, discussed as follows.

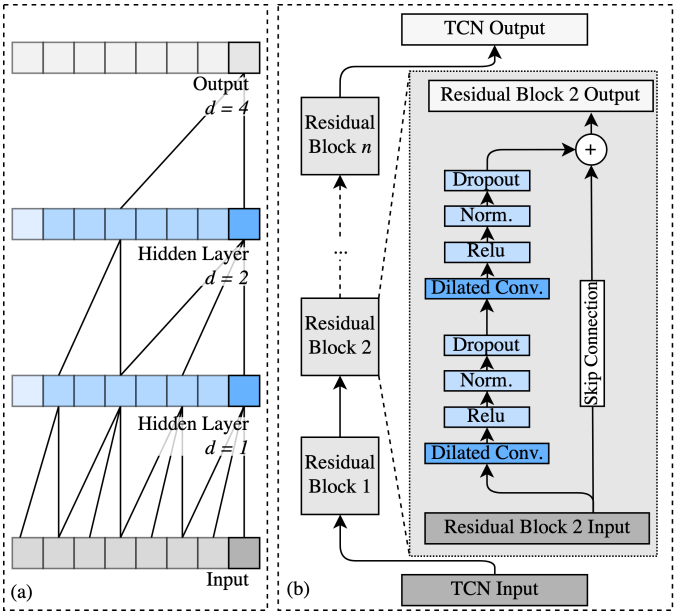


Fig. 6. The TCN DL architecture: (a) causal dilated convolution - dilation rate $d$; (b) the residual block architecture, where the output of one residual block is used as input for the following resiudal block.

*a) Causal Dilated Convolutions:* Causal convolutions ensure that no future information can affect past information - an output at time $t$ is based only on previous inputs $\leq t-1$. TCN applies causal dilated convolution. Without dilation, causal convolution scales the receptive field with the network depth linearly. Hence the considered sequence only grows linearly with each added layer - a substantial impediment when dealing with long sequences [38]. To address this issue, dilated causal convolution is applied - comparable with stride - allowing the receptive field to grow exponentially with the network depth [20]. Dilated convolution provides a method to extend the receptive field, while reducing the number of required parameters compared to simple convolution [39]. Ultimately, the size of the receptive field can be controlled with the dilation rate and the filter-size. Fig. 6-(a) demonstrates the working principle of dilated causal convolution.

*b) Residual Block:* TCN replaces the classical CNN block - convolution, activation, pooling, dropout - with a structural element referred to as *residual block*. A residual connection forks the layer input and performs a set of transformations which are then added to the input again. Such residual skip connections are used to stabilise the network - significant with increasing network depth in regards to the vanishing gradient problem [23]. Residual connections have been proposed by [40], adopted for TCN by [21] and further refined by [20]. As depicted in Fig. 6-(b), two causal dilated convolution operations with non-linearity (e.g. ReLU) are inserted, followed by normalisation to improve generalisation and spatial-dropout, i.e. channel-dropout, for regularisation.

### D. Uncertainty Estimation During Inference

To determine the model uncertainty, during inference, we employ Monte-Carlo dropout, previously known as model averaging. Gal and Ghahramani [41] derive a practical implementation to determine this epistemic uncertainty in DL by employing dropout during inference. Controlled by the dropout rate, a set of neurons are randomly dropped out at each forward pass, causing a different prediction at each guess. Hence, averaging over $K$ forward passes is similar to an ensemble of $K$ different individually learned NN. As the authors demonstrate, this can be interpreted as a Bayesian approximation of the Gaussian process. The predictive mean and the model uncertainty can usually be approximated with a relative small number of forward passes for each sample, i.e. $K \leq 1000$. It is proposed to implement the dropout-layer after each NN layer that contains trainable weights [42]. Therefore, in line with Bayesian Inference, the Monte Carlo Dropout estimates a distribution of the NN weights, rather than determining the optimal weight, to predict a posterior distribution of the target variable.

### E. Model Training and Scoring

Categorical cross-entropy is used as an objective function $L$; $h_c$ denotes the ground-truth, $f_s$ the *softmax* activation function and $p_c$ the networks prediction (Eq. 2).

$$L = -\sum_{c=1}^{C} h_c \log f_s(p_c) \tag{2}$$

The adaptive-moment-estimation (Adam) optimiser is applied [43]; an adaptive learning rate is used. We compare accuracy and f1-Score to evaluate the performance of the model on the test data. An exhaustive explanation of scoring methods used in classification can be found in [44].

## IV. Results

Table II depicts the model performance, evaluated on the test set, i.e. 30 % of the available data; the f1-Score and the joined accuracy for a three class model (I), a four class model (II) and a five class model (III) under consideration of different input sequence lengths, namely 10, 100 and 1000 actuations is shown. As the results indicate, the classifier is capable to distinguish SOH throughout the EMR life as well as towards the critical phase at the EOL. Adding classes affects the overall model performance, but returns a clearer differentiated SOH estimation for the last 25%, 12.5%, and 6.25% of

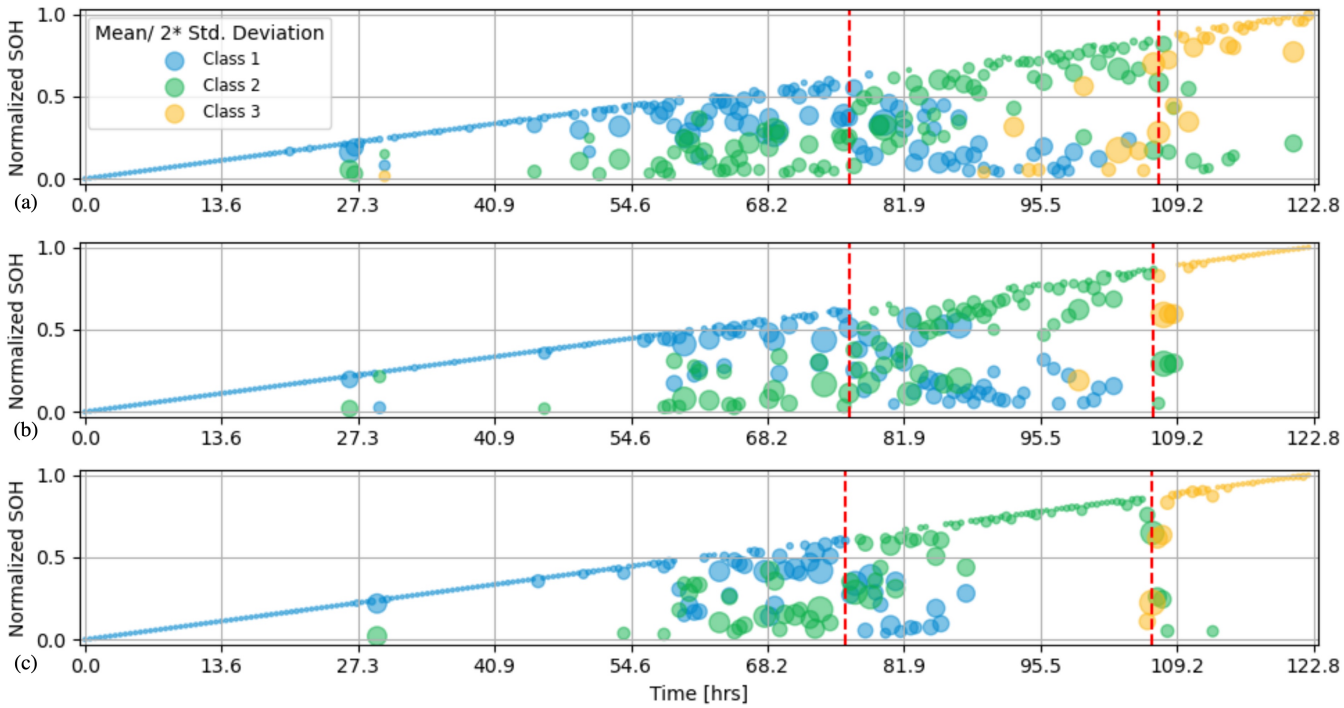


Fig. 7. The normalised SOH *Dev#3* classification results for the 3-Class EMR-SOH CAE-TCN. The dashed red line represents the true class splits. The predicted class is presented by the scatter, where the size resembles the models confidence in the prediction (small - high confidence; large - low confidence); Considered Sequence Length (a) 10 , (b) 100, (c) 1000.

TABLE II
THE EMR-SOH CAE-TCN PIPELINE RESULTS - F1-SCORE PER CLASS.

| Actuations | 10 | | | 100 | | | 1000 | | |
|---|---|---|---|---|---|---|---|---|---|
| Model | I | II | III | I | II | III | I | II | III |
| >50 % | 0.89 | 0.90 | 0.81 | 0.93 | 0.94 | 0.90 | 0.98 | 0.94 | 0.93 |
| <50 % | 0.78 | 0.73 | 0.56 | 0.86 | 0.77 | 0.72 | 0.96 | 0.93 | 0.83 |
| <25 % | 0.86 | 0.62 | 0.51 | 0.92 | 0.66 | 0.59 | 0.98 | 0.97 | 0.83 |
| <12.5 % | - | 0.75 | 0.42 | - | 0.78 | 0.52 | - | 0.98 | 0.93 |
| <6.25 % | - | - | 0.66 | - | - | 0.72 | - | - | 0.98 |
| Accuracy | 0.85 | 0.74 | 0.59 | 0.90 | 0.78 | 0.69 | 0.97 | 0.96 | 0.90 |

the EMR life, if the considered sequence length increases, i.e. considered number of actuations. However, for a shorter sequence length the class separability decreases significantly. The findings support our initial assumption, namely the EMR degradation does not follow a continuous decay, but rather an exponential trajectory, showing only very little indication for degradation in the first $75\%$ of its life. Therefore, treating the initial component degradation as a single SOH class seems reasonable. Increasing the sequence length, improves the model performance in all considered scenarios. However, such an increase comes at the cost of computational efficiency, as the number of model parameters increases.

Exemplarily we present the results for *Dev#3*. Therefore, to calculate the class SOH from the predicted class probabilities we consider the model's mean prediction per class at any point in the EMR life according to Eq. 3.

$$SOH_c(t) = P(H_c|t) * a(t) \tag{3}$$

where, $H_c$ is the class with e.g. $\forall H_c \in [1, 2, 3]$ and $a(t)$ denotes the number of passed actuations at time $t$. As Fig. 7 shows, the discriminatory power of the CAE-TCN increases towards the EOL. For class 3, the model exhibits a good performance for all shown models, displaying high confidence in its prediction for instances not belonging to this class and vice versa.

## V. CONCLUSION

This paper highlights a novel DL pipeline for EMR-SOH classification relying on TCN in combination with CAE for automated feature extraction from high-volume raw multivariate time series data, mitigating the need for expert-knowledge to derive features. The presented architecture is scalable with BDA due to the nature of the chosen DL algorithms; it integrates a computational efficient way to quantify the model uncertainty using Monte-Carlo dropout. The method was trained and evaluated on data aggregated through a developed, complete life-cycle EMR test platform. The results show that the model is able to predict EMR-SOH, aiding maintenance-decision making for such system critical components.

With respect to future work, we aim to advance this approach, fitting the model to a wider range of the recorded sensor data as well as evaluate further methods to determine the optimal data input format. Performance gains will be evaluated and bench-marked against existing DL architectures, with respect to model robustness in the presence of sensor noise, data resolution and under different loading conditions. The presented method will be extended to incorporate RUL forecast.

## VI. Acknowledgment

This work was supported by the EPSRC Centre for Doctoral Training in Embedded Intelligence, UK, under Grant EP/L014998/1, and by Baker Hughes, and the Lloyds Register Foundation under Grant AtRI100015.

## References

[1] M. Pecht and M. Kang, *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*. Wiley-IEEE Press, 2018.

[2] E. Miguelañez-Martin and D. Flynn, "Embedded intelligence supporting predictive asset management in the energy sector," in *Asset Management Conference 2015*, 2015, pp. 1–7.

[3] W. Tang, M. Andoni, V. Robu, and D. Flynn, "Accurately forecasting the health of energy system assets," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1–5.

[4] O. Zaki, M. Dunnigan, V. Robu, and D. Flynn, "Reliability and safety of autonomous systems based on semantic modelling for self-certification," *Robotics*, vol. 10, no. 1, 2021.

[5] X. Zhao, A. Banks, J. Sharp, V. Robu, D. Flynn, M. Fisher, and X. Huang, "A safety framework for critical systems utilising deep neural networks," in *International Conference on Computer Safety, Reliability, and Security*. Springer, 2020, pp. 244–259.

[6] M. Osborne, J. Lantair, Z. Shafiq, X. Zhao, V. Robu, D. Flynn, and J. Perry, "Uas operators safety and reliability survey: Emerging technologies towards the certification of autonomous uas," in *2019 4th International Conference on System Reliability and Safety (ICSRS)*. IEEE, 2019, pp. 203–212.

[7] D. Flynn, W. Tang, D. Roman, R. Dickie, and V. Robu, "Optimisation of hybrid energy systems for maritime vessels," in *9th International Conference on Power Electronics, Machines and Drives 2018*. Institution of Engineering and Technology, 2018.

[8] D. Roman, R. Dickie, V. Robu, and D. Flynn, "A review of the role of prognostics in predicting the remaining useful life of assets," *Safety and Reliability-Theory and Applications*, vol. 135, 2017.

[9] L. C. W. Kong, S. Harper, D. Mitchell, J. Blanche, T. Lim, and D. Flynn, "Interactive digital twins framework for asset management through internet," in *2020 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, 2020, pp. 1–7.

[10] Y. Chen and H. Rangwala, "Attention-based multi-task learning for sensor analytics," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 2187–2196.

[11] R. Wan, S. Mei, J. Wang, M. Liu, and F. Yang, "Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting," *Electronics*, vol. 8, no. 8, p. 876, 2019.

[12] K. Wang, Y. Zhao, Q. Xiong, M. Fan, G. Sun, L. Ma, and T. Liu, "Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals," *Scientific Programming*, vol. 2016, p. 9, 2016.

[13] M. Xia, T. Li, T. Shu, J. Wan, C. W. De Silva, and Z. Wang, "A two-stage approach for the remaining useful life prediction of bearings using deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3703–3711, 2018.

[14] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.

[15] A. Odo, S. McKenna, D. Flynn, and J. Vorstius, "Towards the automatic visual monitoring of electricity pylons from aerial images," in *VISAPP 2020: 15th International Conference on Computer Vision Theory and Applications*, 2020, pp. 566–573.

[16] G. S. Babu, P. Zhao, and X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *International conference on database systems for advanced applications*. Springer, 2016, pp. 214–228.

[17] T. Le, M. T. Vo, B. Vo, E. Hwang, S. Rho, and S. W. Baik, "Improving electric energy consumption prediction using cnn and bi-lstm," *Applied Sciences*, vol. 9, no. 20, p. 4237, 2019.

[18] Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, "Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid cnn with a lstm-ae based framework," *Sensors*, vol. 20, no. 5, p. 1399, 2020.

[19] X. Shao, C.-S. Kim, and P. Sontakke, "Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion cnn–lstm," *Energies*, vol. 13, no. 8, p. 1881, 2020.

[20] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[21] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[22] C. Liu, L. Zhang, and C. Wu, "Direct remaining useful life prediction for rolling bearing using temporal convolutional networks," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2019, pp. 2965–2971.

[23] J. Chen, D. Chen, and G. Liu, "Using temporal convolution network for remaining useful lifetime prediction," *Engineering Reports*, p. 12305, 2020.

[24] H. Sauer, *Modern relay technology*. Huethig, 1986.

[25] L. Kirschbaum, F. Dinmohammadi, D. Flynn, V. Robu, and M. Pecht, "Failure analysis informing embedded health monitoring of electromagnetic relays," in *2018 3rd International Conference on System Reliability and Safety (ICSRS)*. IEEE, 2018, pp. 261–267.

[26] C. R. Bell, "Remote condition monitoring of phase sensitive ac track circuits," in *5th IET Conference on Railway Condition Monitoring and Non-Destructive Testing (RCM 2011)*, 2011, pp. 1–7.

[27] P. G. Slade, *Electrical contacts: principles and applications*, 2nd ed. CRC press, 2017.

[28] J. Liu, M. Zhang, N. Zhao, and A. Chen, "A reliability assessment method for high speed train electromagnetic relays," *Energies*, vol. 11, no. 3, p. 652, 2018.

[29] J. Guo, G. Zhang, Y. Bi, and Y. Li, "Life prediction of automotive electromagnetic relay based on wavelets neural network," *Chemical Engineering Transactions*, vol. 62, pp. 1213–1218, 2017.

[30] A. J. Wileman and S. Perinpanayagam, "A prognostic framework for electromagnetic relay contacts," in *Proceedings of the 2nd Euro-pean Conference of the Prognostics and Health Management Society*, 2014.

[31] A. Ksiazkiewicz, G. Dombek, and K. Nowak, "Change in electric contact resistance of low-voltage relays affected by fault current," *Materials*, vol. 12, no. 13, p. 2166, 2019.

[32] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305, 2012.

[33] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[34] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59.

[35] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *arXiv preprint arXiv:2003.05991*, 2020.

[36] M. Chen, X. Shi, Y. Zhang, D. Wu, and M. Guizani, "Deep features learning for medical image analysis with convolutional autoencoder neural network," *IEEE Transactions on Big Data*, 2017.

[37] M. Maggipinto, C. Masiero, A. Beghi, and G. A. Susto, "A convolutional autoencoder approach for feature extraction in virtual metrology," *Procedia Manufacturing*, vol. 17, pp. 126–133, 2018.

[38] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, p. 523, 2019.

[39] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition. corr abs/1512.03385 (2015)," 2015.

[41] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International conference on machine learning*, 2016, pp. 1050–1059.

[42] R. Seoh, "Qualitative analysis of monte carlo dropout," *arXiv preprint arXiv:2007.01720*, 2020.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

[44] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognition Letters*, vol. 30, no. 1, pp. 27–38, 2009.