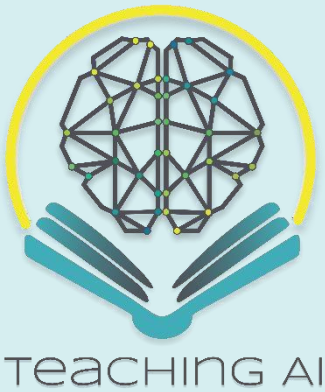


Report WP3-A1:

Searching for Appropriate Databases on the Internet

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency.

Co-funded by the European Union

Result

Searching for appropriate databases on the internet

Related to
WP3-A1: Searching the appropriate databases on internet

Statement of originality

This report contains original unpublished work, except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both.

Disclaimer

Copyright notice

Note

For anyone interested in having more information about the project, please see the website at: https://tai-erasmus.eu/

Document version history

Version	Date	Author	Comment
1.0	9.07.2026.	Jelena Šaković Jovanović, Aleksandar Vujović, Janko Jovanović, UoM	Initial version
2.0	13.07.2026.	Jelena Šaković Jovanović, Aleksandar Vujović, Janko Jovanović, UoM	Version II
3.0	14.07.2026.	Jelena Šaković Jovanović, Aleksandar Vujović, Janko Jovanović, UoM	Final version

Table of contents

1. Introduction

The Teaching Artificial Intelligence (TAI) project is implemented within the Erasmus+ Programme 2024, Key Action 2 - Cooperation Partnerships in Higher Education. The overall aim of the project is to strengthen Artificial Intelligence (AI) education by supporting the development of teaching materials, practical learning resources and research-oriented educational activities. In this context, the collection and analysis of datasets represents an important basis for introducing practical AI and Machine Learning (ML) activities into teaching. Datasets enable students and educators to move from theoretical AI concepts toward hands-on exercises, project-based learning and applied research.

Within this framework, Work Package 3 (WP3) focuses on creating a collection of datasets that can support practical AI teaching, learning and student projects.

This report presents the work carried out within WP3-A1: Searching for Appropriate Databases on the Internet. The main objective of this activity was to identify suitable datasets for different AI applications and data modalities. The search included four main categories of datasets: video, audio/sound, numerical and image datasets**.** Each category was assigned to one partner institution (UNILJ - University of Ljubljana was responsible for collecting video datasets, UoM - University of Montenegro for audio/sound datasets, UNBG - University of Belgrade for image datasets and UNSA - University of Sarajevo for numerical datasets) and the collected information was entered into a shared dataset collection document. UNIRI, as a partner on the project, was not assigned responsibility for collecting datasets, but for reviewing the collected datasets and advising on the structure and formats of the collection.

The purpose of this report is to describe the dataset search methodology, present the collection protocol and provide an overview of the datasets identified during WP3-A1. The results of this activity provide the basis for WP3-A2: Analysis of Databases, where the collected datasets will be further analysed in terms of relevance, quality, format and applicability for AI education.

The University of Montenegro (UoM) is the coordinator for this work package.

2. Dataset search methodology and collection protocol

The methodology for dataset search was defined in accordance with the objectives of WP3-A1, which focuses on identifying appropriate datasets that can support practical teaching and learning of Artificial Intelligence. The main aim was to collect a diverse and balanced set of datasets covering different data modalities and AI application areas. The search methodology was designed to support the identification of relevant, accessible and technically suitable datasets with potential application in further educational and project activities.

2.1 Dataset modalities and partner responsibilities

The search process was organised by assigning one data modality to each partner institution involved in WP3. According to the project proposal, each partner responsible for a particular data modality was expected to collect at least three datasets. However, for WP3-A1, the consortium decided to conduct a substantially broader search and compile a more comprehensive collection. Consequently, each responsible partner-UNILJ (University of Ljubljana), UoM (University of Montenegro), UNBG (University of Belgrade) and UNSA (University of Sarajevo)-identified 30 datasets within its assigned modality. This approach ensured a balanced collection of 120 datasets across four main categories: video, audio/sound, numerical and image datasets. The expanded collection was intended not only to support the subsequent project activities but also to serve as a useful resource for other institutions engaged in similar teaching, research or project activities.

According to the project proposal, UNIRI (University of Rijeka) was not assigned responsibility for dataset collection. Its role was to review the collected datasets and provide advice on the structure and formats of the collection. The distribution of partner responsibilities by modality is presented in Table 2.1.

Table 2.1. Partner responsibilities and number of collected datasets by modality

Partner	Modality	Number of datasets
UNILJ	Video datasets	30
UoM	Audio/Sound datasets	30
UNSA	Numerical datasets	30
UNBG	Image datasets	30
	Total:	**120**

As shown in Table 2.1, the activity resulted in the identification of 120 datasets in total, with an equal number of datasets collected for each modality.

2.2 Dataset search sources

Partners searched for datasets using different publicly available and relevant sources. The main focus was on identifying datasets that are accessible, well documented and suitable for use in Artificial Intelligence education and practical training.

The search included public dataset repositories (Kaggle, Zenodo, OpenML, …), scientific and research platforms, institutional websites, project repositories, open data portals and other relevant online sources. In addition, partners also considered datasets used in previous research activities or datasets referenced in scientific publications, when they were available and suitable for educational purposes.

During the search process, partners paid attention not only to the availability of the datasets, but also to the reliability of the source, clarity of documentation, licensing conditions and relevance of the datasets for specific AI tasks. Preference was given to datasets with clear descriptions, available references, standard formats and sufficient technical information for further analysis and use. An overview of the main sources used for dataset identification is provided in Table 2.2.

Table 2.2 Main dataset search sources

Source type	**Description**
Public dataset repositories	Open platforms and repositories containing datasets for AI and machine learning tasks
Scientific publications	Datasets referenced or introduced in research papers
Institutional websites	Datasets published by universities, laboratories, research centres or public institutions
Open data portals	Publicly available data portals providing structured datasets
Project repositories	Datasets available through previous research or educational projects
Other relevant sources	Additional sources providing technically suitable and accessible datasets

These sources enabled partners to identify datasets from different domains and application areas, supporting the diversity and relevance of the dataset collection.

2.3 Dataset selection criteria

During the search process, partners applied a common set of criteria to support the identification of datasets suitable for further analysis and potential use in the project. The selection was not limited to finding available datasets, but also included checking whether the datasets contained sufficient descriptive, legal and technical information.

The main criteria considered during dataset selection included dataset accessibility, license and usage conditions, source or owner, reference, application area, number of items, dataset size, data structure, data format and available technical characteristics. These criteria were used to assess whether the datasets could be reliably used for AI-related teaching, practical exercises, student projects and further project activities.

Special attention was given to datasets that are clearly documented, technically suitable for processing, relevant to one or more AI application areas and accompanied, where available, by information on licensing and usage conditions. Datasets with standard formats, clear structure and available references were considered particularly useful, as they provide a better basis for further analysis in WP3-A2 and for later implementation in Python-based educational examples. An overview of the main dataset selection criteria and their purpose is presented in Table 2.3.

Table 2.3. Main dataset selection criteria

Criterion	Purpose
Accessibility	To check whether the dataset can be accessed and used by project partners
License and usage conditions	To assess whether the dataset can be used for educational and project purposes
Source or owner	To identify the origin and reliability of the dataset
Reference	To provide traceability and link the dataset to relevant documentation or publication
Application area	To determine the AI-related field in which the dataset can be used
Number of items and size	To assess the scope and technical feasibility of using the dataset
Data structure and format	To evaluate compatibility with AI tools and Python-based processing

These criteria were designed to support the collection of datasets that are diverse in terms of modality, relevant, well documented and suitable for practical use in AI education.

2.4 Data recording and standardisation

All collected information was recorded in a shared dataset collection document available on the project SharePoint platform. The document was prepared in a common format in order to support consistent data recording by all partners and to enable comparison of datasets across different modalities.

All identified datasets were recorded in a shared spreadsheet using a common structure that included descriptive, technical, legal and application-related information. UNIRI reviewed the collected datasets and advised the responsible partners on the structure of the collection and the recording of data formats. This contributed to a more consistent presentation of information across the four data modalities and supported the preparation of the collection for further analysis and subsequent project activities. Where certain information was unavailable, not disclosed by the dataset provider or not applicable, the corresponding fields were left incomplete or marked accordingly.

3. Dataset collection outcome

The main outcome of WP3-A1 is a structured collection of datasets recorded in a shared dataset collection document. The document brings together the results of the search conducted by the partners responsible for the four data modalities and provides a common overview of datasets identified as potentially suitable for AI education, practical exercises and student projects. UNIRI contributed to this process by reviewing the dataset collection and advising the responsible partners on its structure and the recording of dataset formats. In addition to documenting the identified datasets, each partner responsible for a data modality selected and uploaded three datasets to the common project repository. These datasets were either owned by the respective partner institution or made available under licensing and usage conditions that permitted their use in subsequent project activities. They were selected as particularly relevant for the implementation of WP4, which involves the development of Python-based AI examples, and WP5, which focuses on the development of educational materials. In this way, WP3-A1 provided both a broad overview of potentially useful datasets and a concrete set of resources prepared for practical use in the subsequent work packages.

Because the collected datasets belong to different data modalities, they could not all be described in exactly the same way. Some information, such as dataset name, source, license, reference, application area, size and format, was relevant for all datasets. At the same time, additional technical details depended on the type of data. For example, video datasets require information such as duration, resolution and frame rate, while audio datasets require sampling rate, channels and bit depth. Similarly, image and numerical datasets include their own specific characteristics.

For this reason, the dataset collection was organized in a way that combines common descriptive information with modality-specific technical information. This chapter presents these recorded data in a structured form and explains their relevance for further dataset analysis, comparison and selection within WP3 and the following project activities.

3.1 Common information recorded for all datasets

A common set of information was recorded for all datasets in order to ensure comparability across different data modalities. These common fields provide the basic information needed to identify, access, describe and evaluate each dataset.

Table 3.1. Common information recorded for all collected datasets

Information recorded	Purpose and relevance
Dataset description	Provides a short overview of the dataset content and purpose
Modality (video, image, audio, numerical, multimodal)	Shows whether the dataset contains video, audio/sound, numerical or image data
Link / access reference	Enables access to the dataset and verification of the collected information
Source / owner	Identifies the institution, platform, research group or organization responsible for the dataset
License / usage conditions	Defines whether and how the dataset can be used for teaching, research and project outputs
Reference / publication	Provides academic traceability through related papers, reports or official citations
Application area	Shows the potential use of the dataset in AI-related domains
Number of items	Indicates the scale of the dataset and its suitability for exercises or projects
Total size (GB)	Helps estimate storage and computational requirements
Data structure (files, sequences, tabular, time-series, ...)	Describes how the data are organized, for example as folders, files, tables, sequences or annotations
Data format (mp3, mp4, avi, ...)	Identifies the file formats used, such as CSV, JSON, WAV, MP4, JPG or PNG
Compression	Helps assess storage requirements, download size, data quality and preprocessing needs related to compressed dataset file

These common fields make it possible to compare datasets collected by different partners and across different modalities. They also provide the basis for deciding which datasets can be used in later project activities.

3.2 Modality-specific information recorded for different dataset types

In addition to the common information, the shared dataset collection document also includes modality-specific technical characteristics. These fields differ depending on the type of dataset, because video, audio/sound, image and numerical data have different structures, formats and requirements for AI processing.

3.2.1 Video datasets

For video datasets, specific information was recorded in order to describe their visual, temporal and technical characteristics. These data are important for assessing whether a dataset is suitable for AI tasks such as video classification, action recognition, object detection, tracking and temporal analysis.

Table 3.2.1 Specific information recorded for video datasets

Specific information recorded	Purpose and relevance
No. of videos	Indicates the number of video recordings included in the dataset and helps assess its scale and suitability for training, testing or educational demonstrations.
Video length	Provides information on the duration of video recordings, such as minimum, maximum or average length. This is important for understanding temporal complexity and pre-processing needs.
Video resolution	Shows the visual dimensions and quality of the videos. Resolution affects storage requirements, computational complexity and model input preparation.
Frame rate (fps)	Indicates the number of frames per second. This is particularly relevant for motion analysis, action recognition and other temporal AI tasks.
Codec	Provides information on how the video data are encoded. This may influence video decoding, quality, compatibility with processing tools and pre-processing requirements.
Video colour space	Shows how colour information is represented, for example RGB or YUV. This is relevant for visual pre-processing and model compatibility.
Camera motion	Describes whether the videos were recorded with a static or moving camera. This information is useful for understanding scene complexity and possible challenges in object tracking or motion analysis.
Temporal annotations	Indicates whether the dataset includes time-based labels, segments or annotations. This is important for supervised learning tasks involving events, actions or temporal localization.

3.2.2 Audio/sound datasets

For audio and sound datasets, specific information was recorded in order to describe the technical properties of audio signals and the availability of annotations. These characteristics are important for speech processing, sound classification, music analysis and signal processing tasks.

Table 3.2.2 Specific information recorded for audio/sound datasets

Specific information recorded	Purpose and relevance
No. of audio files	Indicates the number of audio recordings included in the dataset and helps assess its size and suitability for training, testing or educational use.
Audio duration	Provides information on the length of audio recordings, such as minimum, maximum or average duration. This is important for estimating the amount of signal data and preprocessing needs.
Sampling rate	Shows how many audio samples are recorded per second, usually expressed in Hz. It affects audio quality, feature extraction and compatibility with signal processing methods.
Channels	Indicates whether the audio is mono, stereo or multi-channel. This information is relevant for preprocessing and for selecting appropriate analysis methods.
Bit depth	Describes the precision of audio samples. It may affect sound quality, dynamic range and the accuracy of audio analysis.
Audio format	Identifies the audio file format, such as WAV, MP3, FLAC or MIDI. This is important for compatibility with audio processing tools and Python libraries.
Temporal annotations	Indicates whether time-based labels, transcripts, segments or event annotations are available. This is important for speech recognition, sound event detection and audio segmentation tasks.

3.2.3 Numerical datasets

For numerical datasets, specific information was recorded in order to describe the structure, dimensionality and quality of tabular, time-series or measurement-based data. These fields are important for regression, classification, forecasting, anomaly detection and other numerical AI tasks.

Table 3.2.3 Specific information recorded for numerical datasets

Specific information recorded	Purpose and relevance
No. of rows (tabular)	Indicates the number of records or observations in the dataset. This helps assess the dataset size and suitability for statistical analysis and machine learning.
No. of features	Shows the number of variables or input attributes available for modelling. This is important for understanding dataset dimensionality and model complexity.
Feature types	Describes the types of variables included in the dataset, such as numerical, categorical, binary or mixed features. This helps determine appropriate pre-processing and modelling techniques.
Missing data (%)	Indicates the proportion of missing values in the dataset. This is important for evaluating data quality and identifying potential pre-processing or imputation needs.
Time-series	Shows whether the dataset contains time-dependent observations. This is relevant for forecasting, trend analysis, anomaly detection and temporal modelling.
Sampling frequency (tabular)	Indicates how often the data were collected or recorded, when applicable. This is particularly important for time-series, sensor-based or monitoring datasets.
Units	Provides information on the units of measurement or target variable. This supports correct interpretation, comparison and analysis of numerical data.

3.2.4 Image datasets

For image datasets, specific information was recorded in order to describe visual quality, image representation and the level of available annotations. These fields are important for computer vision tasks such as image classification, object detection, segmentation and image recognition.

Table 3.2.4. Specific information recorded for image datasets

Specific information recorded	Purpose and relevance
Image resolution	Shows the image dimensions and level of visual detail, such as fixed or variable image size. This affects storage requirements, pre-processing steps and preparation of input data for AI models.
Image color space	Indicates how image information is represented, such as RGB, grayscale, RGB with depth information, multispectral data or palette-indexed images. This is important for image pre-processing, feature extraction and model compatibility.
Bit depth	Provides information on the amount of colour or intensity information stored in each image, for example 8-bit per channel or 16-bit data. This may affect image quality, storage requirements and the type of analysis that can be performed.
Image format	Identifies the image file format or encoding used in the dataset, such as JPEG, PNG, TIFF or other image-related formats. This is important for compatibility with AI tools, image processing libraries and Python-based implementation.
Annotation level	Shows the type and detail of annotations available in the dataset, such as image-level labels, bounding boxes, segmentation masks, keypoints, landmarks or attribute labels. This is especially important for supervised learning and computer vision exercises.

By recording modality-specific information, the project partners ensured that each dataset type was described in a technically meaningful way. This is important because different AI tasks require different data properties. For example, frame rate is relevant for video analysis, sampling rate for audio processing, annotation level for image-based supervised learning, and missing values or feature types for numerical modelling. Therefore, modality-specific information provides an important basis for selecting datasets that are suitable for further analysis and practical implementation in AI education.

3.3 Structure of the shared dataset collection document

The shared dataset collection document was structured as a combination of common information fields and modality-specific technical fields. The common fields were used for all collected datasets and enabled general identification, access verification and comparison across modalities. In parallel, additional modality-specific fields were included in order to describe the technical characteristics of video, audio/sound, image and numerical datasets in a more precise way. The structure of the shared dataset collection document is presented in Figure 3.1.

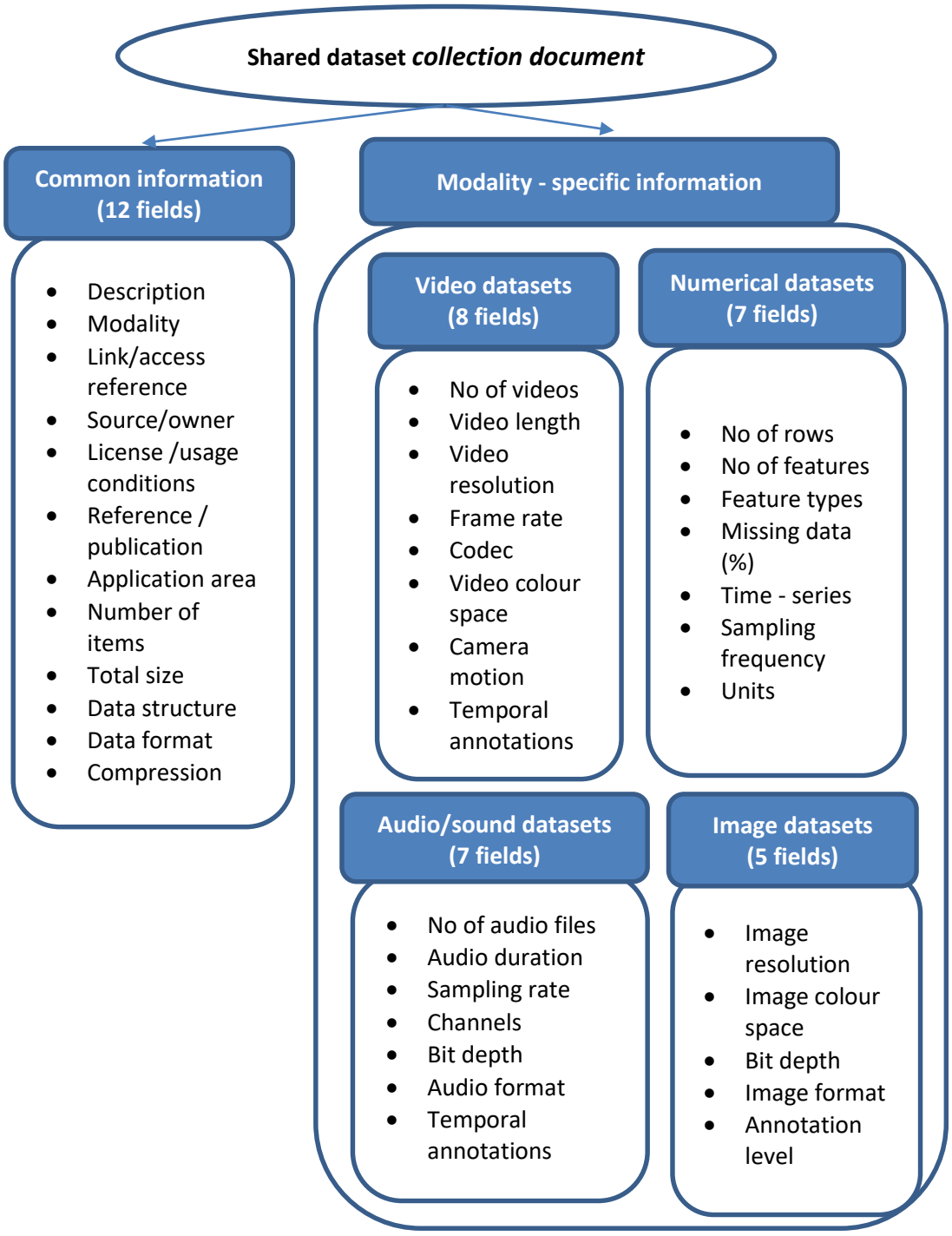


Figure 3.1 Structure of the shared dataset collection document

As shown in Figure 3.1, the shared dataset collection document was organised around a common set of information applicable to all datasets, complemented by specific technical fields for each dataset modality. This structure enabled partners to record dataset information in a harmonised way, while still allowing enough flexibility to capture the specific properties of different types of data.

Figure 3.2 further illustrates the number of recorded information fields by information category. The common information category contains the largest number of fields, since these data are relevant for all collected datasets. The modality-specific categories contain a smaller number of fields, but these fields are essential for describing the technical characteristics of each dataset type.

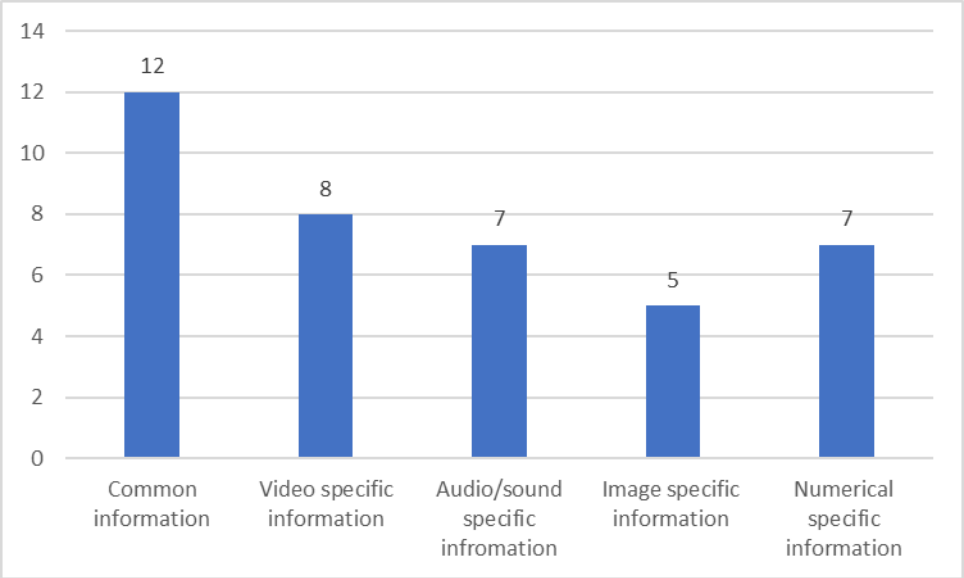


Figure 3.2 Number of recorded information fields by information category

This confirms that the shared dataset collection document was designed to ensure both consistency in data recording and flexibility in describing different data modalities.

4. Selected examples of collected datasets by modality

This chapter presents selected examples from the dataset collection identified during WP3-A1. Instead of reproducing the complete dataset collection in the main body of the report, a limited number of representative examples is shown for each modality in order to keep the report clear and focused. The complete list of collected datasets, including all recorded common and modality-specific information, is provided in Annex 1.

The selected examples are grouped according to the four dataset modalities: video, audio/sound, numerical and image datasets. For each modality, the overview table includes key information needed to identify the dataset, understand its basic content and indicate its potential relevance for AI education and further project activities. A more detailed assessment of dataset quality, technical suitability and applicability will be carried out in the following activity, WP3-A2: Analysis of Databases.

4.1 Selected examples of collected video datasets

Selected information on the collected video datasets is presented in Table 4.1. The table provides an excerpt from the shared dataset collection document and includes only the fields considered most relevant for a general presentation of the collected datasets.

Table 4.1: Selected examples of collected video datasets

No	Dataset description	Source / Owner	License	Reference	Applications	Number of items	Data format
1.	Video dataset for Word-Level American Sign Language (ASL) recognition, which features 2,000 common different words	Dongxu Li and Hongdong Li	Licensed under the Computational Use of Data Agreement (C-UDA).	Li, Dongxu, et al. "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020.	Gesture recognition	12000 videos	mp4
2.	Something Something v2: human interaction with everyday actions.	Qualcomm AI Research	Available for research purposes.	Goyal, Raghav, et al. "The" something something" video database for learning and evaluating visual common sense." Proceedings of the IEEE international conference on computer vision. 2017.	Gesture recognition	220847 videos	webm
3.	UCF101 action recognition data set of realistic action videos, collected from YouTube	University of Central Florida		Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. "Ucf101: A dataset of 101 human actions classes from videos in the wild." arXiv preprint arXiv:1212.0402 (2012).	Action recognition	13320 videos	.avi

The information presented in Table 4.1 represents only a selected summary of the collected data for video datasets. The complete set of recorded information, including both common and video-specific technical characteristics, is provided in Annex 1.

4.2 Selected examples of collected audio/sound datasets

Selected information on the collected audio/sound datasets is presented in Table 4.2. The table provides an excerpt from the shared dataset collection document and includes only the fields considered most relevant for a general presentation of the collected datasets.

Table 4.2 Selected examples of collected audio/sound datasets

No	Dataset description	Source / Owner	License	Reference	Applications	Number of items	Data format
1.	GTZAN is an audio dataset intended for music genre classification. It contains 1,000 audio recordings, each lasting 30 seconds, organized into 10 genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Each genre is represented by 100 audio examples.	Tzanetakis and Cook, 2002	Available for research and education, widely used	Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293–302.	Music genre recognition	1000	WAV audio files (.wav)
2.	IRMAS - is an audio dataset designed for automatic recognition of predominant musical instruments in audio signals. It contains musical audio excerpts annotated with the predominant instrument(s)	Dataset by Juan J. Bosch, Ferdinand Fuhrmann, Perfecto Herrera. Music Technology Group - Universitat Pompeu Fabra	Free of charge for non-commercial use only	Bosch, J. J., Janer, J., Fuhrmann, F., & Herrera, P. "A Comparison of Sound Segregation Techniques for Predominant	Musical instrument recognition	9579	WAV audio files (.wav)

	present. It contains 9,579 audio excerpts divided into training and testing sets, with annotations for 11 instrument classes			Instrument Recognition in Musical Audio Signals", in Proc. ISMIR (pp. 559-564), 2012			
3.	UrbanSound 8K - This dataset contains 8732 labeled sound excerpts (<=4s) of urban sounds from 10 classes: air_conditioner, car_horn, children_playing, dog_bark, drilling, enginge_idling, gun_shot, jackhammer, siren, and street_music. The classes are drawn from the urban sound taxonomy.	Justin Salamon, Christopher Jacoby and Juan Pablo Bello, New York University, Music and Audio Research Laboratory.	UrbanSound8K datasets are offered free of charge for non-commercial use only under the terms of the Creative Commons Attribution Noncommercial License (by-nc), version 3.0:	J. Salamon, C. Jacoby and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research", 22nd ACM International Conference on Multimedia, Orlando USA, Nov. 2014.	Classification and recognition of urban environmental sounds.	8732	WAV audio files (.wav)

The information presented in Table 4.2 represents only a selected summary of the collected data for audio/sound datasets. The complete set of recorded information, including both common and audio/sound-specific technical characteristics, is provided in Annex 1.

4.3 Selected examples of collected numerical datasets

Selected information on the collected numerical datasets is presented in Table 4.3. The table provides an excerpt from the shared dataset collection document and includes only the fields considered most relevant for a general presentation of the collected datasets.

Table 4.3: Selected examples of collected numerical datasets

No	Dataset description	Source / Owner	License	Reference	Applications	Number of items	Data format
1.	Cleveland heart disease dataset: clinical and test attributes used to detect presence of coronary artery disease.	Janosi, Steinbrunn, Pfisterer, Detrano / UCI ML Repository	CC BY 4.0	Detrano, R., et al. "International application of a new probability algorithm for the diagnosis of coronary artery disease." The American Journal of Cardiology 64.5 (1989): 304-310.	Classification (medical diagnosis)	303 records	csv
2.	Direct marketing (phone-call) campaigns of a Portuguese bank; predict client subscription to a term deposit.	S. Moro, P. Cortez, P. Rita / UCI ML Repository	CC BY 4.0	Moro, S., Cortez, P., Rita, P. "A data-driven approach to predict the success of bank telemarketing." Decision Support Systems 62 (2014): 22-31.	Classification	45211 records	csv
3.	California housing block-group data from the 1990 census; predict median house value from demographic/ geographic features.	R. K. Pace & R. Barry (StatLib)	Public domain	Pace, R. Kelley, and Ronald Barry. "Sparse spatial autoregressions." Statistics & Probability Letters 33.3 (1997): 291-297.	Regression	20640 records	csv

The information presented in Table 4.3 represents only a selected summary of the collected data for numerical datasets. The complete set of recorded information, including both common and numerical-dataset-specific technical characteristics, is provided in Annex 1.

4.4 Selected examples of collected image datasets

Selected information on the collected image datasets is presented in Table 4.4. The table provides an excerpt from the shared dataset collection document and includes only the fields considered most relevant for a general presentation of the collected datasets.

Table 4.4 Selected examples of collected image datasets

No	Dataset description	Source / Owner	License	Reference	Applications	No. of items	Data format
1.	CarLogos and CarIDs	Laboratory for Data Analysis and Applied AI, University of Belgrade, School of Electrical Engineering	CC BY 4.0	/	Object Detection, Image Classification, License Plate Recognition (ALPR/OCR), Vehicle Re-Identification, Traffic Monitoring & Surveillance, Automotive Fraud Detection.	700	JPG
2.	Brand logos and icons dataset	/	CC0: Public Domain	/	Image Classification	1991	JPG, PNG
3.	CIFAR-10 Image Classification	University of Toronto, Department of Computer Science	CC0: Public Domain	Robbie Beane. CIFAR-10 Image Classification. https://kaggle.com/competitions/cifar10-mu, 2021. Kaggle.	Image Classification	60000	PNG

The information presented in Table 4.4 represents only a selected summary of the collected data for image datasets. The complete set of recorded information, including both common and image-specific technical characteristics, is provided in Annex 1.

The modality-based overview tables provide a structured representation of the content recorded in the shared dataset collection document. They allow the reader to see which datasets were identified by each partner and how they are distributed across the four dataset modalities.

5. Conclusion

WP3-A1 resulted in the identification and structured recording of datasets across four main data modalities: video, audio/sound, image and numerical datasets. The dataset search was carried out by four WP3 partners, with each partner responsible for collecting 30 datasets within its assigned modality. As a result, a total of 120 datasets were identified and recorded.

UNIRI was not assigned responsibility for collecting datasets within a specific data modality. Instead, its role was to review the collected datasets and advise the responsible partners on the structure of the collection and the recording of dataset formats. This contribution supported a more consistent organisation and presentation of the collected information across the four data modalities.

All collected information was recorded in a shared dataset collection document available on the project SharePoint platform. This document was designed to include both common information fields applicable to all datasets and modality-specific technical fields relevant to particular dataset types. This enabled harmonised data recording across the consortium, while also allowing partners to capture the specific characteristics of different dataset modalities.

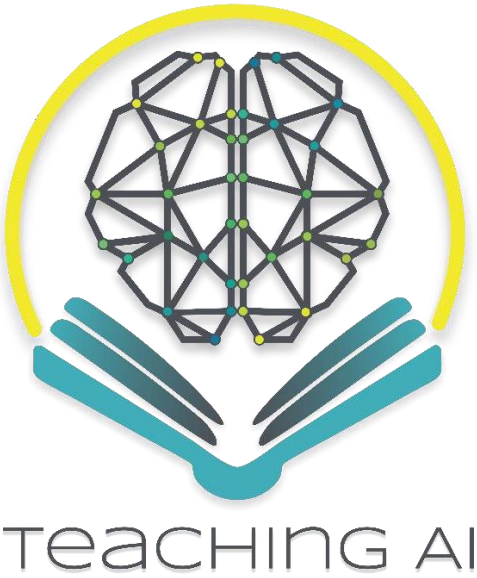
This report therefore provides a structured overview of the dataset search process and presents the main results of WP3-A1 in a clear and readable form. It does not aim to provide a detailed evaluation of the collected datasets, but rather to document the search methodology, partner responsibilities, data recording approach and the identified dataset collection.

The overview prepared within WP3-A1 will serve as the basis for the following WP3 activity, WP3-A2: Analysis of Databases. The next report will focus on a more detailed analysis of the identified datasets, including their relevance, accessibility, technical quality, format, completeness of documentation and applicability for AI education, practical exercises, student projects and further project activities.

6. ANNEX

Annex 1: Shared dataset collection document

The complete shared dataset collection is provided as Annex 1 in a separate file accompanying this report. Annex 1 contains the available descriptive, licensing and usage-related, structural and technical information recorded for all 120 datasets, covering video, audio/sound, numerical and image data modalities.


TEACHING AI

Lead Partner

UNIVERZA V LJUBLJANI
University of Ljubljana

Partners

UNIVERSITAS STUDIORUM SARAJEVOENSIS
UNIVERZITET U SARAJEVU

УНИВЕРЗИТЕТ У БЕОГРАДУ
UNIVERSITY OF BELGRADE

UCG
Univerzitet Crne Gore

uniri

Metadata	Project partner	Dataset description	Modality	Link	Source / Owner	License	Reference	Applications	Number of items	Total size (GB)	Structure	Data format	Compression	No of. videos	Video length	Video resolution	Frame rate (fps)	Codec	Video color space	Camera motion	Temporal annotations
Comment/ example			*video, image, audio, numerical, multimodal*								*files, sequences, tabular, time-series, ...*	*mp3, mp4, ...*			*min/max/median*						
1	UNILJ	Video dataset for Word-Level American Sign Language (ASL) recognition, which features 2,000 common different words	video	https://github.com/dxli94/WLASL	Dongxu Li and Hongdong Li	Licensed under the Computational Use of Data Agreement (C-UDA).	Li, Dongxu, et al. "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020.	Gesture recognition	12000 videos	5.4 GB	files	mp4	Lossy	12000		320x240	25	H.264	RGB		
2	UNILJ	Something Something v2: human interaction with everyday actions.	video	https://www.qualcomm.com/developer/software/something-something-v-2-dataset	Qualcomm AI Research	Available for research purposes.	Goyal, Raghav, et al. "The" something something" video database for learning and evaluating visual common sense." Proceedings of the IEEE international conference on computer vision. 2017.	Gesture recognition	220,847 videos	19.4 GB	files	webm	Lossy	220,847		320x240	12	VP9	RGB		
3	UNILJ	UCF101 action recognition data set of realistic action videos, collected from YouTube	video	https://www.crcv.ucf.edu/data/UCF101.php	University of Central Florida		Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. "Ucf101: A dataset of 101 human actions classes from videos in the wild." arXiv preprint arXiv:1212.0402 (2012).	Action recognition	13,320 videos	6.48 GB	files	.avi	Varies	13,320	1.06s/71.04s/7.21 s (mean)	320x240	25	Varies	RGB		
4	UNILJ	PEViD-UHD: Privacy Evaluation Ultra High Definition Video Dataset	video	https://www.epfl.ch/labs/mmspg/downloads/pevid-uhd/	EPFL	Non-commercial academic use license	Korshunov, Pavel, and Touradj Ebrahimi. "UHD video dataset for evaluation of privacy." 2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX). IEEE, 2014.	Video surveilance	26	825 MB	files	mp4	Lossy	26	13s each	3840 x 2160	30	H.264	RGB		
5	UNILJ	Qualcomm Interactive Video Dataset (QIVD) for interpretation and responding to real-time visual and audio inputs	video, audio	https://www.qualcomm.com/developer/software/qualcomm-interactive-video-dataset-qivd	Qualcomm AI Research	Research purposes only	Pourreza, Reza, et al. "Can Vision-Language Models Answer Face to Face Questions in the Real-World?." arXiv preprint arXiv:2503.19356 (2025).	Understanding with Vision-Language Models	2,900 videos	846 MB	files	mp4	Lossy	2,900	5.1s (average)	640 x 382.29 (average)	30	H.264	RGB		
6	UNILJ	nuScenes: A multimodal dataset for autonomous driving	video, LIDAR, RADAR	https://www.nuscenes.org/	Motional	Free non-commercial use	Caesar, Holger, et al. "nuscenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.	Autonomous driving detection and tracking	1000 scenes	350 GB	files	jpeg, .pcd	Varies	1000 scenes	20 s	1600×900	12	Varies	RGB		
7	UNILJ	EGO4D - Large-scale first-person video dataset of daily-life activities	video, audio	https://ego4d-data.org/	Meta	Research License Agreement	Grauman, Kristen, et al. "Ego4d: Around the world in 3,000 hours of egocentric video." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.	First-person perception	24246 videos	20 TB	files	mp4	Lossy	24246	Varies (min to h)	Varies	30	VP9	RGB		
8	UNILJ	The Kinetics-700: a large scale collection of YouTube video URLs for human action recognition	video	https://huggingface.co/datasets/ataiaydenknalbant/Kinetics-700	Google	Varies (YouTube License)	Carreira, Joao, et al. "A short note on the kinetics-700 human action dataset." arXiv preprint arXiv:1907.06987 (2019).	Action Recognition, Video Understanding	635,000 videos	871 GB	files	mp4	Lossy	635,000	~10 s	Varies	Varies	H.264	RGB		
9	UNILJ	Fall Vision: A Benchmark Video Dataset for Advancing Fall Detection Technology	video	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/75QPKK	Department of Computer Science and Engineering, University of Asia Pacific, Dhaka, Bangladesh	CC0 1.0	Rahman, Nakiba Nuren, et al. "FallVision: A benchmark video dataset for fall detection." Data in Brief 59 (2025): 111440.	Fall Detection	11732 videos	50 GB	files	mp4	Lossy	11,732	~3 s	1280x720, 1920x1080	30	MPEG-4	RGB		
10	UNILJ	Dataset of type-1252 Durum wheat kernels and impurities obtained during the movement of a conveyor belt system	video (+image, features)	https://www.muratkoklu.com/datasets/	KAYA Esra, SARITAS Ismail, Faculty of Technology, Turkey	CC0: Public Domain	Kaya, E., & Saritas, İ. (2019). Towards a real-time sorting system: identification of vitreous durum wheat kernels using ANN based on their morphological, colour, wavelet and gaborlet features. Computers and Electronics in Agriculture, 166, 105016. Link: https://doi.org/10.1016/j.compag.2019.105016	Classification, clustering	4 videos, 9000 wheat kernels, 236 features	1.7 GB	files	avi, tiff, docx, xlsx	no	4	8 s (all 4 videos)	887x1088	8	24 bits RGB (RV24)	RGB		
11	UNILJ	Dataset on UAV RGB videos acquired over a vineyard property of Bodegas Terras Gauda at an early stage of Botrytis cinerea infection in 2021	video	https://zenodo.org/records/7330951	Mar Ariza-Sentís, Sergio Vélez, João Valente, Wageningen University & Research	CC BY 4.0	Ariza-Sentís, Mar, Sergio Vélez, and João Valente. "Dataset on UAV RGB videos acquired over a vineyard including bunch labels for object detection and tracking." Data in Brief 46 (2023): 108848.	object detection & tracking, yield estimation, disease monitoring (Botrytis)	40 videos	21.7 GB	files, annotations	mp4 + png/mots (annotations)	lossy	40	5 s	4096x2160	60	H.264	YUV (typ. 4:2:0), source RGB sensor		
12	UNILJ	GrapeMOTS (UAV Vineyard Multi-Object Tracking)	video	https://zenodo.org/records/10625595	Mar Ariza-Sentís, Sergio Vélez, João Valente, Kaiwen Wang, Zhen Cao, Wageningen University & Research	CC BY	Mar, A.-S., et al. MOTS-annotated UAV Vineyard Dataset Captured Using Multiple Perspectives to Avoid Leaf Occlusion for Object Detection and Tracking. Zenodo, 7 Feb. 2024, https://doi.org/10.5281/zenodo.10625595.	object detection and tracking, avoid leaf occlusion	11 videos	87,4 GB	files	mp4 + png (frames/annotations)	standard	11	25 to 30 seconds	1920x1080 (Full HD) + 3840x2160 (4K)	30	MP4	RGB		
13	UNILJ	FieldSAFE: multimodal dataset for obstacle detection in agriculture	video(stereo, thermal, web, 360° cameras), LiDAR, radar, localization	https://vision.eng.au.dk/fieldsafe/	Aarhus University	CC BY 4.0	Kragh, Mikkel Fly, et al. "Fieldsafe: dataset for obstacle detection in agriculture." *Sensors* 17.11 (2017): 2579.	obstacle detection	1 stream per sensor (4 cameras + LiDAR + radar)	NA	files, annotations	ROS bag / raw sensor streams	partially (web, 360°), partially raw (stereo, thermal)	1 continuous stream per sensor (camera) of a 2 hour event - mowing the grass, synchronized	2 h	stereo camera: 1024x544, web camera: 1920x1080, 360° camera: 2048x833, thermal camera: 640x512	stereo camera: 10 fps, web camera: 20 fps, 360° + thermal camera: 30 fps	not specified (web camera most likely H.264 or MJPEG)	stereo: left camera RGB, right camera grayscale, web camera: RGB, 360° camera: RGB, thermal camera: grayscale		
14	UNILJ	UAV videos for mango yield estimation	video	https://ieee-dataport.org/documents/mangouav-video-dataset	Indian Institute of Information Technology Sri City	NA	https://dx.doi.org/10.21227/r8g8-et87	fruit counting and orchard monitoring	3 videos	93 MB	files	mp4	standard	3	3 to 4 minutes per video	1280x720	NA	NA	RGB		
15	UNILJ	EV2 Video Dataset: RGB and thermal videos (	video, thermal sequence	https://www.kaggle.com/datasets/alcor-lab/ev2-video-dataset	Alcor Lab, Department of Computer, Control, and Management Engineering A. Ruberti, University of Roma	CC BY-NC 4.0	Project Edge Vision against Varroa (EV2) https://alcorlab.diag.uniroma1.it/projects/ev2	pest detection, classification, monitoring	306 + 301 videos	199,19 GB	files	mkv	yes for RGB, raw for thermal	306 infested bees + 301 non infested bees videos	various, up to 15 seconds	7680x2350	15	V_MPEG4/ISO/AVC	YUV 4:2:0 + thermal		
16	UNILJ	TeaWeed-Action: A Vision-Based Dataset for Weeding Behavior Recognition in Tea Plantations	video	https://ieee-dataport.org/documents/teaweed-action-vision-based-dataset-weeding-behavior-recognition-tea-plantations	Guangdong University of Petrochemical Technology, Maoming, Nanjing Agricultural University, Jiangsu Normal University, Xuzhou	NA	Han R, Liang X, Shu L, Jing X, Yang F and Tian R (2025) TeaWeeding-Action: a vision-based dataset for weeding behavior recognition in tea plantations. Front. Plant Sci. 16:1722007. doi: 10.3389/fpls.2025.1722007	action recognition, behavior analysis, agricultural robotics	108 videos	41 GB	files	mp4	NA	108	NA	854x489, 852x480, 1280x720	NA	NA	RGB		
17	UNILJ	CBVD-5(Cow Behavior Video Dataset)	video	https://www.kaggle.com/datasets/fandaoerji/cbvd-5cow-behavior-video-dataset	College of Electronic Information Engineering, Inner Mongolia University	CC0: Public Domain	Li, K., Fan, D., Wu, H. et al. A new dataset for video-based cow behavior recognition. Sci Rep 14, 18702 (2024). https://doi.org/10.1038/s41598-024-65953-x	behavior recognition, tracking, livestock monitoring	687 + 200 videos	11.64 GB	files, annotations	mp4	lossy	687 + 200	10 seconds	1920x1080	25	Advanced Video Coding	YUV 4:2:0		
18	UNILJ	Raw Oil Palm Fruit Video Dataset	video	https://www.scidb.cn/en/detail?dataSetId=3776689052d5483095ad8e13aeaad4dc	Bina Nusantara University (Indonesia)	CC BY 4.0	Suharjito, Naftali, M.G., Hugo, G. et al. Oil Palm Fruits Dataset in Plantations for Harvest Estimation Using Digital Census and Smartphone. Sci Data 12, 972 (2025). https://doi.org/10.1038/s41597-025-05227-x Suharjito, Junior, F.A., Koeswandy, Y.P. et al. Annotated Datasets of Oil Palm Fruit Bunch Piles for Ripeness Grading Using Deep Learning. Sci Data 10, 72 (2023). https://doi.org/10.1038/s41597-023-01958-x	ripeness detection, fruit counting, yield estimation	440 videos	2.65 GB	files	mp4	standard	440	from 8 to 91 s	320x640	30	NA	RGB		
19	UNILJ	UAV monocular RGB video dataset za SLAM, SfM in 3D rekonstrukcijo	video	https://zenodo.org/records/14658376	Wageningen University & Research	CC BY 4.0	Wang, Kaiwen, et al. "GrapeSLAM: UAV-based monocular visual dataset for SLAM, SfM and 3D reconstruction with trajectories under challenging illumination conditions." *Data in Brief* 60 (2025): 111495.	SLAM, navigation, 3D reconstruction, robotics	24 videos	40 GB	files, annotations	mov (mpeg-4)		24	from 1 min to 7 min 20 s	3840x2160	30 fps original, downsampled to 10 fps	H.264 / AVC (avc1)	YUV 4:2:0		
20	UNILJ	A database of videos of traffic on the highway. The video was taken over two days from a stationary camera overlooking I-5 in Seattle, WA.	video	https://www.kaggle.com/datasets/aryashah2k/highway-traffic-videos-dataset	Statistical Visual Computing Lab University of California, San Diego Antoni Chan	CC0: Public Domain	[1] A. B. Chan and N. Vasconcelos, "Probabilistic Kernels for the Classification of Auto-Regressive Visual Processes". Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, 2005. [2] A. B. Chan and N. Vasconcelos, "Classification and Retrieval of Traffic Video using Auto-Regressive Stochastic Processes". Proceedings of 2005 IEEE Intelligent Vehicles Symposium, Las Vegas, June 2005.	The video were labeled manually as light, medium, and heavy traffic, which correspond respectively to free-flowing traffic, traffic at reduced speed, and stopped or very slow speed traffic.	254 videos	92.45 MB	files	avi	Lossy	254	5 s	320x240	10	MS MPEG-4 Video v1 (MPG4)	YUV (typ. 4:2:0)		
21	UNILJ	The videos were shot around cars in daylight from all sides. Each video is up to 100 seconds long. Cars are divided into 2 types: with damage and without damage.	video	https://www.kaggle.com/datasets/tapakah68/videos-around-cars	Roman Kucev	Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)		Car damage recognition	1200 videos	574.57 MB	files	MPEG-4		1200	varying (up to 100 s)	2400x1080	30	mp42 (isom/mp42)	YUV (typ. 4:2:0)		

22	UNILJ	This dataset contains videos of people doing workouts. The name of the existing workout corresponds to the name of the folder listed.	video	https://www.kaggle.com/datasets/hasyimabdillah/workoutfitness-video	Hasyim Abdillah	CC BY-NC-SA 4.0		Exercise recognition	652 videos	4.9 GB	files	various (590 x mp4, 62 x mov)		652	min: 1 s, max: 228 s, median: 5 s	1280x720 (321 videos), 1920x1080 (307 videos), 1920x1440 (8 videos), 640x360 (6 videos), 1080x608 (5 videos), 988x556 (2 videos), 320x240 (1 video), 1280x714 (1 video)	min: 23.97 fps, max: 60 fps, median: 29.97 s	h264	YUV typ. 4:2:0 (643 videos), YUVJ 4:2:0 (9 videos)		
23	UNILJ	Videos of annotated traffic lights	videos	https://www.kaggle.com/datasets/mbornoe/lisa-traffic-light-dataset	Morten Borno Jensen and Mark Philipsen	CC BY-NC-SA 4.0	Jensen MB, Philipsen MP, Møgelmose A, Moeslund TB, Trivedi MM. Vision for Looking at Traffic Lights: Issues, Survey, and Perspectives. I E E E Transactions on Intelligent Transportation Systems. 2016 Feb 3;17(7):1800-1815. Available from, DOI: 10.1109/TITS.2015.2509509 Philipsen, M. P., Jensen, M. B., Møgelmose, A., Moeslund, T. B., & Trivedi, M. M. (2015, September). Traffic light detection: A learning algorithm and evaluations on challenging dataset. In intelligent transportation systems (ITSC), 2015 IEEE 18th international conference on (pp. 2341-2345). IEEE.	Trafic light detection	13 daytime clips and 5 nighttime clips (43,007 frames and 113,888 annotated traffic lights)	4.2 GB	files	single frames and a csv file with annotations		18 (collection of frames)	varying number of frames	1280 x 960					
24	UNILJ	Video database for fire and smoke detection	videos	https://www.kaggle.com/datasets/chrisfilo/firesense	Chris Gorgolewski	Attribution 4.0 International (CC BY 4.0)	K. Dimitropoulos, P. Barmpoutis, N. Grammalidis, "Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection", IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), Vol. 25, No. 2, February 2015, pp. 339-351.	Fire and smoke detection	For flame detection 11 positive and 16 negative videos are provided, while for smoke detection, 13 positive and 9 negative videos are provided.	1.23 GB	files	avi	Lossy	49 videos	varying	352 x 288	various	XVID	YUV (typ. 4:2:0)		
25	UNILJ	HWID12 (Highway Incidents Detection Dataset)	videos	https://www.kaggle.com/datasets/landrykezebou/hwid12-highway-incidents-detection-dataset	Landry Kezebou	Freely available for research and education purposes only	Landry Kezebou, Victor Oludare, Karen Panetta, James Intriligator, and Sos Agaian "Highway accident detection and classification from live traffic surveillance cameras: a comprehensive dataset and video action recognition benchmarking", Proc. SPIE 12100, Multimodal Image Exploitation and Learning 2022, 121000M (27 May 2022); https://doi.org/10.1117/12.2618943	Highway incident detection and classification from live surveillance footage.	2782	12.03 GB	files	mp4		2782	varying (3 to 8 seconds)	1280 x 720	30	isom (isom/iso2/avc1/mp41)	YUV (typ. 4:2:0)	car camera	
26	UNILJ	This dataset contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios.	videos	https://www.kaggle.com/datasets/vafaeii/kth-action-recognition-dataset	Amir Vafaei	CC BY-NC 4.0	Schuldt, Christian, Ivan Laptev, and Barbara Caputo. "Recognizing human actions: a local SVM approach." Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.. Vol. 3. IEEE, 2004.	Action recognition	600	1.16 GB	files	avi	Lossy	600 videos	varying (around 20 s)	160 x 120	25	DX50	YUV (typ. 4:2:0)		
27	UNILJ	Dataset for Word-Level Arabic sign language (ArSL)	videos	https://www.kaggle.com/datasets/yousefdotpy/karsl-502	Sidig, Ala Addin		Sidig, Ala Addin I., et al. "KArSL: Arabic sign language database." ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 20.1 (2021): 1-19.	Sign words recognition	75300 videos	24.99 GB	files	mp4		75300 videos	min: 0.6 s, max: 4.166 s, mean: 1.66 s	1920x1080	30	h264	YUV (typ. 4:2:0)		
28	UNILJ	Underwater tracking benchmark dataset	videos	https://www.kaggle.com/datasets/landrykezebou/uot32-underwater-object-tracking-dataset	Landry Kezebou	For non-commercial research and educational purposes	L. Kezebou, V. Oludare, K. Panetta and S. S. Agaian, "Underwater Object Tracking Benchmark and Dataset," 2019 IEEE International Symposium on Technologies for Homeland Security (HST), Woburn, MA, USA, 2019, pp. 1-6, doi: 10.1109/HST47167.2019.9032954. K. Panetta, L. Kezebou, V. Oludare, and S. S. Agaian, "Comprehensive Underwater Object Tracking Benchmark Dataset and Underwater Image Enhancement with GAN," IEEE Journal of Oceanic Engineering, June 2021, doi:10.1109/JOE.2021.3086907. early access: https://ieeexplore.ieee.org/document/9499961	Underwater object tracking	32 videos (24241 frames)	3.76 GB	files	mp4		32 videos	varying	854 x 480	30		YUV (typ. 4:2:0)		
29	UNILJ	Extension of YouTube-8M with human-verified temporal labels for 5-second video segments.	video, audio	https://research.google.com/youtube8m/	Google Research	CC BY 4.0 license	Abu-El-Haija, Sami, et al. "Youtube-8m: A large-scale video classification benchmark." arXiv preprint arXiv:1609.08675 (2016).	Temporal concept localization, segment-level video understanding, video classification	230K to 237K human-verified 5-second segments	N/A	files	tensorflow, TFRecord file	N/A	230K to 237K human-verified 5-second segments	229.6 s	N/A	1	N/A	N/A		Human-verified labels at 5-second segment level
30	UNILJ	CMU-MOSEI Multimodal Opinion Sentiment and Emotion Intensity dataset. 6 Emotions (happiness, sadness, anger, fear, disgust, surprise)	video, audio	https://github.com/CMU-MultiComp-Lab/CMU-MultimodalSDK/tree/main/mmsdk/mmdatasdk/dataset/standard_datasets/CMU_MOSEI	CMU MultiComp Lab	N/A	Zadeh, AmirAli Bagher, et al. "Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018.	sentiment analysis, emotion recognition	23453 video segments	30 GB	files	csd, hdf5	N/A	23453 video segments	7 s	N/A	30	N/A	N/A	moving camera	Visual and acoustic modalities are aligned to words by interpolation

Metadata	Project partner	Dataset description	Modality	Link	Source / Owner	License	Reference	Applications	Number of items	Total size (GB)	Structure	Data format	Compression	Image resolution	Image color space	Bit depth	Image format	Annotation level			
Comment/ example			*video, image, audio, numerical, multimodal*								*files, sequences, tabular, time-series, ...*	*mp3, mp4, ...*			*min/max*						
1	UNBG	CarLogos and CarIDs	image	https://drive.googl	Laboratory for Data Analysis and Applied AI, University of Belgrade, School of Electrical Engineering	CC BY 4.0	/	Object Detection, Image Classification, License Plate Recognition (ALPR/OCR), Vehicle Re-Identification, Traffic Monitoring & Surveillance, Automotive Fraud Detection.	700	2.02 GB	Files group in folders + XLSX file for annotations	JPG	Uncompressed JPEG	2252 x 4000	RGB	8-bit per channel	JPEG	Image-level labels			
2	UNBG	Brand logos and icons dataset	image	https://www.kaggl	/	CC0: Public Domain	/	Image Classification	1991	30 MB	1810 PNG images and 145 images in JPG formats	JPG, PNG	Lossy / Lossless	120x120 or 400x400	RGB	8-bit per channel	JPEG	Image-level labels			
3	UNBG	CIFAR-10 Image Classification	image	https://www.kaggl	University of Toronto, Department of Computer Science	CC0: Public Domain i	Robbie Beane. CIFAR-10 Image Classification. https://kaggle.com/competitions/cifar10-mu, 2021. Kaggle.	Image Classification	60000	136.73 MB	50,000 images in training set and 10,000 images in testing set.	PNG	Lossless	32x32	RGB	8-bit per channel	PNG	Image-level class labels using folders names			
4	UNBG	ImageNet (ILSVRC)	image	https://www.imag	Stanford Vision Lab, Princeton University	Custom (Non-commercial)	J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848	Image Classification, Object Detection	14200000	150	Folder-per-class	Tar / Raw Images	Uncompressed JPEG	224x224 or 256x256	RGB	8-bit per channel	JPEG	Image-level labels, Bounding boxes			
5	UNBG	COCO (Common Objects in Context)	image	https://cocodatas	COCO Consortium	CC BY 4.0	https://doi.org/10.48550/arXiv.1405.0312	Object Detection, Instance Segmentation, Captioning	~328k images (~200k labelling)	~25 GB	Separate Train/Val/Test folders + JSON annotations	JSON + Images	Uncompressed JPEG	Variable (avg 640x485)	RGB	8-bit per channel	JPEG	Object instances (Polygons), Keypoints, Stuff (Background)			
6	UNBG	MNIST - database of handwritten digits	image	https://www.kaggl	Yann LeCun, Corinna Cortes	CC BY-SA 3.0	LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86, no. 11 (1998): 2278-2324.	Handwritten Digit Recognition, Classification	70,000 images (60k train / 10k test)	~11 MB (compressed)	IDX file format (binary matrix)	Binary (IDX)	GZip	28 x 28 pixels	Grayscale	8-bit (0-255)	Proprietary binary (export to PNG)	Image-level class labels (0-9)			
7	UNBG	ChestX-ray14	image	https://nihcc.app.	NIH Clinical Center	Public Domain	Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers, "ChestX-ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2097-2106	Medical Image Classification, Disease Detection	112,120 images from 30,805 pacients	~42 GB	12 compressed tar arhiva + CSV metadata	CSV (annotation) + Images	TAR	1024 x 1024	Grayscale (RGB)	8-bit per channel	PNG	Image-level text-mined labels (14 categories)			
8	UNBG	Fashion-MNIST (Zalando's article images intended as a direct drop-in replacement for the original MNIST database)	image	https://github.co	Zalando Research	MIT	Xiao, Han, Kashif Rasul, and Roland Vollgraf. "Fash	Image Classification, Benchmark	70,000 images	0,03 GB	70,000 files (Split Train/Test IDX files)	Binary (IDX)	GZip	28 x 28 pixels	Grayscale Images	8-bit per channel	Proprietary binary (export to PNG)	Image-level text-mined labels (10 categories: 0 T-shirt/top, 1 Trouser, 2 Pullover, 3 Dress, 4 Coat, 5 Sandal, 6 Shirt, 7 Sneaker, 8 Bag, 9 Ankle boot)			

9	UNBG	CelebA (CelebFaces Attributes) - Large-scale face attributes dataset containing extensive celebrity face images.	image	https://mmlab.ie.c	MMLAB, Chinese University of Hong Kong	Non-commercial research	Liu, Ziwei, Ping Luo, Xiaogang Wang, and Xiaoou Tang. "Deep learning face attributes in the wild." In Proceedings of the IEEE international conference on computer vision, pp. 3730-3738. 2015.	Face Attribute Recognition, Generative Modeling (GANs/Diffusion)	202,599 images	1.5 GB	Flat directory + text annotation files	TXT (Annotations)	Lossy JPEG	178x218 (cropped)	RGB	8-bit per channel	JPEG	40 attribute binary annotations, 5 landmark locations			
10	UNBG	KITTI - Real-world stereo, optical flow, and 2D/3D object tracking data captured from a driving platform.	image	https://www.cvlib	Karlsruhe Institute of Technology & Toyota Technological Institute Chicago	CC BY-NC-SA 3.0	Geiger, Andreas, Philip Lenz, Christoph Stiller, and Raquel Urtasun. "Vision meets robotics: The kitti dataset." The international journal of robotics research 32, no. 11 (2013): 1231-1237.	3D Object Detection, Autonomous Driving, Visual Odometry	15,000 frames total (7,481 train, 7,518 test)	12 GB (Object set)	Separated directories for image_2, velodyne, and labels	TXT (Labels), Binary (LiDAR)	Lossless PNG	1242x375 (approx)	RGB	8-bit per channel	PNG	2D/3D Bounding boxes (text matrices)			
11	UNBG	PASCAL VOC 2012 - Historical benchmark dataset providing standardized image sets for object class recognition	image	http://host.robots.	PASCAL VOC Consortium	Public Domain / Open Research	Everingham, Mark, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. "The pascal visual object classes (voc) challenge." International journal of computer vision 88, no. 2 (2010): 303-338.	Object Detection, Semantic Segmentation	11,530 images (27,450 object instances)	2.6 GB	XML metadata paired with source folders (JPEGImages, SegmentationClass)	XML	Lossy JPEG	Variable (~500x375)	RGB	8-bit per channel	JPEG	Bounding box coordinates, pixel-level segmentations			
12	UNBG	SVHN (Street View House Numbers) - Real-world image dataset for developing machine learning and object recognition algorithms	image	http://ufldl.stanfor	Stanford University / Google	Non-commercial research usage	Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, Andrew Y. Ng Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011.	OCR, Digit Classification, Object Localization	600,000+ digits	2.1	Digit structs mapped inside .mat binaries	MATLAB (.mat)	Lossless (PNG/Mat)	32x32 (cropped) or Full scenes	RGB	8-bit per channel	PNG	Character-level bounding boxes and digit labels			
13	UNBG	The Oxford-IIIT Pet Dataset: A 37-category pet dataset with roughly 200 images for each breed class	image	https://www.robo	Visual Geometry Group, Oxford University	CC BY-SA 4.0	Parkhi, Omkar M., Andrea Vedaldi, Andrew Zisserman, and C. V. Jawahar. "Cats and dogs." In 2012 IEEE conference on computer vision and pattern recognition, pp. 3498-3505. IEEE, 2012.	Fine-grained Classification, Semantic Segmentation	7,349 images	0.8	Separated folders for images and annotations (XML/PNG masks)	XML / PNG	Lossy JPEG	Variable	RGB	8-bit per channel	JPEG	Breed labels, Head-ROI bounding boxes, Pixel-level trimap segmentations			
14	UNBG	STL-10: Image recognition dataset inspired by CIFAR-10 but modified for unsupervised deep learning	image	https://cs.stanford	Stanford University (Adam Coates)	Open Source Research	Adam Coates, Honglak Lee, Andrew Y. Ng An Analysis of Single Layer Networks in Unsupervised Feature Learning AISTATS, 2011.	Unsupervised Learning, Self-Supervised Benchmarking	113,000 images	2.5	Flat binary arrays split into train, test, and unlabeled matrices	Binary format	Uncompressed raw binary	96x96	RGB	8-bit per channel	Raw bytes	Image-level labels (10 classes) for labeled partition			
15	UNBG	Places365: A repository of scene-centric images mapping environment categories	image	http://places2.csai	MIT CSAIL	CC BY 4.0	B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million Image Database for Scene Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.	Scene Classification, Contextual Visual Modeling	~10 million	105	Folder-per-category schema + text metadata indices	TXT	Lossy JPEG	256x256 or High-Res native	RGB	8-bit per channel	JPEG	Image-level scene category labels (365 categories)			
16	UNBG	LFW (Labeled Faces in the Wild): Database of face photographs designed for studying unconstrained face recognition	image	https://www.kaggl	University of Massachusetts, Amherst	Free for non-commercial research	Huang, Gary B., and Erik Learned-Miller. "Labeled faces in the wild: Updates and new reporting procedures." Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep 14, no. 003 (2014): 17.	Face Verification, Face Clustering	13,233 faces	0.18	Folders named by individual identities containing JPEG files	Text labels	Lossy JPEG	250x250	RGB	8-bit per channel	JPEG	Name identities (5,749 unique people labels)			
17	UNBG	Caltech 101: Classic object category dataset consisting of pictures representing distinct item classes	image	https:// data.caltech.edu/ records/ mzrjq-6wc02 | https://www.kaggl e.com/datasets/i mbikramsaha/calt ech-101	California Institute of Technology	Creative Commons Attribution 4.0	Fei-Fei, Li, Rob Fergus, and Pietro Perona. "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories." In 2004 conference on computer vision and pattern recognition workshop, pp. 178-178. IEEE, 2004.	Object Classification, Early Model Benchmark	9,146 images	0.13	Category folder layout	Folder structured	Lossy JPEG	Variable (~300x200)	RGB/Grayscale mixed	8-bit per channel	JPEG	Category labels (101 target objects + 1 background class)			
18	UNBG	EuroSAT - Land use and land cover classification challenge utilizing Sentinel-2 satellite observation data	image	https:// github.com/ phelber/EuroSAT | https://zenodo.or g/records/771181 0#.ZAm3k-zMKEA	German Research Center for Artificial Intelligence (DFKI)	MIT License	Helber, Patrick, Benjamin Bischke, Andreas Dengel, and Damian Borth. "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, no. 7 (2019): 2217-2226.	Remote Sensing, Earth Observation, Semantic Scene Mapping	27,000 patches		Directory structure separated by land type	2 classes	Folder structured	Uncompressed TIFF / Lossy JPEG	64x64	13 bands (Multispectral) / RGB	16-bit unsigned (TIFF) / 8-bit (JPEG)	TIFF / JPEG	Land cover patch category labels (10 classes)			
19	UNBG	Cityscapes Dataset - High-fidelity video sequences recording urban street views from automotive cockpits	image	https://www.citys	Daimler AG, Max Planck Institute, TU Darmstadt	Custom Non-Commercial Academic	M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.	Semantic Segmentation, Instance Segmentation, Autonomous Navigation	25,000 frames total		Split hierarchy by city, partitioned into leftImg8bit and gtFine	12	PNG (Masks/Images)	Lossless PNG	2048x1024	RGB	8-bit per channel	PNG	Pixel-level dense annotations, instance IDs (30 classes)			
20	UNBG	SUN RGB-D - indoor scene spatial dataset containing depth orientation visuals for scene comprehension models	image	https://rgbd.cs.pri	Princeton University Vision Group	Free for academic use	Song, Shuran, Samuel P. Lichtenberg, and Jianxiong Xiao. "Sun rgb-d: A rgb-d scene understanding benchmark suite." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 567-576. 2015.	3D Object Detection, Indoor Room Layout Estimation	10,335 pairs		Frame-indexed folders holding image directories and MATLAB orientation matrices	35	MATLAB / PNG metadata	Lossy JPEG / Lossless PNG	Variable	RGB + Depth metrics	8-bit (RGB) / 16-bit (Depth)	JPEG + PNG	3D Oriented bounding boxes, semantic labels			
21	UNBG	Comprehensive Cars (CompCars) - Large-scale dataset for fine-grained car model tasks	image	https://mmlab.ie.c	Multimedia Laboratory (MMLAB), Chinese University of Hong Kong	Non-commercial research	Linjie Yang, Ping Luo, Chen Change Loy, Xiaoou Tang. A Large-Scale Car Dataset for Fine-Grained Categorization and Verification, In Computer Vision and Pattern Recognition (CVPR), 2015.	Fine-grained Vehicle Classification, Attribute Prediction	186,726 images		Folders grouped by structural design year or source viewpoints	15	TXT (Attributes)	Lossy JPEG	Variable (High-res native)	RGB	8-bit per channel	JPEG	Car make/model hierarchies, viewpoints, 5 attribute annotations			
22	UNBG	PlantVillage Dataset - Open access health inventory mapping leafy images across agricultural yield species	image	https://github.co	Penn State University (David Hughes, Marcel Salathé)	CC0 1.0 Public Domain	Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." Frontiers in plant science 7 (2016): 1419.	Agricultural Computer Vision, Plant Disease Diagnostics	54,306 images		Categorized class subfolders	0.83	Folder structured	Lossy JPEG	256x256 standardized profiles	RGB	8-bit per channel	JPEG	Image-level crop-species disease class tags (38 target labels)			
23	UNBG	DeepFashion - Large-scale apparel repository detailing visual descriptors across multi-pose consumer settings	image	https://mmlab.ie.c	Multimedia Laboratory (MMLAB), Chinese University of Hong Kong	Non-commercial research license	Liu, Ziwei, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1096-1104. 2016.	Fashion Landmark Detection, Cross-Domain Category Retrieval	800,000+ entries		Segmented zip archives mapping multi-task attributes via indices	35	TXT indices	Lossy JPEG	Variable	RGB	8-bit per channel	JPEG	Category tags, attribute descriptors, clothing bounding frames, spatial landmarks			
24	UNBG	WIDER FACE - Face detection benchmark containing images selected from the publicly available WIDER dataset	image	http://shuoyang12	The Chinese University of Hong Kong	Open Research License	Yang, Shuo, Ping Luo, Chen-Change Loy, and Xiaoou Tang. "Wider face: A face detection benchmark." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5525-5533. 2016.	Face Detection, Complex Occlusion Localization	32,203 images (393,703 human faces labeled)		Target event category directories (e.g., Parade, Meeting) + text label maps	3.4	TXT labels	Lossy JPEG	Variable	RGB	8-bit per channel	JPEG	Facial boundary box matrices with occlusion/pose tags			
25	UNBG	MPII Human Pose Dataset - State-of-the-art benchmark for evaluation of articulated human pose estimation	image	https://www.mpi-i	Max Planck Institute for Informatics	CC BY 4.0	Andriluka, Mykhaylo, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. "2d human pose estimation: New benchmark and state of the art analysis." In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, pp. 3686-3693. 2014.	Human Pose Estimation, Activity Recognition	~25,000 images (over 40,000 people profiles)		Unpacked image directories paired with centralized MATLAB annotations	12.5	MATLAB (.mat)	Lossy JPEG	Variable	RGB	8-bit per channel	JPEG	16 joint coordinates, body orientation angles, activity class indices			
26	UNBG	KITTI-360: subsequent large-scale evolution of driving telemetry offering surround-view environments	image	https://www.cvlib	Karlsruhe Institute of Technology & MPI Informatics	CC BY-NC-SA 4.0	Liao, Yiyi, Jun Xie, and Andreas Geiger. "Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 3 (2022): 3292-3310.	3D Scene Understanding, Semantic Instance Mapping, Novel View Synthesis	Over 320,000 frames		Structured log-sequence folders split by camera stream arrays	150	PLY (3D Mesh arrays)	Lossless PNG	1400x1400 (Fisheye) / 1418x1398	RGB	8-bit per channel	PNG	Unified 2D/3D semantic instance segmentations			
27	UNBG	LSUN (Large-Scale Scene Understanding): Large-scale dataset for scene understanding and generative model benchmarks	image	https://github.co	Princeton University / Fisher Yu	Free Research Usage	Yu, Fisher, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop." arXiv preprint arXiv:1506.03365 (2015).	Image Generation, Large-scale Scene Classification	~10 million images across 10 scene & 20 object classes		Category-specific LMDB binary files	400	LMDB	Lossy JPEG byte streams stored inside LMDB	Variable native	RGB	8-bit per channel	LMDB encoded	Scene category indexing			
28	UNBG	COCO-Stuff: Augmentation of original Microsoft COCO schema tracking contextual background 'stuff' textures	image	https://github.co	University of Edinburgh (Holger Caesar)	CC BY 4.0	H. Caesar, J. Uijlings, V. Ferrari, In Computer Vision and Pattern Recognition (CVPR), 2018. https://doi.org/10.48550/arXiv.1612.03716	Panoptic Segmentation, Scene Context Parsers	164,000 images		Flat mask asset files matching base COCO filename index tracks	25	JSON / PNG	Lossless PNG	Matches base COCO	RGB / Palette indexed	8-bit per channel	PNG	Comprehensive pixel-level annotations covering 91 thing and 91 stuff classes			
29	UNBG	ADE20K Dataset - High-density dataset for scene parsing, segmentation, and object localization.	image	https://ade20k.csa	MIT CSAIL Vision Group	CC BY-NC-SA 4.0	Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso and Antonio Torralba, "Scene Parsing through ADE20K Dataset," Computer Vision and Pattern Recognition (CVPR), 2017.	Semantic Segmentation, Panoptic Scene Parsing	25,214 images		Folders partitioned into index mappings split by training/validation paths	3.8	Folder structured	Lossy JPEG / Lossless PNG labels	Variable	RGB	8-bit per channel	JPEG / PNG	High-density dense pixel-level class masks, object part decompositions			
30	UNBG	Stanford Cars Dataset: Fine-grained car image collection classified down to specific model iterations and design eras	image	https:// github.com/ cyizhuo/ Stanford_Cars_dat aset | https://www.kaggl e.com/datasets/e duardo4jesus/stan ford-cars-dataset	Stanford AI Lab (Jonathan Krause)	Custom Academic Research Usage	Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, "3D Object Representations for Fine-Grained Categorization", 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013.	Fine-grained Image Classification	16,185 images		Split train/test folders paired with a central metadata file (car_devkit)	2	MATLAB (.mat)	Lossy JPEG	Variable (High-res)	RGB	8-bit per channel	JPEG	Bounding box dimensions mapped with fine make/model indices			

Metadata	Project partner	Dataset description	Modality	Link	Source / Owner	License	Reference	Applications	Number of items	Total size (GB)	Structure	Data format	Compression	No. of audio files	Audio duration	Sampling rate	Channels	Bit depth	Audio format	Temporal annotations
Comment/ example			*video, image, audio, numerical, multimodal*								*files, sequences, tabular, time-series, …*	*mp3, mp4, …*			*min/max/median*	*Hz*				

1	UoM	GTZAN is an audio dataset intended for music genre classification. It contains 1,000 audio recordings, each lasting 30 seconds, organized into 10 genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Each genre is represented by 100 audio examples.	audio / sound	https://www.kaggl	Tzanetakis and Cook, 2002	Available for research and education, widely used	Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293–302. Yadav, A. (2025). Music genre classification using audio features: A machine learning approach with the GTZAN dataset. IOSR Journal of Computer Engineering (IOSR-JCE), 27(3, Ser. 3), 05–16.	Music genre recognition	1000	~1.2GB	files	WAV audio files (.w	Uncompressed	1000	30 sec	22,050 Hz	Mono	not specified	WAV	Not provided	
2	UoM	IRMAS - is an audio dataset designed for automatic recognition of predominant musical instruments in audio signals. It contains musical audio excerpts annotated with the predominant instrument(s) present. It contains 9,579 audio excerpts divided into training and testing sets, with annotations for 11 instrument classes	audio / sound	https://www.upf.e	Dataset by Juan J. Bosch, Ferdinand Fuhrmann, Perfecto Herrera. Music Technology Group - Universitat Pompeu Fabra	Free of charge for non-commercial use only	Bosch, J. J., Janer, J., Fuhrmann, F., & Herrera, P. "A Comparison of Sound Segregation Techniques for Predominant Instrument Recognition in Musical Audio Signals", in Proc. ISMIR (pp. 559-564), 2012	Musical instrument recognition	9,579	~11GB	files	WAV audio files (.w	Uncompressed	9579	3 s for training files; 5-20 s for testing files.	44100 KHz	Stereo	16 bit	WAV	Not provided	
3	UoM	UrbanSound 8K - This dataset contains 8732 labeled sound excerpts (<=4s) of urban sounds from 10 classes: air_conditioner, car_horn, children_playing, dog_bark, drilling, enginge_idling, gun_shot, jackhammer, siren, and street_music. The classes are drawn from the urban sound taxonomy.	audio / sound	https://urbansoun	Justin Salamon, Christopher Jacoby and Juan Pablo Bello, New York University, Music and Audio Research Laboratory.	UrbanSound8K datasets are offered free of charge for non-commercial use only under the terms of the Creative Commons Attribution Noncommercial License (by-nc), version 3.0:	J. Salamon, C. Jacoby and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research", 22nd ACM International Conference on Multimedia, Orlando USA, Nov. 2014.	Classification and recognition of urban environmental sounds.	8732	~6GB	Files + tabular metadata	WAV audio files (.w	Uncompressed	8732	up to 4 s	Variable - 8,000 Hz		n/a	WAV	Yes - start and end times are provided for each sound excerpt.	
4	UoM	VS13 - is an audio-video dataset designed for vehicle speed estimation using video and/or sound recordings. It contains recordings of 13 different vehicles passing by a camera at known constant speeds, ranging from 30 to 105 km/h. The dataset includes video files, extracted audio recordings, and annotations with vehicle speed and passing time, making it suitable for research in signal processing, machine learning, and intelligent transportation systems.	audio / sound+video	https://slobodan.u	Slobodan Djukanović, Nikola Bulatović, Ivana Čavor, Faculty of Electrical Engineering, University of Montenegro (@inproceedings{djukanovic2022dataset, author = "Slobodan Djukanovi\'{c} and Nikola Bulatovi\'{c} and Ivana \v{C}avor", title = "A dataset for audio-video based vehicle speed estimation", booktitle = "2022 30th Telecommunications Forum (TELFOR)", year = "2022", pages = "1-4" })	The dataset is publicly available and freely accessible for research purposes, but no formal dataset license is explicitly specified.	S. Djukanović, N. Bulatović, and I. Čavor, "A dataset for audio-video based vehicle speed estimation," in 2022 30th Telecommunications Forum (TELFOR), Belgrade, Serbia, 2022, pp. 1–4, doi: 10.1109/TELFOR56187.2022.9983773.	Vehicle speed estimation using audio-video data	Number of items: 400 audio-video recordings Number of video files: 400 Number of audio files: 400 extracted audio files	~15 GB	files	MP4 video files, WAV audio files, and TXT annotation files	video files are compressed in MP4 format audio files are provided as WAV files	Number of video files: 400 Number of audio files: 400	10 s	44,100 Hz	Not explicitly specif	32-bit float PCM	WAV	Each annotation file contains the vehicle speed and the pass-by-camera instant, given in seconds from the beginning of the recording.	
5	UoM	Sopele sounds - Sopele music dataset is an audio dataset created for automatic music transcription of traditional Croatian woodwind instruments called sopele. It contains WAV recordings of individual tones of small and great sopela, their polyphonic tone combinations, and real traditional music pieces accompanied by corresponding PDF music sheets. The recordings were acquired using a mobile phone microphone at a sampling rate of 44.1 kHz. The dataset is intended for developing, training and evaluating AMT algorithms for monophonic and polyphonic traditional woodwind music.	audio / sound	https://www.scien	University of Rijeka, Faculty of Engineering, Department of Computer Engineering, Croatia	CC BY 4.0	A. Skoki, S. Ljubic, J. Lerga, I. Štajduhar, Automatic music transcription for traditional woodwind instruments sopele, Pattern Recognition Letters, Volume 128, 2019, pp. 340–347, https://doi.org/10.1016/j.patrec.2019.09.024 Skoki, A., Ljubic, S., Lerga, J., & Štajduhar, I. (2020). Sopele music dataset. Data in Brief, 28, 104840. https://doi.org/10.1016/j.dib.2019.104840	Automatic music transcription (AMT), music information retrieval, pitch detection, tone classification, development and evaluation of machine learning algorithms for transcription of traditional woodwind music.	Combined tones: 36 Real pieces: 15 Single tones: 38		files	Audio WAV files and PDF music sheets; raw and analysed data	Uncompressed WAV audio files		21–38 s per single	44,100 Hz	Single-tone and monophonic real-piece recordings are mono, while polyphonic combinations of small and great sopela are provided as stereo files.		WAV		
6	UoM	Xeno-canto is an online audio database and sound archive that provides access to wildlife sound recordings from around the world, especially bird sounds. The recordings are contributed by a global community of amateur and professional recordists and include metadata such as species, location and recording information. The dataset is widely used in bioacoustics, bird sound classification, biodiversity monitoring and machine learning research for automatic recognition of animal vocalizations.	audio / sound	https://xeno-canto	Xeno-canto Foundation	CC (mixed: BY, BY-NC-SA, CC0)	Vellinga, W.-P., Planque, R. "The Xeno-canto collection and its relation to sound recognition and classification." CLEF (2015).	Bird species ID, bioacoustics	800,000+ recordings	several TB (full)	files + metadata	mp3 / wav	lossy / PCM	800,000+ clips	seconds to minutes	44100 (typical)	mono / stereo	lossy / 16-bit	mp3 / wav	yes (species labels)	
7	UoM	FMA - Free Music Archive Dataset - s a large-scale audio dataset for music analysis and Music Information Retrieval (MIR). It contains Creative Commons-licensed music tracks with metadata, genre labels, tags, and pre-computed audio features. The dataset is suitable for tasks such as music genre classification, music recommendation, tagging, and audio-based music analysis.	audio / sound	https://github.co	GitHub (M. Defferrard et al., EPFL)	CC BY 4.0	Defferrard, M., et al. "FMA: A Dataset for Music Analysis." Proc. ISMIR (2017).	Music classification, MIR, tagging	106,574 tracks	~900 GB (full); 7.2 GB (small)	files + csv	mp3	lossy	106,574 clips	30 s (small) / full	44100 Hz	stereo	lossy (mp3)	mp3	yes (genre/metadata)	
8	UoM	ESC-50 - 2,000 environmental sound recordings labeled across 50 balanced classes (animals, nature, human, urban, interior).	audio / sound	https://github.co	GitHub (K. J. Piczak)	CC BY-NC 3.0	Piczak, K. J. "ESC: Dataset for Environmental Sound Classification." Proc. ACM Multimedia (2015): 1015-1018.	Environmental sound classification	2,000 clips	~600 MB	files	wav	no (PCM)	2,000 clips	5 s	44100 Hz	mono	16-bit	wav	no temporal annotations; class labels are provided at track level	
9	UoM	LibriSpeech - 1,000 hours of read English audiobook speech from LibriVox, segmented and aligned for ASR benchmarking.	audio / sound	https://www.open	OpenSLR (V. Panayotov, D. Povey)	CC BY 4.0	Panayotov, V., et al. "LibriSpeech: an ASR corpus based on public domain audio books." Proc. ICASSP (2015): 5206-5210.	Automatic speech recognition	~1,000 hours	60 GB	files + transcripts	flac	Yes – lossless FLAC compression	~292,000 utterances	avg ~12 s	16000 Hz	mono	16-bit	flac	yes (utterance transcripts)	
10	UoM	Nsynth - 305,979 musical notes from 1,006 instruments, each a 4-second sample annotated with pitch, velocity and instrument family.	audio / sound	https://magenta.t	Google Magenta	CC BY 4.0	Engel, J., et al. "Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders." Proc. ICML (2017).	Audio synthesis, instrument classification	305,979 notes	~70 GB	files + json	wav	no (PCM)	305,979 clips	4 s	16000 Hz	mono	16-bit	wav	yes (pitch/instrument labels)	
11	UoM	VoxCeleb - Real-world speaker recordings extracted from YouTube interviews of celebrities; standard speaker verification benchmark.	audio / sound	https://www.robo	Visual Geometry Group, University of Oxford	CC BY 4.0	Nagrani, A., Chung, J. S., Zisserman, A. "VoxCeleb: a large-scale speaker identification dataset." Proc. Interspeech (2017).	Speaker recognition/verification	1.28M utterances (V1+V2)	~270 GB	files	wav / m4a (aac)	lossy/lossless	1,281,762 utterances	avg ~8 s	48,000 Hz	mono	16-bit	wav / m4a	yes (speaker labels)	
12	UoM	MedleyDB - 122 royalty-free multitrack recordings with per-stem audio and melody/instrument annotations.	audio / sound	https://medleydb.	Music and Audio Research Lab (MARL), NYU	CC BY-NC-SA 4.0	Bittner, R., et al. "MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research." Proc. ISMIR (2014).	Melody extraction, source separation	122 multitracks	~30 GB	files (stems + mix)	wav	no (PCM)	122 songs (+stems)	3-5 min	44100 Hz	stereo	16-bit	wav	yes (melody/instrument)	
13	UoM	AudioSet Ontology- AudioSet Ontology is a hierarchical vocabulary of sound event classes developed by Google. It organizes a wide range of everyday sounds, including human and animal sounds, musical instruments, music genres, vehicles, environmental sounds, and other acoustic events. It is used together with the AudioSet dataset for audio event classification, sound tagging, and machine learning research in general-purpose audio recognition.	audio / sound	https://research.g	Google Research	Free of charge for non-commercial use only Attribution 4.0 International	Giacomelli, S., Giordano, M., Rinaldi, C., & Graziosi, F. "AudioSet - Tools: a Phyton framework taxonomy-aware AudioSet curation and reproducible audio research", EURASIP Journal on Audio Speech and Music Processing, pp. 1-29, 2026	Audio event detection and sound classification	2000000	~2.4 GB pre-computed features	files	CSV text files and pre-extracted 128-dimensional audio features in TensorFlow Record format	Uncompressed	2000000	Total length 5800 h (single files 10 sec lenght)	16 KHz	mono	8 bit	CSV text files and pre-extracted 128-dimensional audio features in TensorFlow Record format		
14	UoM	FSD50K - is an open dataset of human-labeled sound events based on audio clips from Freesound. It contains 51,197 audio clips, totaling over 100 hours of audio, annotated with 200 sound classes from the AudioSet Ontology. The dataset is suitable for sound event classification, audio tagging, environmental sound recognition, and machine learning research in general audio analysis.	audio / sound	https://fsanngtato	Music Technology Group at Universitat Pompeu Fabra	Free of charge for non-commercial use only Attribution 4.0 International	Fonseca, E., et al.,FSD50K: An Open Dataset of Human-Labeled Sound Events. *IEEE/ACM Transactions on Audio, Speech, and Language Processin, 2005*	Sound Event Recognition	51197	34,5 GB	files	WAV audio files (.wav)	Uncompressed	51197	Total length 108 h (single files from 0,3 to 30 sec lenght)	44,1 KHz	mono	16 bit	WAV audio files (.w		

15	UoM	MUSAN - is an audio corpus containing music, speech, and noise recordings. It is mainly used for speech and audio processing tasks, especially for training and evaluating systems in speaker recognition, speech recognition, speech enhancement, and noise robustness research.	audio / sound	https://openslr.org/17/	Center for Language and Speech Processing, The Johns Hopkins University	Free of charge for non-commercial use only Attribution 4.0 International	Snyder, D., Chen, G., & Povey, D. "MUSAN: A Music, Speech and Noise Corpus", pp. 1-4, 2015	Musical, sspeach and noise recognition	Noise - 929 files Music - 660 files Speech - 427 files	~11GB	files	WAV audio files (.wav)	Uncompressed	Noise - 929 files Music - 660 files Speech - 427 files	Speech - total length 60h Music - total length 42h (single files form 10 sec to several minutes) Noise - total length 6 h (single files from 1 to 30 sec)	16 KHz	mono	16 bit	WAV		
16	UoM	MagnaTagATune - is a music audio dataset for Music Information Retrieval (MIR), especially automatic music tagging and music similarity research. It contains short MP3 music clips from the Magnatune label, together with human-generated tag annotations collected through the TagATune game. The tags describe musical content such as instruments, genre, vocals, mood, and other audio characteristics.	audio / sound	https://mirg.city.a	City University London	Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported	Law, L., West, K., Mandel, M., Bay, M. & Downie, J.	Benchmarking tool in Music Information Retrieval (MIR) and Audio Machine Learning	25,863	3 GB	file	MP3 audio file	Compressed	25,863	Total length 209 h (single files 29 sec lenght)	16 KHz	mono	Decoding dependant	MP3 audio file		
17	UoM	CSTR VCTK Corpus is an English multi-speaker speech dataset containing recordings from 110 speakers with various accents. Each speaker reads approximately 400 sentences selected to provide broad phonetic coverage. The corpus includes audio recordings, corresponding text transcriptions and speaker-related information. It was originally developed for speaker-adaptive and multi-speaker text-to-speech synthesis and can also support speaker identification, accent analysis and speech-recognition research.	audio / sound	https://datashare.ed.ac.uk/items/30e7453c-9ea8-48b4-8e18-f96d0dc62928	Centre for Speech Tech	CC BY 4.0	Yamagishi, J., Veaux, C., and MacDonald, K. (2019).	Multi-speaker text-to-spec	44,455 utterances – mic1 configuration	10.94 GB compress	Audio files organised by speaker, with corresponding text transcriptions and speaker metadata	WAV audio files and TXT transcriptions	ZIP archive; unco	44,455 – mic1 confi	Approximately 44 h	48,000 Hz	mono	16 bit	WAV	Text transcriptions,	
18	UoM	MTG-Jamendo - is a large open music dataset for automatic music tagging. It contains over 55,000 full music tracks from Jamendo, available under Creative Commons licenses, with 195 tags covering genre, instruments, and mood/theme categories. The dataset is suitable for music classification, auto-tagging, music recommendation, and Music Information Retrieval research.	audio / sound	https://mtg.github.io/mtg-jamendo-dataset/	Music Technology Group (MTG) at Universitat Pompeu Fabra	Creative Commons License	Bogdanov, D., Won, M., Tovstogan, P., Porter, Al.,	Build and evaluate machine learning models	55000	508 GB	file	MP3 audio file	Compressed	55000	Total length 3777 h (single files form 30 sec to 10 min lenght)	44.1 KHz raw 16 KHz pre-computed	mono	16 bit	MP3 audio file		
19	UoM	Heartbeat Sounds - is an audio dataset containing stethoscope recordings of heart sounds, originally used for a machine learning challenge focused on classifying heartbeat anomalies. The recordings come from two sources: public recordings collected through the iStethoscope Pro iPhone app and clinical recordings collected in hospitals using the DigiScope digital stethoscope. The dataset is used for heart sound classification, anomaly detection, and biomedical audio analysis.	audio / sound	https://www.kaggle.com/datasets/kinguistics/heartbeat-sounds	Kaggle (Bentley et al., PASCAL CHSC)	Open (research use)	Bentley, P., et al. "The PASCAL Classifying Heart Sounds Challenge (CHSC2011)." (2011).	Heartbeat classification, anomaly detection	~830 recordings	~200 MB	files + labels	wav / aif	no (PCM)	~830 files	1-30 s	varies (4000 / 44100)	mono	16-bit	wav / aif	yes (S1/S2 locations)	
20	UoM	TAU Urban Acoustic Scenes 2020 Mobile is an audio dataset for acoustic scene classification. It contains 10-second audio recordings from 10 urban acoustic scenes, such as airport, shopping mall, metro station, pedestrian street, public square, traffic street, tram, bus, metro, and urban park. The dataset is used for computational auditory scene analysis, environmental sound recognition, and machine learning classification of everyday urban soundscapes.	audio / sound	https://zenodo.org/records/3819968	Zenodo / Tampere University (DCASE)	Non-commercial research (custom)	Heittola, T., Mesaros, A., Virtanen, T. "TAU Urban Acoustic Scenes 2020 Mobile." DCASE 2020 / Zenodo.	Acoustic scene classification	23,040 segments	~30 GB	files + meta	wav	no (PCM)	23,040 clips	10 s	44100 Hz	mono (mobile)	24-bit	wav	yes (scene + device)	
21	UoM	Google Speech Commands - One-second utterances of 35 short command words (yes, no, up, down, etc.) from thousands of speakers; keyword-spotting benchmark.	audio / sound	https://www.tens	Google / TensorFlow	CC BY 4.0	Warden, P. "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition." arXiv:1804.03209 (2018).	Keyword spotting, on-device ASR	~105,000 clips	2.3 GB	files (by word)	wav	no (PCM)	105,829 clips	1 s	16000 Hz	mono	16-bit	wav	yes (word labels)	
22	UoM	BirdVox-DCASE-20k is an audio dataset for bird audio detection and bioacoustic classification. It contains 20,000 ten-second mono WAV recordings collected by autonomous recording units near Ithaca, New York, USA. Around half of the recordings contain at least one bird vocalization, such as song, call, or chatter. The dataset is used for bird sound detection, acoustic event detection, machine listening, and bioacoustic monitoring	audio / sound	https://zenodo.org/records/1208080	Zenodo / BirdVox (NYU)	CC BY 4.0	Lostanlen, V., et al. "BirdVox-DCASE-20k: a dataset for bird audio detection in 10-second clips." Zenodo (2018).	Bird audio detection (bioacoustics)	20,000 clips	~13 GB	files + csv	wav	no (PCM)	20,000 clips	10 s	44100 Hz	mono	16-bit	wav	yes (presence/absence labels)	
23	UoM	RAVDESS - The Ryerson Audio-Visual Database of Emotional Speech and Song is a multimodal audio-visual dataset for emotion recognition. It contains speech and song recordings performed by 24 professional actors, expressing different emotions such as calm, happy, sad, angry, fearful, surprise, and disgust. The dataset includes audio-only, video-only, and audio-video files, and is widely used for research in speech emotion recognition, facial expression analysis, and affective computing.	audio / sound	https://doi.org/10.5281/zenodo.1188976	Zenodo (Ryerson University, Livingstone & Russo)	CC BY-NC-SA 4.0	Livingstone, S. R., Russo, F. A. "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)." PLoS ONE 13.5 (2018): e0196391.	Speech emotion recognition	2,452 audio-only files	~24.8 GB (full A/V)	files	wav	no (PCM)	2,452 (audio-only)	3-5 s	48000 Hz	stereo	16-bit	wav	yes (emotion labels)	
24	UoM	CREMA-D - is a multimodal audio-visual dataset containing emotional speech clips from 91 actors, used for speech and facial emotion recognition. 7,442 emotional speech clips from 91 actors of diverse ethnicities, 6 emotions and 4 intensity levels.	audio / sound	https://github.com/CheyneyComputer Science/CREMA-D	Cheyney Computer Science (Cao et al.)	Open Database License (ODbL)	Cao, H., et al. "CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset." IEEE Trans. Affective Computing 5.4 (2014): 377-390.	Speech emotion recognition	7,442 clips	~1.6 GB (audio)	files	wav	no (PCM)	7,442 clips	2-3 s	16000 Hz	mono	16-bit	wav	yes (emotion labels)	
25	UoM	AudioSet is a large-scale audio dataset developed by Google Research for audio event recognition and sound classification. It contains over 2 million human-labeled 10-second audio clips extracted from YouTube videos. The dataset covers a wide range of sound event categories, including human sounds, animal sounds, musical instruments, music genres, natural sounds, environmental sounds, and everyday acoustic events. The classes are organized using a hierarchical audio event ontology.	audio	https://research.g	Google Research	CC BY 4.0 (labels)	Gemmeke, J. F., et al. "Audio Set: An ontology and human-labeled dataset for audio events." Proc. ICASSP (2017): 776-780.	Audio tagging, sound event detection	~2,000,000 clips	~2.4 TB (audio)	csv labels + YouTube IDs	(YouTube source)	lossy	2,084,320 clips	10 s	varies (~44100)	varies	lossy	mp4/webm (source)	yes (clip-level weak labels)	
26	UoM	LJ Speech Dataset is a public-domain English speech dataset designed mainly for text-to-speech (TTS) and speech synthesis research. It contains 13,100 short audio clips of a single female speaker reading passages from 7 non-fiction books. Each audio clip has a corresponding text transcription. The clips are between 1 and 10 seconds long, with a total duration of about 24 hours of speech. The audio was recorded by the LibriVox project, and the dataset is widely used for training and evaluating speech synthesis models.	audio / sound	https://keithito.co	Keith Ito (LibriVox public-domain recordings)	Public Domain	Ito, K., Johnson, L. "The LJ Speech Dataset." (2017). https://keithito.com/LJ-Speech-Dataset/	Text-to-speech synthesis	~24 hours	2.6 GB	files + metadata	wav	no (PCM)	13,100 clips	1-10 s	22050 Hz	mono	16-bit	wav	yes (transcripts)	

27	UoM	Free Spoken Digit Dataset (FSDD) is a small open audio/speech dataset containing recordings of spoken digits from 0 to 9. It is often described as an audio version of MNIST, because it is simple and suitable for beginners in audio classification and speech recognition. The recordings are stored as WAV files, sampled at 8 kHz, and trimmed to contain minimal silence at the beginning and end. The dataset is used for spoken digit recognition, audio classification, speech recognition experiments, and machine learning education.	audio / sound	https://github.co	GitHub (Jakobovski et al.)	CC BY-SA 4.0	Jackson, Z., et al. "Free Spoken Digit Dataset (FSDD)." GitHub (2018).	Spoken digit classification, teaching	3,000 recordings	~10 MB	files	wav	no (PCM)	3,000 clips	<1 s	8000 Hz	mono	16-bit	wav	yes (digit labels)	
28	UoM	MAESTRO (MIDI and Audio Edited for Synchronous TRacks and Organization) is a music dataset developed by Google Magenta for piano music transcription, generation, and audio synthesis. It contains about 200 hours of paired audio and MIDI recordings of virtuoso piano performances from the International Piano-e-Competition. The audio and MIDI files are finely aligned, with approximately 3 ms accuracy, and each musical piece includes metadata such as composer, title, and year of performance. The dataset is widely used for automatic piano transcription, music generation, expressive performance modeling, and Music Information Retrieval.	audio / sound	https://magenta.t	Google Magenta	CC BY-NC-SA 4.0	Hawthorne, C., et al. "Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset." Proc. ICLR (2019).	Music transcription, generation	~200 hours	~120 GB (wav)	files (audio + MIDI)	wav + midi	no (PCM)	1,276 performances	min to several min	44100/48000	stereo	16-bit	wav + midi	yes (aligned MIDI)	
29	UoM	MUSDB18 is a music audio dataset designed for music source separation research. It contains 150 full-length music tracks from different genres, with both the complete stereo mixture and separated source stems. The main stems are vocals, drums, bass, and other accompaniment. The dataset is divided into 100 training tracks and 50 test tracks, with a total duration of about 10 hours. The compressed version is distributed in STEMS format (.mp4), where each file contains several stereo audio streams. It is widely used as a benchmark for evaluating music source separation algorithms.	audio / sound	https://sigsep.gith	Zenodo / SiSEC (Z. Rafii et al.)	CC BY-NC-SA 4.0	Rafii, Z., et al. "The MUSDB18 corpus for music separation." Zenodo (2017). https://doi.org/10.5281/zenodo.1117372. Stöter, F.-R., Liutkus, A., & Ito, N. (2018). The 2018 Signal Separation Evaluation Campaign. In Latent Variable Analysis and Signal Separation. Springer.	Music source separation	150 tracks (4 stems each)	~30 GB (HQ)	files (stems)	stem.mp4 / wav (HQ)	lossy/lossless	150 tracks	avg ~4 min	44100 Hz	stereo	16-bit	stem.mp4 / wav	yes (stem tracks)	
30	UoM	ICBHI 2017 Respiratory Sound Database is a Biomedical audio dataset containing 920 annotated lung auscultation recordings from 126 subjects, with respiratory cycle annotations including crackles and wheezes; used for respiratory sound classification and lung disease detection.	audio / sound	https://bhichallen	Aristotle Univ. Thessaloniki & Univ. of Coimbra	Free for research	Rocha, B. M., et al. "An open access database for the evaluation of respiratory sound classification algorithms." Physiol. Meas. 40.3 (2019): 035001.	Respiratory disease / lung sound classification	920 recordings	~3.7 GB	files + annotations	wav	no (PCM)	920 files	10-90 s	varies (4000-44100)	mono	16-bit	wav	yes (cycle/event timestamps)	

Metadata Comment/ example	Project partner	Dataset description	Modality *video, image, audio, numerical, multimodal*	Link	Source / Owner	License	Reference	Applications	Number of items	Total size (GB)	Structure *files, sequences, tabular, time-series, …*	Data format *mp3, mp4, …*	Compression	No. of rows (tabular)	No. of features	Feature types	Missing data (%)	Time-series	Sampling frequency (tabular)	Units
1	UNSA	Cleveland heart disease dataset: clinical and test attributes used to detect presence of coronary artery disease.	numerical	https://archive.ics.uci.edu/dataset/45/heart+disease	Janosi, Steinbrunn, Pfisterer, Detrano / UCI ML Repository	CC BY 4.0	Detrano, R., et al. "International application of a new probability algorithm for the diagnosis of coronary artery disease." The American Journal of Cardiology 64.5 (1989): 304-310.	Classification (medical diagnosis)	303 records	40 KB	tabular	csv	no	303	13	integer + categorical	~0.2% (few missing)	no	NA	diagnosis (0-4)
2	UNSA	Direct marketing (phone-call) campaigns of a Portuguese bank; predict client subscription to a term deposit.	numerical	https://archive.ics.uci.edu/dataset/222/bank+marketing	S. Moro, P. Cortez, P. Rita / UCI ML Repository	CC BY 4.0	Moro, S., Cortez, P., Rita, P. "A data-driven approach to predict the success of bank telemarketing." Decision Support Systems 62 (2014): 22-31.	Classification	45,211 records	0.5 MB	tabular	csv	no	45211	16	integer + categorical	0% (unknowns coded)	no	NA	subscribed yes/no
3	UNSA	California housing block-group data from the 1990 census; predict median house value from demographic/geographic features.	numerical	https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html	R. K. Pace & R. Barry (StatLib)	Public domain	Pace, R. Kelley, and Ronald Barry. "Sparse spatial autoregressions." Statistics & Probability Letters 33.3 (1997): 291-297.	Regression	20,640 records	1.4 MB	tabular	csv	no	20640	8	real	0%	no	NA	median house value (USD)
4	UNSA	Building-energy simulation of 12 building shapes; predict heating and cooling loads from 8 design parameters.	numerical	https://archive.ics.uci.edu/dataset/242/energy+efficiency	A. Tsanas & A. Xifara, University of Oxford / UCI ML Repository	CC BY 4.0	Tsanas, A., Xifara, A. "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools." Energy and Buildings 49 (2012): 560-567.	Regression (building energy)	768 records	70 KB	tabular	xlsx	no	768	8	real	0%	no	NA	heating/cooling load (kWh/m2)
5	UNSA	Hourly responses of a gas multisensor device deployed in an Italian city, with reference pollutant concentrations.	numerical	https://archive.ics.uci.edu/dataset/360/air+quality	S. De Vito et al., ENEA / UCI ML Repository	CC BY 4.0	De Vito, S., et al. "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario." Sensors and Actuators B 129.2 (2008): 750-757.	Regression, sensor calibration	9,358 records	0.5 MB	tabular, time-series	csv	no	9358	15	real	yes (coded -200)	yes	hourly	CO, NOx, benzene (mg/m3, ppb)
6	UNSA	Hourly ambient variables (temperature, pressure, humidity, exhaust vacuum) and net electrical output of a CCGT plant.	numerical	https://archive.ics.uci.edu/dataset/294/combined+cycle+power+plant	P. Tufekci & H. Kaya / UCI ML Repository	CC BY 4.0	Tufekci, P. "Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods." Int. J. of Electrical Power & Energy Systems 60 (2014): 126-140.	Regression (energy/power)	9,568 records	0.5 MB	tabular	xlsx	no	9568	4	real	0%	no	NA	electrical output (MW)
7	UNSA	Surface-defect measurements of stainless-steel plates classified into 7 fault types.	numerical	https://archive.ics.uci.edu/dataset/198/steel+plates+faults	Semeraro, CRS4 (Sardinia) / UCI ML Repository	CC BY 4.0	Buscema, M., Terzi, S., Tastle, W. "Steel Plates Faults." UCI Machine Learning Repository (2010). https://doi.org/10.24432/C5J88N.	Classification (defect detection)	1,941 records	0.2 MB	tabular	csv/txt	no	1941	27	real + integer	0%	no	NA	fault type (7 classes)
8	UNSA	Smartphone accelerometer/gyroscope signals (30 subjects) processed into 561 features for 6 activities.	numerical	https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones	D. Anguita et al., University of Genova / UCI ML Repository	CC BY 4.0	Anguita, D., et al. "A public domain dataset for human activity recognition using smartphones." Proc. ESANN (2013).	Classification (activity recognition)	10,299 records	26 MB	tabular	txt	no	10299	561	real (normalized)	0%	yes (derived from 50 Hz)	window 2.56 s	activity (6 classes)
9	UNSA	Hourly PM2.5 concentration from the US Embassy in Beijing (2010-2014) with co-located meteorological variables.	numerical	https://archive.ics.uci.edu/dataset/381/beijing+pm2+5+data	X. Liang et al., Peking University / UCI ML Repository	CC BY 4.0	Liang, X., et al. "Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating." Proc. Royal Society A 471.2182 (2015): 20150257.	Time-series forecasting (air quality)	43,824 records	1.8 MB	tabular, time-series	csv	no	43824	11	real + categorical	yes (missing PM2.5)	yes	hourly	PM2.5 (ug/m3)
10	UNSA	Demographics, account and service usage of 7,043 fictional telco customers (IBM sample); predict customer churn.	numerical	https://www.kaggle.com/datasets/blastchar/telco-customer-churn	IBM Sample Data Sets (via Kaggle)	CC BY-NC-ND 4.0	IBM. "Telco Customer Churn (IBM Cognos Analytics sample)." IBM / Kaggle (2018).	Classification (churn prediction)	7,043 records	1 MB	tabular	csv	no	7043	21	real + categorical	~0.16% (11 in TotalCharges)	no	NA	churn yes/no
11	UNSA	Prices and attributes (carat, cut, color, clarity, dimensions) of nearly 54,000 round diamonds.	numerical	https://github.com/tidyverse/ggplot2/blob/main/data-raw/diamonds.csv	Tidyverse / ggplot2 (H. Wickham)	MIT (package) / open data	Wickham, Hadley. "ggplot2: Elegant Graphics for Data Analysis." Springer-Verlag (2016). diamonds dataset.	Regression, EDA, teaching	53,940 records	3 MB	tabular	csv	no	53940	10	real + categorical	0%	no	NA	price (USD)
12	UNSA	Fuel consumption (miles-per-gallon) and engine/vehicle attributes for 1970s-80s automobiles.	numerical	https://www.openml.org/d/196	Carnegie Mellon StatLib (R. Quinlan)	Public domain	Quinlan, J. R. "Combining instance-based and model-based learning." Proc. Tenth Int. Conf. on Machine Learning (1993): 236-243.	Regression	398 records	20 KB	tabular	csv	no	398	7	real + integer + categorical	yes (6 in horsepower)	no	NA	mpg
13	UNSA	Trip-level records of New York City yellow taxis (pickup/dropoff time & location, distance, fares, payment).	numerical	https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page	NYC Taxi & Limousine Commission (NYC Open Data)	Public domain (NYC Open Data)	NYC Taxi & Limousine Commission. "TLC Trip Record Data." NYC Open Data (updated monthly).	Regression, time-series, demand analysis	millions of trips per month	0.05-0.5 GB / month (Parquet)	tabular, time-series	parquet/csv	Snappy (parquet)	millions/month	18	real + categorical	minimal	yes	per-trip	fare (USD), distance (mi)
14	UNSA	World Development Indicators: ~1,400 development indicators for 217 economies and aggregates, time series since 1960.	numerical	https://datatopics.worldbank.org/world-development-indicators/	The World Bank	CC BY 4.0	World Bank. "World Development Indicators." Washington, DC: The World Bank (annual).	Macro-economic analysis, panel/time-series	~1,400 indicators x 217 economies	0.2 GB (full)	tabular, time-series	csv	no	383000	66	real + categorical	yes (sparse)	yes	annual	various (per indicator)
15	UNSA	Daily country-level COVID-19 cases, deaths, testing, vaccinations and policy indicators (2020-2024).	numerical	https://github.com/owid/covid-19-data	Our World in Data (OWID)	CC BY 4.0	Mathieu, E., Ritchie, H., et al. "A global database of COVID-19 vaccinations." Nature Human Behaviour 5 (2021): 947-953.	Time-series, epidemiology, dashboards	~430,000 records	0.1 GB	tabular, time-series	csv	no	430000	67	real + categorical	yes (sparse)	yes	daily	cases, deaths, vaccinations

16	UNSA	Country-level CO2 and greenhouse-gas emissions, energy and GDP indicators, with long historical coverage.	numerical	https://github.com/owid/co2-data	Our World in Data (Global Carbon Project)	CC BY 4.0	Global Carbon Project; Our World in Data. "CO2 and Greenhouse Gas Emissions Dataset." (annual).	Climate analysis, time-series	~50,000 records	20 MB	tabular, time-series	csv	no	50000	79	real + categorical	yes (older years)	yes	annual	CO2 (Mt), per-capita (t)	
17	UNSA	Detailed and summary listings, calendars and reviews for Airbnb in cities worldwide (price, location, availability, reviews).	numerical	http://insideairbnb.com/get-the-data/	Inside Airbnb (M. Cox)	CC BY 4.0	Cox, Murray. "Inside Airbnb: Adding data to the debate." insideairbnb.com (quarterly snapshots).	Regression, geospatial, market analysis	tens of thousands of listings/city	5-100 MB / city	tabular, time-series	csv	gzip	~40,000/city	75	real + categorical	yes (reviews, price)	yes (calendar)	daily (calendar)	price/night (local currency)	
18	UNSA	114,000 Spotify tracks across 125 genres with audio features (danceability, energy, tempo, valence, etc.) from the Web API.	numerical	https://huggingface.co/datasets/maharshipandya/spotify-tracks-dataset	maharshipandya (Hugging Face / Spotify Web API)	Open (CC0 / Database license)	Pandya, Maharshi. "Spotify Tracks Dataset." Hugging Face Datasets (2022).	Regression, classification, recommendation	114,000 records	20 MB	tabular	csv/parquet	no	114000	20	real + categorical	minimal	no	NA	popularity (0-100)	
19	UNSA	Oil temperature and 6 power-load features of electricity transformers in China (2016-2018); benchmark for long-sequence forecasting (Informer).	numerical	https://github.com/zhouhaoyi/ETDataset	H. Zhou et al. (ETDataset, Informer paper)	CC BY 4.0	Zhou, H., et al. "Informer: Beyond efficient transformer for long sequence time-series forecasting." Proc. AAAI 35.12 (2021): 11106-11115.	Time-series forecasting (energy)	17,420 records (ETTh1, hourly)	2.5 MB (4 files)	tabular, time-series	csv	no	17420	7	real	0%	yes	hourly (ETTh) / 15 min (ETTm)	oil temperature (degC)	
20	UNSA	Cumulative table of Kepler Objects of Interest with transit and stellar parameters; predict exoplanet disposition.	numerical	https://exoplanetarchive.ipac.caltech.edu/	NASA Exoplanet Archive (Caltech/IPAC)	Public domain (NASA)	NASA Exoplanet Archive. "Kepler Objects of Interest (Cumulative KOI table)." Caltech/IPAC.	Classification (exoplanet vetting)	9,564 records	8 MB	tabular	csv	no	9564	49	real + categorical	yes (some parameters)	no	NA	disposition (candidate/false positive)	
21	UNSA	Hourly electricity consumption (MW) from PJM Interconnection regions in the eastern US, spanning several years.	numerical	https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption	Rob Mulla (data from PJM Interconnection)	CC0 (Public Domain)	PJM Interconnection LLC / R. Mulla. "Hourly Energy Consumption." Kaggle (2018).	Time-series forecasting (load)	~145,000 records/region	20 MB	tabular, time-series	csv	no	145366	2	real	0%	yes	hourly	energy consumption (MW)	
22	UNSA	Records of terrorist incidents worldwide (1970-2020) with location, attack type, weapons, targets and casualties.	numerical	https://www.start.umd.edu/gtd/	START, University of Maryland	Free for non-commercial research	LaFree, G., Dugan, L. "Introducing the Global Terrorism Database." Terrorism and Political Violence 19.2 (2007): 181-204.	Classification, geospatial, time-series	~200,000 records	0.15 GB	tabular, time-series	csv/xlsx	no	200000	135	real + categorical	yes (many fields)	yes	per-event	casualties (count)	
23	UNSA	~130,000 wine reviews scraped from WineEnthusiast: country, variety, winery, price and a 100-point score.	numerical	https://www.kaggle.com/datasets/zynicide/wine-reviews	zackthoutt (data from WineEnthusiast)	CC BY-NC-SA 4.0	Thoutt, Zack. "Wine Reviews." Kaggle (2017).	Regression, NLP, classification	~130,000 records	50 MB	tabular	csv	no	129971	13	real + categorical	yes (price, region)	no	NA	points (80-100), price (USD)	
24	UNSA	Detailed attributes (skill, physical, positional, wage, value) of football players from the FIFA video-game series.	numerical	https://www.kaggle.com/datasets/stefanoleone992/fifa-23-complete-player-dataset	S. Leone (data from SoFIFA.com)	CC BY-NC-SA 4.0	Leone, Stefano. "FIFA 23 Complete Player Dataset." Kaggle (2022).	Regression, clustering, EDA	~18,000 players	30 MB	tabular	csv	no	18000	89	real + integer + categorical	yes (some attributes)	no	NA	overall rating, value (EUR)	
25	UNSA	Accepted and rejected peer-to-peer loan applications (2007-2018) with borrower, loan and repayment attributes.	numerical	https://www.kaggle.com/datasets/wordsforthewise/lending-club	Lending Club (via Kaggle)	CC0 (Public Domain)	Lending Club. "Lending Club Loan Data (2007-2018)." Kaggle.	Classification (credit risk), regression	~2,260,000 records (accepted)	1.5 GB	tabular	csv	gzip	2260668	151	real + categorical	yes (many sparse cols)	no	NA	loan status (default/paid)	
26	UNSA	Eurostat annual national accounts: GDP and main aggregates for EU member states and partners, time series.	numerical	https://ec.europa.eu/eurostat/databrowser/product/view/nama_10_gdp	Eurostat (European Commission)	CC BY 4.0 (Eurostat reuse policy)	Eurostat. "GDP and main components (nama_10_gdp)." European Commission (annual).	Macro-economic analysis, time-series	~50,000 observations	15 MB	tabular, time-series	tsv/csv	gzip	50000	10	real + categorical	yes (some countries/years)	yes	annual	GDP (million EUR / PPS)	
27	UNSA	Hourly meteorological and air-quality data for Sarajevo (full year): time, wind, humidity, pressure, precipitation, temperatures (Sarajevo & Bjelasnica), temperature inversion, with PM10 and PM2.5 concentrations as targets.	numerical	https://www.fhmzbih.gov.ba/	Federalni hidrometeoroloski zavod FBiH (FHMZ)	Institutional data (FHMZ FBiH) - restricted use	Federalni hidrometeoroloski zavod FBiH. Meteoroloski i PM podaci za Sarajevo. (Masinski fakultet UNSA, istrazivacki/PhD podaci).	PM10/PM2.5 air-quality prediction, time-series forecasting	17,257 records	1.07 MB	tabular, time-series	csv	no	17257	11	real + integer	0%	yes	hourly	PM10 / PM2.5 (ug/m3)	
28	UNSA	High-performance concrete mix designs (cement, blast furnace slag, fly ash, water, superplasticizer, coarse & fine aggregate, age) with measured compressive strength. Local working copy (xls/xlsx + source paper).	numerical	https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength	Prof. I-Cheng Yeh, Chung-Hua University / UCI ML Repository	Free reuse with citation retained (Yeh)	Yeh, I-Cheng. "Modeling of strength of high-performance concrete using artificial neural networks." Cement and Concrete Research 28.12 (1998): 1797-1808.	Regression (materials / civil eng.)	1,030 records	66 KB (xlsx)	tabular	xlsx / xls	no	1030	8	real	0%	no	NA	compressive strength (MPa)	
29	UNSA	Synthetic/enriched loan-applicant dataset (age, gender, education, income, employment, home ownership, loan amount, intent, interest rate, credit score, prior defaults) with loan approval label.	numerical	https://www.kaggle.com/datasets/taweilo/loan-approval-classification-data	Kaggle (taweilo; based on the Credit Risk Dataset)	Open data (Kaggle - see dataset page)	Lao, T. (taweilo). "Loan Approval Classification Data." Kaggle (2024). Synthetic data based on the Credit Risk Dataset.	Classification (loan / credit approval)	45,000 records	3.6 MB	tabular	csv	no	45000	13	real + integer + categorical	0%	no	NA	loan status (approved/denied)	
30	UNSA	Residential house listings with price and 12 attributes: area, bedrooms, bathrooms, stories, main road, guest room, basement, hot-water heating, air conditioning, parking, preferred area, furnishing status.	numerical	https://www.kaggle.com/datasets/yasserh/housing-prices-dataset	Kaggle (M. Yasser H.)	CC0 / Open (Kaggle - see dataset page)	Yasser H., M. "Housing Prices Dataset." Kaggle (2022).	Regression (house price prediction)	545 records	30 KB	tabular	csv	no	545	12	integer + categorical	0%	no	NA	price (currency unit)	