Intelligent Traction Inverter in Next Generation Electric Vehicles: The Health Monitoring of Silicon-Carbide Power Modules

Carmelo Pino, Alessandro Sitta, Giulia Castagnolo, Angelo A. Messina, Salvatore Coffa, Michele Calabretta, Fabio Scotti, *Senior Member, IEEE,* Angelo Genovese, *Senior Member, IEEE,* Vincenzo Piuri, *Fellow, IEEE,* Concetto Spampinato, Francesco Rundo

***Abstract*—In automotive and industrial domains, the "health monitoring" or "condition monitoring" of electronic devices is gradually playing a key role in manufacturing processes and innovation roadmaps. The concept of health monitoring is often related to the so-called "residual lifetime" of the monitored system. In this work, the authors have designed a deep learning system for the health monitoring of power devices in Silicon Carbide (SiC) technology used in the Traction Inverter Systems of the latest generation electric cars. A Temporal Fusion Transformer embedding such layers of Temporal Convolutional Network with a Multi-Head Attention block for the robust lifetime assessment of SiC power devices, is proposed. Specifically, the designed system predicts such future samples of the ON-state voltage between drain and source of the low-side part of the SiC power module Vds_{LS}, in half-bridge configuration. Extensive literature confirmed that the Vds_{LS} signal can be efficiently used as a robust predictive device-degradation marker. Through the learning of the temporal feature relationships at different scales and the intelligent selection of relevant input features, the proposed solution will discard unnecessary input dynamics building a multi-step predictive model of the Vds_{LS} signal, significantly more performing than the existing state-of-the-art architectures. The proposed deep pipeline has been tested on several ACEPACK™ DRIVE SiC power modules delivered by STMicroelectronics, with an average error of about 0.2%, confirming the effectiveness of the proposed system.**

***Index Terms*—Temporal Fusion Transformer, SiC Power Module, Device Health Monitoring**

INTRODUCTION

SILICON Carbide (SiC) power devices and modules are recently finding a wide application in the realization of power systems for the industrial and automotive fields [1], [2]. This extensive use of SiC technology devices is mainly triggered by their interesting properties and resistance to operating modes both at high temperature and at high switching frequency [3]. Thanks to the extreme compactness and high-efficiency of the SiC power devices, automotive designers have greater flexibility in vehicle integration, cooling system design, battery pack sizing, etc. Anyway, the growing use of SiC-based electronic power systems in various automotive applications - with special focus to electric vehicles - have made their reliability a main priority for the automotive industries [4].

Area reduction and higher temperature capability require to SiC power devices package to withstand higher power density with respect to Silicon (Si) counterparts [5]. Operative conditions, like device temperature swing during active cycle, can trigger interface degradation like wire bonding lift-off, which impact directly device on-state drain-source resistance [6].

If a component or a subsystem of a power electronic system shows an abnormal behavior, it may compromise the correct working of the whole system. For automotive applications, these undesired behaviour can have a significant impact on traction-system safety resulting in an impact on the overall safety of the car. For instance, in hybrid electric vehicles, a fault in a SiC power devices-based systems can certainly have an impact on the car cooling system and therefore on the whole electric car-engine performance with a consequent impact to the driving safety.

In this contribution we mainly refer to modular power systems in SiC technology. SiC power modules are key systems embedded in the latest generation of electric cars as they are the electronic heart of the vehicle's traction inverter [3]. The worldwide exponential increase in vehicle electrification puts the design of cost-effective and robust traction inverters into focus. Consequently, the implementation of robust SiC power modules equipped with a health-monitoring system that promptly reports degradation processes is of considerable strategic importance both for the market and for improving safety in the automotive field. Intelligent health monitoring system for SiC power devices to check time-by-time the degradation level of the device has been investigated by the authors of this contribution.

Starting from a careful examination of the prior art, we have investigated the electric car development scenario and the related necessity to monitor the state of degradation and the residual lifetime of the SiC power modules embedded in the car traction system.

This work was supported by the projects: "First and euRopEAn siC eigTh Inches pilOt liNe (REACTION)" - HORIZON 2020 Program under Grant Nr. 783158 and "Edge-AI" - HORIZON-KDT-JU-2021-2-RIA-Focus-Topic-1

Francesco Rundo, Carmelo Pino, Giulia Castagnolo, Alessandro Sitta, Salvatore Coffa and Michele Calabretta are with STMicroelectronics ADG R&D Power and Discretes Division, Catania, IT , (e-mail: francesco.rundo@st.com, carmelo.pino@st.com, giulia.castagnolo@st.com, alessandro.sitta@st.com, michele.calabretta@st.com, salvo.coffa@st.com).

Angelo Alberto Messina is with STMicroelectronics IPA Division, Catania, IT, (e-mail: angelo.messina@st.com).

Concetto Spampinato is with PerceiveLab, University of Catania, Catania, IT, (e-mail: concetto.spampinato@unict.it).

Angelo Genovese, Vincenzo Piuri, Fabio Scotti are with Department of Computer Science, University of Milan, IT, (e-mail: angelo.genovese@unimi.it, vincenzo.piuri@unimi.it, fabio.scotti@unimi.it).

Through artificial intelligence techniques, specifically through architectures based on composite deep learning backbones, the authors correlated the dynamics of a specific voltage signal of the power modules with the residual lifetime of the power device, thus allowing an early diagnosis of degradation useful for performing any device recovery operations or device replacement.

This proposed scientific contribution is organized as follows: in the section "Introduction to the Traction Inverter and Power Modules", a summary overview of the working-mode of the traction inverter and the role of the power modules in SiC technology will be discovered, contextualizing a specific application and specifying the testing methodology from which the validated dataset was obtained. In the "Related Works" section, the main methods proposed in the scientific literature for the predictive characterization of the residual lifetime of a power device will be analysed.

In the "Materials and Methods" section, the proposed pipeline and the learning phase of the designed deep learning model will therefore be described in detail. Finally, in the "Experimental Results" section, the experimental results obtained in the tested application scenario related to the prediction of the residual lifetime of the SiC Power modules will be presented. The layout of the present contribution will be completed by the section "Conclusion and Future works" which will include a discussion of the results obtained as well as the research road-map of the proposed solution.

INTRODUCTION TO TRACTION INVERTER AND SIC POWER MODULES

As introduced, the application scenario of this contribution is the traction inverter system of the electric car. Electric and hybrid vehicles require a traction inverter to convert DC energy from the high-voltage battery to the three-phase AC energy used to drive the traction motor. Traction inverters are typically capable of transferring power in the 20 to 100 kW range, with switching voltages in the 200 V to 800 V range and currents in the hundreds of amperes [7]. The traction inverter function is accomplished using six switches embedded in the high-side and low-side parts (and for each phase) of the internal power device. In Fig. 1, an overall diagram of a modern electric vehicle embedding a traction inverter system is reported [7].

The weight and power losses inside this component as well as the considered implemented battery voltage, have a direct impact on road performance, mileage and reliability of the whole vehicle. The main part of inverter is the power module, which is made by several semiconductor devices. It reproduces the desired electric topology and allows electric power transformation, which is the product of current and voltage with high conversion efficiency. As reported in Fig. 1, the traction inverter embeds such electronic sub-systems. In Fig. 2 a detail of a traction inverter system embedding the ACEPACK^{TM} DRIVE power modules, is reported. The control is implemented by the automotive-grade SPC58 series microcontroller delivered by STMicroelectronics. ACEPACK^{TM} DRIVE represents the leading SiC power product in electric traction, delivered by STMicroelectronics [7].

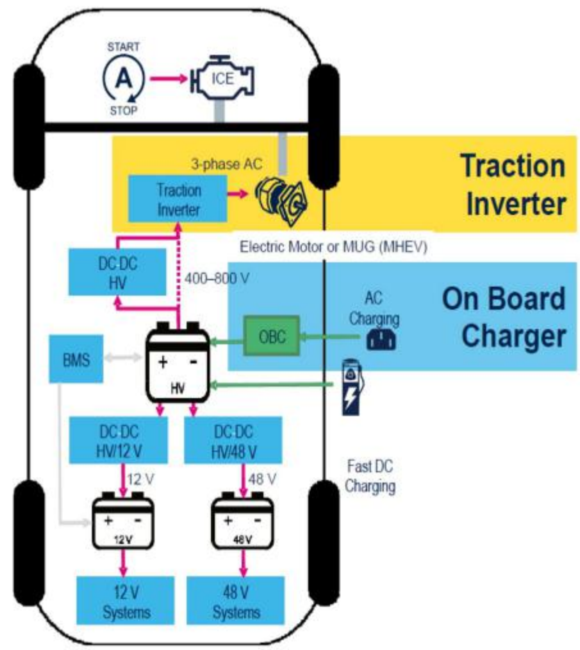


Fig. 1. Electrical Vehicle sub-systems with a detail of the Traction Inverter

In order to boost the performances of entire inverter, it is of paramount importance to optimize power module efficiency. In this framework, SiC device-based power modules are spreading in the automotive market [8]. When Silicon Carbide (SiC) semiconductors are used as switches, the overall system efficiency is improved by allowing higher operating temperatures and switching frequencies [9]. On the other hand, as mentioned before, SiC power devices should handle mechanical and thermal stress during normal operations. Indeed, these converters must handle high power and currents along with the associated Electro Magnetic Compatibility (EMC) challenges as well as provide fail-safe operation to ensure dependability and safety for the driver and passengers. More details about electo-mechanics features of the ACEPACK^{TM} DRIVE devices.

The ACEPACK^{TM} DRIVE power modules for traction inverter applications shows a minimized stray inductance and boosted speed capabilities. The used topology of the power module is a "Sixpack" type. Each switch is made by different SiC MOSFETs paralleled. Liquid coolant directly in contact with the pin-fins of module's baseplate ensures an excellent thermal management and therefore durability. Moreover, dedicated components and processes were designed to optimize the whole product performances. As instance, silver sintering die attach material was selected to guarantee a regular bond line thickness and to optimize reliability and thermal resistance thanks to its excellent properties with respect to more conventional soft soldering alloy [10]. Moreover, active metal brazed substrates were used due to the high thermal conductivity of their dielectric layers (Si^3N^{34}) and due to the high reliability performance of metal-dielectric joint. In Fig. 3 we have reported the electrical scheme of the used

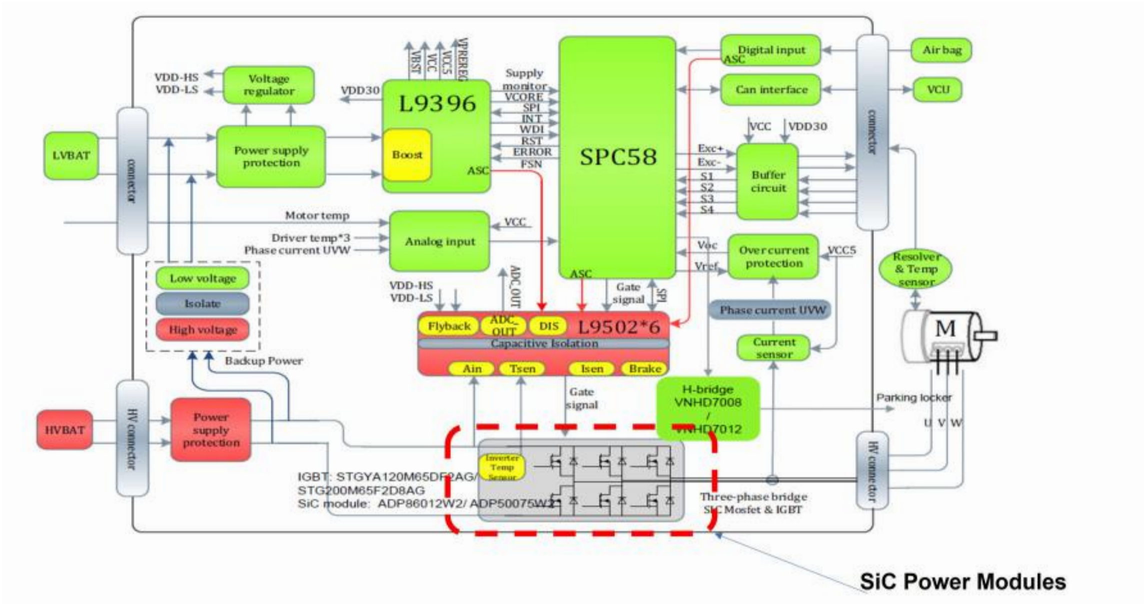


Fig. 2. Overview Scheme of the Traction Inverter System embedded in a Electric Vehicle

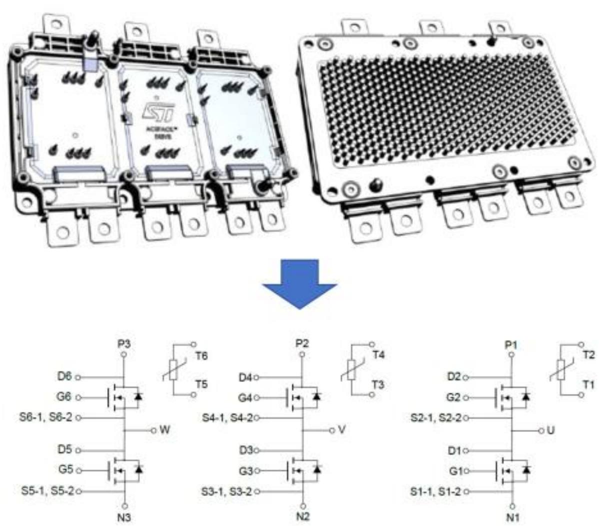


Fig. 3. ACEPACK™ DRIVE SiC Power Module with a detail of its electrical scheme

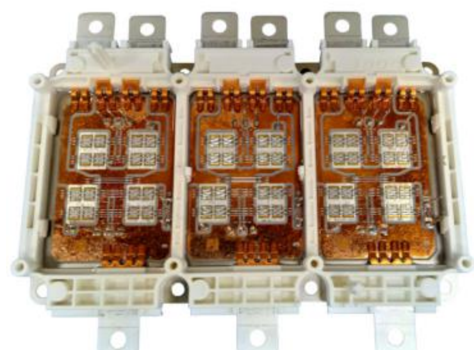


Fig. 4. ACEPACK™ DRIVE SiC Power Module

ACEPACK™ DRIVE SiC power device with a detail of the related 3D representation of the device. A real representation of the tested power module is reported in Fig. 4.

In order to characterize the electrical and structural robustness of the SiC power module, several tests must be performed, as described by standards guidelines like AQG 324 [11]. Although of interest for the present scientific contribution, the validation methodology of the SiC power module is the so-called Power Cycling [11]. It is an accelerated aging methodology of the power device aimed at highlighting the reliability performance over time due to thermal stress generated by device self-heating. In fact, power cycling is an active test, in which the device is cyclically heated by DC current and chilled by a liquid coolant. Device is heated by conduction losses and, indeed, thermal stress is generated at material's interfaces due to the mismatch of coefficients of thermal expansion. The main parameters that determine stress induced on power module are the dies' temperature swing, the maximum dies' temperature and the current forcing time [12]. Test procedure and execution are regulated by AQG 324 guideline [11].

A critical topic of power cycle is the monitoring of electric quantities (drain current and drop voltage) and temperature reached by devices under test [11]. For electric parameters, dedicated instruments measure in real time voltages *Vgs* (i.e. the Gate-Source voltage) and *Vds* (Drain-Source Voltage) and current of the drain terminal of the device (*Id*). These parameters are controlled cycle-by-cycle in order to detect if end-of-life criteria, for the tested SiC power modules, are reached. During the cooling phase, in the case of SiC device, body-diode source-drain voltage (*Vsd*) is also measured. Thanks to the well-known relationship between *Vsd* and temperature, the device's temperature is estimated as "virtual junction temperature" [13], [14]. The picture of power cycle bench considered

Fig. 5. Power Cycle Test-bench

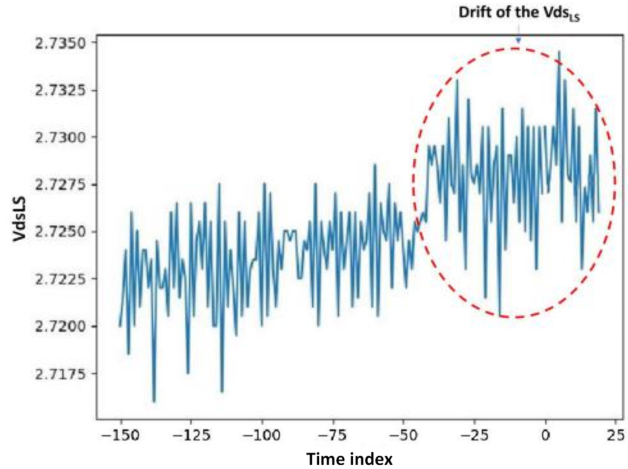


Fig. 6. A time-windows of the Vds_{LS} with detail of the signal-drift

for this activity is depicted in Fig. 5. In the specific framework, power cycling is employed considering temperature swing of 100 °C, maximum die temperature of 150 °C and current on-time of 3.5 s. Device self-heating is accomplished driving the Vgs to positive value, much higher than threshold value, and "virtual junction temperature" is monitored during the off-state, applying a negative voltage of -8 V. Negative value is needed to ensure that sensing current flows across the body diode. The tested SiC power modules are configured in Half-bridge configuration in which we can identify a low-side and high-side part respectively [7]. For each module, one switch from low-side is placed in series with a switch of different phase placed in the high-side part of the device.

In Fig. 3 an ACEPACK^{TM} DRIVE SiC Power Module in half-bridge configuration is reported. We identify the low-side area in the part of the module closest to the battery terminals.

At the end of the power cycling, the SiC power device will be brought to the end-of-life using the aforementioned accelerated aging approach. The end-of-life of the power module is characterized by a degradation of the interface which can be seen by a significant increase in the resistance between drain and source Rds_{ON}. Specifically, a significant drift of the Rds_{ON} equal to or greater than 5 % is considered a confirmation of electrical death of the device with consequent impossibility of use or recovery. Considering that the practical measurement of the Rds_{ON} involves the measurement of voltage and current in the bridges of the module, the voltage between the drain and source of the bridge (low-side or high-side i.e. Vds_{LS} or Vds_{HS} respectively) is often used as a proportional end-of-life marker, as it is directly proportional to the Rds_{ON}. In Fig. 6 an instance of the Vds_{LS} with a detail of the drift in correspondence with the end-of-life of the device is shown.

During this phase the characteristic thermo-electrical variables of the device will be collected by the Power Cycle test-bench (see Fig. 5) software system. Specifically, the following information will be collected during the power cycling of the power module: the voltage between drain and source of the low-side and high-side switches respectively (Vds_{LS} or Vds_{HS}), the junction temperature and the drain current.

Therefore, from this validation system based on power cycling induced accelerated aging, we were able to collect the dynamic of the Vds_{LS} or Vds_{HS} throughout the testing process. Obviously, these voltages are closely related to the degradation process which is generated in "accelerated" modality to validate the robustness of the power module within acceptable times.

Therefore, the goal of this scientific contribution is to design an artificial intelligence model capable of generalizing, in a real application-automotive context, the correlation between Vds_{LS} or Vds_{HS} and power module end-of-life obtained in a power cycling scenario. The following sections, after a brief examination of the state of the art, will present the implemented deep model and the experimental results collected after its validation.

Related Work

The identification of such predictive degradation markers (and then, residual lifetime estimators) for electronic devices, has been extensively investigated in scientific literature due to the significant impact that such discovery would have in the industrial fields. As widely treathed in [15], Artificial Intelligence is also largely used in power electronics, by applying existent models and solutions to specific use cases. Even if some of these solutions are applied to electronic devices of various kinds, the reported methods can certainly be extended to power devices and therefore to the case herein analyzed . An interesting overview of the lifetime-estimation methods for power converters is reported in [16]. The authors preliminary investigated different kinds of failures as well as the physical mechanisms that generate these failures. Subsequently, they investigated the problem of predicting the residual lifetime of power converters both at "component level" and at "system level". At the component level, were investigated the use of three types of lifetime prediction methods: Random Failures Lifetime Prediction Methods, Wear-out Failures Lifetime Prediction Systems (both Model-driven and Data-Driven) and finally Hybrid Lifetime Prediction Methods. The latter category include the usage of both statistical approaches and such machine learning solutions based on the use of Artificial Neural Networks, Fuzzy-based methods, Support Vector Machine (SVM) and so on. The authors concluded for higher efficiency and precision of the hybrid methods than the others. At system level, they analyzed three different approaches: Reliability Block Diagrams (RBD), Fault-Tree Analysis (FTA), and Markov Chains. In an advantages/disadvantages comparison, the FTA method turns out to be the most efficient as it has the least drawbacks. More details in [16].

Another interesting review of real-time lifetime prediction methods for SiC technology power devices is reported in [17]. The authors preliminary analyse online and offline system lifetime modeling solutions identifying some predictive markers of degradation of the power device in SiC technology such as the gate leak current, the drain leak current, the body diode forward culture, the ON-state resistance of drain and source, the ON-state voltage between drain and source etc... Therefore, the authors analyzed the classical methods both at "component level" and at "System level". With reference to the former, they analyzed PoF (Physics of Failures) methods based on the physical study of failures: Gate oxide breakdown, Bond-wire fracture, solder fatigue and so on. Obviously, these approaches require a very careful study of the physics of failures and often have the disadvantage of not being very easily to be generalized to different scenario/devices. Furthermore, in [17] some data-driven methods both based on classical statistical approaches (linear regression models, Werner or Gamma Processes) and on Artificial Intelligence (AI) based solutions (genetic algorithms, deep learning, artificial neural networks, etc..) were analyzed. In [17] the authors has highlighted the efficiency of AI-based methods due to fewer drawbacks with high-generalization capability. The survey reported in [17] concluded the lifetime-estimation methods analysis with "system-level" approach from which the extended version of Fault Tree Analysis (FTA) named GO-FLOW seems to be very effective for the lifetime-prediction. More detail in [17]. Further details about the methods of health monitoring reported in the mentioned surveys can be found in [18]–[24].

An interesting approach based on the use of deep learning recurrent architecture is reported in [25]. The authors proposed a custom framework, named Deep RACE (DEEP learning Reliability Awareness of Converters at the Edge) for real-time reliability modeling and lifetime prediction of high-frequency MOSFET power electronic converters. They learned the designed model by using the dataset provided by NASA and tracking the ON-state resistance of drain-source. The proposed AI model is based on the usage of a stack of Long-Short-Term-Memory cells properly connected. The authors collected interesting results about the lifetime-prediction of the MOSFET power device as they obtained an average miss prediction of 8.9% with 26 ms as average processing time for the underlying designed EDGE hardware.

A very interesting approach based on the usage of deep learning enhanced by self-attention mechanism is presented in [26]. The authors proposed a deep architecture boosted with such attention mechanism. Basically they designed ad-hoc convolutional neural networks embedding attention methodologies based on classical $(Queries, Keys, Values)$ processing with a positional embedding vectors i.e. a classical Multi-Head Self Attention processing. The test of the implemented model were carried out by using the NASA's Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) datasets. The performance of the method reported in [26] confirmed that the proposed network achieves very promising performance in the lifetime-prediction with a best accuracy in terms of RMSE of 12.98 % with respect to a 12.61% of classical Convolutional Network or 16.14% of LSTM based solution.

In Chadha et al. [27] recent deep learning-based architectures have been investigated. The authors proposed a shared temporal transformer architecture embedding an attention layer ables to detect such temporal degradation patterns to be used for retrieving a shared correlation across the features range. They applied the implemented architectures to the public dataset C-MAPSS released by NASA. The performance of this transformer-based solution is very interesting as with this solution has been reached a performance of 10.9% as best RMSE in the used dataset.

As mentioned, in our state-of-the art analysis we found several works which investigated the usage of LSTM based architectures for implementing effective electronic device lifetime-predictor. More specifically, a recent and promising solution has been presented by Zhang et al in [28]. The authors implemented a deep LSTM network with ad-hoc data fusion block able to estimate the remaining useful lifetime (RUL) by using compressed data. The experimental results confirmed the effectiveness of the proposed approach as it showed a best performance in terms of RMSE of 10.24% in (C-MAPSS) dataset. Applied to the same dataset and with similar performances, we find other solutions based on machine learning [28] and specifically on the use of Multi-layer Percetrons (MLP) architecture [29], Support Vector Regression (SVR) [30] , Relevance Vector Regression (RVR) [31] , CNN based regression method [32], deep convolution neural network based method [33] and plain vanilla/ensemble deep LSTM [34], [35]. In Qin et al. [36] a further investigation of the deep learning based methods for RUL prediction were reported. As well known from scientific literature, Capsule-Network shows a very interesting ability to learn entities'hierarchical relationships between high-dimensional vector embeddings. By leaveriging this capability, the authors of [36] proposed a slow-varying dynamics-assisted temporal Capsule-Network architecture to simultaneously learn the slow-varying dynamics of the input data merged with temporal in order to perform an accurate RUL estimation. The proposed method has been applied to the public (C-MAPSS) database showing a very promising performance of 11.73 ± 0.31 as RMSE metric. More detail in [36].

Although the aforementioned methods, tested on industrial systems or discrete electronics, can in principle also be extended to power devices in both silicon and silicon carbide technologies, we have specifically investigated some deep learning-based methodologies directly tested on power devices. On this regard, Heydarzadeh et al. [37] proposed a robust solution, in order to predict the RUL of a power MOSFET based on data retrieved from a thermal aging process of the monitored device. More in detail, the authors proposed a Bayesian Interference estimator which learned the correlation between the ON-state resistance of drain-source of the power MOSFET with respect to the device degradation and then with the RUL. The reported graphics performance seems very promising. Furthermore, such authors investigated the health monitoring of power devices even in noisy scenarios or in the presence of measurement data perturbations [38].

In the state of the art, another aspect investigated by the scientific community is the optimization of the input variables needed to retrieve a robust RUL prediction of power devices. On this regard, Dusmez et al [39] proposed a custom designed accelerated thermal-aging testing-scenario applied to multiple discrete power MOSFETs simultaneously. The so-collected experimental data are specifically related to ON-state resistance of drain-source as the failure early marker of MOSFET degradation. The remaining useful lifetime (RUL) of degraded power MOSFETs were estimated through classical least-squares algorithm run on experimental data filtered by Kalman Filter. With respect to prior-art, the main advantage of that method is that it does not require to measure the junction temperature for retrieving RUL estimation. In a preferred configuration they were able to retrieve a RUL estimation of power device with an average error ranging from 3.5 to 4.5 %. Another similar approach based on monitoring of ON-state resistance of drain-source of discrete power MOSFET is reported in [40], [41]. Further deep learning methods for RUL prediction of power devices can be found in [42]–[44].

Specifically applied to power modules, there are some scientific contributions [45]–[48],

In any case, particularly performing deep architectures in terms of time-series forecasts have recently been presented in the scientific literature. The first was presented in [49] i.e. N-BEATS. It is composed of a backbone that embeds backward and forward residual processing with a very deep stack of fully-connected layers. In some time-series forecast benchmark-competitions, N-BEATS was able to improve forecast accuracy by 11% over a statistical benchmark. Another interesting pipeline based on deep learning is the N-HiTS described in [50]. N-HiTS is similar in some aspects to N-BEATS. It performs local nonlinear projections through basis functions across multiple regression blocks. Each block consists of a multilayer perceptron (MLP), which learns features both in backward and forward processing. The MLP blocks are grouped in stacks specifically trained to learn specific dynamic of the input signal. Finally, the pipeline named DeepAR is discussed [51]. DeepAR is a deep system suitable to produce accurate probabilistic time-series forecast based on training of an auto-regressive recurrent deep backbone over several time-series. More details in [51].

From the careful analysis of the above methods it is clearly evident how the prediction of the residual lifetime in an electronic system is particularly complex and can be faced at the "component level" by building a model which characterizes the physics of the device to be monitored including therein the concept of failures or at "system level" by applying statistical approaches which however have significant limitations (see the FTA method considered among the best performing at the system level).

Among the investigated methods, those based on artificial intelligence (both machine and deep learning) showed particularly performing results both using classic fully connected architectures (MLP, regressors, etc..) and recurrent, convolutive, hybrid networks or networks based on modern Transformer architectures embedding self-attention layers. However, these methods have been applied to discrete electronics (MOSFETs, IGBTs, power MOSFETs) with promising results obtained on tests carried out on public datasets.

Finally, from the analysis of the known art carried out up to now, the careful study that the scientific community has made of the so-called predictive indicators-markers of physical degradation of the device has emerged, noting that among these, the ON-state resistance of power device (and therefore consequently the ON-state voltage between drain and source) would seem to be the most promising.

However in the real application of the electric car, object of this contribution, there is the need to monitor the residual lifetime of a power module in SiC technology which, as described in the introductory section, includes 4 SiC power MOSFETs for each bridge, for a total of 8 power SiC MOSFETs connected in half-bridge configuration, which interact with each other for the electro-thermal function of the module in the traction-inverter of the electric vehicle.

From the analysis of the state-of-the-art emerged that the deep learning methods have significantly improved the health monitoring and RUL prediction systems of the electronic devices. With reference to discrete power devices, solutions based on deep learning which include the use of backbones based on MLP, LSTM, Convolutional Networks -with and without attention mechanism- as well as the recent architectures based on Transformers, have shown to have excellent performance. Some authors have investigated health monitoring with reference to SiC power modules [44]-[47] using machine learning methods or physical-heuristic models for the analysis of degradation over time. However, the analyzed solutions showed some drawbacks that the authors have tried to address in the method herein proposed. Specifically, many of the proposed systems have been validated on discrete power components and with public datasets (C-MAPSS). Many proposed methods are based on the predictive analysis related to the dynamics of the ON-state resistance between drain and source (Rds_{ON}) which for its measurement requires the sampling of both a voltage (Vds) and a current (I_d) with consequent introduction of further measurement noise [37]. Furthermore, the (Rds_{ON}) is strongly influenced by the gate-driving voltage of the power MOSFET device, which in SiC technology is deliberately set significantly higher than the threshold voltage to accelerate dynamic switching. However, the most significant challenge for health monitoring methods is related to the prediction of the RUL on power modules (specifically in SiC technology for traction inverters in latest generation electric cars). As specified in the introductory part of this work, the power module contains 8 parallel power devices for each switch, each having specific characteristics in terms of electrical and physical parameters (threshold voltage, channel width, channel length, etc. ...). Therefore, determining a predictive system for the entire power module is particularly complex since the input marker used and the prediction system have to take into account the impact that each power discrete will have in the overall characterization of the residual lifetime of the entire module. Finally, the proposed RUL prediction system should be scalable and dynamically configurable as it would be appropriate to validate any boost in performance that the simultaneous processing of other co-variate input signal could bring about. A further drawback of the state-of-the-art methods is related to the classical issue embedded in the time-series forecasting scenario i.e. the intelligent pruning of the input data. In fact, considering that the statistical distribution of the input time-series ((Rds_{ON}), (Vds_{ON}), etc...) is unknown a-priori, it is not possible to discriminate the input data components having a significant impact on the output-prediction with respect to those that could be neglected in the processing phase. This involves a computational overload and an inefficiency of the entire prediction system. In this context, the objective of this work is to propose an innovative pipeline based on a deep learning architecture called "Temporal Fusion Transformer" which will process the data of the ON-state voltage drain-source sampled from the low-side or high-side interface of the module, in order to characterize its residual lifetime in a real application context (traction inverter of the electric car) by using data collected in a thermal-electrical stress-aging process of the device through power cycling (see previous section). The authors will show that both in terms of performance and robustness in relation to the input data, the proposed method outperforms the approaches presented in scientific literature based on deep learning methodologies. The designed system is also configurable as it allows to extend the input data allowing the monitoring of cross-correlations between multi-modal input data with respect to the RUL prediction target. It has been tested on real data sampled from electrical-thermal stress procedures of the power module and therefore it is also reliable in terms of model fitting compared to the real application scenario. It is based on the sampling of a single variable (voltage) although, as mentioned, it can be extended to other input data. It has been validated on various power modules of the same product family (ACEPACK DRIVE) confirming high performance in health monitoring of the entire electronic module. Moreover, through ad-hoc intelligent blocks the proposed architecture has been able to pre-process the input data weighting more about the input components that best contribute to the overall predictive performance of the system. The designed architecture is also compressible with Knowledge Distillation techniques, currently undergoing further scientific investigations.

Material and Methods

As introduced in the previous sections, the goal of this scientific contribution is the prediction of the residual useful lifetime (RUL) of SiC power modules in an automotive application scenario, specifically oriented to the electric car traction inverter system. From the analysis of the related work (see previous section) we defined as input of the proposed system, one of the most used predictive marker, i.e. the ON-state drain-source voltage Vds_{LS} of low-side part. As illustrated in the introductory section, the power module is composed of a switch that regulates the low-side and the high-side of the device in half-bridge configuration. Obviously it is enough that only one of these bridges degrades to impact the functioning of the entire module.

Therefore, in order to optimize the proposed prediction system, we have implemented the monitoring of only the ON-state drain-source voltage of the low-side part of the module Vds_{LS} considering the same voltage of the high-side section as a covariate variable to be passed as exogenous data and verifying in the testing phase the impact that this importing has in the RUL prediction. Obviously, it is possible to invert the data sequence considering Vds_{HS} as input and Vds_{LS} as covariate variable without affecting in any way the predictive performances of the proposed system.

As detailed in the introduction section of this contribution, the RUL of the analyzed SiC power module is determined by analyzing the drift of the predicted future dynamics of the same input variable i.e. Vds_{LS}. Therefore, we reduced the prediction of the RUL of the SiC power module to a classic problem of multi-horizon time-series forecasting [52], [53].

Multi-horizon time series forecasting refers to the capability of predicting multiple time-steps samples or patterns of an input time series which includes both observed and past samples as well as such feature information. This represents a key challenge in time series forecasting by machine learning approaches. Several deep backbones based on auto-regressive models, recurrent neural network (RNN) and transformers (as applied in this work) have been used for multi-horizon time series forecasting. Specifically, our target is to predict multi-horizon future samples of the input variable i.e. the ON-state voltage between drain and source of the analyzed power module (Vds_{LS}). Specifically, we have splitted the input variable Vds_{LS} as past and known input samples and we have tested the covariate signal Vds_{HS} (ON-state voltage between drain and source - high-side part of the silicon carbide power module) in the overall designed RUL predictor. In order to correlate the predicted output with RUL of the analyzed model, we considered a threshold-based binary classification system. When the predicted future samples of the Vds_{LS} drift with respect to the fixed threshold (5 % of its initial value), the analyzed power module can be considered degraded. Therefore, the time range that elapses from the moment of the prediction to the overflow of 5% of the Vds_{LS}, is represented by the remaining usefull life of the device. On the other hand, in cases in which the predicted Vds_{LS} samples (in the selected prediction time window which varies from a minimum of 5 samples to a maximum of 50 samples) do not show drifts greater than the pre-set threshold, the system does not detect any start of degradation, therefore no problem in terms of RUL. The same classification scenario has been tested when we have used Vds_{HS} as covariate input. Anyeay, the first step was the design of the deep backbone.

After careful investigation, the authors of this contribution have selected the use of a deep architecture with a backbone based on Temporal Fusion Transformer (TFT). The TFT is an attention-based deep neural network usually applied for multi-horizon data forecasting [54]. TFT is able to reach high-performance with respect to classical deep architectures usually employed for 1D time-series forecasting such as LSTM, Temporal Convolutional Networks and so on [54]. This performance is achieved because the TFT includes the following internal processing blocks: (1) A Static Covariate Encoder Block suitable to encode context vectors of additional input data to be used to boost the overall prediction performance of the TFT; (2) A Gated Residual Network combined with a Variable Selection Layer suitable to optimze the impact of irrelevant input data; (3) A Sequence-to-Sequence Encoding/Decoding Block to perform a locally encoding/decoding of the input data both the so called "past input data " as well as the the known future input data; (4) A Temporal Self-Attention Decoder to learn long-term dependencies embedded in the input data. Moreover, the proposed TFT backbone has been updated with respect to the original ones proposed in [54] as the Sequence-to-Sequence block has been improved by using the Temporal Convolutional Neural Networks instead of the classical LSTM [54], [55]. More detail about TFT architecture in [54]. The proposed SiC power module predictor will be described in the next section.

THE PROPOSED DEEP LEARNING SYSTEM

In Fig. 7 an overview of the SiC power module RUL predictor -based on the usage of the enhanced TFT model- is reported. Each of the reported blocks will be described in the following-subsections.

Input Block

As introduced, the input of the proposed system (as reported in Fig. 7) is the ON-state voltage drain-source of the low-side bridge Vds_{LS} of the analyzed SiC Power module. This signal is splitted partly as "Past input" and partly as "Known future samples" [54]. If we indicate with $\xi_{i,t}$ the vector containing the input variable, it can be composed as follow

$$\xi_{i,t} = [\phi^{i,t}_{Vds_{LS}}, \psi^{i,t}_{Vds_{LS}}] \tag{1}$$

where $\phi^{i,t}_{Vds_{LS}} \in \Re^{m_s}$ represents the part of the input considered as "past input", which we supposed be sampled at each step, while $\psi^{i,t}_{Vds_{LS}} \in \Re^{m_p}$ represents the part of input considered as "known input". The so collected input data will be passed to the "Variable Selection Blocks (VSB)". Anyway, before to introduce and describe che VSB, it is needed to introduce and describe the Gated Residual Network block (GRN) as it will be embedded in the VSB.

Gated Residual Network Layer

The analytic correlation (often nonlinear) between inputs and system output is often unknown or partially known, so that it is very challenging to select the key variables with respect to the negligible ones. Non-linear processing of the input data is not always necessary and sometimes simpler linear processing models may be enough. Not knowing a-priori the statistical correlation between input and output and the need to apply non-linear processing to the input data requires the implementation of the Gated Residual Network [54].

The following Fig. 8 shows the working diagram of the implemented GRN layer.

The GRN offers adaptable depth and network complexity. It performs non-linear processing of the input data as well as context information as follow:

$$GRN_\omega(\xi, c) = LayerNorm(m_i + GLU_\omega(\eta_1)) \tag{2}$$

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega} \tag{3}$$

$$\eta_2 = ELU(W_{2,\omega}i + W_{3,\omega}c_i + b_{2,\omega}) \tag{4}$$

$$GLU_\omega(\gamma) = \sigma(W_{4,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega}) \tag{5}$$

In the set of Eqs (2)-(5) ELU represents the Exponential Linear Unit activation function described in [56] while η_1 and $\eta_2 \in \Re^{d_{model}}$ are intermediate layers, the function "LayerNorm" is a normalization layer as described in [57] while ω represents an index for weight sharing. Finally, with the activation "Gated Linear Units" (GLUs) [58] we are able to suppress any parts of the designed architecture not needed for the given dataset. Specifically, as reported in Eq. (5), we have defined ad-hoc GLU function by embedding classical sigmoid activation σ applied to a weighted $W_{(.),\omega}$ and biased $b_{(.)}$ of the input γ properly multiplied (Hadamard product) with similar factor. d_{model} is the hidden state size. In this way, the activation GLU will be suitable to control the level of application of the GRN related to the original input m_i performing a proper selection of the input data [54].

Variable Selection Block

As introduced, the specific contribution of the input data to the output dynamic are usually unknown. For this reason, it is normally not possible to drop the contribution of input variables that are poorly correlated with the output target. The proposed deep learning system has been designed to provide an intelligent a-priori variable selection through the use of the block "Variable Selection Block" (VSB) applied to both input data and static and time-dependent covariates data. The target of the VSB is clearly to retrieve information about which input variables are most significant for the prediction problem as well as to remove any unnecessary noisy inputs.

Often, in classical time-series forecast based problems the input datasets contain discriminative features with less predictive capability which therefore impact the computational complexity and sometimes negatively the performance of the entire predictive architecture. The VSB block has the purpose of significantly reducing the contribution of these

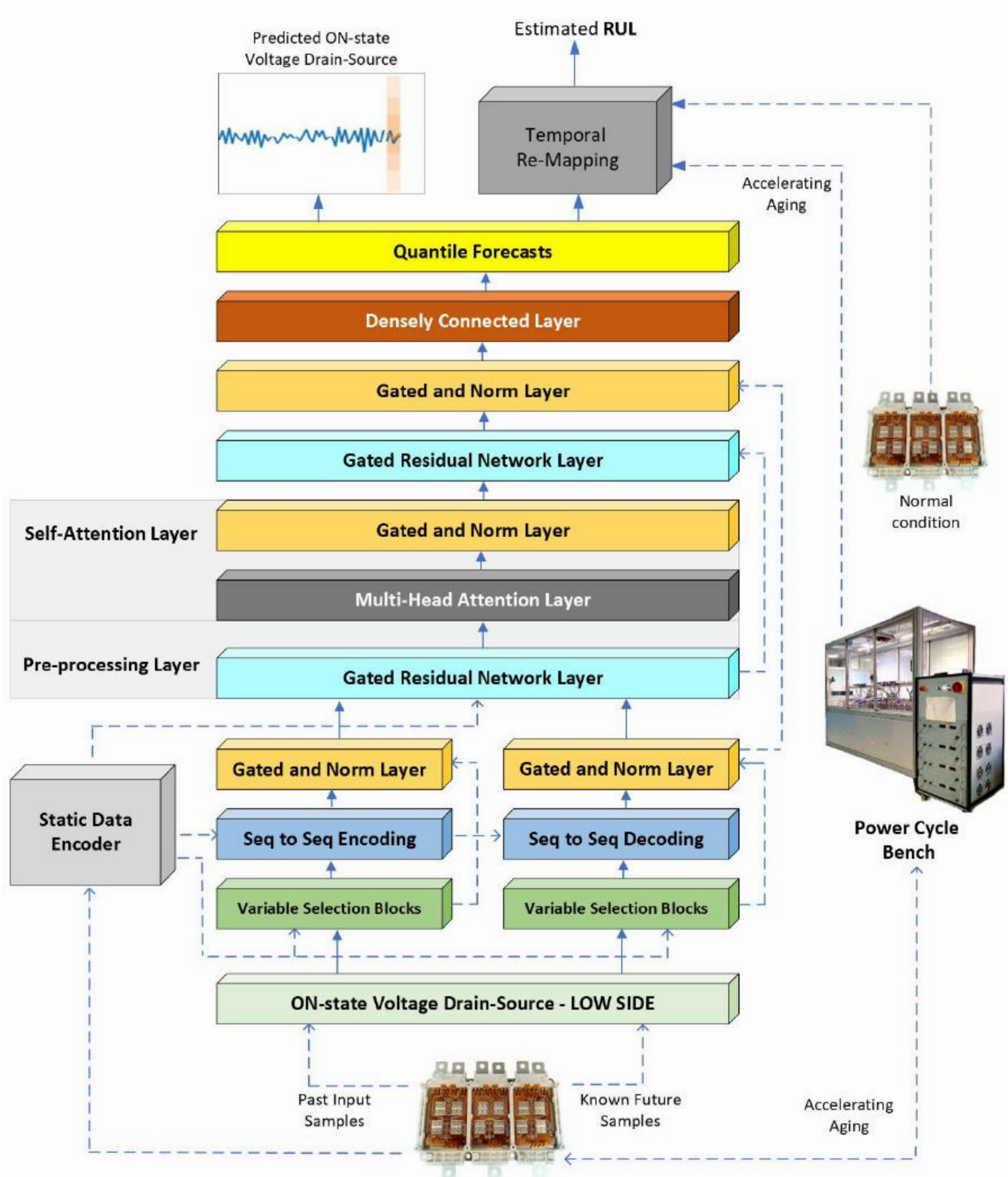


Fig. 7. The overview scheme of the proposed SiC Power Module Lifetime Predictor

features embedded in the input data by weighting more those which instead have a significant contribution in the predictive performance of the model [54]. The following Fig. 9 reports an instance of the VSB diagram.

We now introduce the working flow of the VSB. Preliminary embeddings [59] for categorical variables and linear transformations for continuous variables will be applied. The target is to tranform input variables into vector of dimension d_{model} as requested by the following layers of VSB.

As described in Fig. 9, we indicated with $\phi_t^j \in \Re^{d_{model}}$ the transformed input of the $j-th$ variable at time t and then Φ_i the whole set of so computed variables, while with Φ_F we have denoted the flattened representation of Φ_i, having length l_x. We have also setup the possibility of processing external context vector c_i. The VSB works as follow. Preliminary, a weighting of the flattened representation of the embedded input data (together with context data c_i if any) is made by using a GRN block followed by a softmax processing:

$$\Phi_F^{\xi} = SoftMax(GRN(\Phi_F, c_i)) \tag{6}$$

The context data c_i will be obtained from the static covariate encoding as described in the next section [54]. For each processing step, the designed model performs additional non-

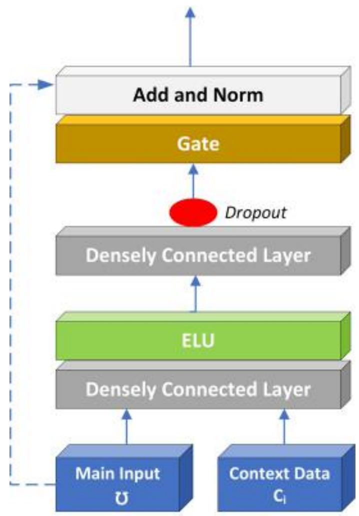


Fig. 8. Gated Residual Network block diagram

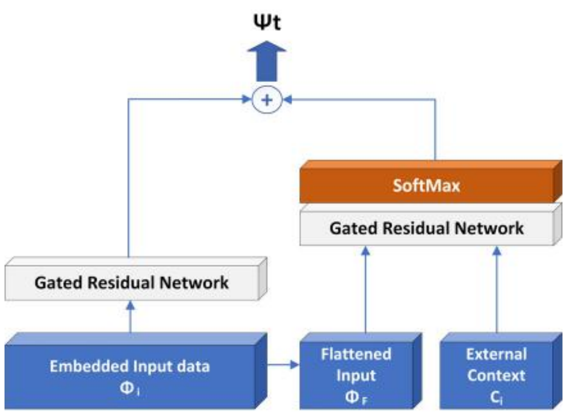


Fig. 9. Variable Selection block diagram

linear processing by analyzing the ϕ_t^j data through a further GRN processing with a weight shared accross all time steps.

$$\tilde{\Phi}_t^j = GRN_{\tilde{\Phi}_t^j}(\Phi_t^j) \tag{7}$$

At this point, the output of the block with SoftMax as reported in equation (6) will be used to weight and suitably combine the so processed input data $\tilde{\Phi}_t^j$ thus performing a selection of variable properly represented by the following equation which represents the output Ψ_t of the block VSB extension.

$$\Psi_t = \sum_{j=1}^{l_x} \phi_F^{\xi,(j)} \tilde{\Phi}_t^{(j)} \tag{8}$$

As described in Fig. 9, such input of the VSB are related to the Static Data Encoder block as detailed in the overview diagram of the proposed pipeline (see Fig. 7). The goal of this block is to integrate more information from external meta-data. These external data will be integrated to the temporal decoder part (by means of a GRN pre-processing) where static variables could improve the overall decoding performance. Moreover, such input data will be used to enrich the input features of the temporal decoder [54]. In this proposal, we have checked as external data the usage of Vds_{HS}. More details on the next section reporting the experimental results.

The so weighted output of VSB will be fed as input of the designed sequence-to-sequence encoding/decoding layer.

Sequence to Sequence Encoding/Decoding

The sequence-to-sequence (Seq2Seq) encoding/decoding of the pre-processed input data have been designed for providing intelligent assessment of the temporal sequential data. Leveraging local correlation between time-series data, we are able to construct discriminative features suitable to improve the overall performance of the downstream deep system embedding attention-based processing [54]. Due to proposed discrimination between input ”past” data and input ”future” samples, the implemented pipeline embeds a deep layers for encoding ”past” data $\Psi_{t-k:t}$ and ”decoding” future samples $\Psi_{t+1:t_{max}}$. In [54] was proposed the usage of LSTM as basic Encoder/Decoder backbone of the Seq2Seq layer. In this work the authors replaced the LSTM with a Temporal Convolutional Neural Network (TCN) as backbone of the Encoder/Decoder of the Seq2Seq layer [55]. Bai et al. [55] introduced TCN to replace the classical recurrent architectures in time-series forecast tasks. More in detail, TCNs are deep architecture embeddings ”causal convolutions” i.e. processing of input with hypothesis that no information “leakage” occurs from future data to the past ones. Basically, TCN can take an input 1D sequence of any length, mapping that to such output sequence of the same length. The performance of TCNs in time-series forecasting is better than classical deep architectures including recurrent network and linear regressors [55].

Considering the split that we have assumed in our system between ”future” and ”past” data, the TCN is certainly the best backbone to perform an encoding/decoding embedding of the input data, obtaining better performances than the architecture originally proposed by the authors which included LSTMs (as confirmed in the experimental results section) [54], [55]. To consider potential imprinting of such external metadata we embedded the context data coming from the Static Covariate Encoders to initialize the structure of the TCN. We indicate with $\Phi(t, n)$ the set of features generated by the designed TCNs layers:

$$\tilde{\Phi}(t, n) = LayerNorm(\psi_{t+n} + GLU(\Phi(t, n))) \tag{9}$$

The output of the Seq2Seq encoding/decoding will be fed as input of a further pre-processing layer which performs a GRN processing as described in the previous section. The GRN processing of the $\tilde{\Phi}(t, n)$ features will be passed to the Multi-Head Attention layer.

The Multi-Head Attention Layer

Attention mechanism is used by all Transformer-based architectures to identify complex relationships among the input data [60]. Anyway, an issue on the input-data long-term relationship learning has been highlighted [54]. To overcome this issue, such authors proposed the usage of the Multi-Head

Attention Layer (MHAL) [54]. In the MHAL, the input is projected into several representation sub-spaces using various "attention heads" based on the classical processing performed by the use of Query/Key/Value weight matrices [54], [60], [61]. By means of the implemented MHAL, we are able to improve the learning capability of the overall proposed system especially in the input data long-term dependencies.

A mathematical representation of the MHAL processing $\Theta_{MHAL}(Q, K, V)$ can be formalized as follow:

$$\Theta_{MHAL}(Q, K, V) = \tilde{H} W_H \tag{10}$$

$$\begin{aligned} \tilde{H} &= \tilde{A}(Q, K) V W_V = \\ &= [1/H \sum_{h=1}^{m_H} A(Q W_Q^{(h)}, K, W_K^{(h)}] V W_V = \\ &= 1/H \sum_{h=1}^{m_H} Attention(Q W_Q^{(h)}, K, W_K^{(h)}, V W_V) \end{aligned} \tag{11}$$

where in (11) we have defined classical attention mechanism variables as follow: Values $V \in \Re^{N x d_V}$, keys $K \in \Re^{N x d_{attn}}$ and queries $Q \in \Re^{N x d_{attn}}$ while $A()$ represents a normalization factors, $W_K^{(h)} \in \Re^{d_{model} x d_{attn}}$, $W_Q^{(h)} \in \Re^{d_{model} x d_{attn}}$, are learnable weights for the queries, keys and values and $W_V \in \Re^{d_{model} x d_{attn}}$ represents weights shared across all attention heads, while in Eq. (10) the term $W_H \in \Re^{d_{attn} x d_{model}}$ performs a final linear mapping of the attention-weighted features [54], [60], [61].

Through the mathematical representation of the MHAL reported in Eqs. (10)-(11), we are able to boost the classical head-attention mechanism as we can learn - for each heads - specific dynamic of the input features to be combined and weighted as per Eq. (10).

Finally, as reported in Figs. 7-9, further "Gated and Norm" (GN) and GRN processing will be applied to the output of the MHAL block for preserving causal correlation between features and attention-based processing [54]. We indicate with $\Psi'(t, n)$ the output of the GN and GRN processing of the MHAL output.

The so processed features will be fed to the Densely Connected Layer for final processing and time-series forecasting. This layer performed a quantile time-series forecasting by parallel prediction of data at different percentiles such as $10^{th}, 50^{th}, 90^{th}$ for each time-step. To do that, we have defined ad-hoc loss function as reported:

$$\mathcal{L}(\Omega, W) = \sum_{y_t \in \Omega} \sum_{q \in \Xi} \sum_{\tau=1}^{\tau_{max}} \frac{Q_L(y_t, \tilde{y}(q, t-\tau, \tau), q)}{M_{\tau_{max}}} \tag{12}$$

$$Q_L(y_t, \tilde{y}, q) = q(y - \tilde{y})_+ + (1-q)(\tilde{y} - y) \tag{13}$$

$$\tilde{y}(q, t, \tau) = W_q \Psi'(t, \tau) + \mu_q \tag{14}$$

where the set Ω represents the training data statistical domain embeddings M samples, while W represents the weights of the implemented boosted temporal fusion transformer, Ξ is the set of output quantiles while with "$(.)_+$" we have defined the "$max(0; .)$" operator. In the Eqs set (10)-(14) the term $\tilde{y}(q, t, \tau)$ represents ad-hoc quantile forecasts of the input time-series computed as linear transformation of the GN/GRN processing of the MHAL output. The term y_t represent the target. Finally, the term W_q and μ_q represents linear coefficients for the specified quantile q. More detail about quantile forecast in [54].

EXPERIMENTAL RESULTS

The aim of this section is to report the experimental results obtained from the application of the enhanced temporal fusion transformer boosted with the convolutional Seq2Seq block, for the prediction of the ON-state Vds_{LS} trend for the characterization of the residual lifetime of the power module in SiC technology. Below we will preliminary describe the phase of data collection, therefore, of training and benchmarking of the designed architecture.

The Dataset collection

Performance analysis is carried out on dataset containing six sessions of Power Cycling data-acquisition applied to six different ACEPACK DRIVE SiC Power Modules. For each power-cycling session, we sampled about 400k samples of ON-state Vds_{LS} and Vds_{HS} as well as some further variables such as temperature, currents, etc.

As previously reported, the predictive marker we want to learn with the proposed system is the ON-state Vds_{LS} dynamic. We collect the input ON-state Vds_{LS} by using a classical CSV format (with ad-hoc software wrapper we have implemented to convert data from the native format of the Power Cycle bench-system). As introduced, through the power-cycling procedure we are able to bring the power module to the end of its life and therefore to characterize, through this "accelerated aging" procedure, the correlation between the residual lifetime of the device and the trend of the ON-state Vds_{LS}.

Some details about the power-cycling procedure we have designed to collect dataset used to learn our proposed deep model. Power cycles are employed considering temperature swing of 100 °C, maximum die temperature of 150 °C and current on-time of 3.5 s. Device self-heating is accomplished driving the V_{GS} to positive value, much higher than threshold value, and "virtual junction temperature" is monitored during the off-state, applying a negative voltage of - 8 V. Negative value is needed to ensure that sensing current flows across the body diode. For each tested power modules, one switch from low-side is placed in series with a high-side switch of different phase. The procedure is repeated until the device reaches the end of life characterized by an exponential and significant increase in the ON-state Vds_{LS}, specifically, with an average increase that exceeds the threshold of 5% compared to the initial value of this quantity. In Fig. 6 a classic trend of the ON-state Vds_{LS} which over time, during the accelerated aging of Power Cycling, progressively increases its value, has been reported.

A preliminary pre-processing of the sampled ON-state Vds_{LS} is made by filtering the collected signal in order to delete samples after the end of life of the device. The so pre-processed dataset contains about 418.700 samples splitted for training validation and test sets as shown in Table I.

TABLE I
INPUT DATASET SPLIT CONFIGURATION

	Training-set	Validation-set	Test-set
Samples	279.134	69.783	69.783

In Fig. 10 and Fig. 11, such instances of the collected and pre-processed input ON-state Vds_{LS} and Vds_{HS} are reported. In the mentioned figures, the abscissa axes report the time progression during the power cycling phase. In the Vds snapshot shown in each figures, the part of the signal corresponding to negative abscissa values refers to the Vds samples considered as input while the part of the samples with abscissa values greater than or equal to zero represents the signal part to be predicted.

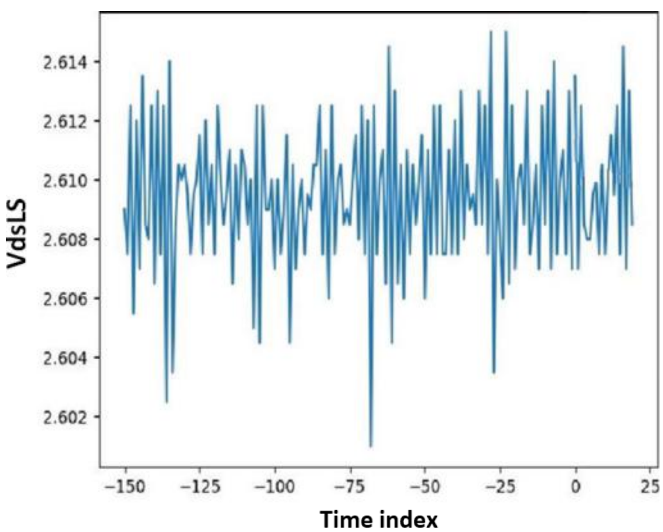


Fig. 10. Dynamic of the pre-processed ON-state Vds_{LS}

In the next section , the training phase of the implemented deep system will be detailed.

Training Procedure

We have trained and validated the designed proposed solution by using the collected dataset as described in the previous section. Moreover, we have compared the performance of our proposed solution with respect to other deep systems usually applied for time-series forecasting applications (see section "Related Works"). Specifically, we have compared our deep learning solution with the following deep models: MLP [29], LSTM [28], [34], [35], N-BEATS [49], N-HiTS [50], DeepAR [51], TFT-LSTM [54].

For all the tested architectures, we have performed a phase of hyperparameters optimization through random search over a predefined dynamic-space, using the same number of iterations across all tested backbones. According to our tests, we have

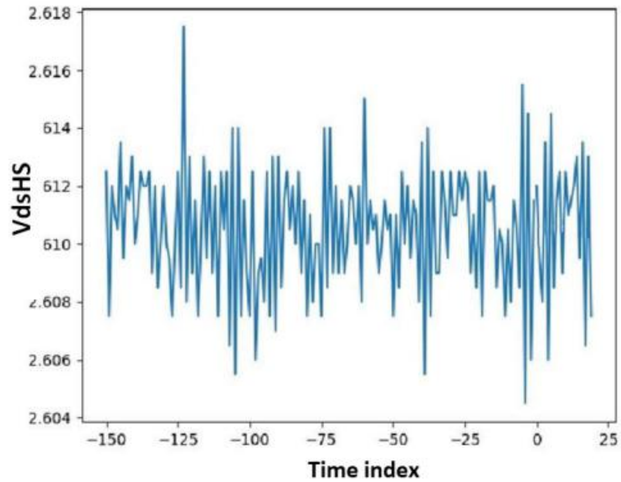


Fig. 11. Dynamic of the pre-processed ON-state Vds_{HS}

noticed that the best temporal-window of input signal to be used as input size of the downstream deep model is greater than 150 samples, so that, we selected 150 as optimal size of the input ON-state Vds_{LS} signal. More in detail, we did many tests tuning the length of the input temporal-window depth verifying the impact on the overall performance the designed pipeline. Specifically, we have noticed that for input temporal depth lower than 150 samples, the performance drastically dropped (we have a drop that oscillates between 15% - 20% for temporal-depth between 100 and 150). Input temporal-depth lower than 100 do not allow the system to predict with acceptable accuracy, considering that we are dealing with automotive applications that require high reliability. By increasing the temporal-depth with values higher than 150, we noticed a little significant performance improvement to justify the computational cost deriving from an increased temporal-depth. Specifically, we noted - for each number of prediction steps (see Table III) - a modest average increase of 1.2% for input temporal depth equal to 200, of 1.56% for a length of 250. For temporal depths greater than 300 , basically there are no further significant performance gains in Vds_{LS} time-series forecasting. Therefore, considering the target of distilling our pipeline on an automotive-grade microcontroller embedded in the latest generation Engine Control Unit of electric car, we opted as a trade-off solution, the use of an input temporal depth of 150 as it represents the best solution considering the performance/computational cost ratio.
As reported in Figs. 10-11, the signal samples corresponding to a value of abscissa lower than zero are considered as input i.e. 150 samples as sliding temporal window of the input signal. We have tested several output window prediction-size in a range from 5 to 50 samples. For each model, the training was carried out for 150 epochs. Our proposed Temporal Fusion Transformer boosted with Temporal Convolutional Networks as Seq2Seq embedding (TFT-TCN) has been developed in Pytorch environment running in a Multi-Core INTEL Server

with a NVIDIA GPU RTX A6000 with 48 GB of video memory. In this phase, we have collected experimental results of our proposed solution (TFT-TCN) compared with respect to classical TFT (embedding LSTM as Seq2Seq layer) [54] and TFT with external static input fed with ON-state Vds_{HS} (TFT-EI), MLP [29], LSTM [28], [34], [35], N-BEATS [49], N-HiTS [50], DeepAR [51]. All the benchmark tests have been executed in the same accelerated underlying hardware. More details about each model configuration are reported in Table II.

For performance evaluation classical and commonly employed metrics for timeseries forecasting have been used. Specifically we have used the following benchmarks metrics: Mean-Abosolute Error (MAE), Root-Mean-Squared Error (RMSE), computed as follow:

$$e_i = A_i - F_i \tag{15}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|e_i^2|} \tag{17}$$

In Eqs (15)-(17) we have defined as A_i the actual value (label) i.e. the real ON-state Vds_{LS} collected during the Power Cycling while with F_i the predicted value by the tested deep model. In Table II we have also defined the Loss function used in the tested models. For more details about the mathematical formulation of the used Loss function, please refer to related cited reference.

We have configured as ON-state Vds_{LS} cut-off threshold related to the end-of-life of the tested SiC power module, an increase in average value equal to or greater than 5% with respect to its initial value. In the following Table III we have reported the benchmark performance comparisons related to all the tested deep pipelines including the proposed ones. In the "Steps" column we have reported the collected test-results with respect to different temporal depth prediction steps of the ON-state Vds_{LS} i.e. from 5 to 50 Vds_{LS} samples of prediction.

As highlighted from a careful reading of the results reported in Table III, the proposed architecture (TFT-TCN) outperforms the other pipelines in all of the tests and with the variation of the temporal prediction steps. Both with reduced temporal depth of prediction (5 samples) and with the increase of these up to the maximum of 50 samples of prediction of the ON-state Vds_{LS}, our architecture showed high predictive performances in order to the various metrics adopted for the performance comparison. Moreover, our solution outperforms the previous solution based on TFT but with LSTM as Seq2Seq layer, thus demonstrating the very promising capability of TCN - compared to LSTM - for performing embeddings and learning of the input signal features. Even the use of an exogenous input signal (ON-state Vds_{HS}) used as a static covariate input, has improved the performance of the TFT in some tests (see TFT-EI), it has not significantly increased the prediction capabilities of the system in general terms.

Ablation Study

In this section we have analyzed the effects of ad-hoc ablation of such sub-systems embedded in the proposed pipeline. Specifically, we have tested the impact of the Variable Selection Block (VSB) and the Gated Residual Network (GRN) layers on the overall performance of the designed solution. In the previous section as reported in Table III, we highlighted the effects deriving from the replacement of the LSTMs with the TCN. We also verified the effects on the performance of activating or deactivating the Static Data Encoder block referred to the covariate inputs (Vds_{HS}). We proceed with a first ablation study for evaluating the impact of the VSB on the overall performance of the delivered architecture in relation to the various prediction steps (see Table III).
As mentioned previously, the target of the VSB is to retrieve information about which input variables are most significant for the analyzed prediction problem as well as to remove any unnecessary noisy inputs. We directly fed the Vds_{LS} samples into the Seq2Seq Encoding/Decoding blocks without VSB pre-processing. The performance of the so modified deep architecture is reported in the following Table IV.
Similarly to the VSB, we tested the proposed system without the GRN blocks which have the goal of preliminary estimating the statistical correlation between the input data and the output data. For both ablation studies we had to adjust the dimension of the intermediate data accordingly.
The results of these ablation tests are reported in Table IV (TFT-TCN-VSB: The proposed architecture without VSB layers; TFT-TCN-GRN: The proposed architecture without GRN layers; TFT: The native implementation of the temporal fusion transformer [54]; TFT-EI: The native implementation of the temporal fusion transformer with external input; TFT-TCN: The proposed solution). From an examination of the results, the importance of both VSB and GRN pre-processing appears evident. It can be seen that the proposed architecture dropped in performance significantly as the prediction time depth increases due to the absence of a pre-processing that carefully weights input data components that contribute most to the predictive performance of the model.

About the RUL, we have extended our "experimental results" section by adding the performance in terms of RUL prediction. The following section has been added to the revised manuscript.
After that we also proceeded to analyze the performance of our pipeline in terms of prediction of the Remaining Useful Lifetime (RUL). Specifically as introduced, a drift of 5% of the Vds_{LS} -compared to its initial value- corresponds to a clear marker of device degradation with consequent risk for the traction system of the electric vehicle. Therefore, our system will discriminate (from a thermo-electrical and structural point of view) a healthy power module if the prediction of the Vds_{LS} (according to the considered steps i.e. the selected prediction time depth between 5 and 50) does not show any drift or shows a drift within the critical alert threshold. Conversely, the RUL estimation is triggered in the scenario in which a drift exceeding the critical threshold of 5% appears in the Vds_{LS} prediction time window. In this case, the prediction

TABLE II
OPTIMAL HYPERPARAMETERS CONFIGURATION. TFT COLUMN REFERS TO ALL TESTED VERSIONS (TFT, TFT-EI, TFT-TCN).

	N-BEATS	**N-HiTS**	**TFT**	**DeepAR**	**LSTM**	**MLP**
Network Parameters						
Dropout Rate	0.1	0.1	0.1	0.1	0.1	0.1
weight decay	1e-2	1e-2	0.1			
Number of Heads	-	-	4	-	-	-
Training Parameters						
Minibatch Size	128	128	128	128	128	128
Learning Rate	3e-2	0.001	0.001	0.001	0.001	0.001
Loss Function	Quantile	NormalD.	Quantile	MASE	MAE	NormalD.

TABLE III
ON-STATE Vds_{LS} PREDICTION PERFORMANCE BENCHMARKS (%)

Steps	**Metrics**	**N-BEATS**	**N-HiTS**	**TFT**	**TFT-EI**	**DeepAR**	**LSTM**	**MLP**	**TFT-TCN**
	MAE	0.23	0.32	0.51	0.33	0.36	0.42	0.80	**0.21**
5	RMSE	0.34	0.47	0.62	0.44	0.41	0.52	0.90	**0.30**
	MAE	0.23	0.51	0.23	0.81	0.34	0.52	0.73	**0.14**
10	RMSE	0.41	0.72	0.91	0.93	0.47	0.62	0.77	**0.26**
	MAE	0.18	0.28	0.42	0.23	0.24	0.67	0.68	**0.15**
20	RMSE	0.27	0.64	0.83	0.44	0.32	0.81	0.72	**0.16**
	MAE	0.72	0.47	0.62	0.63	0.81	0.73	0.14	**0.14**
30	RMSE	0.61	0.71	0.53	0.82	0.62	0.81	0.91	**0.25**
	MAE	0.82	0.42	0.52	0.53	0.53	0.26	0.33	**0.15**
50	RMSE	0.55	0.63	0.96	0.72	0.64	0.37	0.34	**0.27**

time-length of the Vds_{LS} up to the time instant of the drift will determine the residual RUL (temporally re-mapped, as explained below).

Therefore, we proceeded to validate our system on a dataset of 18 ACEPACK DRIVE power modules previously tested in power cycling and with the same configurations described in this work. We splitted, for each module, the related dataset of dimensions similar to the ones used to make learning of our pipeline (see Table I) proceeding to specific read-out moments in which we sampled 150 samples and generated the Vds_{LS} in inference to the our model pre-trained with the dataset described in Table I. The read-out was performed every 500 power-cycling cycles, therefore, for each module we performed an average of 850 test sessions of the model in inference. Specifically, in terms of RUL, we have classified the prediction pertaining to class "0" where the forecasted Vds_{LS} does not show drift beyond the critical threshold. Consequently, class "1" refers to a prediction in which the Vds_{LS} shows a drift beyond the critical threshold of 5%. We tested the method for each prediction time-depth (steps) reporting the results in terms of accuracy, in the following Table V. As evident from the results, the method performs very well also in relation to the prediction of the RUL and for each of the covered temporal depths.

Obviously, it is necessary in the real application to re-modulate the prediction time measurement unit adapting it to the real operating scenario of the SiC Power module in the "Traction Inverter" application. In fact, the time prediction of our residual lifetime marker of the device i.e. the ON-state Vds_{LS} refers to an "accelerated aging" scenario deriving from the use of aggressive power cycling applied to the power module. This approach is needed to obtain a dynamic model of lifetime in acceptable times and without waiting for the end-of-life of the device under normal operating conditions, which could also take years.

For this reason, our solution assumes a sort of "Twin-Testing" of the SiC Power Module. Specifically, in our Power

TABLE IV
ABLATION STUDY BENCHMARKS (%)

Steps	Metrics	TFT-TCN-VSB	TFT-TCN-GRN	TFT	TFT-EI	TFT-TCN
5	MAE	0.58	0.53	0.51	0.33	**0.21**
	RMSE	0.68	0.65	0.62	0.44	**0.30**
10	MAE	0.32	0.33	0.23	0.81	**0.14**
	RMSE	0.94	0.96	0.91	0.93	**0.26**
20	MAE	0.51	0.51	0.42	0.23	**0.15**
	RMSE	0.89	0.87	0.83	0.44	**0.16**
30	MAE	0.75	0.71	0.62	0.63	**0.14**
	RMSE	0.84	0.85	0.53	0.82	**0.25**
50	MAE	0.68	0.71	0.52	0.53	**0.15**
	RMSE	0.99	0.99	0.96	0.72	**0.27**

TABLE V
RUL PREDICTION PERFORMANCE

Steps	Metrics	N-BEATS	N-HITS	DeepAR	LSTM	TFT-TCN
5	Accuracy	0.9523	0.9440	0.9190	0.8809	**0.9761**
	Sensitivity	0.9090	0.9000	0.8500	0.7727	**0.9545**
	Specificity	0.9677	0.9596	0.9435	0.9193	**0.9838**
10	Accuracy	0.9345	0.9222	0.8988	0.8630	**0.9642**
	Sensitivity	0.8863	0.8636	0.8181	0.7500	**0.9318**
	Specificity	0.9516	0.9435	0.9274	0.9032	**0.9758**
20	Accuracy	0.9272	0.9142	0.8761	0.8452	**0.9488**
	Sensitivity	0.8818	0.8590	0.8181	0.7272	**0.9181**
	Specificity	0.9435	0.9338	0.9112	0.8870	**0.9596**
30	Accuracy	0.9071	0.8928	0.8630	0.8309	**0.9392**
	Sensitivity	0.8772	0.8454	0.7545	0.6954	**0.9054**
	Specificity	0.9193	0.9096	0.9016	0.8790	**0.9516**
50	Accuracy	0.8940	0.8777	0.8452	0.8059	**0.9321**
	Sensitivity	0.8227	0.8000	0.7272	0.6454	**0.9000**
	Specificity	0.9193	0.9048	0.8870	0.8629	**0.9435**

Cycing bench, in addition to the accelerated aging under-stress device, another SiC Power Module of the same family is placed in parallel (in our case ACEPACK DRIVE) which however is tested in ON/OFF switching not accelerated (or slightly accelerated) and compatible with the normal switching dynamics in a traction inverter application. During the power cycling phase (accelerated aging) the instants in time will be sampled in which the ON-state Vds_{LS} will increase by pre-defined thresholds such as 0.15%. and then 0.30%, 0.45% and so on. Simultaneously, the time instants referred to the same ON-state Vds_{LS} increments but referred to the second device in normal switching will be sampled. The latter will obviously take place in temporal moments subsequent to those of the device in Power Cycling, thus defining a direct linear proportionality relationship between the temporal dynamics of the lifetime of the two analyzed devices. Specifically, if we indicate with t^k_{PC} the time instants "k" of average increase of the ON-state Vds_{LS} of the pre-set thresholds (0.15% 0.30%, etc..) and with t^k_{NC} the same ones referring to the device in normal condition, we will get the following time correlation:

$$t^k_{NC} = \rho^k * t^k_{PC} \tag{18}$$

Therefore, by calculating the average of these proportional coefficients ρ^k (total number: N_t):

$$\rho = 1/N_t \sum_{i=1}^{N_t} \rho^i \tag{19}$$

we will obtain that the prediction steps shown in Table III must be related to the time unit ρ, determined by the time remapping due to the difference in the working-condition between Power Cycling/accelerated aging and normal condition of the device in the real application. In our tests performed on ACEPACK DRIVE, we estimated that one prediction step corresponds to approximately $\rho = 3.5$ days. Thus, for example, a 5-step prediction indicates a prediction of the end-of-life of the power device performed 17.5 $(3.5 * 5)$ days in advance. The goal of this temporal re-modulation block (Temporal Re-mapping) is precisely to map the temporal unit of the "sample" acquired under thermo-electrical stress (power cycling) into a temporal quantity that is applicable to a typical "normal" power module funcionality in the vehicle's traction inverter. As introduced, the accelerated aging of power cycling allows to train the predictive model of Vds_{LS} in a scenario of thermo-electrical stress and in acceptable times. Vds_{LS} forecasting in an accelerated environment has to be time-mapped in a usual scenario free from such stress. To do this, we therefore sampled the Vds_{LS} drift time instants for the same ACEPACK DRIVE power modules, both in a thermo-electrical stress scenario (Power cycling) and in a normal working scenario [7]. Therefore, for each mentioned drift quantiles of the Vds_{LS} (t^k_{PC}), sampled in power cycling and used as a learning dataset of the proposed pipeline, we identify the corresponding timing of the same drift (t^k_{NC}) in the module driven under normal thermo-electric conditions (see datasheet of the ACEPACK DRIVE [7]), thus calculating the proportionality factor between these time-instants averaged between the various measurements at different drifts (ρ^k). This temporal re-mapping factor will allow, in the event of a drift of the Vds_{LS} beyond the critical threshold (5%) to determine the effective residual useful lifetime.

In Fig. 12, we have reported such instances of the prediction made by out TFT-TCN system and related to Vds_{LS} (overlayed with real signal). In Fig. 13 the loss curves of our proposed architecture with respect to the covered temporal-depth prediction is reported. In Fig. 14 the loss curves of the tested deep architectures are reported.

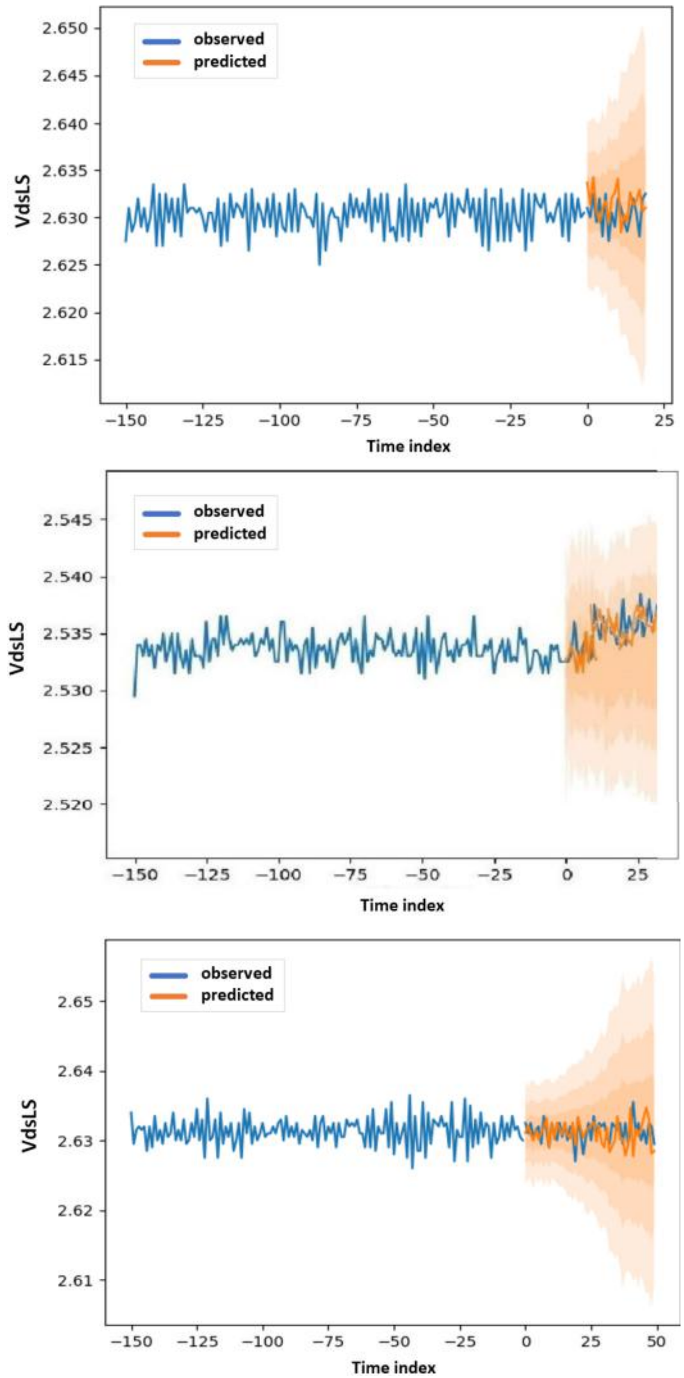


Fig. 12. Overlay diagram of the predicted Vds_{LS} with respect to the real collected signal. A detail of quantile forecasting is reported.

Conclusion and Future works

In this paper the authors have proposed an health monitoring system for SiC power devices in such real automotive application scenario. The proposed system processes as input signal the dynamics of the ON-state voltage Vds_{LS} sampled by the switch of the SiC Power Module in half-bridge configuration. Through the remarkable performance in prediction of the proposed deep pipeline, the designed system is able to early predict the end-of-life of the device, therefore the residual lifetime (the system was able to predict till up to 50 samples of Vds_{LS} equal to several weeks in advance in the real application). Real applications for this pipeline can be traced back to the Traction inverter systems of the latest generation electric cars although the proposed method can be extended

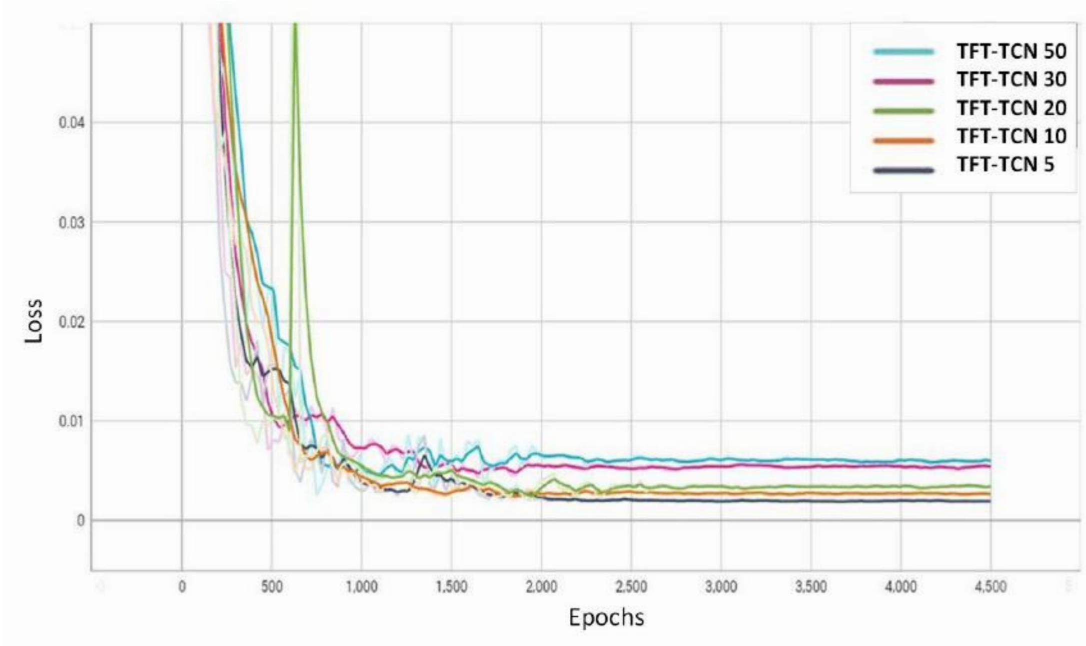


Fig. 13. TFT-TCN loss curves related to different temporal-depth prediction

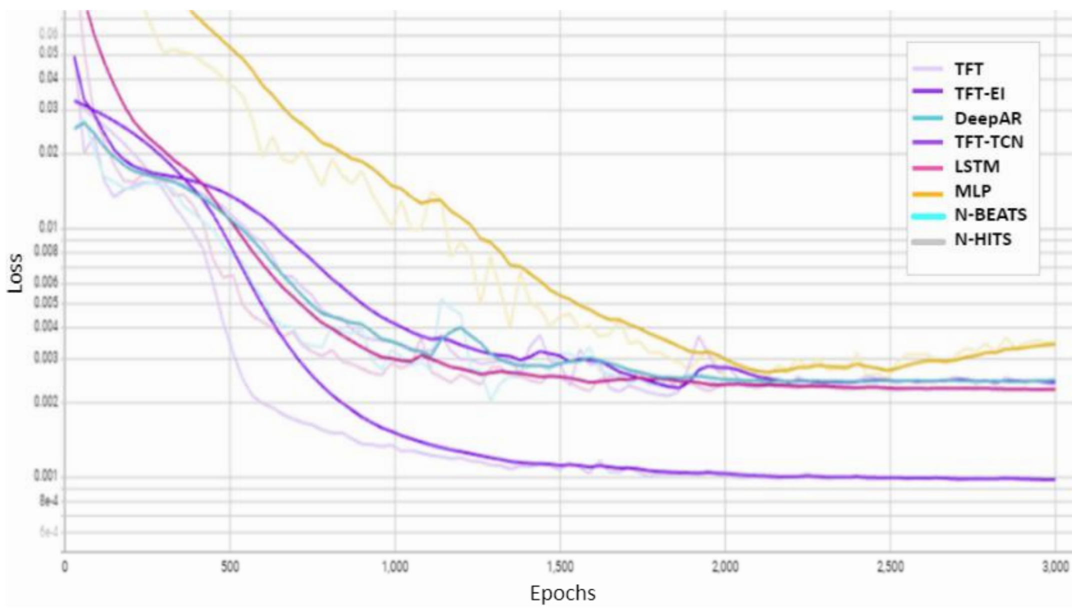


Fig. 14. Loss dynamics of the tested deep architectures

to industrial applications of any nature where it is necessary to monitor the state of degradation (or the residual lifetime) of the eletronic devices. We have implemented a temporal fusion transformer that includes a Seq2Seq encoding/decoding layer based on the use of temporal convolutional networks, thus contributing to the construction of discriminating features and robust correlations in the input time-series samples with respect to the output. By means of the GN and GRN blocks and of the MHAL block, we were able to enhance the capability of the self-attention blocks in reconstructing the correlations between the input and output samples and to discard the sample-dynamics that have no significant impact on the final prediction.

In our research and development activity we are now working on a feature-based Knowledge Distillation algorithm that allows to significantly compress and optimize the backbone of our temporal fusion transformer so as to make it deployable on automotive-grade embedded systems included in the control units of modern electric cars.

Acknowledgment

The authors have to thank the R&D laboratories of the ADG Power and Discretes Division of STMicroelectronics, Catania(Italy) for making available the SiC Power Modules of the ACEPACK^{TM} DRIVE family and for sharing the Power Cycling testing data. This work was supported in part by the project REACTION "first and euRopEAn siC eigTh Inches pilOt liNe"-Horizon 2020 Program under Grant 783158, and in part by the project "Edge-AI" - HORIZON-KDT-JU-2021-2-RIA-Focus-Topic-1.

REFERENCES

[1] M. N. Yoder, "Wide bandgap semiconductor materials and devices," *IEEE Transactions on Electron Devices*, vol. 43, no. 10, pp. 1633–1636, 1996.

[2] P. G. Neudeck, R. S. Okojie, and L.-Y. Chen, "High-temperature electronics-a role for wide bandgap semiconductors?" *Proceedings of the IEEE*, vol. 90, no. 6, pp. 1065–1076, 2002.

[3] S. Hazra, A. De, L. Cheng, J. Palmour, M. Schupbach, B. A. Hull, S. Allen, and S. Bhattacharya, "High switching performance of 1700-v, 50-a sic power mosfet over si igbt/bimosfet for advanced power conversion applications," *IEEE transactions on Power Electronics*, vol. 31, no. 7, pp. 4742–4754, 2015.

[4] F. B. A. Sangwongwanich, "Monte carlo simulation with incremental damage for reliability assessment of power electronics," *IEEE Transactions on Power Electronic*, vol. 36, no. 7, pp. 7366–7371, 2021.

[5] C. Herold, J. Franke, R. Bhojani, A. Schleicher, and J. Lutz, "Requirements in power cycling for precise lifetime estimation," *Microelectronics Reliability*, vol. 58, pp. 82–89, Mar. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0026271415302821

[6] C. Durand, M. Klingler, D. Coutellier, and H. Naceur, "Power Cycling Reliability of Power Module: A Survey," *IEEE Transactions on Device and Materials Reliability*, vol. 16, no. 1, pp. 80–97, Mar. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7377071/

[7] STMicroelectronics, "Acepack driver website: https://www.st.com/en/power-modules-and-ipm/acepack-drive.html," 2022.

[8] J. Reimers, L. Dorn-Gomba, C. Mak, and A. Emadi, "Automotive Traction Inverters: Current Status and Future Trends," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3337–3350, Apr. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8636258/

[9] M. Calabretta, A. Sitta, S. M. Oliveri, and G. Sequenzia, "Silicon carbide multi-chip power module for traction inverter applications: thermal characterization and modeling," *IEEE Access*, pp. 1–1, 2021.

[10] T. Herboth, M. Guenther, A. Fix, and J. Wilde, "Failure mechanisms of sintered silver interconnections for power electronic applications," in *2013 IEEE 63rd Electronic Components and Technology Conference*, 2013, pp. 1621–1627.

[11] "ECPE Guideline AQG 324 - Qualification of Power Modules for Use in Power Electronics Converter Units in Motor Vehicles," ECPE European Center for Power Electronics, Nuremberg, DE, Datasheet, May 2021. [Online]. Available: https://www.ecpe.org/research/working-groups/automotive-aqg-324/

[12] R. Bayerer, T. Herrmann, T. Licht, J. Lutz, and M. Feller, "Model for Power Cycling lifetime of IGBT Modules - various factors influencing lifetime," in *5th International Conference on Integrated Power Electronics Systems*, Mar. 2008, pp. 1–6.

[13] T. Funaki and S. Fukunaga, "Difficulties in characterizing transient thermal resistance of sic mosfets," in *2016 22nd International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*. IEEE, 2016, pp. 141–146.

[14] U. Scheuermann and R. Schmidt, "Investigations on the vce (t)-method to determine the junction temperature by using the chip itself as sensor," in *Proc. PCIM Eur.*, 2009, pp. 802–807.

[15] S. Zhao, F. Blaabjerg, and H. Wang, "An overview of artificial intelligence applications for power electronics," *IEEE Transactions on Power Electronics*, vol. 36, no. 4, pp. 4633–4658, 2020.

[16] S. Rahimpour, H. Tarzamni, N. V. Kurdkandi, O. Husev, D. Vinnikov, and F. Tahami, "An overview of lifetime management of power electronic converters," *IEEE Access*, vol. 10, pp. 109 688–109 711, 2022.

[17] Z. Ni, X. Lyu, O. P. Yadav, B. N. Singh, S. Zheng, and D. Cao, "Overview of real-time lifetime prediction and extension for sic power converters," *IEEE Transactions on Power Electronics*, vol. 35, no. 8, pp. 7765–7794, 2020.

[18] H. Malorie, B. Pascal, A. Bruno, C. Guy, and R. Hubert, "Test bench and data analysis towards an on-line health monitoring for emerging power modules," in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 2018, pp. 1531–1536.

[19] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1314–1326, 2016.

[20] M. Heydarzadeh, S. Dusmez, M. Nourani, and B. Akin, "Bayesian remaining useful lifetime prediction of thermally aged power mosfets," in *2017 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2017, pp. 2718–2722.

[21] A. Alghassi, S. Perinpanayagam, and M. Samie, "Stochastic rul calculation enhanced with tdnn-based igbt failure modeling," *IEEE Transactions on Reliability*, vol. 65, no. 2, pp. 558–573, 2016.

[22] K. Le Son, M. Fouladirad, and A. Barros, "Remaining useful life estimation on the non-homogenous gamma with noise deterioration based on gibbs filtering: A case study," in *2012 IEEE Conference on Prognostics and Health Management*, 2012, pp. 1–6.

[23] H. Boudali and J. Dugan, "A continuous-time bayesian network reliability modeling, and analysis framework," *IEEE Transactions on Reliability*, vol. 55, no. 1, pp. 86–97, 2006.

[24] X. Qu, Q. Liu, H. Wang, and F. Blaabjerg, "System-level lifetime prediction for led lighting applications considering thermal coupling between led sources and drivers," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 6, no. 4, pp. 1860–1870, 2018.

[25] M. Baharani, M. Biglarbegian, B. Parkhideh, and H. Tabkhi, "Real-time deep learning at the edge for scalable reliability modeling of si-mosfet power electronics converters," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7375–7385, 2019.

[26] Y. Liu and X. Wang, "Deep & attention: A self-attention based neural network for remaining useful lifetime predictions," in *2021 7th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, 2021, pp. 98–105.

[27] G. S. Chadha, S. R. B. Shah, A. Schwung, and S. X. Ding, "Shared temporal attention transformer for remaining useful lifetime estimation," *IEEE Access*, vol. 10, pp. 74 244–74 258, 2022.

[28] Y. Zhang, P. Hutchinson, N. A. J. Lieven, and J. Nunez-Yanez, "Remaining useful life estimation using long short-term memory neural networks and deep fusion," *IEEE Access*, vol. 8, pp. 19 033–19 045, 2020.

[29] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[30] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011.

[31] M. E. Tipping, "The relevance vector machine," in *NIPS*, 1999.

[32] G. Sateesh Babu, P. Zhao, and X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *Database Systems for Advanced Applications*, S. B. Navathe, W. Wu, S. Shekhar, X. Du, X. S. Wang, and H. Xiong, Eds. Cham: Springer International Publishing, 2016, pp. 214–228.

[33] "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.

[34] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2017, pp. 88–95.

[35] Y. Cheng, J. Wu, H. Zhu, S. W. Or, and X. Shao, "Remaining useful life prognosis based on ensemble long short-term memory neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[36] Y. Qin, C. Yuen, Y. Shao, B. Qin, and X. Li, "Slow-varying dynamics-assisted temporal capsule network for machinery remaining useful life estimation," *IEEE Transactions on Cybernetics*, pp. 1–15, 2022.

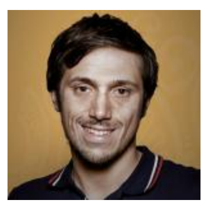
[37] M. Heydarzadeh, S. Dusmez, M. Nourani, and B. Akin, "Bayesian remaining useful lifetime prediction of thermally aged power mosfets," in *2017 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2017, pp. 2718–2722.

[38] S. Zhao, Y. Peng, F. Yang, E. Ugur, B. Akin, and H. Wang, "Health state estimation and remaining useful life prediction of power devices subject to noisy and aperiodic condition monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–16, 2021.

[39] S. Dusmez, H. Duran, and B. Akin, "Remaining useful lifetime estimation for thermally stressed power mosfets based on on-state resistance

variation," *IEEE Transactions on Industry Applications*, vol. 52, no. 3, pp. 2554–2563, 2016.
[40] S. Dusmez, M. Heydarzadeh, M. Nourani, and B. Akin, "A robust remaining useful lifetime estimation method for discrete power mosfets," in *2016 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2016, pp. 1–6.
[41] Y. Zheng, L. Wu, X. Li, and C. Yin, "A relevance vector machine-based approach for remaining useful life prediction of power mosfets," in *2014 Prognostics and System Health Management Conference (PHM-2014 Hunan)*, 2014, pp. 642–646.
[42] M. Ma and Z. Mao, "Deep recurrent convolutional neural network for remaining useful life prediction," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2019, pp. 1–4.
[43] H. Yamasaki, K. Miyazaki, Y. Lo, A. K. M. Mahfuzul Islam, K. Hata, T. Sakurai, and M. Takamiya, "Power device degradation estimation by machine learning of gate waveforms," in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2020, pp. 335–338.
[44] J. R. Regatti, H. You, H. Wang, J. Hall, A. Schnabel, B. Hu, J. Zhang, J. Wang, and A. Gupta, "A discussion of artificial intelligence applications in sic mosfet device operation," in *2021 IEEE International Electric Machines & Drives Conference (IEMDC)*, 2021, pp. 1–5.
[45] M. Nazar, A. Ibrahim, Z. Khatir, N. Degrenne, and Z. Al Masry, "Remaining Useful Lifetime estimation for Electronic Power modules using an analytical degradation model," in *Fifth European Conference of the PHM Society 2020 (The Prognostics and Health Management)*, vol. 5, no. 1, Turin (virtual conference), Italy, Jul. 2020, p. 10. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03152503
[46] M. Hologne-Carpentier, B. Allard, G. Clerc, and H. Razik, "Discussion on classification methods for lifetime evaluation of a lab-scale SiC MOSFET power module," in *ELECTRIMACS 2022*, Nancy, France, May 2022. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03683095
[47] Z. Ni, X. Lyu, O. P. Yadav, and D. Cao, "Review of sic mosfet based three-phase inverter lifetime prediction," in *2017 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2017, pp. 1007–1014.
[48] C. Olivares, R. Rahman, C. Stankus, J. Hampton, A. Zedwick, and M. Ahmed, "Predicting power electronics device reliability under extreme conditions with machine learning algorithms," 2021. [Online]. Available: https://arxiv.org/abs/2107.10292
[49] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," 2019. [Online]. Available: https://arxiv.org/abs/1905.10437
[50] C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, and A. Dubrawski, "N-hits: Neural hierarchical interpolation for time series forecasting," 2022. [Online]. Available: https://arxiv.org/abs/2201.12886
[51] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "Deepar: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207019301888
[52] J.-H. Böse, V. Flunkert, J. Gasthaus, T. Januschowski, D. Lange, D. Salinas, S. Schelter, M. Seeger, and Y. Wang, "Probabilistic demand forecasting at scale," *Proc. VLDB Endow.*, vol. 10, no. 12, p. 1694–1705, aug 2017. [Online]. Available: https://doi.org/10.14778/3137765.3137775
[53] C. Capistrán, C. Constandse, and M. Ramos-Francia, "Multi-horizon inflation forecasts using disaggregated data," *Economic Modelling*, vol. 27, no. 3, pp. 666–677, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0264999310000076
[54] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207021000637
[55] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *ArXiv*, vol. abs/1803.01271, 2018.
[56] D. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2016. [Online]. Available: http://arxiv.org/abs/1511.07289
[57] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016. [Online]. Available: https://arxiv.org/abs/1607.06450
[58] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," 2016. [Online]. Available: https://arxiv.org/abs/1612.08083
[59] Y. Gal and Z. Ghahramani, "A theoretically grounded application of dropout in recurrent neural networks," in *NIPS*, 2015.
[60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6000–6010.
[61] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, *Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting*. Red Hook, NY, USA: Curran Associates Inc., 2019.

Carmelo Pino is member of Artificial Intelligence for Modeling and Predictive Reliability Team in STMicroelectronics (ADG - R&D Power and Discretes) where he works as Advanced Research Senior Engineer. He worked as research assistant at the University of Catania, Italy, where he completed his Master's degree and Ph.D. studies. From 2014 and currently is member of the Pattern Recognition and Computer Vision Laboratory at the University of Catania and from 2020 to 2022 has been a research assistant at INAF (The National Institute for Astrophysics), Catania, Italy. His research interests include the areas of artificial intelligence applied to medical data processing, detection and segmentation in endoscopic video imaging systems, visual-knowledge ontology modelling, processing of radio-astronomical images, temporal series analysis, etc.



Alessandro Sitta received the M.S. degree in mechanical engineering and the Ph.D. degree in system, energetic, computer and telecommunication engineering from the University of Catania, Italy, in 2016 and 2021, respectively. Since 2017, he has been a Research Engineer with the Automotive and Discrete Group, Research and Development Department, STMicroelectronics, Catania, Italy. His research interests include power semiconductor device thermo-mechanical modeling, reliability estimation, and material and device experimental characterization.

Giulia Castagnolo received the M.S. degree in Computer Engineering at the University of Catania in 2020. Currently, she is a PhD Student at the University of Catania and a member of Artificial Intelligence for Modeling and Predictive Reliability Team in STMicroelectronics (ADG - R&D Power and Discretes) where she works as Advanced Research Junior Engineer. Her research interests lie in the area of machine and deep learning methods applied in different context such as video action recognition and sensors data analysis. Actually she is working on the application of continual learning approaches.

Angelo Messina was graduated with a M.Sc. degree in Electronics Engineer in 1997 from the University of Catania, after he got his Ph.D. in "Advanced Technologies for the Photonics and the Opto-electronics and Electromagnetic Modelling" and an M. Sc. degree in Physics from the University of Messina, where he worked four years in the former research team of the "Physics of Matter and Electronic Engineer" Department. He is appointed independent scientific evaluator for Regional, Italian and EU R&D&I funded programs, scientific conferences reviewer or chair and he is coauthor of tens of publications. He is appointed associated for research to CNR-IMM from 2017. Since 1999 he works in STMicroelectronics, being now Portfolio Manager of several EU R&D&I Funded Project, working now in the "Public Affairs for Italy" Department.

Salvatore Coffa was born in Carlentini, Italy, in 1962. He received the Basic and Ph.D. degrees in physics from the University of Catania, Catania, Italy, in 1985 and 1991, respectively. During more than 30 years of research activity Salvatore Coffa has achieved several important results in various fields and, more specifically, a large expertise in the field of technology transfer from basic research ideas to prototypes and then to products and applications. This expertise has been build up combining advanced research work (within or in cooperation with university, research labs, small/medium enterprises) and application to technologies and products within STMicroelectronics. He has innovated front-end and back-end technologies in the field of power devices introducing new Si power structures (using thrench, thin wafers, etc.) and power structures in semiconductors like SiC and GaN. SiC Power devices are now in full mass production within STMicroelectronics. He has authored more than 250 publications on international refereed journals and holds more than 50 patents.

Michele Calabretta received the M.S. degree in mechanical engineering and the Ph.D. degree in mathematics applied to engineering from the University of Catania, Catania, Italy, in 2005 and 2009, respectively. From 2005 to 2008, he was a Research and Development Engineer with Ferrari S.p.A, Maranello, Italy, and later a Simulation Team Leader, from 2008 to 2011, with Automobili Lamborghini, Sant'Agata Bolognese, Italy. He has been involved in design and multi-physics simulation (CAD/CAE) and material characterization. Since 2011, he has been with STMicroelectronics, Catania, where he is currently an Advanced Reliability & Physical Analysis Sr. Manager within the Automotive and Discrete Group (ADG), Research and Development Department. His research interests include semiconductor packaging and reliability aspects, lifetime modeling, and multi-physics simulations.

Fabio Scotti (Senior Member, IEEE) received the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003. He was an Assistant Professor at the Department of Information Technologies, Università degli Studi di Milano, Italy (2002-2015). He was an Associate Professor at the Department of Computer Science, Università degli Studi di Milano, Italy (2015-2020). He is a Full Professor at the Università degli Studi di Milano, Italy since 2020. Original results have been published in over 150 papers in international journals, proceedings of international conferences, books, book chapters, and patents. His current research interests include biometric systems, machine learning and computational intelligence, signal and image processing, theory and applications of neural networks, three-dimensional reconstruction, industrial applications, intelligent measurement systems, and high-level system design. He is an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Open Journal of Signal Processing. He is serving as Book Editor (Area Editor, section Less-constrained Biometrics) of the Encyclopedia of Cryptography, Security, and Privacy (3rd Edition), Springer. He has been an Associate Editor of the IEEE Transactions on Information Forensics and Security, Soft Computing (Springer) and a Guest Coeditor for the IEEE Transactions on Instrumentation and Measurement.

Angelo Genovese (S'12-M'15-SM'22) received the Ph.D. degree in computer Science from the Università degli Studi di Milano, Milan, Italy, in 2014. He has been an Associate Professor in Computer Science with the Università degli Studi di Milano since 2022. He has been a Visiting Researcher with the University of Toronto, Toronto, ON, Canada. Original results have been published in over 60 papers in international journals, proceedings of international conferences, books, and book chapters. His current research interests include signal and image processing, three-dimensional reconstruction, artificial intelligence for medical imaging, industrial and environmental monitoring systems, and design methodologies and algorithms for self-adapting systems. He is an IEEE Senior Member. Dr. Genovese is an Associate Editor of the Journal of Ambient Intelligence and Humanized Computing (Springer) and Array (Elsevier).

Vincenzo Piuri (Fellow, IEEE) received the Ph.D. degree in computer engineering from Politecnico di Milano, Italy, in 1989. He is a Full Professor of computer engineering at the University of Milan, Italy, since 2000. He has also been an Associate Professor at Politecnico di Milano, Italy, and a Visiting Professor at The University of Texas at Austin, USA, and a Visiting Researcher at the George Mason University, USA. His research interests include artificial intelligence, computational intelligence, intelligent systems, machine learning, pattern analysis and recognition, signal and image processing, biometrics, intelligent measurement systems, industrial applications, digital processing architectures, fault tolerance, dependability, and cloud computing infrastructures. Original results have been published in more than 400 articles in international journals, proceedings of international conferences, books, and book chapters. He is a Fellow of the IEEE, a Distinguished Scientist of ACM and a Senior Member of INNS. He is IEEE Region 8 Director (2023-24). He has been IEEE Vice President for Technical Activities (2015), IEEE Director, President of the IEEE Systems Council and the IEEE Computational Intelligence Society, Vice President for Education of the IEEE Biometrics Council, Vice President for Publications of the IEEE Instrumentation and Measurement Society and the IEEE Systems Council, and Vice President for Membership of the IEEE Computational Intelligence Society. He is an Associate Editor of the IEEE Transactions on Cloud Computing. He has been the Editor-in-Chief of the IEEE Systems Journal (2013-19) and an Associate Editor of IEEE Transactions on Computers, IEEE Transactions on Neural Networks, IEEE Transactions on Instrumentation and Measurement, and IEEE Access.

Concetto Spampinato is currently an associate professor at the University of Catania, Italy. He is also a courtesy faculty member at the Center for Research in Computer Vision University of Central Florida. In 2014, he created and currently leads the Pattern Recognition and Computer Vision Laboratory (PeRCeiVe Lab). His research interests lie in the area of machine learning and its application in multiple domains from medical image analysis to autonomous robot navigation. He is associate editor for Computer Vision and Image Understanding, IEEE Transactions of Multimedia and Machine Vision and Applications journals as well as area chair for multiple top-tier conferences (including CVPR 2022).

Francesco Rundo is a Senior Technical Staff Research Member at STMicroelectronics of Catania. He is a member of the Automotive R&D Power and Discretes Division of STMicroelectronics. He received his degree in Computer Science Engineering and Ph.D. in "Applied Mathematics for Technology" from the University of Catania. He was a member of the National Group of Physics-Mathematics of the National Institute of High Mathematics (InDAM) at the University of Bologna. He is currently team leader and project leader regarding to the development of Artificial Intelligence based solutions (hardware and software) for automotive, industrial and medical applications. Currently, he is working on the development of AI-based systems for Silicon and Silicon-Carbide device health-monitoring and modeling. He has co-authored more than 100 contributions in international journals, conference proceedings contributions, book chapters, posters, abstracts, lectures. He serves as Reviewer and Guest Editor (or Lead Guest Editor) of several special issues organized by such key-editors in the field of Computer Science such as Springer, Elsevier, Frontiers, MDPI. He serves as Associate Editor for IET Image Processing, IET Networks, Hindawi Applied Computational Intelligence and Soft Computing as well as Associate RT Editor for Frontiers in Computer Science and Topic Editor for MDPI Electronics. He is a member of the Computer Science Ph.D. Committee at the Department of Mathematics and Computer Science of the University of Catania as well as for the National Ph.D. of Artificial Intelligence. He is also co-inventor of several international patents as well as member of several international conference program committees, chair and program chair of such ICCV and ECCV special sessions. His main research interests include advanced bio- inspired models, advanced and perceptual deep learning, embedded systems for deep learning algorithms, advanced deep learning and mathematical modeling for automotive, industrial and healthcare fields.