Cognitive Electronic Unit for Assisted Ultrasound: Preliminary Results and Future Perspectives

Emanuele De Luca[1], Emanuele Amato[2,3], Vincenzo Valente[4], Marianna La Rocca[2,3], Tommaso Maggipinto[2,3], Roberto Bellotti[2,3] and Francesco Dell'Olio[1,*]

[1]*Micro Nano Sensor Group, Politecnico di Bari, Bari, Italy*
[2]*Dipartimento Interateneo di Fisica, Università degli Studi di Bari Aldo Moro, Bari, Italy*
[3]*Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Bari, Bari, Italy*
[4]*Predict s.r.l., Bari, Italy*

Abstract

This paper reports on the preliminary results of an ongoing research activity aimed at developing a cognitive electronic unit for assisted echocardiography. The paper discusses the selection of the suitable hardware and the design of a neural network model optimized for small datasets. A data capture unit has been implemented to facilitate the collection of ultrasound data in collaboration with clinical professionals. The chosen hardware supports real-time image processing and data transmission, while the neural network is intended for image classification. Initial results indicate the potential of the cognitive electronic unit we are developing to reduce inter-operator variability and enhance diagnostic precision in echocardiography. Further data collection and model refinement are still ongoing.

Keywords

Assisted ultrasonography, artificial intelligence, embedded systems

1. Introduction

Ultrasound is one of the most widely used diagnostic techniques due to its numerous advantages, including cost-effectiveness, safety (being radiation-free), and the ability to be performed in real time [1]. Additionally, ultrasound demonstrates significant versatility, applicable to almost any part of the body except for bones, lungs, and sections of the intestine.

However, performing an ultrasound scan correctly necessitates substantial training and years of experience. This requirement creates a considerable workload for specialists and complicates the repeatability of diagnostic examinations. Consequently, the expertise and skills of the sonographer profoundly influence the performance of an examination, leading to substantial inter-operator variability.

Recent advancements in artificial intelligence techniques have enabled the exploration of both robotic and assisted ultrasound applications to address these critical issues [2]. In the robotic scenario, a robotic arm performs ultrasound examinations under the guidance of artificial intelligence algorithms. These algorithms manage various functions, including planning the scanning path, adjusting the arm's movement in space and the probe's pressure on the scanned body area, and completing the scan. Robotic ultrasound systems can be classified into three categories based on their level of autonomy: teleoperated [3], semi-autonomous [4], and fully autonomous systems [5]. Teleoperated systems involve robotic arms piloted by an operator, aiming to reduce or prevent musculoskeletal disorders in physicians and facilitate remote diagnosis. Semi-autonomous systems autonomously perform some tasks, such as positioning the probe in the body region of interest, but still require operator intervention to conduct the examination. Fully autonomous systems can independently plan and execute the ultrasound scan, acquiring the necessary images without sonographer intervention, thereby allowing the physician to

CPS Workshop, September 16, 2024, Alghero, IT
*Corresponding author.
francesco.dellolio@poliba.it (F. Dell'Olio)
https://mnsensor.poliba.it/team/prof-francesco-dellolio/ (F. Dell'Olio)
0000-0001-9874-5008 (F. Dell'Olio)

review the examination subsequently. In the assisted ultrasound scenario, a sonographer performs the examination while being guided by an algorithm in the movement of the probe. Although robotic ultrasound remains primarily within the research domain, some applications of assisted ultrasound have already reached the market [6]. Notable advancements in the echocardiographic domain employ deep learning algorithms to guide non-expert operators. This technology includes real-time image quality assessment systems and determines the optimal probe movements, enabling operators to obtain transthoracic echocardiographic images without the necessity of external tracking systems.

The objective of the ongoing research activity reported in this paper is to develop a cognitive unit that enhances ultrasound imaging by assisting operators with real-time guidance and assessment. This unit aims to improve image quality and provide valuable feedback during the scanning process. The integration of such a cognitive unit with commercial ultrasound scanners can potentially aid in reducing inter-operator variability, benefiting both trainees and general practitioners.

2. Hardware setup

The selection of appropriate hardware for executing deep learning algorithms was guided by a thorough literature review focused on evaluating the performance of edge devices in artificial intelligence applications. Given the parallel development of software and the absence of specific algorithms to test, it was crucial to base hardware choices on established benchmarks and comparative analyses available in the literature.

The initial comparative analysis included devices commonly tested in image processing applications, such as the ASUS Tinker Edge R, Raspberry Pi 4, Google Coral Dev Board, NVIDIA Jetson Nano, and Arduino Nano 33 BLE. The evaluation considered inference speed and accuracy across various network models. Findings indicated that the Google Coral Dev Board demonstrated superior performance in continuous computational applications for models compatible with the TensorFlow Lite framework. The NVIDIA Jetson Nano ranked closely behind, offering greater versatility and the capability to train models due to its GPU presence [7].

Further analysis narrowed the focus to GPU-equipped devices. Literature indicated that the NVIDIA Jetson Nano exhibited better image processing performance compared to the Jetson TX2, GTX 1060, and Tesla V100 in convolutional neural network applications [8]. Within the NVIDIA Jetson series, the Jetson Orin Nano was identified as significantly outperforming both the Jetson Nano and Jetson AGX Xavier for video processing tasks using convolutional neural network models developed in PyTorch and optimized with NVIDIA's Torch-TensorRT SDK [9].

Based on these insights, the NVIDIA Jetson Orin Nano was selected for its robust performance. Table 1 summarizes the main features on the NVIDIA Jetson Orin Nano.

Table 1
NVIDIA Jetson Orin Nano Technical Specifications.

CPU cores	6
CUDA cores (GPU)	1024
Tensor cores (GPU)	32
TOP/s	40
Power (W)	7-15

Figure 1 shows the experimental setup we are using for both ultrasound data collection and the inference phase. A small inertial measurement unit (IMU) on the market was selected to attach to the ultrasound probe to utilize the probe's spatial position during inference.

After conducting preliminary tests on both online and offline ultrasound video processing through deep learning applications, we started to evaluate the performance of the Hardware setup by developing a simple neural network model that was trained on a public dataset [10], without taking into account the probe's spatial position. The neural network classified echocardiographic images into 'Apical projection

with 2-chamber view (2CH)', 'Apical projection with 4-chamber view (4CH)', and 'Unknown', with the latter including all non-classifiable images. This network was designed to classify automatically and in real time the observed cardiac projection.

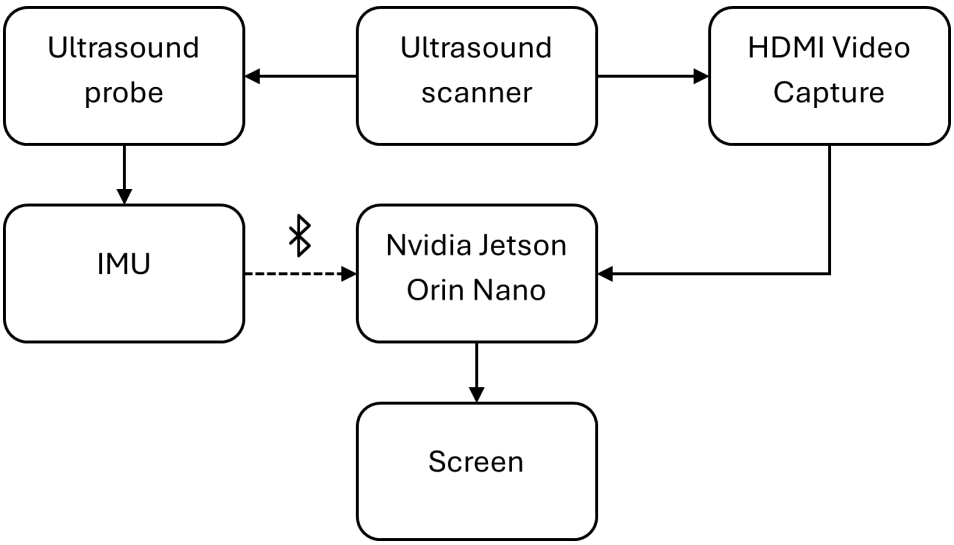

Figure 1: Hardware setup.

3. Neural network model and data capture

The neural network model, presented in figure 2, is designed to balance computational complexity and generalization capacity, making it suitable for smaller datasets. Its architecture comprises several layers, each performing specific operations to extract features and classify images. Original images were in NIfTI format and different in size, so they were converted into JPG format and resized to a resolution of 64x64x3. The resizing of the images was carried out to keep the inference computational cost lower. Each pixel is represented on 8 bits.

The model begins with a convolutional layer that applies 8 filters of size 3x3 on each input image (64x64x3). This convolutional layer uses a stride of (1,1) and 'same' padding to maintain the output dimensions equal to the input. Following the convolution, a Rectified Linear Unit (ReLU) activation function [11] introduces non-linearity into the model, essential for learning complex data relationships. To mitigate overfitting, a dropout layer with a 30 % dropout rate is applied, randomly deactivating neurons during training.

The second stage involves another convolutional layer, this time with 16 filters of size 3x3, maintaining the same stride and padding settings. The ReLU activation function is used again to ensure non-linearity. Subsequently, a max pooling layer with a pool size of (2,2) and a stride of (2,2) reduces the spatial dimensions of the input while preserving the most significant features.

In the third stage, the model includes a third convolutional layer with 32 filters of size 3x3, followed by a max pooling layer with identical configurations to the previous one. This further reduces the spatial dimensions of the input, ensuring the model focuses on the most prominent features.

The output from the previous layers is then flattened into a one-dimensional vector, preparing it for the fully connected layers. The first fully connected layer consists of 32 units, with a ReLU activation function to introduce further non-linearity. The final fully connected layer, the output layer, comprises 3 units, corresponding to the number of classes in the classification problem. This layer uses a softmax activation function to assign probabilities to each class.

The choice of a simpler model, such as the one described, is motivated by the limited availability of training data. More complex models like ResNet [12] or Vision Transformers (ViTs) [13] are more prone to overfitting and require extensive computational resources, which are not ideal for scenarios

with smaller datasets. The modular structure of this simple CNN model allows for easy integration of additional components, such as Long Short-Term Memory (LSTM) layers [14], for further enhancements if necessary. This flexibility and efficiency make the model well-suited for the targeted application. It is also important to notice that untill now, there has been no need to quantize the neural network model.

Data collection is a critical aspect in the development of the artificial intelligence model. The available datasets are insufficient to meet the research objectives, necessitating the acquisition of new data in collaboration with clinical professionals. To streamline this process, a comprehensive data capture unit has been implemented. This unit is capable of recording ultrasound screens, capturing the spatial position of the probe, and receiving anonymized data from the ultrasound scanner. It also facilitates the secure transmission of these data for further analysis. The unit incorporates components that automatically receive, process, and anonymize data from the ultrasound scanner. A dedicated graphical user interface (GUI) has been developed to support the data collection process. Despite the automation provided by this system, the labelling of data still requires significant back-office effort from clinical staff members.

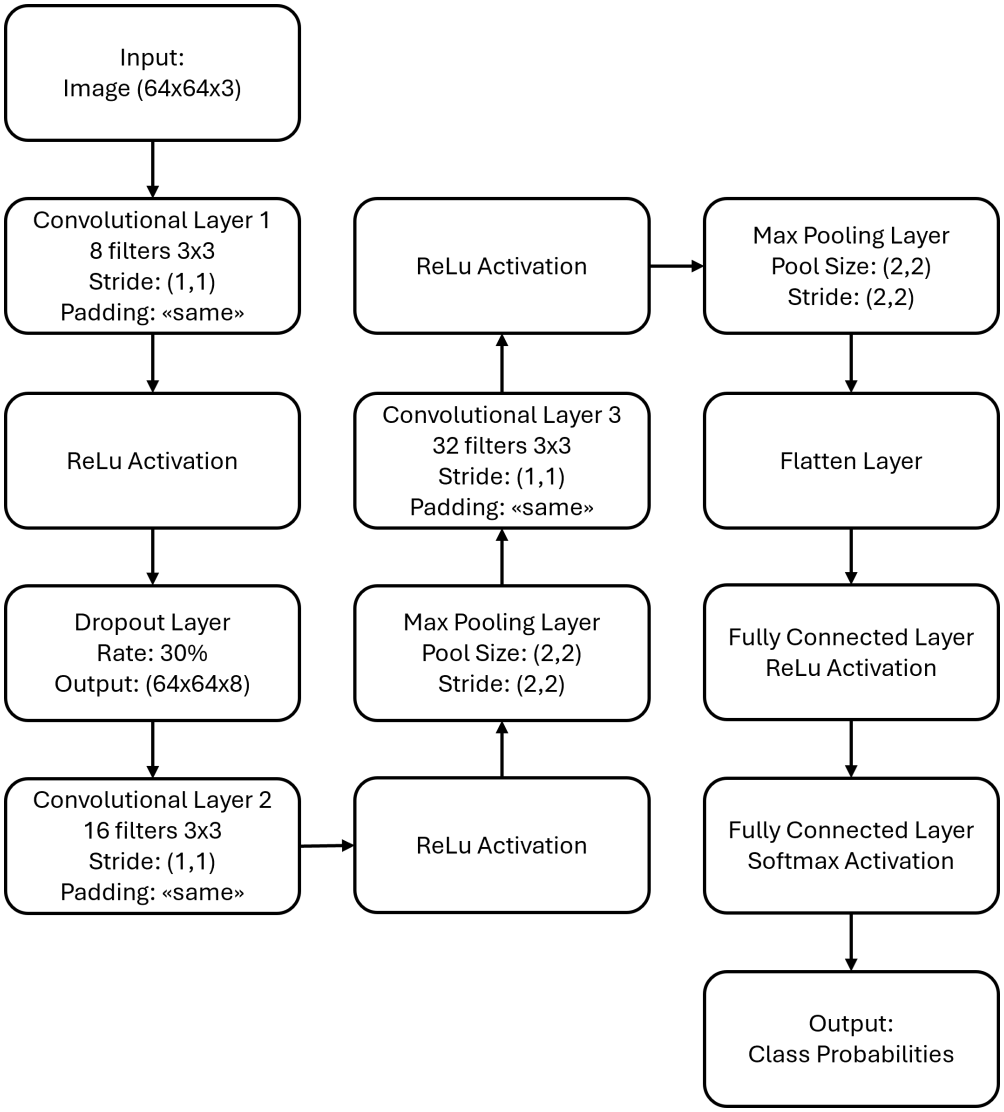

Figure 2: Neural Network Model.

4. Preliminary experimental results

In order to verify the performance of the neural network on the embedded hardware, we acquired the ultrasound video stream during an echocardiography performed on a volunteer.

Using the neural network, we classified the cardiac projection contained in each frame in real time. The output of the network was printed on a monitor, superimposed on the ultrasound image. The spatial information collected by inertial sensor was not used. The performance of the neural network is shown in Table 2, while the confusion matrix on test set is reported in figure 3.

Table 2
Neural Network Classification Report. The overall classification accuracy is of 85 %. The high performance in the classification of Unknown type images is probably due to the fact that they are very different from those belonging to the other classes.

	precision	recall	f1-score
2CH	0.86	0.79	0.82
4CH	0.81	0.88	0.85
Unknown	1.00	0.97	0.99
Accuracy			0.85
Macro Average	0.89	0.88	0.89
Weighted Average	0.85	0.85	0.85

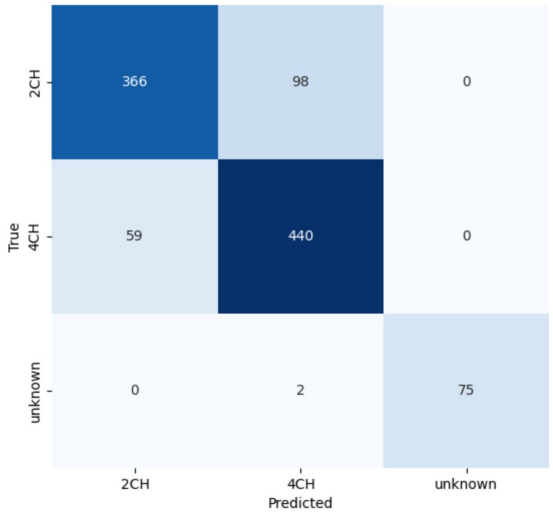

Figure 3: Confusion Matrix On Test Set. 1040 images were used.

During the test, the CPU and its temperature remained stable around 55.5 °C, and the RAM usage was approximately 70 %. The average inference time for a single frame was measured at 13.74 ± 2.48 ms, confirming the hardware's capability for the intended application. It should be noted that, untill now, it was not necessary to use the GPU to perform the tests.

5. Conclusions

The research is in its preliminary stages, but the initial findings establishes a foundation for developing a system capable of supporting more precise and reliable echocardiographic diagnostics. The next essential step involves the comprehensive collection of a dataset and collaboration with sonographers for accurate frame labelling. Strategies for incorporating accelerometer and gyroscope data will be

explored to enhance visual assistance for probe movements. Future efforts will focus on integrating software advancements with the chosen hardware to improve overall system performance.

References

[1] Z. Jiang, S. E. Salcudean, N. Navab, Robotic ultrasound imaging: State-of-the-art and future perspectives, Medical Image Analysis 89 (2023) 102878. URL: https://linkinghub.elsevier.com/retrieve/pii/S136184152300138X. doi:10.1016/j.media.2023.102878.

[2] R. Tenajas, D. Miraut, C. I. Illana, R. Alonso-Gonzalez, F. Arias-Valcayo, J. L. Herraiz, Recent Advances in Artificial Intelligence-Assisted Ultrasound Scanning, Applied Sciences 13 (2023) 3693. URL: https://www.mdpi.com/2076-3417/13/6/3693. doi:10.3390/app13063693.

[3] K. Mathiassen, J. E. Fjellin, K. Glette, P. K. Hol, O. J. Elle, An Ultrasound Robotic System Using the Commercial Robot UR5, Frontiers in Robotics and AI 3 (2016). URL: http://journal.frontiersin.org/Article/10.3389/frobt.2016.00001/abstract. doi:10.3389/frobt.2016.00001.

[4] B. Mathur, A. Topiwala, S. Schaffer, M. Kam, H. Saeidi, T. Fleiter, A. Krieger, A Semi-Autonomous Robotic System for Remote Trauma Assessment, in: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), IEEE, Athens, Greece, 2019, pp. 649–656. URL: https://ieeexplore.ieee.org/document/8941790/. doi:10.1109/BIBE.2019.00122.

[5] Q. Huang, J. Lan, X. Li, Robotic Arm Based Automatic Ultrasound Scanning for Three-Dimensional Imaging, IEEE Transactions on Industrial Informatics 15 (2019) 1173–1182. URL: https://ieeexplore.ieee.org/document/8472788/. doi:10.1109/TII.2018.2871864.

[6] A. Narang, R. Bae, H. Hong, Y. Thomas, S. Surette, C. Cadieu, A. Chaudhry, R. P. Martin, P. M. McCarthy, D. S. Rubenson, S. Goldstein, S. H. Little, R. M. Lang, N. J. Weissman, J. D. Thomas, Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use, JAMA Cardiology 6 (2021) 624. URL: https://jamanetwork.com/journals/jamacardiology/fullarticle/2776714. doi:10.1001/jamacardio.2021.0185.

[7] S. P. Baller, A. Jindal, M. Chadha, M. Gerndt, DeepEdgeBench: Benchmarking Deep Neural Networks on Edge Devices, in: 2021 IEEE International Conference on Cloud Engineering (IC2E), IEEE, San Francisco, CA, USA, 2021, pp. 20–30. URL: https://ieeexplore.ieee.org/document/9610432/. doi:10.1109/IC2E52221.2021.00016.

[8] J. Jo, S. Jeong, P. Kang, Benchmarking GPU-Accelerated Edge Devices, in: 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, Busan, Korea (South), 2020, pp. 117–120. URL: https://ieeexplore.ieee.org/document/9070647/. doi:10.1109/BigComp48618.2020.00-89.

[9] H. V. Pham, T. G. Tran, C. D. Le, A. D. Le, H. B. Vo, Benchmarking Jetson Edge Devices with an End-to-End Video-Based Anomaly Detection System, in: K. Arai (Ed.), Advances in Information and Communication, volume 920, Springer Nature Switzerland, Cham, 2024, pp. 358–374. URL: https://link.springer.com/10.1007/978-3-031-53963-3_25. doi:10.1007/978-3-031-53963-3_25, series Title: Lecture Notes in Networks and Systems.

[10] S. Leclerc, E. Smistad, J. Pedrosa, A. Ostvik, F. Cervenansky, F. Espinosa, T. Espeland, E. A. R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, J. Dhooge, L. Lovstakken, O. Bernard, Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography, IEEE Transactions on Medical Imaging 38 (2019) 2198–2210. URL: https://ieeexplore.ieee.org/document/8649738/. doi:10.1109/TMI.2019.2900516.

[11] A. F. Agarap, Deep Learning using Rectified Linear Units (ReLU), 2018. URL: https://arxiv.org/abs/1803.08375. doi:10.48550/ARXIV.1803.08375, version Number: 2.

[12] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 770–778. URL: http://ieeexplore.ieee.org/document/7780459/. doi:10.1109/CVPR.2016.90.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words:

Transformers for Image Recognition at Scale, 2020. URL: https://arxiv.org/abs/2010.11929. doi:10.48550/ARXIV.2010.11929, version Number: 2.

[14] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Computation 9 (1997) 1735–1780. URL: https://direct.mit.edu/neco/article/9/8/1735-1780/6109. doi:10.1162/neco.1997.9.8.1735.