# Implementation of ANN for PMSM interturn short-circuit detection in the embedded system

KOZOVSKÝ, M.; BUCHTA, L.; BLAHA, P.

IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society

eISBN: 979-8-3503-3182-0

DOI: https://doi.org/10.1109/IECON51785.2023.10312642

Accepted manuscript

# Implementation of ANN for PMSM interturn short-circuit detection in the embedded system

1st Matus Kozovsky
*Central European Institute of Technology*
*Brno University of Technology*
Brno, Czech Republic
matus.kozovsky@ceitec.vutbr.cz

2nd Ludek Buchta
*Central European Institute of Technology*
*Brno University of Technology*
Brno, Czech Republic
ludek.buchta@ceitec.vutbr.cz

3rd Petr Blaha
*Central European Institute of Technology*
*Brno University of Technology*
Brno, Czech Republic
petr.blaha@ceitec.vutbr.cz

***Abstract*—The problem of condition monitoring and fault detection in powertrain systems becomes more critical with the increasing use of fail-operational systems. These systems are essential in the automotive industry, robotics, and other industrial applications. One of the critical features of such a system is recognizing the fault and suppressing its influence. The paper describes a feed-forward artificial neural network-based diagnostic of interturn short-circuit faults in a dual three-phase permanent magnet synchronous motor. The paper focuses on using multi-layer perceptron network (MLPN) and convolutional neural network (CNN) for interturn short-circuit detection and, more importantly, their real implementation into the automotive AURIX TC397 microcontroller. The paper presents the achieved neural network inference times as well as data preprocessing computation time. The behavior of the artificial neural networks (ANNs) is tested on an experimental configurable multi-phase permanent magnet synchronous motor (PMSM) with the possibility to emulate interturn short-circuit fault using prepared winding taps. The paper includes the essential aspects that should be respected during ANN design and implementation into the microcontroller.**

***Index Terms*—Neural network, fault detection, diagnostic, PMSM, motor**

## I. Introduction

Fail-safe and fail-operational powertrain systems are frequently discussed in recent publications. Reliable propulsion systems with high efficiency are essential in industrial applications, robotics, and transport applications, including the automotive industry. Fail-safe and fail-operational systems are usually closely related to the redundancy in the system. This presumption leads to complex systems with a more complicated control strategy. One of the promising structures is a dual three-phase propulsion system which, assuming correctly chosen motor parameters and control algorithm, can operate under fault conditions [1], [2]. The diagnostic system must also be integrated into the control algorithm in such a case. The control algorithm should use diagnostic data and adjust the control strategy depending on the detected fault.

An extensive study related to fault diagnostic and fault-tolerant control [3] shows recent trends and the most common methods used for fault detection. The study describes mechanical, magnetic, and electrical failures in permanent magnet synchronous (PMS) motors. Fault detection methods are divided into signal analysis fault detection methods, model-based fault detection methods, and artificial intelligence-based methods.

Electrical faults are mainly detected from stator currents waveforms. These are often transformed into rotor coordinates. Signal analysis is divided into frequency domain methods and time domain methods. The major drawback of frequency-based methods using fast Fourier transformations (FFT) is the necessity of stationary signals without transients. This condition may not be fulfilled in some applications.

Moreover, methods may require long measurements, extending the minimal time needed for successful fault detection. On the other hand, methods based on time domain analysis are not very common. These methods use residual analysis between complex motor models and actual motor currents. They combine analytical and model-based approaches.

Model-based methods use analytical motor models with modifications and extensions that allow reconstruction of the machine behavior under the fault. Models suitable for fault modeling are often in stator coordinates, respecting the mechanical construction of the machine [4]. Fitting model parameters to actual motor behavior can provide relevant features for fault detection. Extended Kalman filters are often used in model-based fault detection methods. These detection methods can provide faster reactions compared to signal analysis methods.

Artificial intelligence-based methods can combine benefits from both previously mentioned methods. For instance, the artificial neural network (ANN) can use data preprocessed using some form of signal analysis. Input features can provide the full spectrum provided by FFT or extracted fault-related harmonics [5], [6]. ANN working with extracted features mainly use multiple fully connected layers. These networks are referred to as multi-layer perceptron networks (MLPN). On the other hand, convolutional layers typically used in

The work has been performed in the project AI4CSM: Automotive Intelligence for/at Connected Shared Mobility No 101007326/8A21013. The work was co-funded by grants of Ministry of Education, Youth and Sports of the Czech Republic and Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU). The work was supported by the infrastructure of RICAIP that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857306 and from Ministry of Education, Youth and Sports under OP RDE grant agreement No CZ.02.1.01/0.0/0.0/17_043/0010085.

a modern neural network can preprocess data and extract features without complex data preprocessing [7]–[10]. ANN networks can also contain recurrent layers that provide memory to the neural networks. These layers are beneficial in reconstructing time series signals. From a mathematical point of view, the internal structure of the neural network can provide functionality comparable to motor modeling [11]. One of the benefits of neural networks is the possibility of using unsupervised learning. Authors in [12] use big datasets to prove the possibility of detecting the fault without manual feature extraction. This method can simplify the preparation process of the neural networks; however, the ANN inference time is significantly higher using complex ANN structures.

The already performed study [3] shows the importance of real-time detection algorithms. For real-time detection, it is necessary to consider the target time for fault detection. Publications often present solutions using transfer learning methods [13] or neural networks containing huge amounts of trainable parameters [14]. Presented neural networks reach excellent precision; however, publications do not even estimate the presented neural network inference time. For this reason, deep neural networks with millions of configurable parameters presented in [13], [14] can be used only using an additional powerful computational platform.

Authors of [15] integrated complex 2D convolutional neural network into an embedded control system and thus show a real possibility of implementation. Reached results show that even smaller ANN can provide sufficient fault detection results. However, the inference time of the presented neural network is in the order of seconds. Achieved inference time can limit the use of the presented solution for real industrial or automotive applications. Because fault detection time should be as short as possible, some publications propose detection time for interturn short-circuits in the range of several tenths of milliseconds to protect the motor against thermal damage.

The results achieved using neural networks presented by recent publications show that their use can bring many advantages. However, it is also necessary to consider the real possibilities of implementation and inference time in the final hardware.

## II. Interturn Short-Circuit Fault Detection Using Neural Networks

A stator winding fault is one of the most common faults occurring in electric drives. According to the [3], 36 % of all motor faults are related to stator winding faults. The situation is even worse in high-voltage applications. Stator winding faults appear here, even in 66 % of cases. This phenomenon is primarily caused by significant insulation stress and gradual material degradation due to vibrations, temperature, and high voltage stress combined with modern semiconductor elements, which can generate steeper switching edges. The most typical fault is the interturn short-circuit fault.

This paper presents ANNs prepared for interturn short-circuit fault detection. The structure of the neural networks considers using dual three-phase electric motors, which can continuously operate under various winding faults to get as close to real fail-operational applications as possible. The dual three-phase motor has two isolated three-phase stars in the stator. Despite the electrically isolated windings, mutual magnetic couplings must be considered. The interturn short-circuit in one sub-system is detectable in the healthy sub-system thanks to mutual coupling. Considering the assumptions, fault detection using separated neural networks for individual subsystems is not possible. Initially performed experiments clearly show that one complex neural network needs to be used to evaluate faults for dual three-phase PMSM.

The studied literature and previous research focused on fault modeling [4] and fault suppression [16] show that the most suitable feature for fault detection is the 2nd harmonics component of the motor dq currents or voltages depending on actual motor speed. The paper demonstrates the use of MLPN and convolutional neural network (CNN) for interturn short-circuit detection and their implementation into the embedded system.

### A. Tested neural networks

The first prepared neural network is MLPN. The neural network uses filtered values of currents and voltages in dq coordinates as inputs. The actual motor load information is provided thanks to these data. The following input is the continuously extracted and filtered second harmonic. The last input is the actual motor speed measured by the incremental encoder and reconstructed using an angle-tracking observer. Designed filters are easily implementable and can be calculated in each motor control period $T_s$. The control algorithm and pulse-width modulation (PWM) operate with a frequency of 10 kHz.

$$y_{dq(k+1)} = KT_s\omega_e(x_{dq(k)} - y_{dq(k)}) + y_{dq(k)} \tag{1}$$

(1) describes used discrete filters. Variable $\omega_e$ denotes the actual electrical motor speed, and coefficient $K$ sets the dynamics of the filter. The overall filter dynamic is proportional to the motor speed to decrease the time needed for detection in higher-speed areas. In this area, a high short-circuit current can cause additional thermal damage in a shorter time. To improve the suppression of higher frequencies, the third order of this filter is used. Variables $x_{dq(k)}$ and $y_{dq(k)}$ represent the actual input value and the actual state of the filter, respectively. Voltages and currents in dq coordinates for both sub-systems are used as inputs.

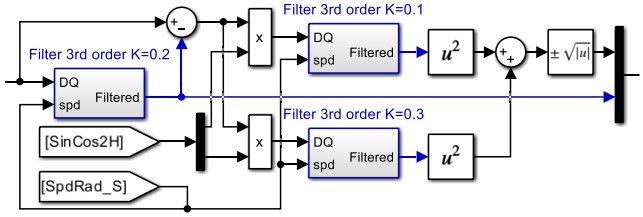


Fig. 1: Implemented filters in MATLAB/Simulink

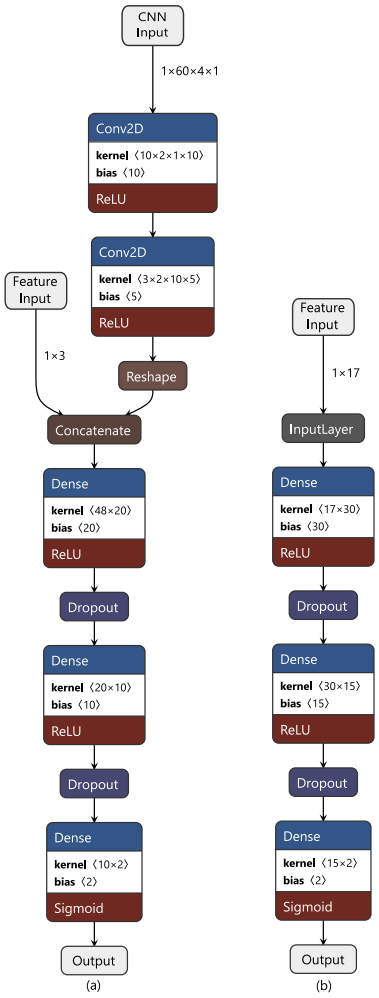


Fig. 2: (a) structure of the convolutional neural network, (b) structure of the multilayer perceptron network

Second harmonic components are extracted using sine and cosine signals of specific frequency. The difference between the actual and filtered values is multiplied by these signals and filtered subsequently. The whole process of extraction features for one set of dq variables is shown in Figure 1. Prepared MLPN uses 17 inputs for this reason. The neural network structure is visible in the following Figure 2 (b). The total number of trainable parameters in the presented MLPN is 1037.

The second prepared neural network uses convolutional layers to extract features from signals. Convolutional layers are suitable for extracting related features from arrays of various dimensions. The most common use of convolutional neural networks is in image processing. If convolutional networks are used in motor fault detection, time waveforms are typically converted into 2D images. However, 2D images are strongly speed dependent. Mentioned data preprocessing is usable; however, deeper neural networks with many convolutional filters are needed to extract relevant features.

The designed convolutional neural network uses voltages and current amplitudes from three electrical revolutions as

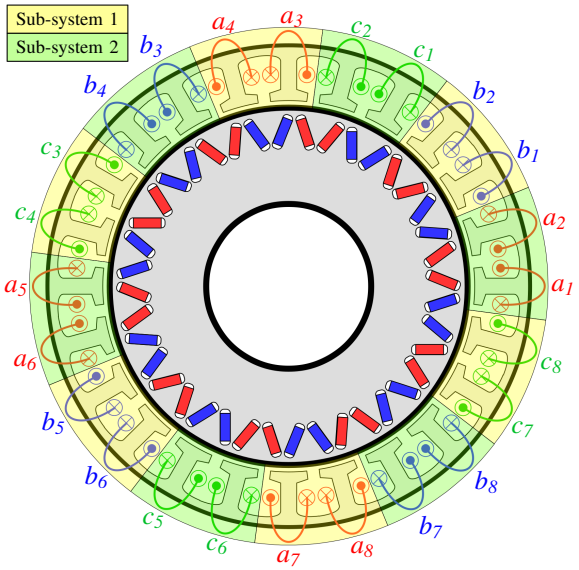


Fig. 3: The experimental motor structure and assignment of individual coils into sub-systems

input for the convolutional part. Four vectors are created considering two sub-systems, two voltage vectors, and two current vectors. Each vector has 60 elements (20 elements per electrical revolution). These vectors are compounded into a 60x4 2D array. The first convolutional layer uses ten kernels with a size of 10x2. The layer is configured to use the same kernels independently on signals corresponding to individual sub-systems. This functionality is achieved using the stride parameter. Values are set to 5 and 2. The second convolutional layer can extract the features reflecting differences between the sub-system. Output from the convolutional layers is concatenated with the motor speed and rotor position sine and cosine signals. Extracted features are connected to fully connected layers. This part of the CNN network has a similar number of trainable parameters to the first MLPN. A convolutional neural network contains 515 trainable parameters in the convolutional part. The total number of trainable parameters in the whole network is 1727.

Periodically measured input data are oversampled according to rotor position, reducing the speed dependency of the input data pattern. In lower speed areas, samples for a specific position are calculated using a mean value from corresponding time samples. This method is used up to 7500 rpm. Missing samples for speeds over 7500 rpm are calculated using a linear approximation. The described preprocessing algorithm is more computationally demanding. On the other hand, simpler convolutional layers structure can be used. The structure of the designed convolutional neural network is shown in Figure 2 (a).

## III. Test Platform and Achieved Results

All tests use the experimental motor to get as close to real applications as possible and structures that can provide fail-operational behavior. This machine has configurable winding, which allows using various internal structures. The motor was

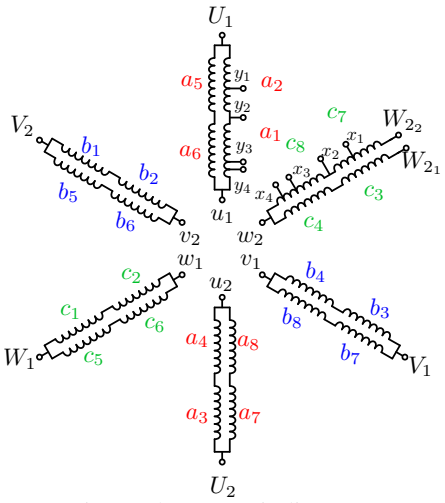


Fig. 4: The experimental motor winding arrangement, winding taps $y_{1-4}$ and $x_{1-4}$ can be used to emulate interturn short-circuit fault

connected as a dual three-phase motor with interlaced sub-systems and electrically isolated middle points. The assignment of the individual motor coils to the sub-systems can be seen in figure 3. Moreover, figure 4 shows the electrical connection and prepared winding taps $y_{1-4}$ and $x_{1-4}$, which are used for interturn short circuit emulation. The figure respects the number of turns and the position of winding taps. Seven turns form each motor coil; four coils form each phase winding.

Short-circuit can be created permanently during the experiment or switched by a fault insertion unit (FIU). FIU consists of LEV200A4NAF relay and thyristor module STT800N16P55XPSA1 with a control board. Detailed motor parameters are summarized in Table I. One of the most important parameters to achieve fail-operational behaviour is the characteristic current of the motor. Characteristic current is lower than maximal continuous current, and the motor can continuously operate under active short-circuit (ASC). This operation mode can be used to reduce fault current. The amplitude of the fault current depends on actual motor speed and the number of short-circuited turns. Figure 5 shows the measured values of short circuit current. Figure 6 shows the whole experimental setup.

### A. *Implementation of neural networks into the microcontroller*

Data for learning and validation were measured on the experimental motor. The training dataset contains the long-term operation. The required speed was randomly changing with various slopes. This process provided data with various load conditions. New waveforms in the training dataset were derived from measured data adding a virtual shift of the coils and the rotor position by $\pm 120$ electrical degrees. Motor sub-systems were also virtually switched to each other. Data for validation were measured during gradual acceleration under the constant load using a dynamometer.

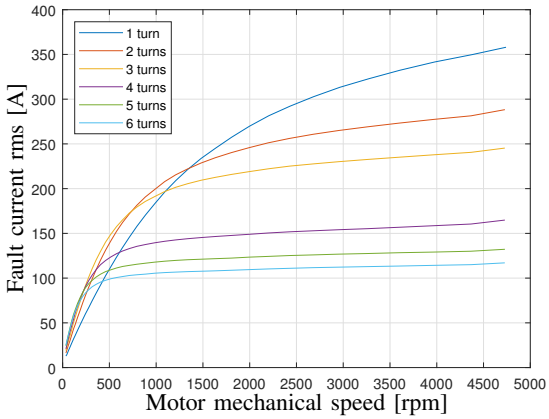


Fig. 5: Motor fault current depending actual motor speed for various inter-turn short-circuit depths including resistance and inductance of additional wires and FIU ( 0.6 Ω 0.25 µH)

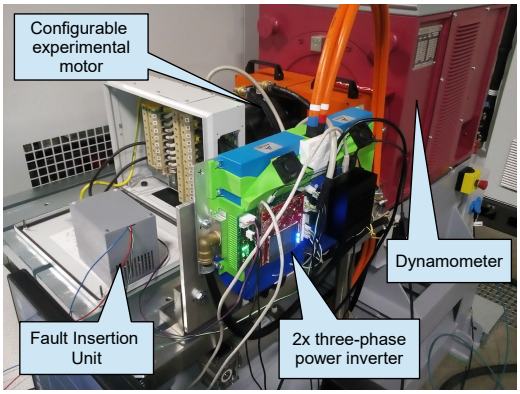


Fig. 6: Dynamometer and experimental PMSM under test

Both neural networks were prepared using TensorFlow. The trained neural networks were stored in TensorFlow format and subsequently imported into MATLAB 2022b using *importTensorFlowLayers* function. MLPN is supported by MATLAB deep network toolbox. However, the concatenation and reshape layers used in CNN are not fully compatible with MATLAB layers. For this reason, the default *depthConcatenationLayer* was replaced by *concatenationLayer* with dimension 1 after importing CNN into MATLAB. An additional dimension

TABLE I: Experimental motor parameters

| Name | Symbol | Value | Unit |
|---|---|---|---|
| Nominal DC-link voltage | $U_{DC}$ | 400 | V |
| Used DC-link voltage | $U_{DC}$ | 200 | V |
| Maximum continues motor current | $I_s$ | 107 | A |
| Nominal motor power | $I_s$ | 30.16 | kW |
| Winding resistance | $R_s$ | 6.5 | mΩ |
| Back-EMF constant | $\psi_m$ | 0.0117 | Vs/rad |
| Winding inductance | $L_{u-v}$ | 180 | µH |
| Nominal speed | $\omega_n$ | 8000 | rpm |
| Maximum speed | $\omega_M$ | 10500 | rpm |
| Nominal power | $P_c$ | 30160 | W |
| Number of pole pairs | $Pp$ | 10 | - |
| Characteristic current | $I_{ch}$ | 81 | A |

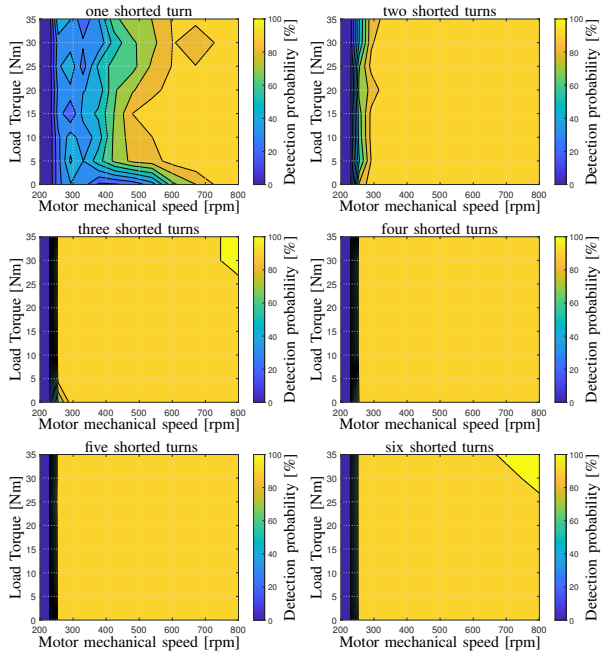


Fig. 7: The output value of the multiplayer perceptron neural network (probability of the fault). Measured for various operating points at steady state.

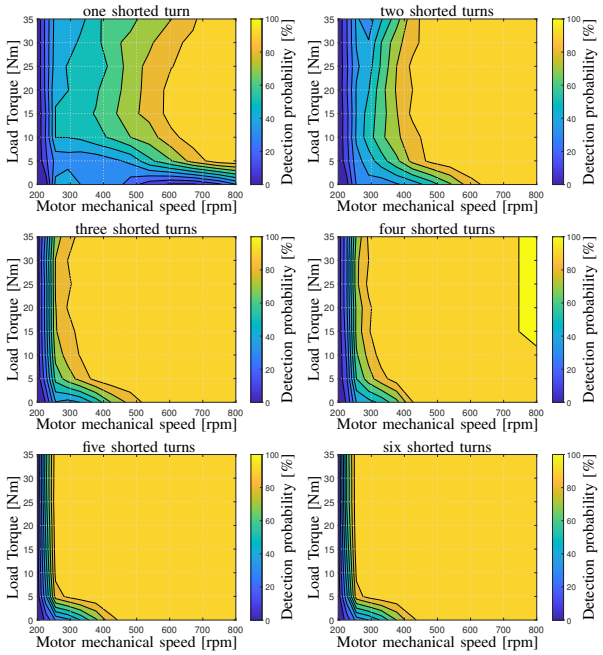


Fig. 8: The output value of the convolutional neural network (probability of the fault). Measured for various operating points at steady state.

problem was solved by adding *flatten* layer between *reshape* layers and *concatenationLayer*.

The imported neural networks were used in a Simulink scheme. The scheme can be used directly for ANN inference or code generation processes. Simple C/C++ source codes can be generated using MATLAB Coder Interface for Deep Learning Libraries. Another option is to generate CUDA source code using GPU Coder Interface for Deep Learning Libraries. AURIX TC397 microcontroller is used in the power inverter. This microcontroller support floating point operations, and an ANN can be used without quantization. For this reason, C/C++ code was used for the integration in the microcontroller.

## IV. Experimental results

The previously mentioned experimental setup was used to test the created neural networks. Both implemented neural networks were able to detect an interturn short-circuit in the nominal speed area. However, additional tests were performed to validate neural network behavior in the whole operational range. These tests show 100 % fault detection accuracy for the electrical speed of the motor over 1500 rad/s. This area is critical because of possible high short-circuit currents.

Tests for electrical speed in the range from 800 rad/s to 1500 rad/s are shown in Table II. MLPN reaches an accuracy of over 99.8 % in this speed range. The convolutional network reaches slightly lower performance. However, both neural networks are applicable for fault detection. In this speed range, short-circuit current can cause additional thermal damage. Nevertheless, required fault detection time is lower.

The least critical is the low-speed area. The motor can work continuously without the risk of further damage, even in the case of an interturn short circuit in this area. The probability of the fault detected by neural networks is shown in Figure 7 for MLPN and Figure 8 for CNN.

One of the most important results of implementing a neural network on the target hardware is the inference time of the neural network and the time required for data preprocessing. Figure 9 shows measured computational times. An important finding is that the calculation of a convolutional neural network is significantly longer compared to MLPN. Despite the relatively small number of trainable parameters in convolutional layers, it is necessary to consider that each parameter is used multiple times during the inference process. In contrast to the convolutional network, the number of trainable parameters in MLPNs indicates the number of necessary multiplications during the inference process.

TABLE II: Accuracy of neural network detection for electrical motor speeds above 1200 rad/s

| ANN | Shorted turns | Health state detection [times] | Fault state detection [times] | success rate [%] |
|---|---|---|---|---|
| MLPN | 0 | 104967 | 33 | 99.96 |
| | 1 | 135 | 104865 | 99.87 |
| | 2 | 0 | 105000 | 100 |
| | 3 | 0 | 105000 | 100 |
| | 4 | 0 | 105000 | 100 |
| | 5 | 0 | 105000 | 100 |
| | 6 | 0 | 105000 | 100 |
| CNN | 0 | 15881 | 289 | 98.21 |
| | 1 | 333 | 15837 | 97.94 |
| | 2 | 1 | 16169 | 99.99 |
| | 3 | 0 | 16170 | 100 |
| | 4 | 0 | 16170 | 100 |
| | 5 | 0 | 16170 | 100 |
| | 6 | 0 | 16170 | 100 |

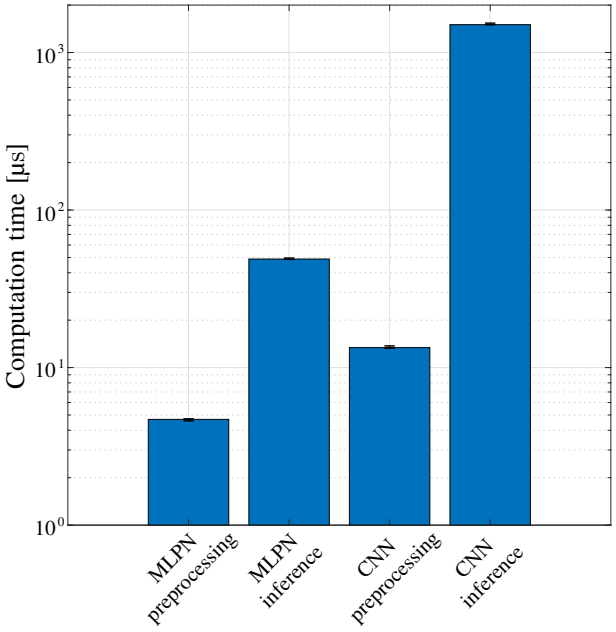


Fig. 9: Comparations in computation time (microcontroller Aurix TC397, MATLAB 2022b generated C code (C99), built with AURIX™Development Studio)

## V. Conclusion

The results presented in this article show that it is possible to implement different kinds of ANNs for interturn fault detection in advanced powertrain systems. Two ANNs are demonstrated, one is MLPN, and the second one is CNN. The measured inference times of both neural networks are short enough, giving the space to detect the interturn short circuit fault before it causes irreversible thermal damage. One can argue that faulty alarms can be very harmful in automotive applications, but it is necessary to keep in mind that the multi-phase configuration of the motor for which the ANNs were prepared enables us to continue with the operation due to implemented fail-operational control strategy just with reduced maximum torque.

The paper shows the usage of the AURIX microcontroller. Its advantage is multicore construction giving the option to distribute the computation of different inverter tasks to different cores. It is expected that one core can be dedicated to diagnostic tasks. AURIX TC397 is a powerful embedded microcontroller designed for high safety demanding applications. Measured results thus show that often presented large ANNs with millions of trainable parameters cannot be deployed in such devices without the support of an edge computational accelerator. This is primarily due to the long inference time needed to evaluate large ANNs but also due to the lack of memory for storing such large ANNs directly in the microcontroller.

## References


[1] H. T. Canseven and A. Unsal, "Performance Improvement of Fault-Tolerant Control for Dual Three-Phase PMSM Drives Under Inter-Turn Short Circuit Faults," in *IECON 2021 47th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, oct 2021, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9589578/

[2] M. Kozovsky and P. Blaha, "Double three-phase PMSM structures for fail operational control," *IFAC-PapersOnLine*, vol. 52, no. 27, pp. 1–6, 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2405896319326825

[3] T. Orlowska-Kowalska, M. Wolkiewicz, P. Pietrzak, M. Skowron, P. Ewert, G. Tarchala, M. Krzysztofiak, and C. T. Kowalski, "Fault Diagnosis and Fault-Tolerant Control of PMSM DrivesState of the Art and Future Challenges," *IEEE Access*, vol. 10, pp. 59 979–60 024, 2022.

[4] M. Kozovsky, L. Buchta, and P. Blaha, "Interturn short circuit modelling in dual three-phase PMSM," in *IECON 2022 48th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, oct 2022, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9968364/

[5] S. Moosavi, A. Djerdir, Y. Ait-Amirat, and D. Khaburi, "ANN based fault diagnosis of permanent magnet synchronous motor under stator winding shorted turn," *Electric Power Systems Research*, vol. 125, pp. 67–82, aug 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378779615000917

[6] L. S. Maraaba, Z. M. Al-Hamouz, A. S. Milhem, and M. A. Abido, "Neural Network-Based Diagnostic Tool for Detecting Stator Inter-Turn Faults in Line Start Permanent Magnet Synchronous Motors," *IEEE Access*, vol. 7, pp. 89 014–89 025, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8740936/

[7] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, nov 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7501527/

[8] D. T. Hoang and H. J. Kang, "A Motor Current Signal-Based Bearing Fault Diagnosis Using Deep Learning and Information Fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3325–3333, jun 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8788665/

[9] L. S. Maraaba, A. S. Milhem, I. A. Nemer, H. Al-Duwaish, and M. A. Abido, "Convolutional Neural Network-Based Inter-Turn Fault Diagnosis in LSPMSMs," *IEEE Access*, vol. 8, pp. 81 960–81 970, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9079811/

[10] L. Wen, X. Li, L. Gao, and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, jul 2018. [Online]. Available: http://ieeexplore.ieee.org/document/8114247/

[11] Y. Huang, C.-H. Chen, and C.-J. Huang, "Motor Fault Detection and Feature Extraction Using RNN-Based Variational Autoencoder," *IEEE Access*, vol. 7, pp. 139 086–139 096, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8835037/

[12] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, may 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7386639/

[13] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, apr 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8432110/

[14] M.-Q. Tran, M.-K. Liu, Q.-V. Tran, and T.-K. Nguyen, "Effective Fault Diagnosis Based on Wavelet and Convolutional Attention Neural Network for Induction Motors," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9666871/

[15] S. Lu, G. Qian, Q. He, F. Liu, Y. Liu, and Q. Wang, "In Situ Motor Fault Diagnosis Using Enhanced Convolutional Neural Network in an Embedded System," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8287–8296, aug 2020.

[16] M. Kozovsky, L. Buchta, and P. Blaha, "Compensation methods of interturn short-circuit faults in dual three-phase PMSM," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, oct 2020, pp. 4833–4838. [Online]. Available: https://ieeexplore.ieee.org/document/9254734/