TinyML: Applied AI for Development

ABSTRACT

Artificial intelligence (AI) will likely be an instrumental part of progress towards the United Nations' Sustainable Development Goals (SDGs). However, its adoption and impact are limited by the immense power consumption, strong connectivity requirements and high costs of cloud-based deployments. TinyML is a new technology that allows machine learning (ML) models to run on low-cost, low-power microcontrollers, circumventing many of these issues. We believe that TinyML has a significant role to play in achieving the SDGs and facilitating scientific research in areas such as environmental monitoring, physics of complex systems and energy management. To broaden access and participation and increase the impact of this new technology, we present an initiative that is creating and supporting a global network of academic institutions working on TinyML in developing countries. We suggest the development of additional open educational resources, South–South academic collaboration and pilot projects of at-scale TinyML solutions aimed at addressing the SDGs.

Authors: Marco Zennaro, ICTP/UNESCO; Brian Plancher, Harvard University; Vijay Janapa Reddi, Harvard University

Challenges with Machine Learning in Developing Countries

Machine learning has a huge potential to tackle societal issues in diverse fields that include agriculture, conservation and healthcare. A recent study [1] highlights the influence of AI on all aspects of sustainable development, in particular on the 17 Sustainable Development Goals (SDGs) and 169 targets internationally defined in the 2030 Agenda for Sustainable Development. The study shows that AI can act as an enabler for 134 targets through technological improvements, but it also highlights the challenges of AI on some of the targets. When considering AI and societal outcomes, the study shows that AI can benefit 67 targets, but it also warns about the issues related to the implementation of AI in countries with different cultural values and wealth.

Most of today's applications integrating Internet of Things (IoT) and Machine Learning (ML) rely on a cloud-centric architecture, which means that IoT nodes are required to send collected data to the cloud for ML inference. This necessitates strong connectivity and massive computational resources which are not available in most developing countries or in remote regions of the world. These leads to five main challenges preventing ML from addressing the SDGs:

1. **Connectivity:** Applications such as outdoor object-detection using high-definition video require more bandwidth than is available in many developing countries.
2. **Energy:** When a device is required to operate in a real-time mode by continuously transmitting data to the cloud it can consume large amounts of energy as wireless communications are the main energy consumer component of an embedded system.

3. **Privacy:** Applications that send data from the point of collection to the cloud may leak private information as data must be transmitted over the internet.
4. **Latency:** Some real-time IoT applications may have very strict requirements in terms of latency, therefore excluding the option to communicate with the cloud over a potentially unreliable communication channel.
5. **Real-World Applicability**: There is a gap in real-world applications of AI systems. AI applications are currently biased towards SDG issues that are more relevant to those nations where most AI researchers live and work. There are only a handful of examples where AI technologies are applied to SDG-related issues in nations without strong AI research. An analysis of highly influential papers in AI finds that they not only favor the needs of research communities and large firms over broader social needs, but also that they take this for granted. The values of performance, efficiency, and novelty are studied in ways that disfavor societal needs, usually without discussion or acknowledgment [2].

What is TinyML

TinyML is a subfield of ML focused on developing models that can be executed on small, real-time, low-power, and low-cost embedded devices [3]. The TinyML process flow is like the classical ML one, except that inference takes place on embedded devices. The TinyML process starts with collecting data from IoT devices, then training the collected dataset in the cloud to extract knowledge patterns, these patterns are then packaged into a TinyML model that considers the target microprocessor's limited resources such as memory and processing power. The resulting model is then deployed on embedded devices where it is used to evaluate new sensor data in real-time without communicating with the cloud. Typically, power requirements are in the mW range and below which enables a variety of use-cases targeting battery operated devices.

TinyML and SDGs

TinyML can overcome the challenges presented above by enabling data to be processed locally on edge devices. For example, for the previously mentioned outdoor object-detection using high-definition video example, data could be processed at the edge, and communication with the cloud would only occur when an object was found. This would save a tremendous amount of energy, require low connectivity, and ensure privacy as no raw information was sent over the internet. If our application also needed to execute a task once an object was found it could do this with low latency. Finally, given the low-cost of TinyML equipment we believe that it can enable AI research in the developing world, fostering local solutions with a small carbon footprint.

Some existing examples of TinyML applications that address SDGs include:

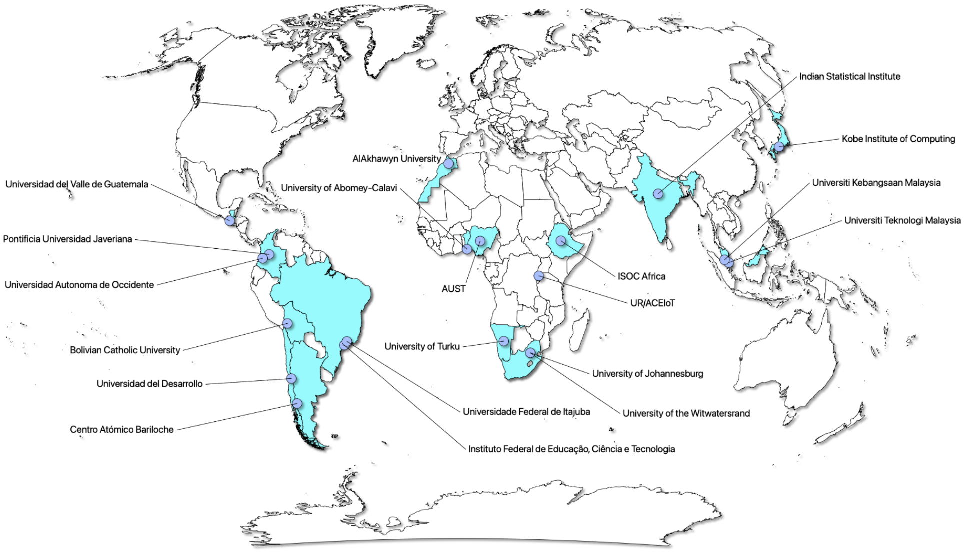
- A prototype rapid water-bone cholera detector kit pluggable into existing taps has been developed at ACEIoT in Rwanda to lower the cost of mass deployment [4]. The prediction of water-bone cholera is done by monitoring water's physicochemical patterns.
- A solution to use TinyML for wildlife conservation has been proposed. The solution uses camera traps and applies TinyML to enable inference at the edge to help track wildlife movement and aid in wildlife conservation [5]. This is important especially for conserving wildlife at risk across Africa.

- A related solution, Smart Wildlife Tracker has also been developed [6]. The trackers are attached to elephant collars. The collars capture real-time elephant movements using a GPS and capture surrounding images so that TinyML can be used to predict events around each animal such as presence of human predators. Audio models are also applied to determine the mood of the elephant and an accelerometer is used to understand the physical behavior and movement of the elephant.
- An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks has been developed to determine the quality of tea [7]. High quality tea can be sold at a much higher price thus benefiting the local community.
- The Ribbit Network was developed to provide accurate readings of carbon dioxide gas through a crowdsourced network of open-source, low-cost, smart sensors [8]. This kind of high quality data will help scientists better understand and predict the impacts of climate change.

The TinyML4D network

To facilitate the adoption of TinyML for SDGs, with support from the STI Unit of UNESCO's International Centre for Theoretical Physics (ICTP) and from Harvard University, in 2021 we built a network of academic institutions, based in Developing Countries, interested in expanding access to Applied Machine Learning by establishing best practices in education [9]. Our aim is to ultimately develop a community of researchers and practitioners focused on both improving access to TinyML education and enabling innovative solutions for the unique challenges faced by Developing Countries [10]. Member institutions are involved in research activities in IoT or AI. They have lab facilities where the hands-on sessions will be held when in-person training courses will be allowed.

A map of the current 20 network members is shown below.



Network activities included:

- Providing 10 Arduino TinyML Kits to each University in the network. The goal has been to create a critical mass of researchers/practitioners in each University. This distribution of hardware was critical as while TinyML hardware is low-cost it is not no-cost and as such can be a barrier to entry for many researchers in Developing Countries.
- Organizing online workshops and seminars on TinyML with lab sessions hosted by and for network members. In 2021 we organized four such activities, ranging from one-hour seminars to a one week-long workshop [11].
- An online forum to consolidate the TinyML community of researchers, educators, and practitioners. We have now over 300 members that share their knowledge on TinyML using Discord.
- Exchange of student projects, lesson plans, real-world deployments, and outreach materials.

The Arduino TinyML Learning Kit that has been shipped to network members.

Students from network member universities are well positioned to excel in the HarvardX Professional Certificate in Tiny Machine Learning [12], a free, four-course massive open online course (MOOC) specialization that dives deeper into the world of TinyML, as well as Coursera's Introduction to Embedded Machine Learning [13].

Policy recommendations

1. Given the potential impact of TinyML on SDGs, universities should cover this new technology in their courses on AI and machine learning and modular open educational resources should be created to enable easier course development.
2. The network of academic institutions in developing countries working on TinyML should be strengthened. More universities can be added and provided with the low-cost hardware needed to begin engaging with this growing field. This would help foster additional South–South academic collaboration.

3. A series of pilot projects on the use of TinyML to access SDGs could be developed to showcase the impact on sustainability of deploying this kind of solution at scale.

References:

[1] Vinuesa, Ricardo, et al (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature communications 11.1: 1-10.
[2] Birhane, Abeba, et al (2021). The values encoded in machine learning research." arXiv preprint arXiv:2106.15590.
[3] Janapa Reddi, V., et. al. (2022) Widening access to applied machine learning with TinyML. *Harvard Data Science Review*, 4.1.
[4] Ooko, Samson Otieno, et al (2021) TinyML in Africa: Opportunities and Challenges. 2021 IEEE Globecom Workshops. Institute of Electrical and Electronics Engineers.
[5] Sheth, Dhruv (2020) EleTect - TinyML and IoT Based Smart Wildlife Tracker. See: https://www.hackster.io/dhruvsheth_/eletect-tinyml-and-iot-based-smart-wildlife-tracker-c03e5a
[6] Verma, Swapnil and Jain, Mausam (2020). Elephant AI. See : https://www.hackster.io/dual_boot/elephant-ai-ba71e9?auth_token=014a2ccd55371e063e43f22973b4c58a
[7] Kimutai, Gibson, et al (2020) An optimum tea fermentation detection model based on deep convolutional neural networks. *Data* 5.2: 44.
[8] Rao, Ravi (2021) TinyML unlocks new possibilities for sustainable development technologies. See: https://www.wevolver.com/article/tinyml-unlocks-new-possibilities-for-sustainable-development-technologies/
[9] TinyMLedu: the tiny machine learning open education initiative. See: http://tinymledu.org
[10] Plancher, Brian and Vijay Janapa Reddi (2022) TinyMLedu: The tiny machine learning open education initiative. Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2.
[11] TinyML4D Academic Network Past Events. http://tinymledu.org/4D/pastEvents
[12] Harvardx Tiny Machine Learning Professional Certificate. See: https://www.edx.org/professional-certificate/harvardx-tiny-machine-learning
[13] Introduction to Embedded Machine Learning. See: https://www.coursera.org/learn/introduction-to-embedded-machine-learning