International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

A Survey Report on TinyML and Small Data – the future of machine learning

Nidhi Sawant[1], Jayant Sawarkar[1], Suraj Shegukar[1], Sanjana Sawant[1], Bhargav Shendge[1], Afsha Akkalkot[2]
[1]TE Students, Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

Abstract:
This survey report provides a comprehensive analysis of the concepts of Small Data and Tiny Machine Learning (TinyML), along with their respective challenges, techniques, and applications. Small data refers to datasets that are relatively limited in size, typically characterized by a low number of samples, sparse feature space, or data scarcity due to privacy concerns or data collection challenges, while TinyML is a field of study in Machine Learning and Embedded Systems that explores the types of models you can run on small, low-powered devices like microcontrollers. It enables low-latency, low power and low bandwidth model inference at edge devices. This report aims to explore the landscape of small data and TinyML, shedding light on the unique challenges they present and the strategies employed to tackle them. The report begins by defining small data and highlighting its significance in domains where large-scale data collection is impractical or unfeasible. It explores the characteristics of small data, such as sparsity, noise, and class imbalance. Subsequently, the report delves into the emerging field of TinyML, which seeks to deploy machine learning models on devices with limited resources, such as IoT devices. It examines the challenges inherent to TinyML, including limitations computational power and memory, and energy constraints, as well as the unique considerations involved in training and deploying models on such devices. Furthermore, the report explores the diverse applications of small data and TinyML across various domains, such as healthcare, marketing, retail, agriculture, and manufacturing. It showcases real-world use cases that leverage small data and TinyML techniques to solve critical challenges and enable intelligent decision-making at the edge. By presenting a comprehensive survey of small data and TinyML, this report provides researchers, practitioners, and decision-makers with valuable insights into the potential and limitations of these fields. It offers a foundation for further exploration, innovation, and advancements in small data analysis techniques and the deployment of machine learning on resource-constrained devices.

Keywords: TinyML, small data, machine learning (ML), low-latency, small, ML model, Internet of Things (IoT), embedded systems, hardware, software, data, computing, maintaining, devices, energy, real-time, humans, analytics.

I. INTRODUCTION

Small Data simply means data that is small in size. Unlike big data, it is small enough for humans to comprehend and work with. It can be considered to be a subset of big data that is packaged and organized by algorithms to make it more actionable for human beings. It focuses on users, customers, and their behavior. **TinyML** incorporates Machine Learning and Artificial Intelligence into small hardware components. It is a rapidly growing field of machine learning technologies and applications that includes dedicated integrated circuits (hardware), algorithms, and software capable of performing on-device sensor data analytics at extremely low power, typically in the mW range and below. **TinyML** is a branch of machine learning and **embedded systems** research that looks into the types of models that can be run on small, low-power devices like micro controllers.

II. SMALL DATA

The overwhelmingly large amount of data produced every day is often poor in quality. Obstacles like missing data, entry or measurement errors, duplicates, irrelevant predictors make it hard to train a model but are very common within large datasets. The large datasets also have to deal with severe class imbalance. Insufficient access to powerful machine learning hardware cannot be used to train big, parameter rich models. They need powerful computing resources. Additionally, we always tend to look at only specific subsets of data. This is where small data comes into picture.

A. What is small data?

Small Data is a dataset small enough in volume and format so that it is accessible, informative, actionable, and comprehensible by people without the use of complex systems and machine for analysis. A formal definition of small data has been proposed by Allen Bonde, former vice-president of Innovation at Actuate - now part of OpenText: "Small data connects people with timely, meaningful insights, organized and packaged – often visually – to be accessible, understandable, and actionable for everyday tasks." Small data did not become established as a stand-alone category until the emergence of big data, and thus represents a derivative of the later.

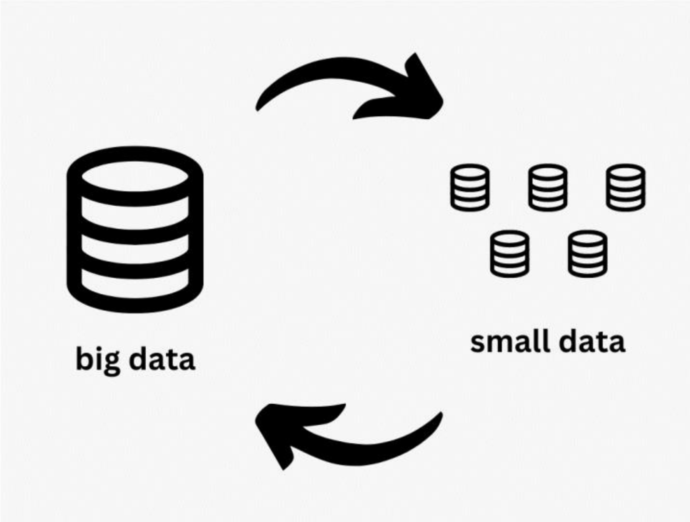

Fig.1.: small data and big data

Small data is a subclass of data deemed modest enough so as to make it accessible, informative, and actionable by people, without the need for overly complex analytical tools. Big data is about machines, while small data is about people specifically, meaningful insights organized and packaged for the derivation of causation's, patterns, and the reasons "why" about people. Small data is usually small enough for humans to understand in terms of volume and format. Though it doesn't have the same level of impact as big data when it comes to the overall business, it has a greater impact on short-term and current decisions.

B. Benefits

And although small data is sparse and noisy, it is gaining popularity due to the cost efficiency and time savings that it offers. Some other benefits of Small Data include, [1] Easier and more actionable, [2] Visualization and inspection, [3] Closer to the end user, [4] Simpler, [5] Better Customer insight, [6] Increased market intelligence, [7] Supply chain management, [8] Data driven innovation

III. TINYML

According to **tinyml.org**, "*Tiny machine learning is broadly defined as a fast-growing field of machine learning technologies and applications including hardware, algorithms, and software capable of performing on-device sensor data analytics at extremely low power, typically in the mW range and below, and hence enabling a variety of always-on use-cases and targeting battery operated devices.*" TinyML is a surfacing field that acts as a junction between embedded systems and ML. TinyML is a tool provider for ML model development and their execution on devices that are considered to be resource-constrained or low-energy devices. TinyML congregates microcontrollers and IoT devices with machine learning to execute on-device analytics by leveraging substantial amounts of data amassed by them.

A. Emergence and upside

The expansion of TinyML over the past years can be ascribed to evolution of embedded systems that act as its foundation. It obviates the need of transmitting data to a central server as it is managed by a standard IoT device over cloud. The IoT device accumulates data and sends it to a central server. This is done over a cloud where hosted ML models provide all the insights. TinyML is pretty advantageous. It has coigns of vantage like fast inference, low latency, data privacy, energy savings, reduced bandwidth and independence of internet connectivity. Additionally, the ability of TinyML to optimize Machine learning models to work on low power and low memory devices like sensors and microcontrollers, and making machine learning responsive in real-time brings it to a high ground. Its objective is to do more with less.

Typically, a micro controller consumes electricity in the milliwatts or micro watts range, whereas a consumer CPU consumes between 65 to 85 watts and a consumer GPU consumes between 200 and 500 watts. This means, a thousand-fold reduction in power use. Because of their low power consumption, TinyML devices can run on batteries for weeks, months, or even years while running machine learning applications.

B. Use cases

The potential and applications of TinyML are extensive. It is used across multiple fields, especially the ones that use IoT networks and data. [1] It can be used in, the agricultural sector for gathering and monitoring real-time livestock data, [2] in retail for monitoring inventories and sending alerts so that out-of-stock situations can be prevented, [3] in manufacturing and industrial predictive maintenance to save downtime and cost of equipment failure. [4] It is also used in health care devices to gather real time patient data, [5] It can be used as a marketing tool and personalization of user experience. Applications based on TinyML enable businesses to comprehend the behaviour of their customers.

C. Workflow

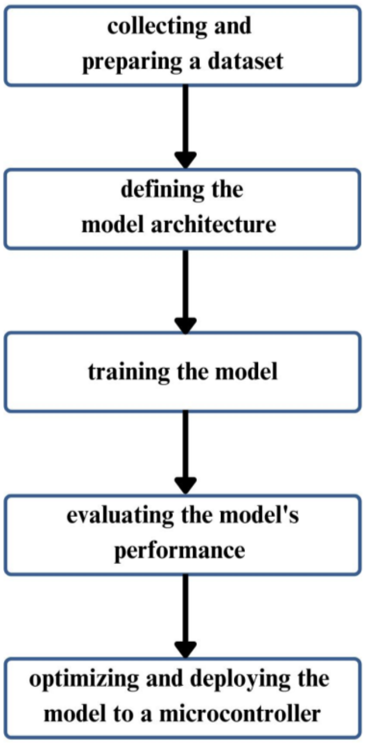

Fig.2.: General workflow for developing a TinyML model

IV. LIMITATIONS OF TINY ML

4.1. Limited Computational Power:

Resource-constrained devices typically have limited computational power, including processing speed, memory, and energy consumption. This limitation restricts the complexity and size of the machine learning models that can be deployed on such devices.

4.2. Limited Memory Capacity:

TinyML models need to fit within the limited memory capacity of the target devices. This constraint poses a challenge when dealing with larger models or datasets. The models often need to be optimized and compressed to reduce memory requirements while maintaining performance.

4.3. Energy Efficiency:

Energy consumption is a critical consideration for resource-constrained devices, as they are often powered by batteries or have limited power budgets. Machine learning algorithms can be computationally intensive, requiring significant energy. Developing energy-efficient models and algorithms is crucial for TinyML applications.

4.4. Limited Training Data:

Training machine learning models usually requires large datasets to achieve good accuracy. However, on resource-constrained devices, the amount of available training data might be limited. This limitation can affect the performance and generalization of the models.

4.5. Limited Communication Bandwidth:

Transferring data between resource-constrained devices and the cloud or other devices can be slow and costly in terms of bandwidth. This limitation affects tasks such as model updates, data synchronization, and communication with other devices, which can be critical for certain applications.

4.6. Model Complexity and Performance Trade-offs:

To fit within the constraints of resource-constrained devices, TinyML models often need to be simplified or pruned, which can lead to a trade-off between model complexity and performance. Balancing model accuracy with resource limitations becomes crucial, and achieving high accuracy might not always be possible.

V. FUTURE SCOPE TINY ML

5.1. Edge Computing:

Edge computing, which involves processing data on or near the devices generating it, is gaining momentum. TinyML plays a crucial role in enabling intelligent decision-making and data analysis at the edge, reducing latency, bandwidth requirements, and dependence on cloud connectivity. The ability to run machine learning algorithms locally on resource-constrained devices opens up opportunities for applications in various fields, including healthcare, agriculture, manufacturing, and smart cities.

5.2. Internet of Things (IoT):

The integration of TinyML with IoT devices has the potential to revolutionize the IoT landscape. By enabling on-device machine learning capabilities, TinyML can enhance data processing, real-time analytics, and decision-making within IoT systems. This leads to increased efficiency, reduced latency, improved privacy and security, and enables more intelligent and autonomous IoT devices and systems.

5.3. Wearable Devices and Healthcare:

TinyML can power advanced analytics and machine learning algorithms on wearable devices, such as fitness trackers, smartwatches, and medical devices. This enables personalized health monitoring, real-time diagnostics, early detection of health issues, and proactive healthcare interventions. The ability to process data locally on wearable devices also addresses privacy concerns associated with transmitting sensitive health data to the cloud.

5.4. Energy Efficiency and Sustainability:

Resource-constrained devices are often powered by batteries or have limited energy budgets. TinyML algorithms can be optimized to minimize energy consumption, making them ideal for energy-efficient applications. For example, in the context of smart buildings, TinyML can help optimize energy usage by analyzing sensor data and controlling energy-consuming devices.

5.5. Robotics and Autonomous Systems:

TinyML can be integrated into robotic systems, enabling them to make intelligent decisions and adapt to changing environments in real-time. This opens up possibilities for autonomous robots, drones, and vehicles that can perform complex tasks locally, reducing the need for continuous communication with a central processing unit.

VI. CONCLUSION

In summary, this survey report explores the concepts of Small Data and Tiny Machine Learning, their challenges, techniques, and applications. It emphasizes the importance of small data in domains where large-scale data collection is impractical, while highlighting the emerging field of TinyML for deploying machine learning models on resource-constrained devices. The report showcases real-world applications across various industries and provides valuable insights for researchers, practitioners, and decision-makers. It lays the foundation for further advancements in small data analysis techniques and the deployment of machine learning at the edge.

VII. REFERENCES

[1]. Chael Chui, James Manyika, Mehdi Miremadi, Nicolaus Henke, Rita Chung, Pieter Nel, and Sankalp Malhotra, 2018, Notes from the AI frontier: Applications and value of deep learning.

[2]. Vuppalapati C., Ilapakurti A., Kedari S., Vuppalapati J., Kedari S., & Vuppalapati R. (2020). Democratization of AI, Albeit Constrained IoT Devices & Tiny ML, for Creating a Sustainable Food Future. 2020 3rd International Conference on Information and Computer Technologies (ICICT).

[3]. S. A. R. Zaidi, A. M. Hayajneh, M. Hafeez and Q. Z. Ahmed, "Unlocking Edge Intelligence Through Tiny Machine Learning (TinyML)," in IEEE Access, vol. 10, pp. 100867-100877, 2022, doi: 10.1109/ACCESS.2022.3207200.

[4]. Wolf, F. (2022, September 21). *Small data: Machines that learn more from less - Mergeflow*. Mergeflow - AI-augmented Analytics for Faster Insights on R&D and Business. https://mergeflow.com/research/small-data

[5]. Van Heeswijk, W., PhD. (2022, December 21). Is “Small Data” The Next Big Thing In Data Science? *Medium*. https://towardsdatascience.com/is-small-data-the-next-big-thing-in-data-science-9acc7f24907f

[6]. Pykes, K. (2023, February 13). What is TinyML? An Introduction to Tiny Machine Learning. https://www.datacamp.com/blog/what-is-tinyml-tiny-machine-learning

[7]. Arun. (2023, May 13). An Introduction to TinyML - Towards Data Science. *Medium*. https://towardsdatascience.com/an-introduction-to-tinyml-4617f314aa79

[8]. Dilmegani, C. (2023). TinyML in 2023: Machine Learning at the Edge. *AIMultiple*. https://research.aimultiple.com/tinyml/