Neural Architecture for Tennis Shot Classification on Embedded System

Ali Dabbous[1[0009-0004-8978-4979]], Matteo Fresta [1[0009-0000-7265-7501]], Francesco Bellotti [1[0000-0003-4109-4675]], and Riccardo Berta[1[0000-0003-1937-3969]]

[1] Department of Electrical, Electronic and Telecommunication Engineering (DITEN)
University of Genoa, Via Opera Pia 11a, 16145 Genova, Italy.
{ali.dabbous, matteo.fresta}@edu.unige.it,
{riccardo.berta, francesco.bellotti}@unige.it

Abstract. Data analysis has become a common practice in professional and amateur sport activities, to monitor the player state and enhance performance. In tennis, performance analysis requires detecting and recognizing the different types of shots. With the advances in microcontrollers and machine learning algorithms, this topic becomes ever more considerable. We propose a 1-D convolutional neural network (CNN) model and an embedded system based on Arduino-Nano system for real-time shot classification. The network is trained through a dataset composed of three different tennis shot types, with 6 features recorded by an inertial device placed on the racket. Results demonstrate that the proposed model is able to discriminate the tennis shots with high accuracy, also generalizing to different users. The network has been deployed on a low-cost Arduino nano 33 IoT model, with an inference time of 65ms.

Keywords: 1-D CNN, Arduino Nano, Embedded device, Tennis shot.

1 Introduction

Thanks to developments in microelectronics, wearable technology [1]–[3] has attracted a lot of attention for research and commercial applications, also in the sport field, where field data collection and processing is essential for identifying flaws and enhance skills. In tennis, this requires detecting the main shots, such as forehand, backhand and serve.

This paper focuses on the classification of various tennis shots. A lot of research work has been carried out in the field of tennis sport activity detection. Several studies have used various machine learning methods to tackle this classification problem [4]–[7]. For instance, Büthe et al. [8] created a pipeline to detect and categorize leg and arm movement, as well as execute gesture recognition for the firing arm using the longest common subsequence. In [9], Hazem and Al-Sadek employed an artificial neural network (ANN) to determine stroke type and forecast how accurate the player will play in subsequent plays based on statistics.

In all these works, data is recorded in the field and processed in the cloud or on the desktop. However, it would be important to process information online by using a low-cost, low-power embedded device capable of handling AI models, in order to

have real-time feedback to the athlete. Here the literature has much less examples, because the processing is not lightweight, while edge devices have strong limitations in terms of computational capabilities and memory availability.
The main goal of this paper is to propose a new model of neural architecture for real-time classification of tennis shots on a micro-controller directly attachable to a tennis racket. The system processes six inertial signal timeseries collected from an IMU sensor that represents the three axes of accelerometers and gyroscopes.
In literature, timeseries are frequently processed through feature engineering in order to provide suitable input to a machine learning model [10]. Deep-learning techniques, on the other hand, tend to extract features directly from the raw data (e.g., image pixels); however, this capability comes at the cost of significantly increase the number of model parameters, increasing memory needs, computing burden, and energy consumption [11]. This increase is critical for low-power edge devices. Therefore, we targeted the development of a deep learning architecture able to fulfill the strict hardware and cost limitations of edge devices.
Particularly, we propose a deep network featuring 1-D convolutional layers to classify tennis shots. 1-D convolutional neural networks (CNNs) can process 1-D time series. 1-D CNN can handle multi-dimensional input tensors, as in our case, by applying the convolution to each single dimension, independent of the other, producing a feature map for each channel. 1-D convolutions just involve scalar multiplications and additions, and have already proved to be particularly effective also in complex classification tasks (e.g., touch modalities [10], [12]). Thus, we consider them particularly promising for real-time classification of tennis shots.
The contributions of this paper are summarized as follow:

- Adopting 1-D CNNs to classify different sport tennis shots with high accuracy.
- Deploying the 1-D CNN classifier on a low-cost Arduino nano IOT micro-controller.
- Analyzing overall performance of the embedded system.

2 Experimental Setup

2.1 Dataset

The dataset employed in this work is made up of 6-axis time-series maps for three distinct tennis shots: forehand, backhand, and serve. To capture these shots, we asked a amateur tennis player to make several repetitions of the three types of shots, using his racket, to which an Arduino Nano IoT [13] device was connected. During each trial, the time-series data from the IMU (LSM6DS3) sensor was recorded. This time-series has six axes: x, y, and z accelerometers and x, y, and z gyroscopes. The total number of samples collected for all shots is 213 divided into 80% for training and 20% for validation. Additionally, 45 separate time-series samples were recorded for the three shots for testing purposes.

2.2 1-D Classifier Network Architecture

Figure 1 shows the architecture of the proposed neural network, which consists of two 1-D convolutional layers, interleaved by two max pooling layers, followed by a dense layer, and a three neuron softmax output layer. The convolutional layers aim at extracting the classification features, while the max pool layers, with non-overlapping receptive field, aim at reducing dimensionality. We discuss the actual hyperparameter values (e.g., number of neurons and of filters) later in the next section describing the training, which is where the actual values were set.

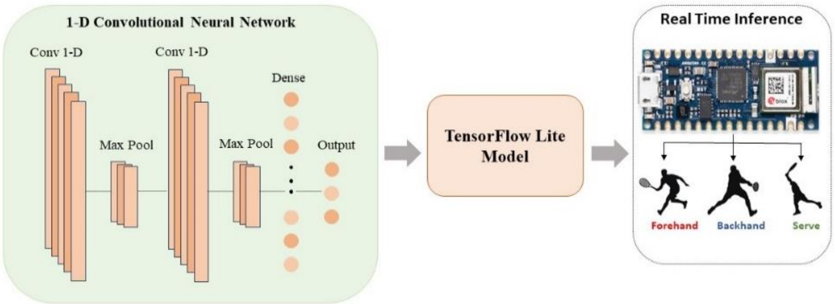

Fig.1. Tennis shot architecture. (Left) 1-D convolutional neural network. (Right) Arduino Nano 33 IoT for model deployment and real time classification.

2.3 Training Strategy

The first main challenge for the training consisted in maintaining consistency in the input shape of the 1-D CNN model, given the different lengths of the shot sequences. To this end, we employed Dynamic Time Warping (DTW) [14] techniques on the samples of each shot. This allowed us to clean up the dataset by identifying and excluding anomalous or poorly created samples. After applying DTW, we calculated the maximum number of time steps across all the samples, which was 20 in our case. For samples with fewer than 20 time-steps, we used padding techniques to maintain homogeneity. This entailed adding more time steps to ensure that all samples had the proper length. As a result, our 1-D CNN model's input shape was specified as (20, 6), where 6 denoted the number of features recorded by each time step.

The actual training procedure then starts with fine-tuning the classifier architecture's hyperparameters, through cross-validation on the training set. A grid of potential candidates was investigated, with a focus on solutions with few parameters and high computing efficiency. The process finally selected a network architecture with two 1-D convolutional layers. The first layer has 8 filters, whereas the second 16. The kernel size in both layers is 3. The Dense layer comprised 32 units. The Rectified Linear Unit (ReLU) activation function is applied to both the convolutional and dense layers. Nevertheless, the total number of trainable parameters was equal to 2,219.

The training strategy's efficacy was assessed by measuring the model's accuracy on the testing set, after 50 epochs.

2.4 Embedded System for Real Time Classification

We deployed the 1-D CNN on an Arduino Nano 33 IoT, with the goal to classify shots in real-time. The Arduino Nano 33 IoT is a cost-effective device that can be easily installed into any tennis racket. It is powered by a SAMD21 Cortex®-M0+ 32-bit low power ARM MCU with a 48MHz CPU. The device includes 256KB of flash memory and 32kB of SRAM, and it operates at a voltage of 3.3V.

We were able to export the trained model in the float32 format suitable with the Arduino Nano 33 IoT after finishing the training of the 1-D CNN using Python and the TensorFlow library. The model was converted to the TFLite format, which is suitable for deployment on low-resource devices, including the Arduino Nano microcontroller. The embedded memory footprint of the proposed neural network is of 14KB.

3 Experimental Results

To evaluate the model's performance, we utilized a test set in which each one of the 45 samples represents an individual shot: 15 forehand strokes (red colour in Fig. 2), 14 backhand (blue), and 16 serves (green). The trained model could correctly classify each shot in the test set.

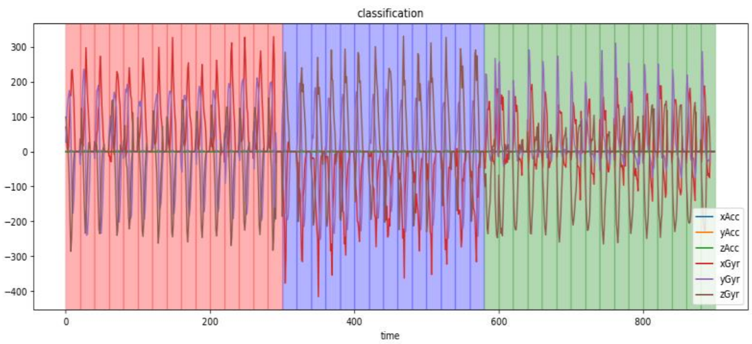

Fig.2. Tennis shots classification results. Red represents the Forehand shot, blue represents the Backhand shot and green represents the serve shot.

As a field-validation step, the model deployed on the Arduino Nano 33 IoT was evaluated with three distinct players, each of whom repeated three separate shots 15 trials. The complete system also included a threshold approach on the accelerometer measurements to assess whether there was movement and thus compute the classification.

Also in this case, the embedded model could detect and classify all the shots, as shown in table 1. The inference time was 65ms, which is suited for real-time performance. These findings show that the proposed network is effective in terms of accuracy, and, advancing the state-of-the-art work [8], [9], has been deployed on an embedded device, executing real time inference.

Table 1. Real time tennis shots classification.

	Forehand	Backhand	Serve	Inference Time
Forehand	45	0	0	**65 ms**
Backhand	0	44	1	
Serve	0	1	44	

4 Conclusion

This paper presented the development of a 1-D CNN to distinguish between different types of tennis shots. The network is capable of discriminating among three types tennis shots, with high accuracy. Advancing the state of art (e.g., [8], [9]) the trained model, converted into a micro model using TFlite Libraries, works in real-time on an embedded device in the field. The memory footprint of the micro model is of 14KB, which is acceptable for the vast majority of current microcontrollers. The inference time for detecting and classifying one shot after applying it by the athletes' in real time was equal to 65ms, which is suited for real time performance, which can be important for a athlete's training. This preliminary work calls for further studies on time series analysis for sport activities. We will keep on resorting computational primitives that have a corresponding hardware implementation using low-power microelectronic devices, with the aim of designing compact and efficient end-to-end systems that can analyze sports activity signals in the field.

References

[1] F. Sakr, F. Bellotti, R. Berta, and A. De Gloria, "Machine Learning on Mainstream Microcontrollers," *Sensors*, vol. 20, no. 9, Art. no. 9, Jan. 2020, doi: 10.3390/s20092638.

[2] S. Coyle, D. Morris, K.-T. Lau, D. Diamond, and N. Moyna, "Textile-Based Wearable Sensors for Assisting Sports Performance," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, Jun. 2009, pp. 307–311. doi: 10.1109/BSN.2009.57.

[3] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. NeuroEngineering Rehabil.*, vol. 9, no. 1, p. 21, Apr. 2012, doi: 10.1186/1743-0003-9-21.

[4] M. Petkovic, W. Jonker, and Z. Zivkovic, "Recognizing Strokes in Tennis Videos Using Hidden Markov Models".

[5] P. Kelly and N. E. O'Connor, "Visualisation of tennis swings for coaching," in *2012 13th International Workshop on Image Analysis for Multimedia Interactive Services*, May 2012, pp. 1–4. doi: 10.1109/WIAMIS.2012.6226750.

[6] D. Connaghan, C. Ó Conaire, P. Kelly, and N. E. O'Connor, "Recognition of tennis strokes using key postures," in *IET Irish Signals and Systems Conference (ISSC 2010)*, Jun. 2010, pp. 245–248. doi: 10.1049/cp.2010.0520.

[7] H. Shah, P. Chokalingam, B. Paluri, N. Pradeep, and B. Raman, "Automated Stroke Classification in Tennis," in *Image Analysis and Recognition*, M. Kamel and A. Campilho, Eds., in Lecture Notes in Computer Science, vol. 4633. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1128–1137. doi: 10.1007/978-3-540-74260-9_100.

[8] L. Büthe, U. Blanke, H. Capkevics, and G. Tröster, "A wearable sensing system for timing analysis in tennis," in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Jun. 2016, pp. 43–48. doi: 10.1109/BSN.2016.7516230.

[9] O. Hazem and A. F. Al-Sadek, "Detection of Tennis Strokes using Wearable Sensor," in *2022 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Sep. 2022, pp. 1–6. doi: 10.23919/SoftCOM55329.2022.9911405.

[10] C. Gianoglio, E. Ragusa, R. Zunino, and M. Valle, "1-D Convolutional Neural Networks for Touch Modalities Classification," in *2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, Nov. 2021, pp. 1–6. doi: 10.1109/ICECS53924.2021.9665576.

[11] E. Ragusa, C. Gianoglio, R. Zunino, and P. Gastaldo, "A Design Strategy for the Efficient Implementation of Random Basis Neural Networks on Resource-Constrained Devices," *Neural Process. Lett.*, vol. 51, no. 2, pp. 1611–1629, Apr. 2020, doi: 10.1007/s11063-019-10165-y.

[12] F. Sakr, H. Younes, J. Doyle, F. Bellotti, A. De Gloria, and R. Berta, "A Tiny CNN for Embedded Electronic Skin Systems," in *Advances in System-Integrated Intelligence: Proceedings of the 6th International Conference on System-Integrated Intelligence (SysInt 2022), September 7-9, 2022, Genova, Italy*, Springer, 2022, pp. 564–573.

[13] "Nano 33 IoT | Arduino Documentation." https://docs.arduino.cc/hardware/nano-33-iot (accessed Jun. 19, 2023).

[14] R. Ma, A. Ahmadzadeh, S. F. Boubrahimi, and R. A. Angryk, "A Scalable Segmented Dynamic Time Warping for Time Series Classification," in *Artificial Intelligence and Soft Computing*, L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, and J. M. Zurada, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 407–419. doi: 10.1007/978-3-030-20915-5_37.