Chapter 5:   Machine learning partners

After completing this chapter, you will understand which partners support Infineon Machine Learning and what capabilities they provide. Details on each partner will be provided in the following chapters.

Table of contents

Document conventions

Convention	Usage	Example
Courier New	Displays code and text commands	`CY_ISR_PROTO(MyISR);` `make build`
Italics	Displays file names and paths	*sourcefile.hex*
[**bracketed, bold**]	Displays keyboard commands in procedures	[**Enter**] or [**Ctrl**] [**C**]
Menu > Selection	Represents menu paths	**File > New Project > Clone**
Bold	Displays GUI commands, menu paths and selections, and icon names in procedures	Click the **Debugger** icon, and then click **Next**.

5.1 Why use a machine learning partner?

As you learned in previous chapters, you can use Keras and TensorFlow to create and train models for your machine learning application or you can use ModusToolbox™ tools. However, in some cases it may be advantageous to use a partner. The three basic approaches to get a machine learning model into ModusToolbox™ for your application are:

- Bring your own model.
- Train your own model.
- Buy your own model.

The best approach will depend on your needs and the needs of your application.

5.1.1 Bring your own model

With this workflow you collect data using your own methods. Next, you use Keras and TensorFlow tools to create and train your model (possibly based on an example model). Then you bring that trained model into the ModusToolbox™ ML Configurator to optimize it for embedded devices and to create the model used by your MCU.

Note: *While this is the most labor-intensive approach, it is entirely in your control.*

5.1.2 Train your own model

In this workflow, you use Imagimob or work with a partner that has tools to help you collect and classify data and tools to help you create and train a model. In some cases, the model creation and training is completely automated (called AutoML) while in other cases, the model creation and training are still user-guided.

The model that you receive will typically already be optimized for embedded – there's no need to use the ModusToolbox™ ML tools in this case. It will either be a pre-compiled C library or source code that you include in your application. The partner will provide an API to interact with their model.
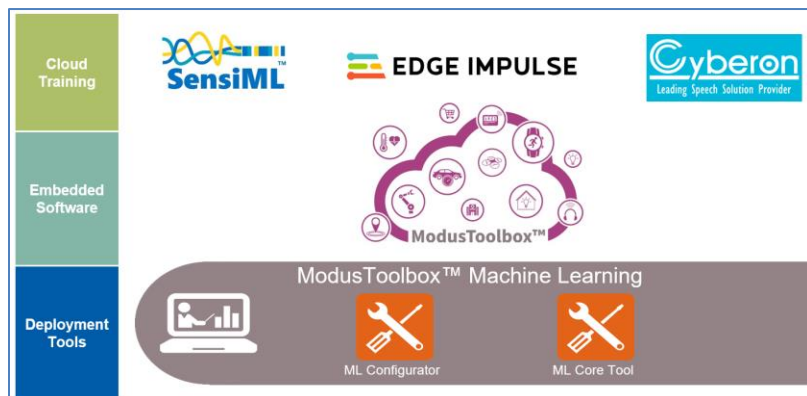
5.1.3 Buy your own model

In this workflow, a partner can provide a completely trained and optimized model that is customized for your needs. This is available for certain well-defined problems. For example, if you want to use machine learning for wake-word detection or even for voice recognition for a small set of words and phrases, you may be able to buy a model that is already customized to the set of words and phrases that you care about. This entirely eliminates the need to collect data – you just provide the list of words or phrases and the partner's ecosystem will do the rest!

As with the "train your own model" approach, you will receive a model that is optimized for embedded. It will either be a pre-compiled C library or source code that you include in your application along with an API to allow you to use it.

5.2   Infineon machine learning partners

Each Infineon machine learning partner provides different capabilities and workflows. Several of our partners are briefly introduced here. Later chapters will go into further detail on each partner.

5.2.1   Cyberon

Cyberon falls under the "buy your own model" category. This partner provides fully trained models for single stage or dual stage keyword detection. Dual stage means that there is an initial trigger keyword followed by additional more specific command keywords. The set of keywords that are recognized is customizable using tools that they provide. Once the keywords are selected, a fully trained model is produced that can be integrated directly into the user's application using a provided API.

5.2.2   SensiML

SensiML is under the category of "train your own model." They provide tools for collecting and labeling datasets and tools for creating and training a model. They allow data collection to be done using a Windows PC or an Android phone. The output is a trained model that is optimized for whichever MCU that you select.

The data collection and labelling tool (Data Capture Lab) allows you to collect data directly from the various sensors sampled by the target MCU. It has features that help automatically label samples and create robust datasets.

The model creation and training tool (Analytics Studio) allows various levels of user interaction including fully automatic, some manual tuning, and full manual control of the pipeline.

Once a model has been completed, you can do detailed analysis of the model's performance both within the tools and running directly on the target MCU. New data collected during testing can be fed back into the model creation flow to help improve your model's accuracy.

5.2.3   Edge Impulse

Edge Impulse is also a "train your own model" partner. Like SensiML, this partner provides methods for collecting and labeling datasets, creating, and training a model, and assists with deploying the model to a target MCU. The main tool is called Edge Impulse Studio. It helps with every step of the process; from data collection through deployment of a completed model to a target MCU.

Edge Impulse Studio runs within a web browser, however, there is also a command line interface (CLI) that runs on Windows, macOS or Linux.

Trademarks
All referenced product or service names and trademarks are the property of their respective owners.