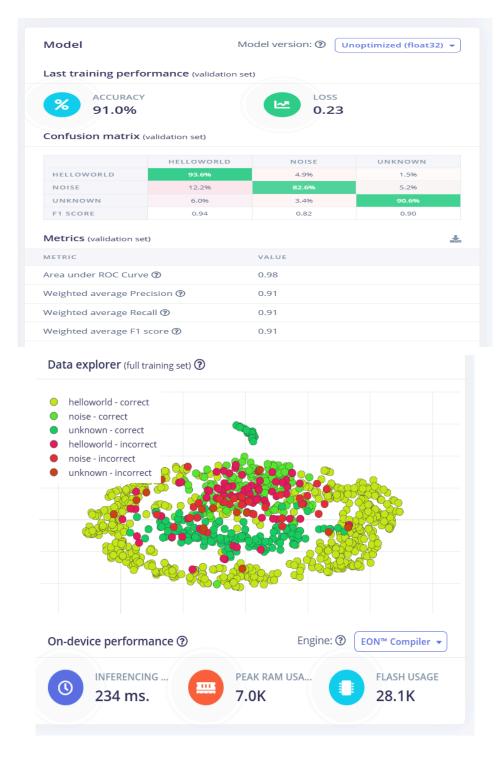Questions
1. Does the model perform as accurately as expected on your smartphone? List a few methods to improve the model's accuracy.

   No. The model did not perform that well on smartphone. There were many misclassifications. The drop in confidence level might be due to the fact that mobile phones don't have the necessary resources dedicated to run the neural network model. The model's accuracy can be improved by making use of various optimizations and techniques like compression, pruning, normalizations etc.

2. When building a model for resource-limited hardware, how do you balance fast inference times with acceptable model accuracy? What trade-offs did you encounter?

   For resource limited hardware balancing of inference time and model accuracy has to be done based on the application. Certain everyday applications like home automation can afford to have comparatively lesser accuracy and slower inference time compared to the ones deployed in ADAS or biomedical application. Further model pruning, quantization, architecture optimization, dimensionality reduction can be employed when building a model. Trade offs associated with model pruning and quantization is that it reduces the accuracy. Reducing dimensionality causes loss in feature and in turn causes accuracy reduction.

3. Include screenshots of the training performance from step 6 of the deployment process.

Model

Model version: ⓘ   Unoptimized (float32) ▾

Last training performance (validation set)

% ACCURACY **91.0%**   **📈** LOSS **0.23**

Confusion matrix (validation set)

	HELLOWORLD	NOISE	UNKNOWN
HELLOWORLD	93.6%	4.9%	1.5%
NOISE	12.2%	82.6%	5.2%
UNKNOWN	6.0%	3.4%	90.6%
F1 SCORE	0.94	0.82	0.90

Metrics (validation set)

METRIC	VALUE
Area under ROC Curve ⓘ	0.98
Weighted average Precision ⓘ	0.91
Weighted average Recall ⓘ	0.91
Weighted average F1 score ⓘ	0.91

Data explorer (full training set) ⓘ

- ● helloworld - correct
- ● noise - correct
- ● unknown - correct
- ● helloworld - incorrect
- ● noise - incorrect
- ● unknown - incorrect

On-device performance ⓘ   Engine: ⓘ   EON™ Compiler ▾

🕐 INFERENCING ... **234 ms.**   🟧 PEAK RAM USA... **7.0K**   🟦 FLASH USAGE **28.1K**

Videos

4. Record and provide links to:
   - The keyword-spotting model working on your smartphone.
     -
     - https://drive.google.com/file/d/11jMq3gn5Q8Gf3GU60SXFrVd3dzHO1_Dg/view?usp=drive_link

- ○
  - ○ The keyword-spotting model working on the embedded Arduino board.
  - ○
  - ○ https://drive.google.com/file/d/11gebK_4U8Qn_Wwl--e1qudg7_J04ppSm/view?usp=drive_link
  - ○

Reflections:

- ○ Share your experience deploying the model to your smartphone and Arduino board. Mention any technical difficulties or interesting observations
- ○
- ○ When deploying the model in smartphone the confidence of classification was rather low than when compared with arduino board. Also in phone there where cases where "hello" or "world" or any word with "o" were classified as hello world.