

Multimedia Systems

Assignment GA3. Environmental Sound Classification

FEUP

1. Goal

This assignment focuses on developing a system for environmental sound classification using the UrbanSound8K dataset. Your task is to build, test, and analyse a model that distinguishes between the ten sound classes defined in the dataset. You will explore practical methods for audio processing and supervised learning, understand how different modelling choices influence results, and follow the official evaluation protocol required by UrbanSound8K.

You may draw on public resources, pre-trained models, or your own implementations. The aim is to study how different approaches behave on real data and to understand their limits within the scope of this course.

In this assignment, you will:

- investigate the task of environmental sound classification with a focus on UrbanSound8K
- select a method and implement it using either classical machine learning or transfer learning
- apply the official 10-fold evaluation protocol of the dataset
- analyse the performance of your system and compare it with at least one alternative method
- present your findings in a short research-style report and provide clear, reproducible code

2. Resources

The following are some important resources to guide your work on environmental sound classification. They include relevant peer-reviewed papers, established datasets, and examples of code repositories. These resources are intended to provide a starting point, but students are encouraged to explore additional materials as needed.

Relevant Papers

- [**"A Dataset and Taxonomy for Urban Sound Research"**](#)
J. Salamon, C. Jacoby, and J.P. Bello, 22nd ACM International Conference on Multimedia (ACM-MM'14), 2014, pp. 1041–1044.
- [**"Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification"**](#)
J. Salamon and J.P. Bello, *IEEE Signal Processing Letters*, Volume: 24, Issue: 3, March 2017.
- [**"Environmental Sound Classification with Convolutional Neural Networks"**](#)
K. J. Piczak, *IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015.

Dataset

- [**UrbanSound8K Dataset**](#) – The mandatory dataset for development and evaluation is **UrbanSound8K**. The dataset includes 8732 labelled clips (<=4s) in 10 classes: air_conditioner, car_horn, children_playing, dog_bark, drilling, engine_idling, gun_shot, jackhammer, siren, and street_music.

3. Code

The following GitHub repositories are examples of projects that may help guide your work. These repositories demonstrate varying levels of quality and serve purely as reference materials. Students are encouraged to explore additional resources and approaches beyond these examples.

- <https://www.kaggle.com/code/salimhammadi07/esc-50-environmental-sound-classification>
- <https://github.com/mariostrbac/environmental-sound-classification>

4. Tasks

Your primary objective is to develop a functional environmental sound classification system while exploring feasible methods for UrbanSound8K.

Task 1: Choice of Method

Select and document your chosen approach. You may choose one of the following paths or combine elements from both:

1. Custom Feature Extraction with Classical Machine Learning

- o Extract features such as MFCCs or other spectral descriptors. You **cannot** use pre-trained models that were trained on UrbanSound8K.
- o Train classical classifiers such as Decision Trees or SVM.
- o **Your role:** design the feature extraction process, inspect the behaviour of the features, and test which classifier works best with them.

2. Transfer Learning, Fine-Tuning, and Data Augmentation

- o Use a pre-trained audio model and adapt it to the UrbanSound8K classes.
- o Fine-tune the model on the training folds.
- o You may also apply data augmentation techniques to improve generalisation.
- o **Your role:** understand how the original model works, explain how you adapted it, and justify any changes.

Note on End-to-End Deep Learning

Training a deep model from scratch is not expected for this assignment due to the limited size of UrbanSound8K and the workload for this course. You may still explore it during early experimentation, but it should not be your main method.

Task 2: Evaluation

Evaluate your system using the official structure of the UrbanSound8K dataset. The dataset is divided into 10 folds, and these folds must be used as provided. Do not reshuffle samples or create new splits. Follow the instructions on the UrbanSound8K page for correct data handling: <https://urbansounddataset.weebly.com/urbansound8k.html>.

Required Evaluation Procedure

For each fold f from 1 to 10:

- train your model on the other 9 folds
- test your model on fold f

At the end, compute the mean accuracy, mean precision, mean recall, mean F1-score, and a mean confusion matrix across the 10 folds.

Analysis of Results

Provide a clear discussion of the results. Include:

- an inspection of the averaged performance metrics
- a description of any limits you observed in your method
- comments on how the dataset structure affected the results

Comparison with an Alternative Approach

Compare your main method with at least one alternative. This may involve:

- a different feature set
- a different classifier
- a different transfer learning setup

Your comparison should focus on:

- changes in accuracy and F1-score
- differences in computational cost
- behaviour on folds that give lower performance

Task 3: Reporting and Code

1. **Report:** Compile your findings in a report formatted as a research paper, with a maximum of 5 double-column pages, adhering to a standard academic style. The report should include:
 - Abstract and Introduction: Briefly describe the sound classification problem and your chosen approach.
 - Methodology: Detail your method selection, feature extraction (if applicable), and any pre-processing steps.
 - Evaluation: Present and discuss your model's performance metrics and insights from your comparative analysis.
 - Conclusion: Summarise the main findings, noting the strengths and potential areas for improvement in your approach.
2. **Code Repository:** Provide a GitHub link containing the code, clear setup instructions and required dependencies, as well as steps to reproduce your methods, including training and fine-tuning.

5. Submission

Submit through Moodle and will include the following components:

1. **Report (PDF)** named GA3_GXX.pdf
2. **Code Repository (GitHub)** with code and README.

6. Evaluation Criteria

Your submission will be assessed on the following criteria:

- Approach and Justification (30%): The appropriateness and rationale behind your choice of methods, including openness to the options you explored.
- Implementation Quality (30%): Correctness, efficiency, and completeness of the code, along with clear instructions in the GitHub repository.
- Evaluation and Analysis (30%): Depth of model evaluation, including comparative insights and critical discussion of limitations.
- Report Quality (10%): Clarity, organisation, and adherence to the 5-page, double-column research paper format.

Note: A penalty of **0.5 points per day** will be applied to late submissions.