universität freiburg

Intelligent
Embedded
Systems Lab

Bowel Sound Pattern Spotting Fine Tuning using Active Learning

Master Project Presentation

M. Sc. Computer Science

Giacomo Mossio
1st July 2024

Background

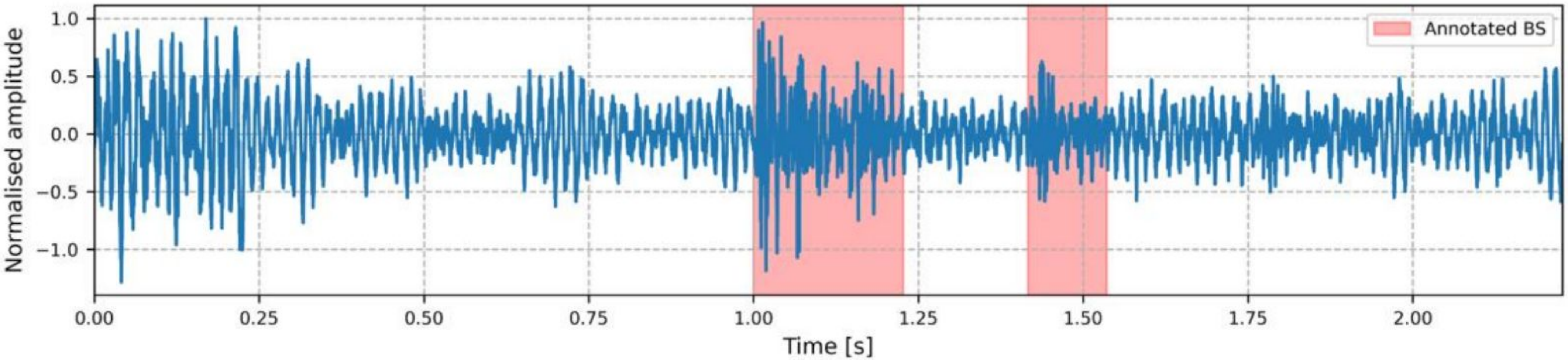
Bowel Sounds (BS) Spotting

- **Digestive disorders** evaluations are lengthy and usually invasive.
- **Auscultation** is a non-invasive option for monitoring gut health.
- **GastroDigitalShirt** [1] enables the recording of bowel sounds.
- **Deep Neural Networks** identify the onset and offset of BS patterns. [2]

GastroDigitalShirt [1]

Introduction to the Project

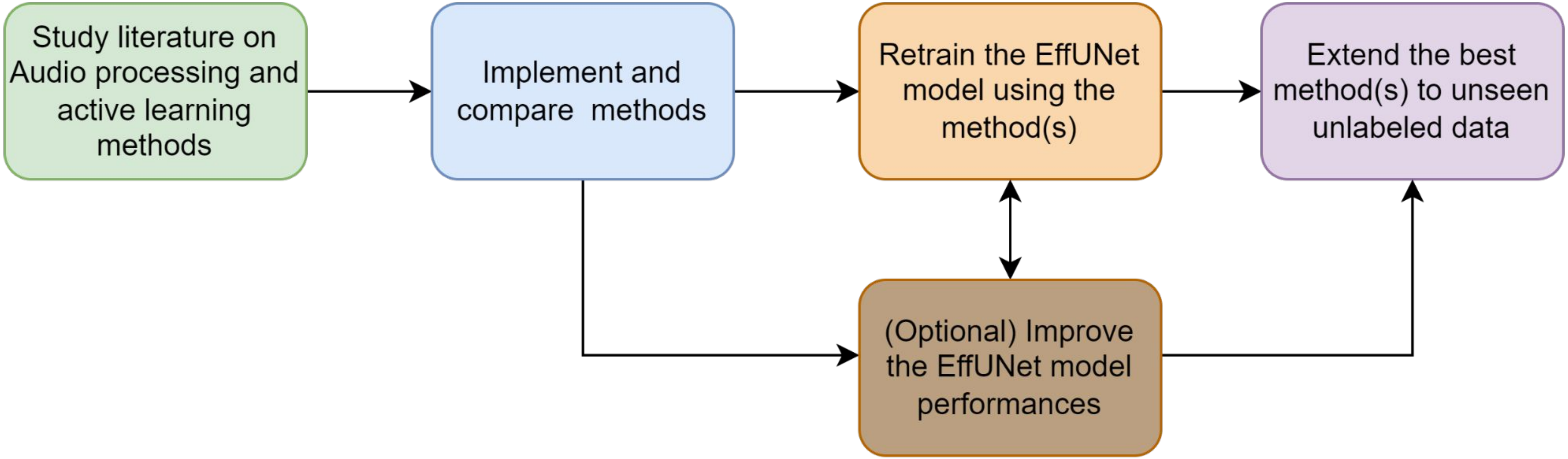
- Deep Neural Networks require **large amount of labeled data**
- 136 hours of audio data were labelled with 11482 BS events [2]
- On average, 1 hour of audio recording required **8-12 hours** for BS labelling [2]
- Various models tested and **EfficientUnet** [3] pretrained with transfer learning yielded the best results



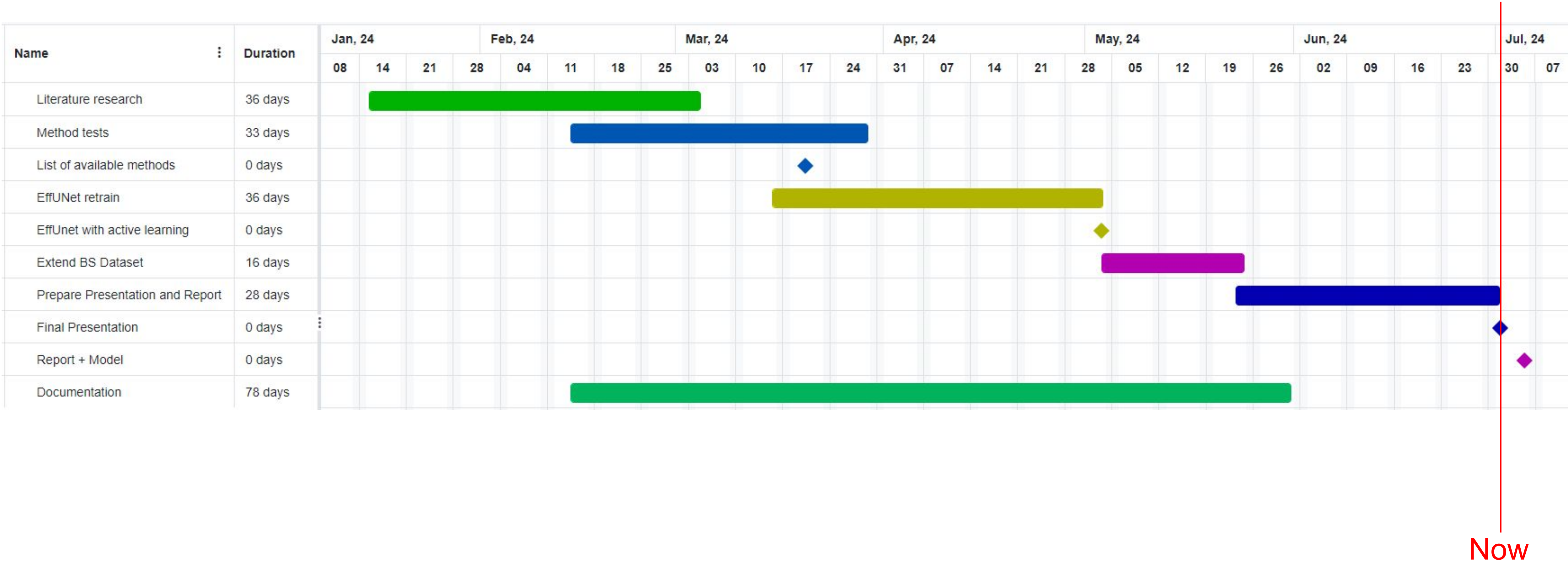
Representation of two expert-annotated BS events [2]

Project Goal and Tasks

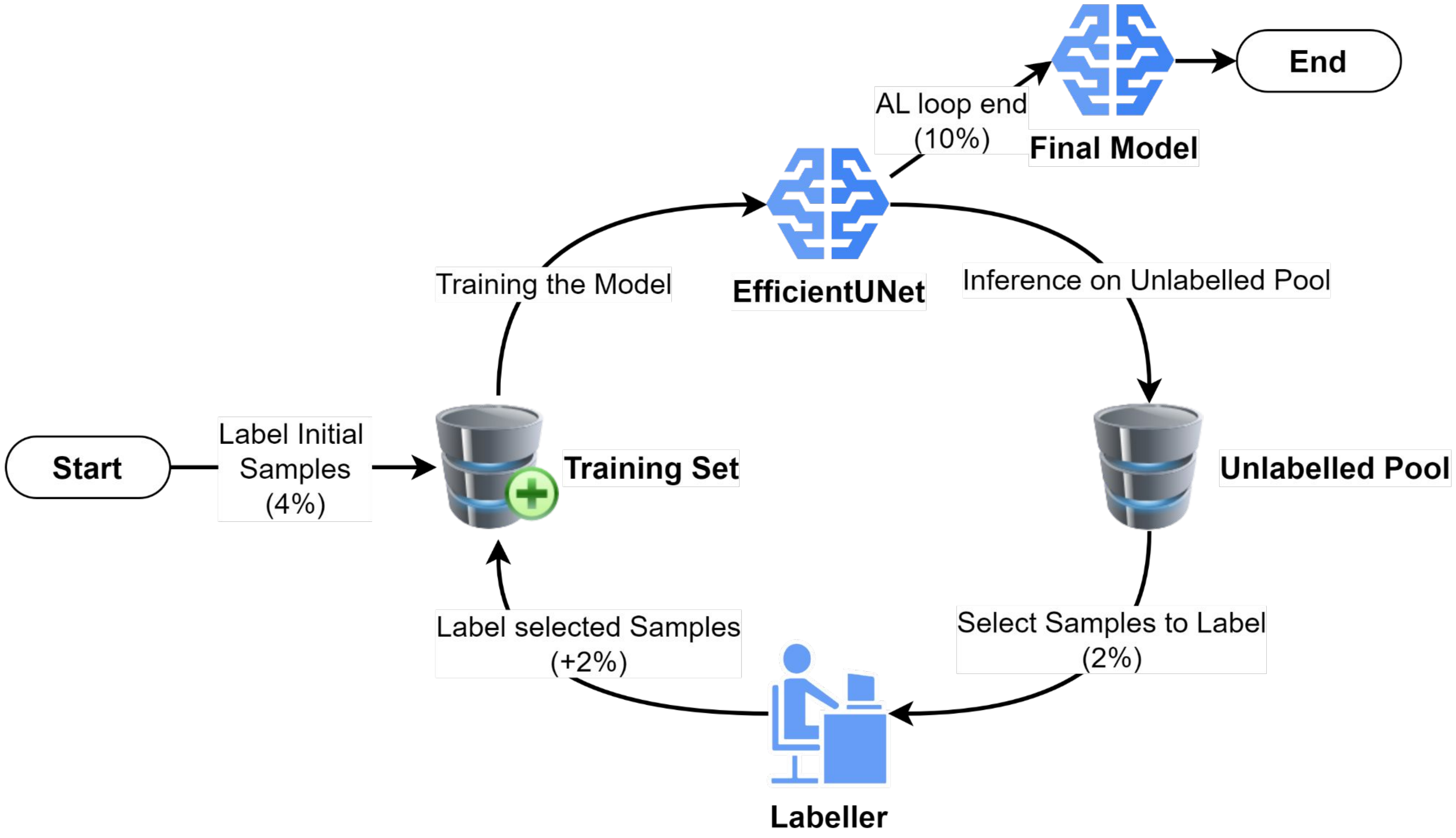
Goal of the project: **Achieve good performances with less labelled data**



Project Plan



Active Learning (AL) Process



(X%): Percentage of Labelled Data

Training Set: The labeled data used for training the model.

EfficientUNet: The deep learning model being trained.

Unlabelled Pool: The pool of unlabeled data from which new samples are selected.

Labeller: The process or person responsible for labeling the selected samples.

Final Model: The model after several iterations of active learning.

Active Learning Details

- **Samples selection process**:
  1. The current model **predicts** on the unlabeled dataset.
  2. The single predictions are a measure of **uncertainty.**
  3. **Heuristic** calculates the **uncertainty** for each sample. [4]
  4. Select **top samples** to label.

- **Initial Sampling**: Can be random or heuristic-based (only if using pretrained models).

Implementation

Built on Top of Existing GastroDigitalShirt Code

- **Language & Framework**: Python with Pytorch Lightning. [5]

Active Learning Integration

- **BAAL Library**: Leveraged Bayesian Active Learning techniques. [6]

Modifications:

- **Data Loader**: Customized to handle active learning queries.
- **Trainer**: Adapted to use BAAL's active learning cycle.
- **Model**: Kept EfficientUnet unchanged for consistency.
- **Pipeline:** Adapted to include Active Learning

Dataset and Model

Dataset:

- **Participants**: 27 (18 healthy and 9 patients with Inflammatory Bowel Disease).
- **Duration**: 136 hours of audio data.
- **Segmentation**: Split into non-overlapping 10-second segments.
- **Transformations**: Converted to Mel Spectrograms with a temporal resolution of 10 ms.
- **Data Imbalance**: BS ratio of 0.89%.

EfficientUnet Model:

- **Architecture**: Combines U-Net with EfficientNet as the encoder.
- **Pretraining**: Initially trained on AudioSet. [7]
- **Input**: Log Mel-spectrograms of shape 128 × 1056.
- **Output**: Binary detection mask indicating Bowel Sounds (BS) or Non-Bowel Sounds (NBS).

Overview of Experiments

- **Active Learning (AL) Parameters**:
  - **AL Start Percentage**: 2%, 4%, 10%, 20%.
  - **AL Step Percentage**: 1%, 2%, 3%, 5%, 10%.
  - **AL Number of Steps**: 3, 4, 5, 8, 10
  - **Heuristic**: Random, Entropy, Margin, BALD.
  - **Training Epochs**: 3, 5, 10, 20

- **Training Strategies:**
  - **From Scratch**: Retrained the model from scratch at each iteration to ensure fresh learning without any bias.
  - **Checkpoints**: Resumed training from the last checkpoint to save time and leverage previously learned features.

- **Model Hyperparameters**: Learning Rate, Loss, Optimizer, Warm-up.

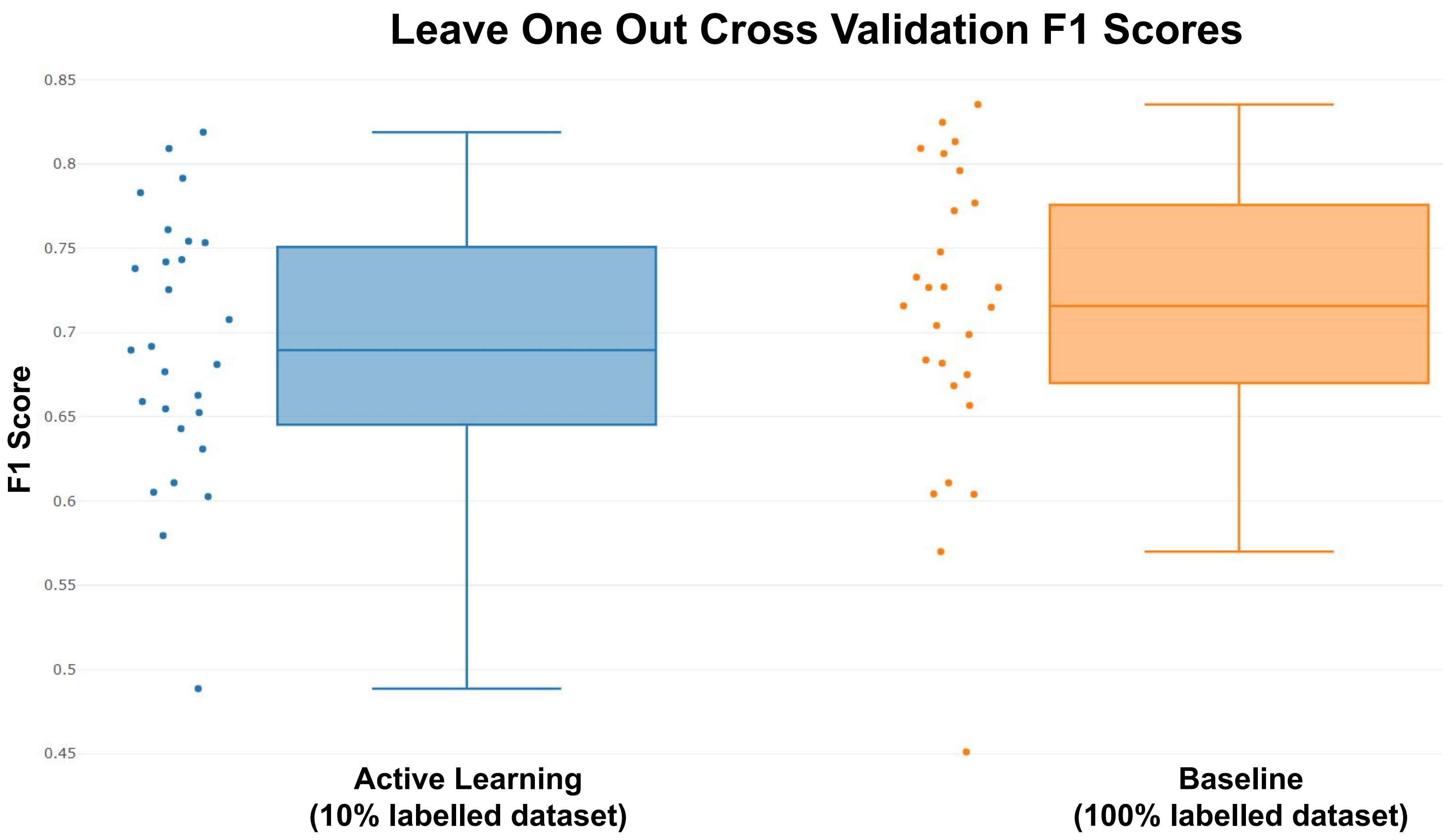
Best Result

Best trade-off configuration found:

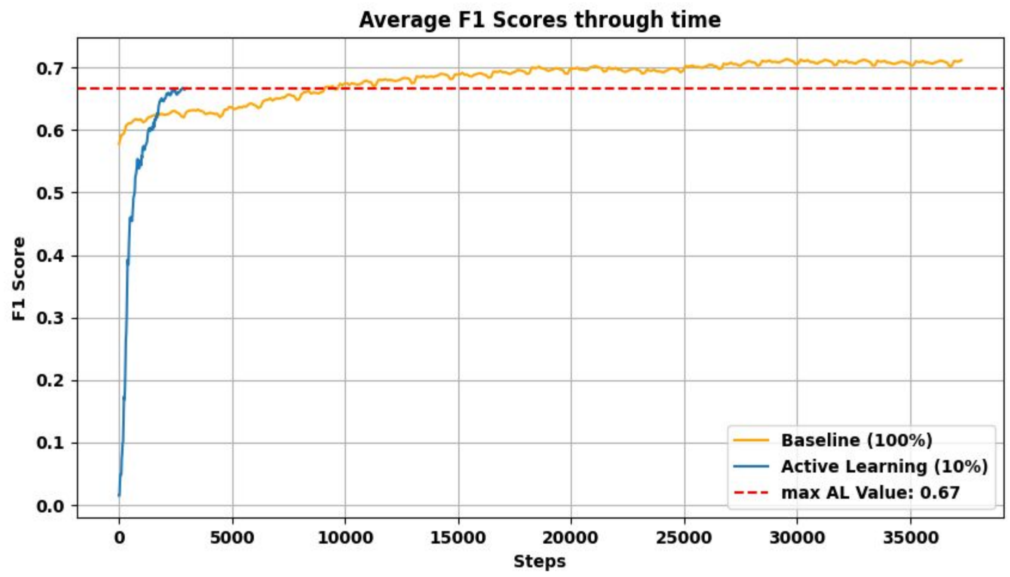
- AL Start Percentage: **4%**.
- AL Step Percentage: **2%**.
- AL Steps: **4**
- **Final Percentage: 10%**
- Heuristic: **Entropy**.
- Strategy: **From Scratch**
- Training Epochs: **10**
- Final Epochs: **25**

EfficientUnet Model (unchanged):

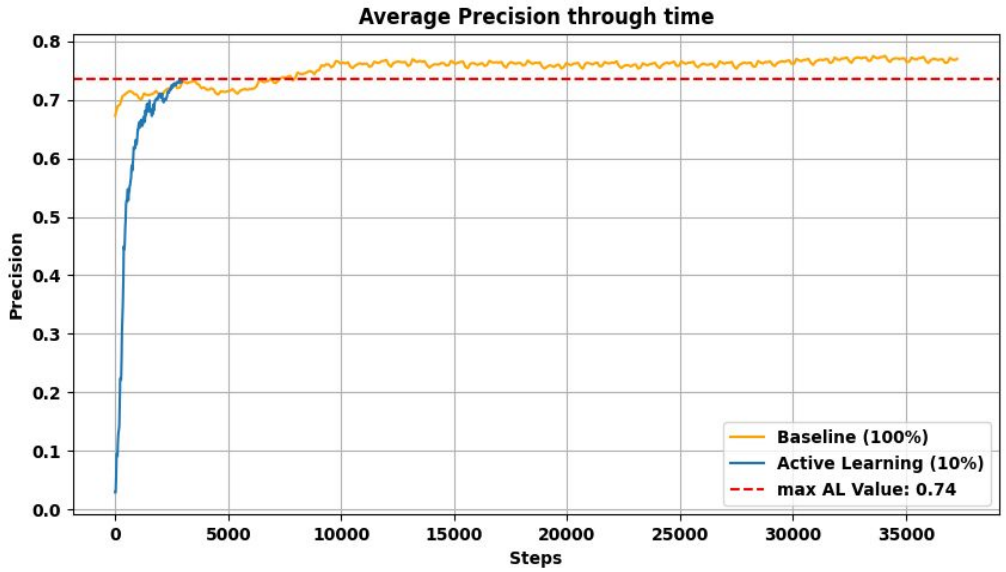
- LR: 0.0001
- Batch size: 32
- Optimizer: ADAM
- Loss: DCEL



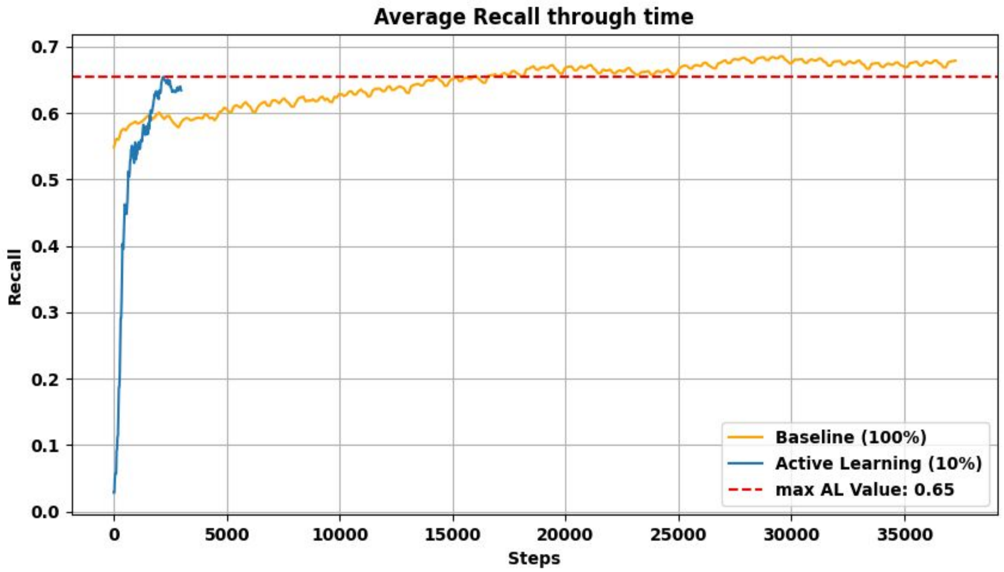
Results



Average F1 Scores through time
F1 Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
5000
10000
15000
20000
25000
30000
35000
Steps
Baseline (100%)
Active Learning (10%)
max AL Value: 0.67

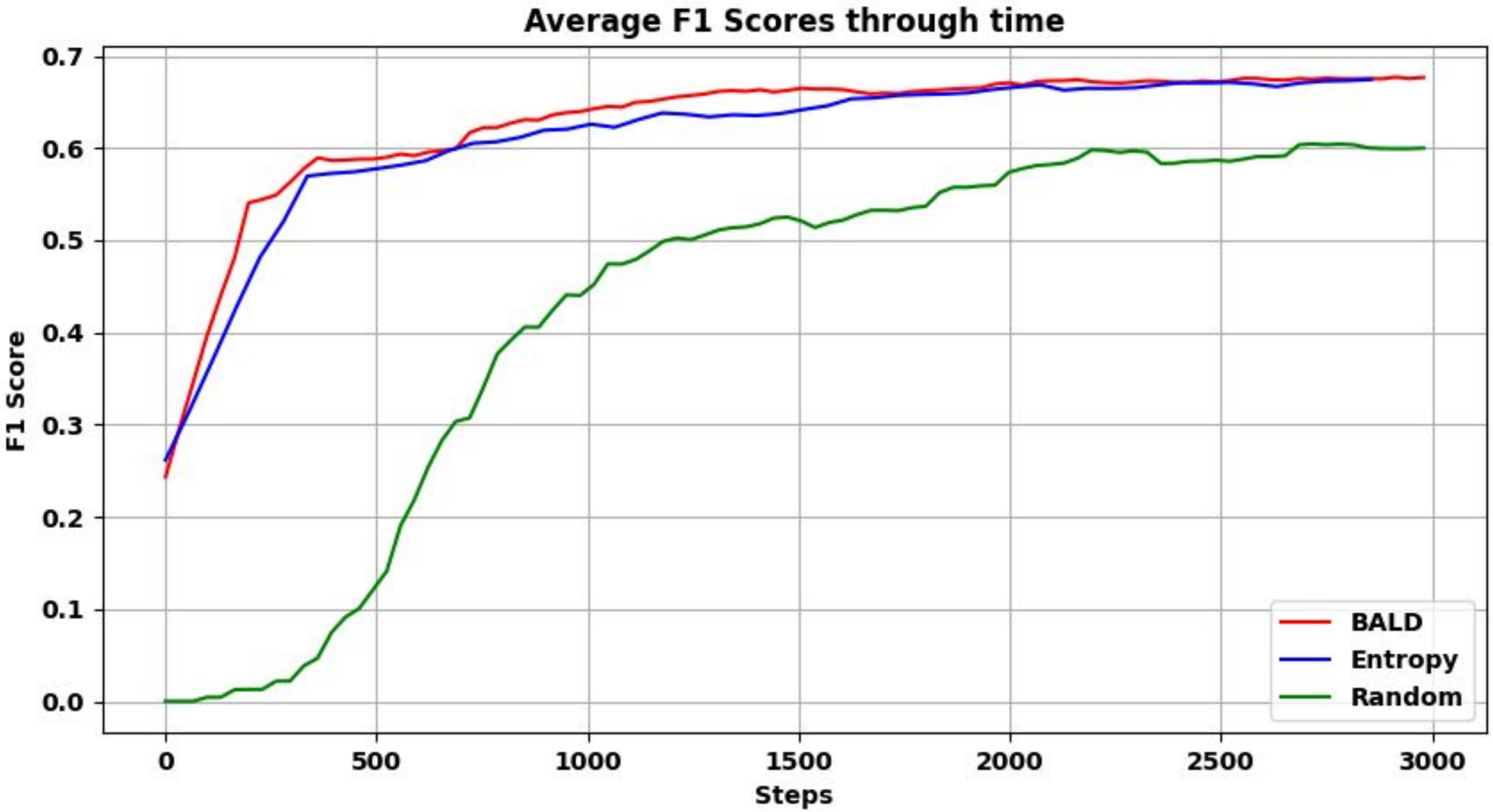


Average Precision through time
Precision
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
5000
10000
15000
20000
25000
30000
35000
Steps
Baseline (100%)
Active Learning (10%)
max AL Value: 0.74

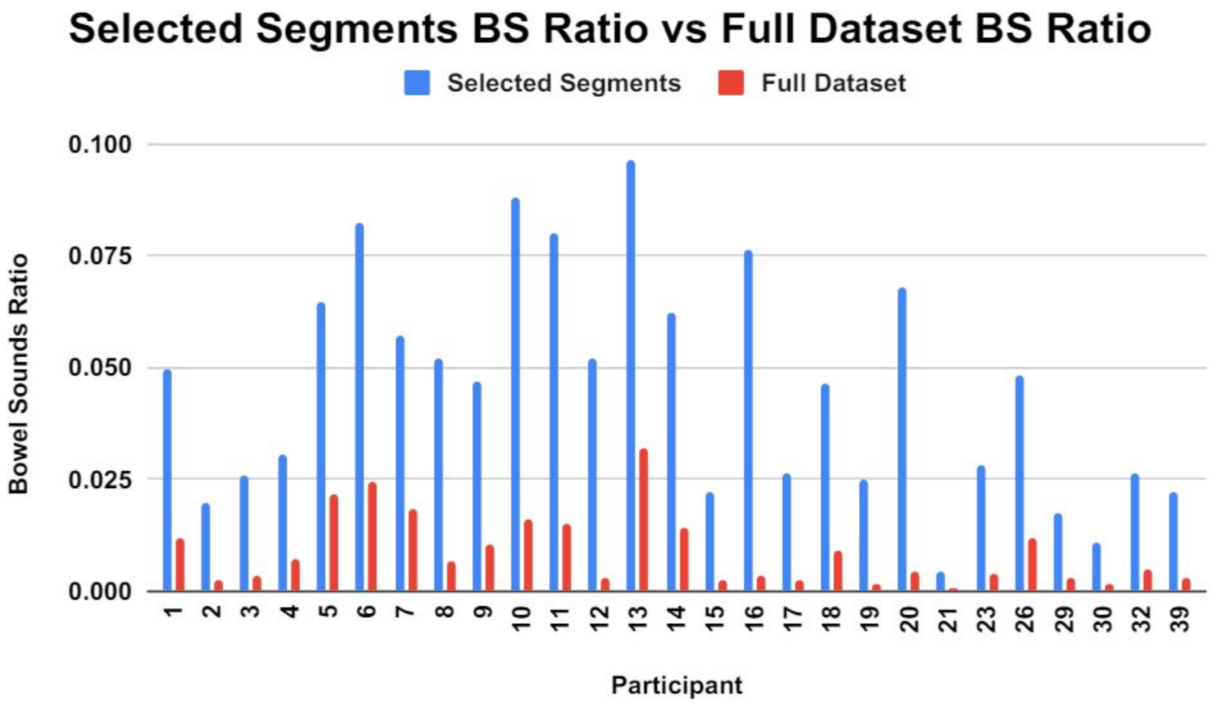


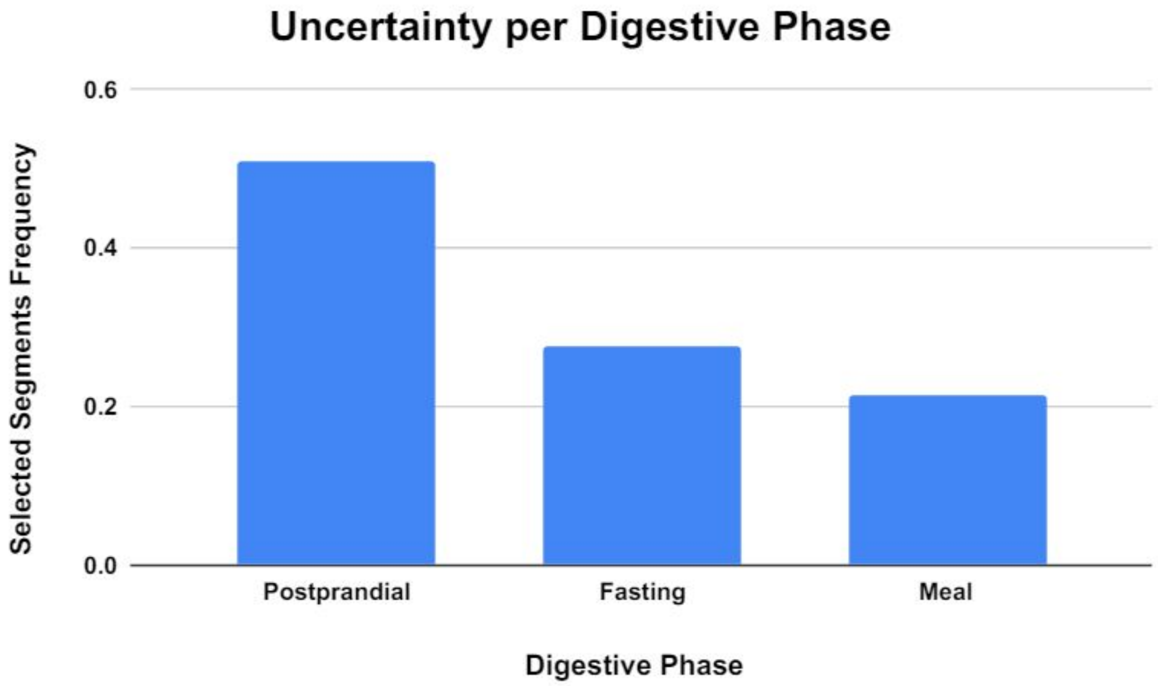
Average Recall through time
Recall
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
5000
10000
15000
20000
25000
30000
35000
Steps
Baseline (100%)
Active Learning (10%)
max AL Value: 0.65

Heuristics Results



Analysis on Selected 10s Audio Segments





Discussion

Why does the method work here?

1. **Unbalanced** dataset with a lot of noise. [2]

2. **High** labelling **costs.** [2]

3. Heuristics based on **Uncertainty.** [4]

4. The method selects segments with **many Bowel Sounds.**

Limitations

1. **Interactive Nature**: Necessity of repeated training and data labelling phases.
2. **Labeling Consistency**: Variability in human labeling can introduce inconsistencies.
3. **Initial Model Performance**: The effectiveness heavily depends on the initial model performance. [4]

Future Works

1. **Experiment with different models**: There might be better models for AL.
2. **Extend to new data**: Labelling new suggested data might further improve model performances.
3. **Implement Crowdsourcing**: Used for efficiently labelling new data. [8]

Conclusions

1. **Key concept:** Not all samples contribute equally to model performance. Samples with the highest uncertainty provide the most information gain.

2. **How it works:** AL algorithms select the most informative and uncertain samples from the dataset.

3. **When to use:** AL is particularly effective in datasets with significant noise or redundant samples.

In this project:

1. **Active Learning Efficiency**: Labelled data need reduced by **90%**.

2. **Time saving**: On average 10 h to label 1 h. The dataset has 136 h of labelled data →
   AL would have saved **1224 h** of data labelling, or **153 days** of 8 h.

References

[1] A. Baronetto, L. S. Graf, S. Fischer, M. F. Neurath, and O. Amft, “GastroDigitalShirt: A Smart Shirt for Digestion Acoustics Monitoring,” in ISWC ’20: Proceedings of the 2020 International Symposium on Wearable Computers, (Virtual Conference), pp. 17– 21, ACM, Sept. 2020

[2] A. Baronetto, L. S. Graf, S. Fischer, M. F. Neurath, and O. Amft, “Multi-scale Bowel Sound Event Spotting in Highly Imbalanced Wearable Monitoring Data,” under review

[3] B. Baheti, S. Innani, S. Gajre and S. Talbar, "Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 2020, pp. 1473-1481, doi: 10.1109/CVPRW50498.2020.00187

[4] Beck, Nathan, et al. "Effective evaluation of deep active learning on image classification tasks." arXiv preprint arXiv:2106.15324 (2021).

[5] https://github.com/Lightning-AI/pytorch-lightning

[6] Atighehchian, Parmida, Frédéric Branchaud-Charron, and Alexandre Lacoste. "Bayesian active learning for production, a systematic study and a reusable library." arXiv preprint arXiv:2006.09916 (2020).

[7] Gemmeke JF, Ellis DPW, Freedman D, Jansen A, Lawrence W, Moore RC, Plakal M, Ritter M. Audio Set: An ontology and human-labeled dataset for audio events. IEEE; 2017. p. 776–780. doi: 10.1109/ICASSP.2017.7952261

[8] S. Hantke, Z. Zhang, and B. Schuller, “Towards Intelligent Crowdsourcing for Audio Data Annotation: Integrating Active Learning in the Real World,” in Proceedings INTERSPEECH 2017, 18th Annual Conference of the International Speech Communication Association, ISCA. Stockholm, Sweden: ISCA, August 2017, pp. 3951–3955.

Questions

?