Abstract—In recent decades, Machine Learning (ML) has become extremely important for many computing applications. The pervasiveness of ultra-low-power embedded devices such as ESP32 or ESP32 Cam with tiny Machine Learning (tinyML) applications will enable the mass proliferation of Artificial Intelligent powered Embedded IoT Devices. In the last few years, the microcontroller device (Espressif ESP32) became powerful enough to be used for small/tiny machine learning (tinyML) tasks. The ease of use of platforms like Arduino IDE, MicroPython and TensorFlow Lite (TF) with tinyML application make it an indispensable topic of research for mobile robotics, modern computer science and electrical engineering. The goal of this paper is to analyze the speed of the Xtensa dual core 32-bit LX6 microprocessor by running a neural network application. The different number of inputs (9, 36, 144 and 576) inputted through the different number of neurons in neural networks with one and two hidden layers. Xtensa LX6 microprocessor has been analyzed because it comes inside with Espressif ESP32 and ESP32 Cam which are very easy to use, plug and play IoT device. In this paper speed of the Xtensa LX6 microprocessor in feed-forward mode has been analyzed.

Keywords—TinyML, Xtensa LX6 microprocessor, Machine Learning, Neural Network, Embedded IoT Device, Espressif ESP32 and ESP32 Cam.

I. INTRODUCTION

The explosive growth in machine learning and miniaturization of electronics has been created new research opportunities. The thought of eco-friendly energy, which stresses the utilization of electric-driven gear [1,2], has been recently in the discussion [3]. Machine Learning already achieved the great success with the multiple core enabled processor, and in recent days researchers thinking to apply small or tiny machine learning application on system on chip (SoC) microcontrollers with less processing power [4].

However, the extensive deployment of the machine learning algorithms will bring a clear rise in the artificial intelligence where a huge amount of processing power will be used due to process small algorithm with high-end multicore processor. Other animals and humans have a brain and a central nervous system to avail the process of information with neural networks. In the different fields, neural network has been successfully implemented, from detecting criminal [5,6] and medical diagnosis [7,8] through image classification [9], to autonomous driving [10].

As part of a long-established custom, practice, or belief, the training of neural network is commonly centralized, and for the computational intensity and big memory requirements it is mostly done on cloud servers. An instance of a particular situation the dataset of deep neural networks it could be more than several dozen terabytes for learning [11,12].

In centralized approach some researcher found disadvantages like data privacy issues [13], between devices and cloud it increased latency and network traffic for that forbid in some determination [14]. It also decreases the reliability, when exchanging of information is important between cloud and device or sensors [15]. And some researcher presents that sometimes the center can be SoC, smartphone and micro data center between source and data [16].

In the most recent decade, many researcher, engineers and professionals have studied on machine learning applications and implement these on microcontrollers, SoC’s and embedded IoT devices. Special libraries [17] and hardware development with ML/AI functions and capabilities [18-20]. For the ESP32 and ESP32-S Series SoCs Espressif developed an IoT Development Framework named ESP-IDF [21] and ESP32 chip-based platform that is used for facial recognition and detection [22].

In addition, many research works have been done, studied and used ESP32 widely like in embedded IoT devices such as LPG gas leakage [23], home monitoring system [24], solar cell current and its battery power prediction [25] and air pollution detect system [26].

In this paper, the Tensilica Xtensa LX6 microprocessor by Cadence is discussed and its neural network capabilities will be analyzed [27].
In Fig. 1, the ESP32 SoC’s processor Xtensa LX6 DPU showing standard, optional, and designer-defined blocks. For neural network acceleration, this SoC does not have any special hardware but some ML frameworks and examples for ESP32 chip has been developed by Espressif. So that I choose a ESP32 WROOM IoT development board by Espressif for the neural network acceleration that inside have a Xtensa LX6 microprocessor SoC.

A SoC or a new class of programmable processor that combines high-performance and industry-standard, software-programmable multi-core CPU is called data plane processing units (DPUs). Xtensa LX6 called as DPU. It’s highly efficient, small and it have low-power 32-bit base architecture. In Fig. 2 an functional block diagram by Espressif is showed for ESP32.

Fig. 2. Espressif Functional Block Diagram for ESP32 chip.

The formation of this study as follows. In section II, a scheme and machine learning usage on ESP32 is discussed previous studies. In section III, some study that deals with neural networks and Xtensa LX6 multi-core typical example or pattern will be described. In section IV, analysis of EPS32 processing unit Xtensa LX6's one, two and both cores speed of feedforward process propagation of neural network. The testing is done and finally in section V result will be analyzed and discussed.

II. THE SCHEME

For analysis of Xtensa LX6 microprocessor for neural network applications, need to perform few necessary steps. An ESP32 development board is essential tools for this. Arduino IDE with Espressif ESP32 board compatible is important. Using Arduino IDE, a neural network will be developed and will implement into an ESP32 for the analyzing its processing power.

It can be observed that Xtensa LX6 microprocessor most are used for the general, including a few neurons. But there is no study report regarding machine learning applications that study, being researched or investigate the speed of running neural networks on a microcontroller in the real-world conditions.

In the next, the different approaches analyzed that already studied by the researchers.

A. Machine Learning on the Cloud

By Chand et al. an ultrasonic sensor with ESP32 has been used to predict the behavior and analyzing the status of a person in a room. With the help of sensor, the ESP32 is used for collecting the data and send it to the cloud. A machine learning algorithm is used to analyze data and give decision whether the situation is normal or abnormal situation [28].

By Fernoaga et al an ESP32 camera has been used for take the images for the neural network’s recognition [29].

Rosato and Masciadri has make a observation system by using an ESP32 that is non-invasive and detect the sitting people by analyzing the data on cloud and the classification of the problem has been solved by Logistic regression algorithm [30].

By Islam et al. designed a module of biosensing with an Analogue Device AD8232, ESP32 as microcontroller to transmit data. A Linux server receive those data and analyze with neural network [31].

By Komarek et al has been presented that MQTT is used in abstraction layer and interface nodes with sensors. Those particular nodes are capable for transmit and receive data to another nodes. A cloud server is used for the analysis [32].

By Zidek et al. optimize data for IoT and accumulate them to the packet and transmit them to the cloud. These packets have multi-value [33].

B. Machine Learning on Xtensa LX6 microprocessor by using ESP32

Very few authors has been studied for ESP32 for some Machine Learning applications.

Kokoulin et al implement a video recognition system that detects the presence of a face or silhouette fragment by using ESP32. For shorten the network traffic and computational load of facial recognition server and got the result of 80% to 90% traffic decrease [34].

Espressif Systems has been developed an open-source ESP-WHO framework that is available on GitHub for recognition and detection [22]. The Cascaded Convolutional
Networks model and new mobile architecture – Mobile Net V2 is based on this ESP-WHO framework [35,36].

C. Machine Learning on Arduino with Arduino IDE

Arduino is an open-source hardware platform and it have many hardware devices that being used for machine learning applications. Some author already implements machine learning applications on Arduino development boards by using Arduino IDE.

Neural network algorithm named C-Mantec that can add new neurons in the leaning process has been implemented by Ortega-Zamorano et al. [37]. Two hidden and one output neuron has been used to control PID controller and DC motor in Arduino Uno based development board, designed and implemented by Rai and Rai [38]. Adhitya et al. used neural networks with two inputs executed from sensor data readings. Whereas, two hidden layers with seven neurons each and two neurons as output layer [39].

III. DECENTRALIZED NEURAL NETWORKS

In this study, the speed of single-machine parallelism in data propagation is tested in neural network feed-forward process. In the hidden layer by different number of neurons in Xtensa LX6 microprocessor has been done testing with it’s one and two cores that was formerly mentioned.

The process will be divided into two same-size groups for data propagation. Simple tasks division can be seen in the Fig. 3 by three method. For the upper half (A) the “core 0" will calculate values and for the lower half (B) another “core 1” will calculate values. After calculating the total (j = A+B) neurons of hidden layer a result will generate (j).

IV. THE FEED FORWARD PROCESS AND THE PROPAGATION SPEED

A. Hardware Preparation

For testing Xtensa LX6 microprocessor propagation speed of neural network process with it both cores, an ESP32 SoC is used. In the market, there are various type of ESP32 boards are available and for the testing purpose “ESP-WROOM-32” ESP32 board has been used that is produced by Espressif and showed in Fig. 4.

B. Software Preparation

Because of opensource there are various way to program ESP32 boards. Two main method for programming ESP32:

- a. Espressif IoT Development Framework “ESP-IDF”
- b. Arduino IDE, ESP32 Arduino Core

Compared with ESP-IDF and Arduino IDE, ESP-IDF low-level programming whereas Arduino IDE has some extra facilities. In this study, for testing Xtensa LX6 microprocessor Arduino IDE is used to measure propagation speed of this process for this feedforward neural network.

C. Measuring the Execution time

For measuring the execution time there are various way. In this study, micros() the simple function is used.

Example Code:

```c
time_start = micros();
//some code
time_end = micros();
time = time_end – time_start;
```

Arduino IDE receive the values through serial monitor when micros() function returns the number of microseconds [40].

![Fig. 4. “ESP-WROOM-32” ESP32 Development Board](image)

![Fig. 5. Pre-activation Layer](image)
In Fig. 5, it can be seen which part is being implemented. In fact, this part is time dependent for measuring in neural networks with each hidden layer. The number of inputs is much higher than the normal neural outputs. Expressed as following equation,

\[ z = b + \sum_{i=1}^{N} a_i w_i \tag{1} \]

Whereas, \[ b = \text{Bias}, \quad w_i = \text{weights}, \quad a_i = \text{input values}, \quad z = \text{a sum of all multiplications of } a_i \]

### D. Experiments and Results

Xtensa LX6 microprocessor on ESP32 board have two cores and for the first time only one core is used for measuring the propagation time. Describes in TABLE I. The FreeRTOS have xTaskCreatePinnedToCore() function which is capable for setting the specific processor core for task. In ESP32 have only two cores and for the first segment of taking measurements with only one core when the function is set to 0. Last argument in the function is set to 1, this argument can measure with only one core when the function is set to 20.

### TABLE I. RESULTS WITH SINGLE CORE IN XTENSA LX6 MICROPROCESSOR ON ESP32

| No | Layer | Operations | Time   |
|----|-------|------------|--------|
|    | Input | Hidden     | μs     |
| 1  | 20    | 1000       | 383    |
| 2  | 40    | 2000       | 681    |
| 3  | 60    | 3000       | 966    |
| 4  | 80    | 4000       | 1258   |
| 5  | 100   | 5000       | 1582   |
| 6  | 120   | 6000       | 1830   |
| 7  | 140   | 7000       | 2203   |
| 8  | 160   | 8000       | 2415   |
| 9  | 180   | 9000       | 2815   |
| 10 | 200   | 10000      | 3071   |

Some additional setup and code are needed for measuring the propagation time in Xtensa LX6 microprocessor on ESP32. For single core, all task has been done in using only one core. But for using both cores, as I mentioned previously Xtensa LX6 microprocessor have two cores, all tasks have to be divided into two parts.

### TABLE II. RESULTS WITH BOTH CORE IN XTENSA LX6 MICROPROCESSOR ON ESP32

| No | Layer | Operations | Time   |
|----|-------|------------|--------|
|    | Input | Hidden     | μs     |
| 1  | 20    | 1000       | 253    |
| 2  | 40    | 2000       | 406    |
| 3  | 60    | 3000       | 547    |
| 4  | 80    | 4000       | 699    |
| 5  | 100   | 5000       | 858    |
| 6  | 120   | 6000       | 976    |
| 7  | 140   | 7000       | 1159   |
| 8  | 160   | 8000       | 1268   |
| 9  | 180   | 9000       | 1482   |
| 10 | 200   | 10000      | 1599   |

Elapsed time ration can be seen in TABLE III. For the number of multiplication and addition operations on Xtensa LX6 microprocessor cores on ESP32 boards, it is obvious that ration is rising. If increasing number of operations has less than the impact is less.

### TABLE III. COMPARISON RESULTS OF XTENSA LX6 MICROPROCESSOR CORES ON ESP32

| No | Time (in μs) | Ratio |
|----|--------------|-------|
|    | One Core     | Two Core |
| 1  | 383          | 253    | 1.51  |
| 2  | 681          | 406    | 1.68  |
| 3  | 966          | 547    | 1.76  |
| 4  | 1258         | 699    | 1.80  |
| 5  | 1582         | 858    | 1.84  |
| 6  | 1830         | 976    | 1.87  |
| 7  | 2203         | 1159   | 1.90  |
| 8  | 2415         | 1268   | 1.90  |
| 9  | 2815         | 1482   | 1.90  |
| 10 | 3071         | 1599   | 1.92  |

In these experiments, the time unit is calculated in microsecond (μs). It’s equal to one millionth (0.000001 or \(10^{-6}\) or the reciprocal of 1,000,000) of a second.
V. DISCUSSION AND CONCLUSION

Xtensa LX6 microprocessor have two cores on ESP32. The main theme of Amdahl’s law is the original execution time divided by an enhanced execution time [41]. In the recent decades, the modern version of Amdahl’s law says that enhancing the fraction $f$ of a computation by a speedup $S$, then I get [41],

$$\text{Speedup}_{\text{enhanced}}(f,S) = \frac{1}{(1-f)+\frac{f}{S}} \quad (2)$$

In the Fig. 6, we can see a chart of comparison results of Xtensa LX6 microprocessor cores on ESP32 from the TABLE III.

![Fig. 6. Comparison results of Xtensa LX6 cores on ESP32](image)

In this study, Xtensa LX6 microprocessor have two cores and from Amdahl's law perspective the result will be analyzed. So that I formulated the Amdahl's law as,

$$S_{\text{latency}}(S) = \frac{1}{(1-p)+\frac{p}{S}} \quad (3)$$

Whereas, $S_{\text{latency}}$ is code execution speedup, $S$ is instead of value 2 because of Xtensa LX6 microprocessor multiple cores, and $p$ is code execution time proportion.

According to this equation (3) in TABLE IV, the values for $p$ are calculated.

| Ratio | Number of Operations | $p$ |
|-------|----------------------|-----|
| 1.51  | 1000                 | 0.68|
| 1.68  | 2000                 | 0.81|
| 1.76  | 3000                 | 0.86|
| 1.80  | 4000                 | 0.89|
| 1.84  | 5000                 | 0.91|
| 1.87  | 6000                 | 0.93|
| 1.90  | 7000                 | 0.95|
| 1.90  | 8000                 | 0.95|
| 1.90  | 9000                 | 0.95|
| 1.92  | 10000                | 0.96|

It is noticeable in the TABLE IV, that the calculated values of $p$ proportion of code execution time is increased and this is obvious.

This is because data transfer between different types of memory has less and less impact on the total run time. If and only if static tasks for processor cores setup [4].

REFERENCES

[1] J.X. Wang, Y.Z. Li, X.D. Liu, C.Q. Shen, H.S. Zhang, K. Xiong. Recent active thermal management technologies for the development of energy-optimized aerospace vehicles in China, Available online, Chin. J. Aeronaut. (2020), https://doi.org/10.1016/j.jca.2020.06.021.
[2] B.J. Brejje, J. Martins, Electric, hybrid, and turboelectric fixed-wing aircraft: a review of concepts, models, and design approaches, Prog. Aero. Sci. 104 (2019) 1–19.
[3] A.R. Gnadh, R.L. Speth, J.S. Sahnis, S.R.H. Barrett, Technical and environmental assessment of all-electric 180-passenger commercial aircraft, Prog. Aero. Sci. 105 (2019) 1–30.
[4] Đokić, K., Martinović, M. and Radišić, B., “Neural networks with ESP32-are two heads faster than one?.” In Conference on Data Science and Machine Learning Applications (CDMA 2020) (p. 141).
[5] Verma, Harsh, Siddharth Lotia, and Anurag Singh. "Convolutional Neural Network Based Criminal Detection." In 2020 IEEE REGION 10 CONFERENCE (TENCON), pp. 1124-1129. IEEE, 2020.
[6] Karamchandani, Sunil, and Ganesh Shukla. "Face Sketch-Image Recognition for Criminal Detection Using a GAN Architecture." International Conference on Information and Communication Technology for Intelligent Systems. Springer, Singapore, 2020.
[7] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," International Journal of Computer Science Issues, vol. 8, pp. 150-154, 2011.
[8] F. Amato, A. López, E. M. Peña-Méndez, P. Vañhara, A. Hampl and J. Havel, Artificial neural networks in medical diagnosis, Elsevier, 2013.
[9] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," 2011.
[10] Van Uytseel, Steven, and Danilo Vasconcellos Vargas. "Challenges for and with Autonomous Vehicles: An Introduction." Autonomous Vehicles. Springer, Singapore, 2021. 1-17.
[11] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
[12] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A. C., 2015. Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), pp.211-252.
[13] Zhao, Jianxin, Richard Motrier, Jon Crowcroft, and Liang Wang. "Privacy-preserving machine learning based data analytics on edge devices." In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 341-346. 2018.
[14] Khelifi, Hakima, Senlin Lou, Bouhakr Noir, Akrem Sellami, Hassine Moungla, Syed Hassan Ahmed, and Mohsen Guizani. "Bringing deep learning at the edge of information-centric internet of things." IEEE Communications Letters 23, no. 1 (2018): 52-55.
[15] Wang, Xiaofei, Yiwen Han, Victor CM Leung, Dusit Niyato, Xueqiang Yan, and Xu Chen. "Convergence of edge computing and deep learning: A comprehensive survey." IEEE Communications Surveys & Tutorials 22, no. 2 (2020): 869-904.
[16] Aazam, Mohammad, and Eui-Nam Huh. "Fog computing micro datacenter based dynamic resource estimation and pricing model for IoT." In 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, pp. 687-694. IEEE, 2015.
[17] G. Hegde, N. Ramasamy and N. Kapre, "CaffeFresso: an optimized library for deep learning on embedded accelerator-based platforms," 2016.
[18] Y. Zhu, M. Mattina and P. Whatmough, "Mobile machine learning hardware at arm: a systems-on-chip (soc) perspective," arXiv preprint arXiv:1801.06274, 2018.
[19] J. Tang, D. Sun, S. Liu and J.-L. Gaudiot, "Enabling deep learning on IoT devices," Computer, vol. 50, pp. 92-96, 2017.
[20] Shawahna, S. M. Sait and A. El-Maleh, "FPGA-Based Accelerators of Deep Learning Networks for Learning and Classification: A Review," IEEE Access, vol. 7, pp. 7823-7859, 2019.
Feed Forward Neural Network in automatic operating temperature and control of DC motor using arduino uno. International Journal of Engineering Trends and Technology, 4(2), 137-140. Rai, N., & Rai, B. (2013). Neural network based closed loop speed and alignment using multitask cascaded convolutional networks. IEEE Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

Ortega-Zamorano, F., Jerez, J. M., Subirats, J. L., Molina, I., & Franco, L. (2014). Smart sensor/actuator node reprogramming in changing environments using a neural network model. Engineering Applications of Artificial Intelligence, 30, 179-188.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitsks cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503.

Rai, N., & Rai, B. (2013). Neural network based closed loop speed control of DC motor using arduino uno. International Journal of Engineering Trends and Technology, 4(2), 137-140.

R. Y. Adhiyia et al., "Comparison methods of Fuzzy Logic Control and Feed Forward Neural Network in automatic operating temperature and humidity control system (Oyster Mushroom Farm House) using microcontroller," 2016 International Symposium on Electronics and Smart Devices (ISESD), 2016, pp. 168-173, doi: 10.1109/ISESD.2016.7886713

[40] Arduino, "micros()." Arduino, [Online]. Available: https://www.arduino.cc/reference/tr/language/functions/time/micros/. [Accessed February 28, 2021]

[41] Hill, Mark D., and Michael R. Marty. "Amdahl's law in the multicore era." Computer 41, no. 7 (2008): 33-38.

[42] Amdahl, Gene M. "Validity of the single processor approach to achieving large scale computing capabilities." In Proceedings of the April 18-20, 1967, spring joint computer conference, pp. 483-485. 1967