Design and Performance Evaluation of an Ultralow-Power Smart IoT Device With Embedded TinyML for Asset Activity Monitoring

Marco Giordano, Nicolas Baumann, Michele Crabolu, Raphael Fischer, Giovanni Bellusci, and Michele Magno, Senior Member, IEEE

Abstract—This article proposes a proof-of-concept device to continuously assess the usage of handheld power tools and detect construction working tasks (e.g., different drilling works) along with potential misusages, e.g., drops, with an energy-efficient architecture design. The designed device is based on Bluetooth low energy (BLE) and NFC connectivity. BLE is used to exchange data with a gateway, whereas NFC has been chosen as an energy-efficient wake-up mechanism. A temperature and humidity sensor is embedded to monitor storage conditions and an accelerometer for tool usage monitoring. The ARM Cortex-M4 core embedded in the BLE module is exploited to process the information at the edge. A Tiny Machine Learning (TinyML) algorithm is proposed to process the data directly on board and achieve low latency and high energy efficiency. The TinyML algorithm has been developed embedded in the proposed device to detect four different usage classes (tool transportation, no-load, metal, and wood drilling). A dataset containing more than 280 min of three-axis accelerations during different activities has been acquired with the device attached to a construction rotary hammer drill and used to train and validate the algorithm. A neural architecture search has been performed to optimize the trade-off between accuracy and complexity, achieving an accuracy of 90.6% with a model size of roughly 30 kB. The experimental results showed an ultralow power consumption in sleep mode of 550 nA and a peak power consumption of 8 mA while running TinyML on the edge. This results in a balanced combination of edge processing capabilities and low power consumption, enabling to obtain a smart Internet of Things (IoT) device in the field with a long lifetime of up to four years in operation and 17 years in shelf mode with a standard 250-mAh coin battery. This work enables a long battery lifetime operation and 17 years in shelf mode with a standard 250-mAh coin battery. It is important not to disrupt the traditional workflow. Thus, SmartTag has to be position-agnostic to allow the users to position it wherever they feel is more convenient. Furthermore, many power tools are corded, hence requiring additional voltage conversion circuits to supply the SmartTag. Considering the technical limitations, lifetime of battery-operated devices poses the biggest challenge preventing the seamless use of IoT as a truly digitalization enabler. The recent wave of IoT [12], [13] and technological advances are enabling the design and development of a new generation of powerful, intelligent, low-energy, miniaturized, lightweight, and wireless sensor devices. Among other wireless technologies, Bluetooth Smart or BLE allows energy-efficient meter-distance communication, achieving a few milliwatt power budget [14]. The wireless capability combined with sensors and on-board signal processing allows the development of a new generation of IoT devices that are used for long-lasting monitoring of industrial machinery, tools, and other devices [5], [15]–[18]. The performance evaluation and prediction of tools’ failure and lifetime is becoming more and more an important problem to address due to the complex and dynamic structure of

I. INTRODUCTION

POWER tools are used in every construction industry in most of the working chain. Assessing power tools’ usage can have a big impact on industry sustainability. Such information can be used to optimize maintenance interventions, enable prolonged tool life cycle, and improve productivity and safety. In fact, a misuse or missed maintenance can interrupt the workflow, or in the worst cases affect the safety of users [1]. In both the cases, it has a huge impact on the productivity and indirectly on the overall sustainability [2]. One of the challenging tasks when designing a product is achieving long operating life, and to achieve such, correct handling and targeted interventions are mandatory to be performed [3]–[5]. The construction industry is increasing its awareness of the positive impact digitalization can have in maintaining a sustainable business [6]–[9]. However, beyond a conservative culture, many technical challenges and limitations still prevent the adoption [10]. Regarding the conservative industry culture, it is important not to disrupt the traditional workflow. Thus, SmartTag has to be position-agnostic to allow the users to position it wherever they feel is more convenient. Furthermore, many power tools are corded, hence requiring additional voltage conversion circuits to supply the SmartTag. Considering the technical limitations, lifetime of battery-operated devices poses the biggest challenge preventing the seamless use of IoT as a truly digitalization enabler. The recent wave of IoT [12], [13] and technological advances are enabling the design and development of a new generation of powerful, intelligent, low-energy, miniaturized, lightweight, and wireless sensor devices. Among other wireless technologies, Bluetooth Smart or BLE allows energy-efficient meter-distance communication, achieving a few milliwatt power budget [14]. The wireless capability combined with sensors and on-board signal processing allows the development of a new generation of IoT devices that are used for long-lasting monitoring of industrial machinery, tools, and other devices [5], [15]–[18]. The performance evaluation and prediction of tools’ failure and lifetime is becoming more and more an important problem to address due to the complex and dynamic structure of
their working operations [19]. Today, sensors are an accurate technology to extract useful information of the performance, condition, and use of machinery and tools [20], [21]. Especially microelectro-mechanical systems (MEMS) inertial sensors [22], [23], as electronic devices designed to measure vibration, acceleration, and orientation with gyroscopes are used in many monitoring applications [5], [17], [18].

Tracking the state of a mobile and relatively small device is of increasing importance, as tools become more complex and expensive. However, the existing sensors alone are not smart enough to extract useful information from the acquired data. Data gathered by current sensors have to be streamed out by a node to be processed further in the cloud. This approach increases the energy required by the communication and the latency of detection and additionally poses security and privacy risks. For this reason, today’s sensor systems mostly focus on monitoring medium- to large-sized stationary machinery or tools [24]. On the other side, designing a truly intelligent IoT node for such tools brings challenges on both the size and the life cycle, mainly due to limited batteries that supply the smart sensor node [25]. This operating lifetime limitation today prevents the widespread adoption of smart and miniaturized IoT sensor nodes for this kind of monitoring tool. In fact, short battery life impacts the usability, as a battery replacement in these circumstances is difficult, expensive, and a time-consuming activity. Such a monitoring application requires a thorough low-power system design to achieving a satisfactory lifetime. The design of those IoT sensor nodes requires careful selection of components to achieve low-power and heavy-duty cycles (i.e., high fraction of inactive state) given the small battery size requirements and the expected long life time of many months or years. Moreover, wireless connectivity plays a crucial role in this application, and a trade-off must be made between power consumption and bandwidth [14].

Such a smart IoT device could in fact be manufactured and stored for a long time before being applied to a tool and, once activated, an operational lifetime of only a few weeks or months could discourage the user to adopt such solutions. The operation itself could then be divided into two different scenarios: one in which the tool, and so the device, is stored on a shelf, and one in which the tool is operated. Moreover, the application scenario of such a device can experience an unpredictable variety of work operations (drilling, hammering, chiseling, grinding, transportation, etc.), accidental drops, and atmospheric circumstances, bringing it to dynamic and harsh operating conditions, as experienced in construction sites. Different quantities should thus be evaluated for each of the scenarios: temperature and humidity can be measured to monitor atmospheric conditions, while vibrations, and so accelerations, can be exploited for activity/operation/drop recognition.

This work builds on top of [26]. In particular, this article presents the design, implementation, and experimental evaluation of a long-lasting SmartTag for remote power tools monitoring with embedded sensors and intelligence at the edge. The main target is the monitoring of industry tools in terms of use and conditions, thus enabling device degradation and utilization accounting. The SmartTag is built around a BLE system on a chip that hosts an ARM Cortex-M4 core, which is particularly suited for onboard information processing and data analytics with inertial sensors. This node lends itself to perform both standard signal processing pipelines, exploiting the optimized CMSIS-DSP [27] libraries, and novel machine learning (ML) and deep learning tasks. In this work, a neural network to perform activity recognition and drilling material classification is proposed and evaluated. A dataset of more than 2700 s of drilling accelerations against different materials has been collected and used to train and validate the introduced model. The neural network has been ported to the device using the TensorFlow Lite for Microcontrollers (TFLM) [28] framework, with enhanced CMSIS-NN [29] kernels.

The main contributions of this article are summarized as follows:

1) hardware–software codesign of a novel smart IoT node capable of recording three-axis accelerations, temperature, and humidity with ML capabilities in an ultralow-power fashion;
2) field data collection of accelerations from the prototype attached to a construction power tool during drilling, no-load, and transportation tasks;
3) network architecture search (NAS) for an efficient and accurate tool activity recognition;
4) porting of the designed ML model at the edge with extreme size optimization of the kernels allowing to run on the battery-powered designed device;
5) analysis of feasibility and performance evaluation on real construction power tool data;
6) experimental evaluation of functionality, power consumption, and battery lifetime.

II. RELATED WORK

A similar approach regarding the application scenario was followed in [5], albeit the authors exploited a different radio technology (Zigbee versus BLE [14]), a different power source (thermoelectric generator (TEG) harvester versus Lithium battery), and they do not provide ambient temperature and humidity measurements.

A tool for device monitoring is presented in [4], but it has a different scope. It is intended to monitor large machinery, as CNCs and lathes, from which it harvests energy in the form of vibration. Such energy scavenging methodology is clearly not viable for handheld devices, due to both their magnitude and on-time of the tool.

CNC machines are also targeted in [30]. The authors assess the tool wear by analyzing forces applied to different CNC axes, as well as drive currents for the axis and the spindle, exploiting a decision tree, a different ML algorithm. Contrarily to SmartTag, the acquisition system is composed of laboratory signal acquisition units, and the sensors were integrated inside the tool and signals could not be acquired with an external device such as the SmartTag.

Zanelli et al. [16] presented a structural health monitoring node with similar capabilities to our design, but it must rely on solar energy harvesting to achieve a satisfactory battery lifetime. This could represent a problem when such energy is not present, as in warehouses or indoor, where the light conditions are very challenging.
A similar neural network approach has been proposed in [31], where the authors classify different types of rocks while drilling. However, the authors elaborate signals from five different physical quantities, while in this work we demonstrate that drilling material classification is possible with only acceleration. Moreover, as previously stated for [32], the sensors used must be integrated in the drill body, and the signals are collected by external equipment, as opposed to SmartTag where the acquisition, as well as the classification, is performed in a compact and battery-powered device attached to the drill body. Moreover, the model proposed in [31] uses only dense layers, while our proposed model takes advantage of the convolutional layers, achieving comparable accuracy with an extremely optimized model size, capable of running on resource-constrained devices.

The work in [33] tackles a similar problem. An angle grinder and a cordless screwdriver are analyzed, on which data are gathered from three-axis accelerometers, gyroscopes, and magnetometers and a series of algorithms are proposed to extract the activity performed. Our work focuses on the material rather than activity recognition, allowing for tool degradation accounting and achieves a similar accuracy with a deep learning model requiring only accelerometer data. This significantly reduces the system complexity, the signal processing overhead, and the power consumption.

In our work, the data gathered by the sensor node are further elaborated at the edge. A highly efficient signal processing pipeline is proposed, featuring the fast Fourier transformation (FFT) and a neural network. The FFT is applied on accelerometer data to detect tool usage; then, if the tool is currently operated, an efficient neural network can be run to classify the drilling material. Previous work demonstrated how accelerometers are a good fit for devices’ health assessment [34]–[36], and the aforementioned signal processing techniques, used isolated, are proven to extract useful information from accelerometer data. In particular, Maruthi and Hegde [22] detected mechanical bearing fault using the FFT, while in [37] and [38] a convolutional neural network (CNN) is exploited to detect bolt-nut alignment. In our work, both FFT and deep learning methods are leveraged to further optimize the energy efficiency of the proposed solution. In particular, the data are preprocessed with FFT so that the neural network, which is an energy-intensive task, is run only when needed.

A good reasoning about wake-up radios can be found in [39] and [40]. The concept has been adopted also in SmartTag as the wake-up is performed via NFC. Wake-up radios are a valid solution for use-case scenarios like the SmartTag: a short-range wireless activation procedure has been chosen to overcome the absence of physical buttons on the final products. Not exposing any mechanical parts improves the reliability, and indirectly the sustainability, of the sensor node, especially in the dusty and harsh environments the SmartTag is thought to operate.

Motion monitoring has also been used to assess workers’ productivity. In particular, [41] reviews productivity assessment methodologies for the construction field, in an industry 4.0 scenario. With the recent advancement of deep learning, a push in the direction of deep activity recognition has been occurring. Most famously in the field of human activity recognition (HAR) [42]–[45], where gyroscope and accelerometer data of either smartphones or smartwatches are used to classify human activity. The current state-of-the-art deep learning method to do so incorporates large and heavy neural networks [45] to process the wearable devices’ data. In the most complex scenarios, even long short-term memory (LSTM) neural networks [45] are being used. Regarding HAR, the application scenario of the SmartTag to classify the drilling tools activity for degradation and utilization accounting is fairly similar. However, the key difference lies in the computational capabilities of the device the neural network is deployed to. While wearables such as smartphones might be capable of inferring a complex recurrent model, the underlying microcontroller unit (MCU) of SmartTag is not capable of inferring a large and recurrent network. Yet this work shows that in cases of simpler activity recognition tasks such as tool monitoring, a feed-forward nonrecurrent CNN can achieve good results while being feasible on an extremely power and computationally constrained device, allowing for ML on the edge benefits.

The rise of tiny machine learning (TinyML), albeit recent, has seen extensive coverage in literature. Neural networks are usually trained on dedicated servers, while inference can be computed on the edge, even on very resource-constrained devices [46]. Recent developments are introducing continuous learning techniques [47], often using custom-designed architectures [48]. TinyML systems are gaining more and more traction, and to support this expansion benchmarking [49], [50] tools have been developed to assess ML performances at the edge.

Sustainability and efficiency were founding principles when designing the SmartTag and key topics in the work of [51]. The authors review several article in the IoT field, with special attention to Smart Cities. An important question regarding the sustainability of IoT nodes themself is risen and has been an important point of reflection while designing the SmartTag.

Concluding, this article presents a novel solution to the problem of asset tracking and usage analysis. An efficient neural network, which uses only accelerometer data, has been introduced for drilling material classification (thus enabling tool degradation and utilization analysis) and has been successfully ported on a resource-constrained device to perform inference on the edge, optimizing accuracy against model size and inference energy. Moreover, an ultralow-power sensor node has been introduced, with an expected battery lifetime of more than four years in operation and 17 years on the shelf powered by a small coin cell battery.

### III. System Architecture

The proposed sensor node was designed to achieve ultralow-power consumption and energy efficiency. A schematic view of the proposed solution can be seen in Fig. 1, while our reference implementation can be seen in Fig. 2. The high-level wireless sensor network architecture is shown in Fig. 3 and a real-world application of the SmartTag on a Combi Hummer drill can be seen in Fig. 4.
A. System Components

The SmartTag architecture was designed to be flexible toward different workloads and allows to easily exchange different system on a chip (SoC) of the same family. The Nordic NRF52 SoC family was selected as the best trade-off in terms of computing power, BLE capabilities, and power management features. Two physical prototypes of the SmartTag were designed, one hosting the Nordic NRF52810 SoC and another with NRF52832. The two SoC, being pin-to-pin-compatible, can be swapped without adding any complexity from the schematic and PCB perspective. Moreover, the Nordic SDK allows writing code seamlessly for both SoC. In particular, the Nordic NRF52810 is preferred when power consumption is the utmost priority, while NRF52832 is the choice when demanding edge ML tasks are desired. This distinction arises from the different power consumption and different memory configurations of the chips: NRF52810 has only 192 kB of flash and 24 kB of RAM, while NRF52832 allows for up to 512 kB of flash and 64 kB of RAM. Both are based on an ARM Cortex-M4, have BLE 5.2, and provide good computing power, with a core clock up to 64 MHz. NRF52810’s operating current can be as low as 35 μA/MHz when running code from Flash, against 60 μA/MHz of the NRF52832. In the System OFF mode, the lowest power mode is as low as 0.5 μA with full RAM retention for both SoCs. These metrics are summarized in Tables IV and V, differentiating between sleep and active mode. Both SoCs boost up to −96-dBm sensitivity for BLE reception, with about 5 mA of current consumption in the RX mode and a configurable TX gain from −20 to +4 dBm.

Two I2C-enabled sensors were then chosen, one for temperature and humidity sensing and one for acceleration measurement, which is the most important sensor for evaluating the use and the condition of tools and machinery. The former is SHTC3 from Sensirion, which features a typical relative humidity accuracy of ±2% and temperature accuracy of ±0.2 °C. It is compatible with the low-voltage rail used in our node, fixed to 1.8 V to optimize power consumption and has an operating current during measurement of around 500 μA. To further reduce the power consumption when not in use, the sensor has been power gated. The accelerometer of selection is IIS2DLPC from ST Microelectronics. It features four different scales of acceleration measurement, from 2G to 16G, with a sensitivity of ±3% mG/digit on all the scales and a noise density as low as 90 μG/√Hz in the 2G range. The power figures are impressive as well, with a current consumption of a few μA for measurements acquired in low-power mode and of 50 nA in the lowest power mode. Since no physical buttons are thought to be present on the final design, to wake up the MCU from deep sleep, going from shelf mode to operating mode, we exploit an NFC chip. NT2H1611F0DTLH from NXP toggles a pin when in the presence of an NFC RF field, from which it harvests the needed energy and exchanges information.
B. Firmware

The first entry point in memory is the Softdevice: it is the proprietary wireless stack of the Nordic semiconductor, provided as a binary package. The MCU boots from the Softdevice, proceeding then to execute the bootloader. The bootloader is based on the Nordic Secure Device Firmware Update (DFU) Bootloader and will be further explained in Section III-C. To summarize, the bootloader checks whether a new app image should be flashed or whether it should start the already present application. Supposing this is the case, the focus will now be put on the application firmware. An overview of the firmware’s flowchart can be found in Fig. 5. Two different operational modes can be distinguished: a shelf mode and an operational mode. In the shelf mode, the device is in a power-off state. This mode will be active from production until the activation of the SmartTag is performed by a potential user wirelessly via NFC. In this mode, the temperature and humidity sensor is power gated, the accelerometer is put in the power-down mode and the MCU is put in the system OFF mode. From this state, it is sufficient to put the device in an NFC field so that the NFC module will assert a pin of the MCU low, waking up the SmartTag. Waking up from System OFF, the lowest power mode of the MCU of choice mandates a system reset. After the reset, the SmartTag will then no longer be powered off and the system will cycle endlessly into the operational mode of the firmware. The operational mode takes care of achieving the lowest power consumption, reading sensors, detecting potentially harmful events, and sending the data to a gateway on request. The operational mode starts with the node advertising itself over Bluetooth, with an advertising period of 7 s to keep the energy footprint as low as possible. Once connected to a gateway, it can either enter the bootloader mode, to update the firmware, or issue sensor readings. For temperature and humidity, single measurements can be issued, while for the accelerometer, either a single measure or a set of measurements at a given sample rate can be acquired. Moreover, there are two asynchronous interrupts, “Drop detection” and “Shake detection.” When either of these occur, the SmartTag is woken up and the current time is registered in a buffer, which can then be sent over BLE to a gateway.

IV. EDGE ML

Edge computing has a series of benefits [52], as lowering latency, which enables applications with strict time requirements. Transmitting only the important features extracted from the data means drastically lowering the bandwidth requirements and on-air collisions. Moreover, wireless communication is usually the most power-consuming activity of a sensor node and limiting the data reflects directly on the battery life
of the node. Clear benefits can also be found in data security and privacy: as only processed data are transmitted, the raw data never leave the sensor node, eradicating the susceptibility to eavesdropping.

Among the many use-cases where TinyML at the edge can bring value, within the scope of this work we defined a narrow, yet very relevant, use-case to experiment with the capabilities of the proposed SmartTag. Specifically, a four-class ML algorithm was designed and ported to detect/distinguish the following working conditions.

1) Transport: The tool is being transported in either a car or hand-carried.
2) No-Load: The tool is running with no load applied.
3) Metal Drilling: Tool used to drill metal.
4) Wood Drilling: Tool used to drill wood.

Altogether, the above classes allow for a more finely grained resolution with regards to how the tool is being used. The TinyML algorithm running at the edge serves as a highly efficient data compression, thus saving battery and wireless bandwidth. Combined with the temperature and humidity measurements, the edge ML estimations can in turn be further processed in the cloud with more advanced ML algorithms obtaining more insights to optimize maintenance, productivity, and overall sustainability of industries like the construction sector.

A. Data Collection and Analysis

An infield accelerometer drilling activity dataset has been acquired to train and validate the ML algorithm. The following type of data was obtained with the SmartTag attached to a professional drilling tool (Hilti TE 30-A36).

1) Transport: The SmartTag was attached to the tool and transported in its carrying box in either the hood of a car or hand-carried. The car was driven in gravel, urban, and motorway paths.
2) No-Load: The tool was running with no load applied (not touching any material) with a 10 mm × 170 mm drill bit. To have a more representative dataset, both the drilling and hammering modes have been included in the no-load data acquisition.
3) Metal: Tool was used to drill a hole into steel material with a 10 mm × 90 mm drill bit. The tool was set to drilling mode since no hammering function is required for this type of material.
4) Wood: Tool was used to drill a hole into wooden material with a 10 mm × 90 mm drill bit. The tool was set to the drilling mode since no hammering function is required for this type of material.

The accelerometer signal was collected from the SmartTag with a sampling frequency of 800 Hz. A total of 45 min, equivalent to 3.04 GB, of raw accelerometer data has been collected.

Fig. 6 depicts the FFT of the superimposed x-, y-, and z-axes of the collected drilling activity accelerometer data, always considering a 448 samples window. The full line represents the mean, while the faded area represents the standard deviation of the FFT dataset. It suggests that the

![Fig. 6. Data analysis of infield drilling experiments in the investigated usage scenarios. The magnitude of the FFT applied on recorded data is plotted against frequency. Full line depicts the mean, and faded area depicts the standard deviation of the superimposed x, y, and z accelerometer axes.](image)

Transport class can be easily distinguished from the other classes with the simple FFT threshold. On the other hand, it is immediately noticeable that the variance of Wood, Metal, and No-load signals is nonnegligible. If one were to overlay the aforementioned three classes, the signals would heavily overlap each other. Thus, it can be argued that a simple thresholding of the FFT frequencies will not suffice to classify between Wood, Metal, and No-load. Furthermore, it can be seen that both Wood and Metal signals display large variance at frequencies where the mean seems distinguishable from each other. On the other hand, No-load shows a clear peak at ~225 Hz with low variance. Thus, one can assume No-load to be simpler (yet still non trivial) to distinguish from Wood and Metal.

To tackle this complex classification task, a neural network is being used. From this analysis, one can derive a system that applies simple FFT thresholding to binary classify between Transport and Other. If the Other class applies, a neural network will perform the challenging task of classifying between the three problematic classes of Wood, Metal, and No-load. This makes sense from a computational cost perspective, as the simple FFT threshold is computationally inexpensive (compared with the neural network inference) and the system is expected to be in a nonworking state (i.e., no drilling activity) most of the time. As soon as the system enters a drilling state (which occurs relatively seldom), the more computationally expensive neural network operation will be performed as in Fig. 5.

B. Neural Network Architecture

In this section, the neural network architecture is introduced. Its task is to classify a sliding window of accelerometer data into the previously mentioned three problematic classes of Wood, Metal, and No-load. The sliding window and thus the model’s input has a shape of (448, 3), corresponding to a vector of length 448 with three channels, namely, the x-, y-, and z-axes of the raw accelerometer data.
It is important to remember that the neural network will have to run on an extremely computationally constrained device, thus making the model size of central importance. To maximize the neural network performance while fitting it on the MCU is a nontrivial task that demands careful architecture tuning and design.

In Fig. 7, the proposed neural network architecture is depicted. It consists of two sequential 1-D convolutions, followed by a fully connected layer that feeds into the final classification head. The deep convolutions allow the model to extract features across the channels in a very parameter-efficient way and have proven throughout the training process to be a very effective architecture choice. The convolution filter depths, kernel size, and the number of necessary fully connected neurons have been evaluated by a hyperparameter sweep using the Weights and Biases [53] tool. From this, one can assess the mentioned parameters in terms of validation loss and number of parameter minimization. Thus, resulting in the architecture of Fig. 7 with a total model size of 93.2 kB.

The derived neural network architecture has been evaluated on a held out balanced test set of 150 accelerometer data windows of 448 samples each. Fig. 8 shows the classification performance metrics of the model in the three classes of interest. On the left side of Fig. 8, the precision and recall of each class are evaluated in a One vs All fashion while overlaying the F1-score isobars. On the right side of Fig. 8, the One vs All receiver operating characteristic (ROC) plot is given with the dashed diagonal indicating the no-skill threshold. It is important to mention that while the test set is balanced in each class, the One vs All evaluation is unbalanced by a ratio of 1:2. Thus, precision and recall are the relevant metrics to focus on, while ROC is depicted for completeness. From model evaluation, one concludes the metrics depicted in Table I.

Finally, the confusion matrix of the test set evaluation is shown in Fig. 9 to demonstrate an unbiased and complete model evaluation. By viewing Table I and Fig. 9, it can be confirmed that from the three classes of interest, Wood and Metal seem to be the most difficult to distinguish. Thus affirming the assumption of Section IV-A that Wood and Metal are harder to distinguish between each other than No-load, based on the FFT data analysis. To conclude, the proposed architecture shows to be very well-suited for the classification task of drilling activities in Wood, Metal, or No-load. It yields an accuracy of 93.3% on the balanced test set, allowing for a more sophisticated analysis of the drilling tool utilization and degradation by incorporating the neural network on the SmartTag.
C. ML MCU Deployment

To deploy the actual neural network on the MCU, the TFLM framework [54], a resource-optimized C porting of Tensorflow, has been integrated into the application firmware running on the SmartTag. The TFLM core runtime takes as little as 80.8 kB on the Cortex-M4 MCU memory (op included) and does not require any OS support. Thus, it can simply be integrated as a standalone library in the SmartTag firmware. Hence, integer quantized neural networks following the TFLM supported operations can now be used to perform inference on the SmartTag, as long as they adhere to the memory size limitations of the MCU. In particular, only Conv2D, FullyConnected, Relu, Reshape, Softmax, Quantize, Dequantize, and Add operations were loaded at runtime, to optimize the memory footprint. As reported in Table II, selecting only the used OPs leads to a 300% improvement on nonvolatile memory footprint, as well as an improvement in RAM usage. Moreover, the kernels offered by TFLM were enhanced with the ones offered by the CMSIS-NN, offering a faster computing time up to 4X over the supported operations [29].

1) Quantization: post training quantization was applied to the aforementioned full precision neural network after training, allowing to convert the 32-bit floating-point weights to 8-bit integers. This has been realized with the TensorFlow Lite converter. The primary reason of this step is to further decrease the model size, which in this case shrinks from the full precision neural network of 93.2 kB down to the quantized size of 30.5 kB, as summarized in Table III. The quantized model achieves a model size decrease by a factor of ∼4, with a rather small tradeoff in accuracy. As reported in Fig. 10, the confusion matrix of the fully int8 quantized neural network shows that the model performance has only been degraded slightly, as it still yields an accuracy of 90.6% on the same test set. A comparison between int8 quantized and float point model can be seen in Table III.

V. POWER CONSUMPTION

The SmartTag has been tested both in the shelf mode and operational mode to assess both power consumption and correct operation. Both FFT and neural network data processing approaches are evaluated in this section.

Table IV reports the current consumption of both the implementations in shelf mode, where all the components are in a power-off state and the node waits to be awoken by NFC. The MCU accounts for the majority of power consumption, while the humidity and temperature sensor has a power consumption in low-power mode comparable with the MCU and has thus been power gated. The accelerometer’s power footprint is one order of magnitude lower than the SoC and has been thus kept powered to reduce system complexity. In Table V, a breakdown of the current consumption in the operating mode for the main hardware components is presented. Again, the SoC accounts for the majority of power consumption, especially in the neural network case.

In Fig. 11, the power profiling of the NRF52810-based sensor node is plotted while executing the typical workload of sensor acquisition, FFT computation, and BLE advertising. In total, 1024 samples are collected from the accelerometer at a sampling rate of 800 Hz, the FFT is run shortly after on the node, and then the BLE advertises. The raw data are reported in shaded blue, while a moving average is reported in orange.

A breakdown of latency and energy for both FFT computation and per inference of the NRF52832-based implementation of the SmartTag can be seen in Table VI. From the numbers, it can be evinced that the FFT is almost 500× faster and less energy hungry than the neural network. Therefore, it is very convenient to use the FFT as a decision mechanism for neural network inference. Moreover, the floating point model has been tested on hardware as well. Data report a speed-up
of 2.8x in favor of the quantized version, at the expense of a decrease in accuracy of only 2.7%.

In Fig. 12, we report the computed expected lifetime of the sensor node against different battery sizes and the number of daily wake-ups. The solid bars refer to the NRF52810-based implementation without any ML task being run, while the hatched part of the plot reports the expected battery lifetime of the NRF52832-based implementation running the neural network described in this work on newly recorded data. Both the NRF52810 and NRF52832 versions achieve more than 17 years in the shelf mode. In operating mode, the difference between running the ML algorithm or not becomes noticeable when it is strictly necessary. To be able to further fine-grain the degradation and utility accounting of the investigated power tool, it might be helpful to differentiate between more classes (e.g., concrete drilling, chiseling). Within this work, to obtain a proof of concept, the analysis was limited to the four classes described in Fig. 6 and the data collected attaching the SmartTag always on the same location of the tool. Thus, future research direction can lie on addressing the generalization capability of the ML algorithm to different attaching positions of the SmartTag on the power tool and investigate which data and what working classes are necessary to optimally estimate degradation and utility analysis of the tools.

VI. CONCLUSION

In this work, an ultralow-power IoT wireless sensor node has been designed, implemented, and evaluated to enable condition and usage monitoring of machinery and mobile power tools in construction fields. The device was tested by collecting real-world data when attached to a professional construction power tool during drilling operations. This article shows that it is possible to obtain a smart node that reaches more than four years of lifetime on a small coin battery of 250 mAh while recording accelerations, ambient storing conditions, and performing inference at the edge via neural network models. The ported TinyML neural network shows that power tool activities, such as different drilling operations, can be correctly classified with a small wireless node attached to the tool with an accuracy of 90.6%. The results of this work enable to obtain more fine-grained utility monitoring of power tools, processing information directly on the edge and preserving long battery life. Such results are possible, thanks to an aggressive power management that brought the node to consume only 3 μW in the sleep mode, and wake up the system only when it is strictly necessary. To be able to further fine-grain the degradation and utility accounting of the investigated power tool, it might be helpful to differentiate between more classes (e.g., concrete drilling, chiseling). Within this work, to obtain a proof of concept, the analysis was limited to the four classes described in Fig. 6 and the data collected attaching the SmartTag always on the same location of the tool. Thus, future research direction can lie on addressing the generalization capability of the ML algorithm to different attaching positions of the SmartTag on the power tool and investigate which data and what working classes are necessary to optimally estimate degradation and utility analysis of the tools.

REFERENCES


[34] Marco Giordano received the B.Sc. degree in electrical engineering from the Politecnico di Torino, Turin, Italy, and the M.Sc. degree in electrical engineering and information technology from ETH Zürich, Zürich, Switzerland, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the Center for Project-Based Learning, Department of Information Technology and Electrical Engineering (D-ITET). His research interests include ultralow-power electronics, tiny machine learning (TinyML), and biomedical devices.
Nicolas Baumann received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from ETH Zürich, Zürich, Switzerland, in 2019 and 2021, respectively. He is currently a Scientific Assistant with the Center for Project-Based Learning, Department of Information Technology and Electrical Engineering (D-ITET), ETH Zürich. His research interests include computer vision and machine learning for autonomous flying and driving on resource-constrained devices.

Michele Crabolu received the Ph.D. degree in bioengineering and robotics from the University of Genoa, Genoa, Italy, in 2017, and the Post-Graduated Diploma degree in machine learning and artificial intelligence from Columbia University, New York, NY, USA, in 2021. He did more than ten clinical validation studies on wearable technologies. He was with the Research Department, Xsens Technologies, Enschede, The Netherlands, as a Lead Engineer, leading research topics about sensor fusion and motion capture. At the time of writing, he is a Technical Project Manager with the Corporate Research and Technology, Hilti Corporation, Schaan, Liechtenstein. He is currently an expert in biomedical embedded wearable technologies, sensor fusion, and TinyML algorithms.

Giovanni Bellusci received the Ph.D. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 2011, and the Executive M.B.A. degree (Hons.) from the London Business School, London, U.K., in 2020. He was the Director of Technology and the CTO at Xsens Technologies, Enschede, The Netherlands, a leading innovator in motion capture and position tracking solutions for applications in industrial, robotics, wearables, and health. He is currently the Head of robotics and mechatronics at Hilti Corporation, Schaan, Liechtenstein, a global supplier of tools and services for the construction industry. He has broad research and development industry experience in diverse technology domains, including the Internet of Things, machine learning, robotics, sensing technologies, and sensor fusion algorithms.

Raphael Fischer received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from ETH Zürich, Zürich, Switzerland, in 2018 and 2020, respectively. He is an Embedded Hard-/Firmware Developer.

Michele Magno (Senior Member, IEEE) received the master’s and Ph.D. degrees in electronic engineering from the University of Bologna, Bologna, Italy, in 2004 and 2010, respectively. He is currently a Senior Scientist and a Lecturer with the Department of Information Technology and Electrical Engineering (D-ITET), ETH Zürich, Zürich, Switzerland, where he has been leading the D-ITET Center for Project-Based Learning since 2020. He is also a Visiting Full Professor with Mid University Sweden, Sundsvall, Sweden. He has authored more than 200 articles in international journals and conferences. Some of his publications were awarded best papers awards in IEEE conferences. His current research interests include smart sensing, low-power machine learning, wireless sensor networks, wearable devices, energy harvesting, low-power management techniques, and extension of the lifetime of battery-operated devices.