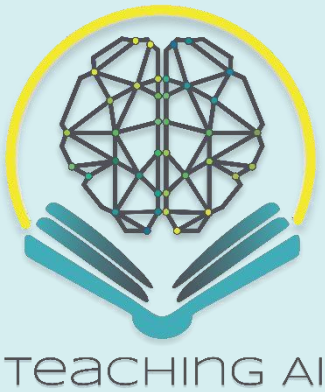


Report WP3-A2:

Analysis of Various Databases

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

Analyses of databases

Related to
WP3-A2: Analysis of various databases

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This report contains original unpublished work, except where clearly indicated otherwise.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

This report contains material, which is the copyright of TAI Consortium Parties. All TAI Consortium Parties have agreed that the content of the report is licensed under a Creative Commons Attribution Non-Commercial Share Alike 4.0 International License. TAI Consortium Parties do not warrant that the information contained in the Report is capable of use, or that use of the information is free from risk and accept no liability for loss or damage suffered by any person or any entity using the information.

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	Prepared by:	Comment
1.0	12.07.2026.	Jelena Šaković Jovanović, Aleksandar Vujović, Janko Jovanović, UoM	Initial version
2.0	16.07.2026.	Jelena Šaković Jovanović, Aleksandar Vujović, Janko Jovanović, UoM	Second version, with contributions from and review by the TAI WP3 partners (UNILJ, UNBG, UNSA, and UNIRI)
3.0	23.07.2026	Jelena Šaković Jovanović, Aleksandar Vujović, Janko Jovanović, UoM	Third version, with contributions from and review by the TAI WP3 partners (UNILJ, UNBG, UNSA, and UNIRI)
4.0	21.08.2026.	UoM, UNILJ, UNBG, UNSA, and UNIRI	Final consolidated version prepared by UoM following contributions, review and verification by the WP3 partners

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 through the development of teaching materials, practical learning resources and research-oriented educational activities. Within this framework, Work Package 3 (WP3) focuses on collecting and analysing datasets that can support practical AI and Machine Learning (ML) teaching, learning and student projects.

This report presents the work carried out within WP3-A2: Analysis of Various Databases. The activity builds on WP3-A1, during which 120 datasets were identified across four data modalities. UNILJ - University of Ljubljana was responsible for 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. UNIRI - University of Rijeka contributed by reviewing the collected datasets and advising on the structure and formats of the collection. The collected information was recorded in a shared dataset collection document, which is provided in Annex 1 and served as the primary data source for the analysis presented in this report.

The purpose of this report is to analyse the relevance, accessibility, conditions of use, formats, technical characteristics and potential educational applicability of the collected datasets. The analysis also considers their main limitations and the level of preparation required before their use in practical exercises, student projects and research-oriented learning activities.

In addition to the separate analysis of the four data modalities, the report provides a comparative assessment of the complete dataset collection, identifies representative datasets suitable for introductory, intermediate and advanced or research-oriented educational activities, and presents the 12 datasets selected and uploaded to the project repository for subsequent project activities.

The findings of this report provide a basis for the further technical preparation of the selected datasets and the development of Python-based educational materials. The report was coordinated and prepared by the University of Montenegro, with contributions from and review by the TAI WP3 partners.

2. Dataset Analysis Methodology

The analysis conducted within WP3-A2 builds on the dataset collection and standardisation process completed within WP3-A1. The shared dataset collection document prepared during WP3-A1 and provided in Annex 1 was used as the primary data source for the analysis. It contains information on 120 datasets, equally distributed across four data modalities: video, image, audio/sound and numerical datasets.

The analysis was designed to assess the applicability of the collected datasets for AI education and research using the descriptive, legal, structural and technical information recorded by the project partners. The methodology combined a separate analysis of each data modality with a comparative assessment across the four modalities. Particular attention was given to the relevance of the datasets for AI and Machine Learning tasks, their accessibility and conditions of use, their documented quality indicators, data formats and technical characteristics, as well as the preparation required for their use in practical educational activities.

The preparation of the report followed a collaborative approach. Each partner responsible for a particular data modality contributed the corresponding dataset information and domain-specific expertise. UoM harmonised these inputs and organised them within a common analytical structure covering all four modalities.

2.1 Scope and Data Source

The analysis covered all 120 datasets identified within WP3-A1. The information recorded in the shared dataset collection document was used to examine dataset relevance, accessibility, conditions of use, format, technical characteristics and potential educational applicability. The complete shared dataset collection document, containing information on all 120 datasets, is provided in Annex 1.

The analysis was based on both common information recorded for all datasets and technical characteristics specific to each data modality. Most of this information was obtained from descriptions provided by the dataset owners or providers in the relevant repositories or accompanying scientific papers. However, the level of available documentation varied across the datasets. Consequently, some information was incomplete, limited or not publicly available. A detailed description of these fields, as well as the dataset collection and standardisation process, is provided in Report WP3-A1.

2.2 Analysis Dimensions and Procedure

The datasets were analysed according to four main dimensions:

1. relevance to AI and Machine Learning tasks and application areas;
2. accessibility and conditions of use;
3. documented quality and technical suitability;
4. potential educational applicability.

Relevance was assessed on the basis of dataset descriptions and stated application areas.

Accessibility and conditions of use were analysed using the available sources, links and licence information. For datasets supporting several AI tasks, one main application category was assigned according to the primary intended use described in the shared dataset collection document. The classification of individual datasets is provided in Annex 2.

Documented quality and technical suitability were assessed based on dataset scale, structure, format and modality-specific technical characteristics. Where available, information on labels or annotations was also taken into account. The assessment relied on information documented by the dataset providers and recorded in the shared dataset collection document and did not include an independent validation of all dataset files or annotations.

Educational applicability was assessed in relation to the suitability of the datasets for practical exercises, student projects and research-oriented activities, taking into account the need for preprocessing, standardisation or the extraction of smaller subsets.

The analysis was conducted separately for video, image, audio/sound and numerical datasets, using the main common and modality-specific parameters recorded in the shared dataset collection document. The modality-specific findings were reviewed by the partners responsible for the corresponding dataset categories before their inclusion in the consolidated report.

Based on these dimensions, datasets were considered particularly suitable when they combined a clearly defined AI application, sufficiently documented data structures and, where relevant, labels or annotations, technically usable formats, manageable preparation requirements, and conditions of use compatible with the intended educational or research activity. Recommendations were adapted to the expected level of use, distinguishing between introductory exercises, intermediate student projects, and advanced or research-oriented activities.

3. Analysis of Collected Datasets

The collected datasets were analysed separately by data modality in order to identify their main characteristics, application potential, accessibility, technical suitability and possible use in AI education. The analysis was based on the information recorded in the shared dataset collection document and focused on the main patterns and differences within each group of 30 datasets.

For each modality, the analysis considers the represented AI application areas, licensing and access conditions, dataset scale and structure, commonly used formats and modality-specific technical characteristics. Where available, information on labels or annotations was also taken into account. These findings are used to identify the main educational opportunities, practical requirements and potential limitations associated with each group of datasets.

The results are presented for video, audio/sound, numerical, and image datasets in the following sections.

3.1 Video Datasets

3.1.1 Main Applications of Video Datasets

The collected video datasets cover a wide range of applications relevant to Artificial Intelligence and computer vision. As shown in Figure 3.1, the largest group consists of datasets intended for human action and behaviour recognition, accounting for 8 datasets or 26.67% of the collection. Agriculture and animal monitoring represents the second largest category with 6 datasets (20.00%), followed by traffic, surveillance and safety-related applications with 5 datasets (16.67%).

Object detection and tracking, autonomous driving and navigation, and multimodal understanding and affect analysis are each represented by 3 datasets (10.00%). General video classification and recognition accounts for 2 datasets, or 6.67% of the collection. This distribution demonstrates that the video datasets support both widely used computer vision tasks and more specialised educational and research applications.

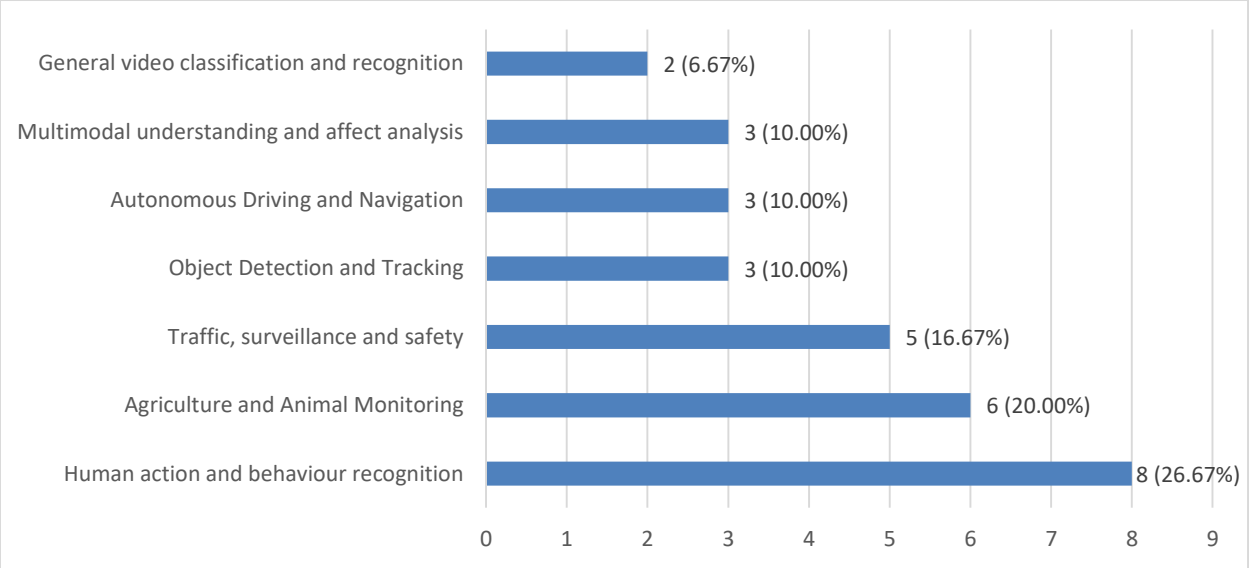


Figure 3.1. Distribution of video datasets by main application

For the purpose of the analysis, each dataset was assigned to one main application category according to its primary intended use. Although some datasets may support more than one AI task, only the dominant application was considered in order to avoid multiple counting. A detailed list of the datasets assigned to each main application category is provided in Annex 2.

3.1.2 Licence and Conditions of Use

The analysis of the licence information shows that the collected video datasets differ considerably in their conditions of use. As presented in Figure 3.2, the largest group consists of datasets distributed under Creative Commons licences, accounting for 12 datasets or 40.0% of the collection. This category includes both attribution-based licences and Creative Commons licences with additional non-commercial, share-alike or no-derivatives conditions.

A further 7 datasets, or 23.3%, are available under separate non-commercial, academic or research-use conditions, while 4 datasets, or 13.3%, are classified as open or public domain. Two datasets, representing 6.7%, are subject to custom or restricted licence arrangements, and licence information was not clearly specified for 5 datasets, or 16.7%.

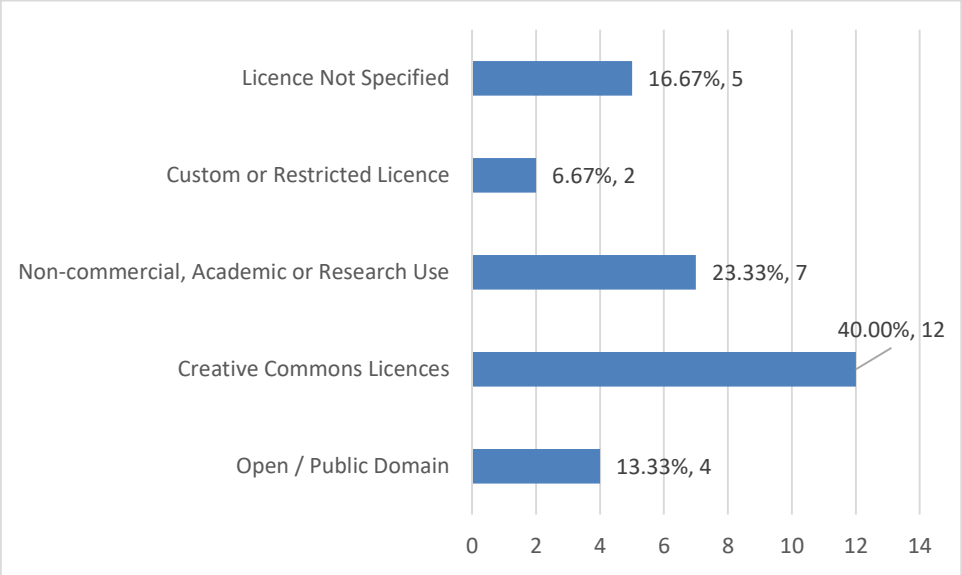


Figure 3.2. Distribution of video datasets by licence and conditions of use

These results indicate that a substantial part of the video dataset collection can be used in higher education and research. However, the precise licence conditions should be checked before datasets or derived materials are incorporated into teaching resources or publicly distributed project outputs.

3.1.3 Dataset Scale: Number of Videos and Total Size

The scale of the collected video datasets was analysed using two complementary parameters: the number of video recordings and the total dataset size. The number of videos indicates the amount of available training and evaluation material, while the total size provides an indication of storage, download and computational requirements.

Number of Videos

As shown in Figure 3.3, the collected datasets differ considerably in the number of video recordings they contain. The classification is based on the number of videos, scenes, clips or video segments reported in the shared dataset collection document. Ten datasets (33.3%) include fewer than 100 videos, while seven datasets (23.3%) contain between 100 and 999 videos. Four datasets (13.3%) include between 1,000 and 9,999 videos, and six datasets (20.0%) contain between 10,000 and 99,999 videos. The remaining three datasets (10.0%) contain 100,000 or more video recordings or segments.

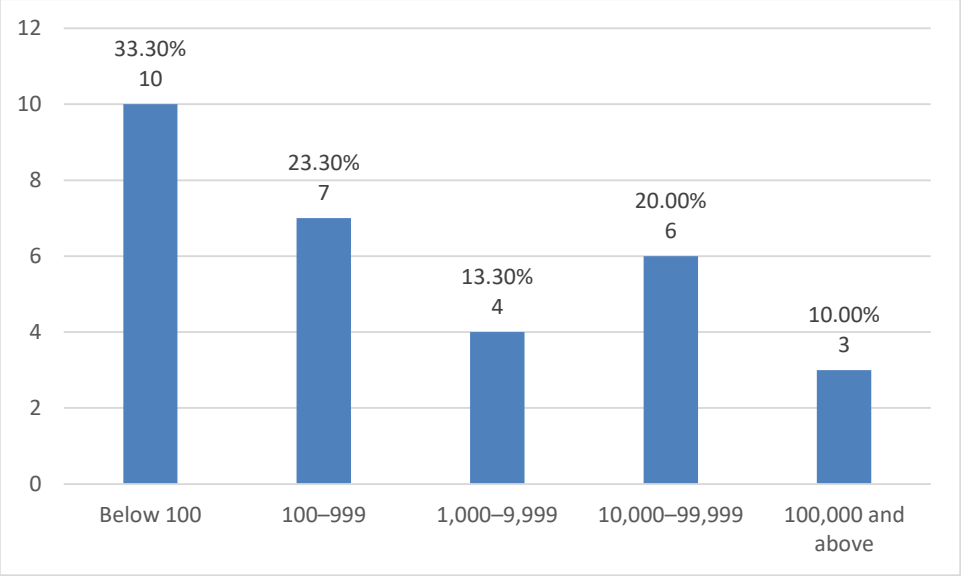


Figure 3.3. Distribution of video datasets by number of videos

The results show that more than half of the datasets contain fewer than 1,000 videos. These collections may be easier to manage in practical classes and student projects. Datasets containing several thousand or hundreds of thousands of recordings provide greater potential for model training and evaluation but generally require more preparation and computational resources.
For multimodal and continuously recorded datasets, scenes, video segments or continuous camera streams were treated as the closest equivalent to individual video units, based on the information available in the dataset collection document.

Total Dataset Size

The total size of the datasets also varies substantially. Ten datasets (33.3%) fall within the range of 10–100 GB, making this the most represented category. Nine datasets (30.0%) have a total size between 1 and 10 GB, while five datasets (16.7%) are smaller than 1 GB. Three datasets (10.0%) range from 100 GB to 1 TB, and one dataset (3.3%) exceeds 1 TB. Size information was not available for two datasets (6.7%).

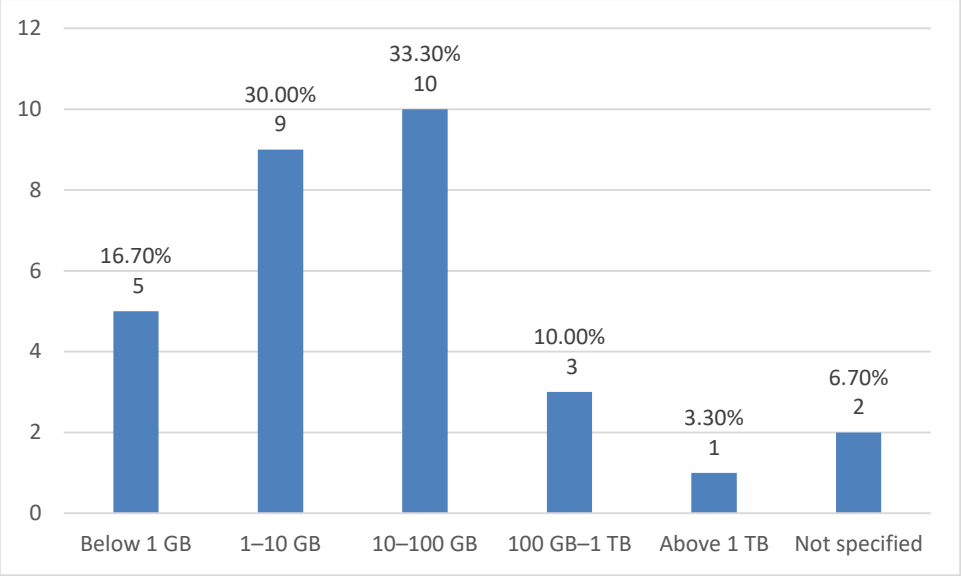


Figure 3.4. Distribution of video datasets by total size

Almost half of the collected datasets are smaller than 10 GB and can therefore be considered relatively manageable for educational use. Datasets between 10 and 100 GB may still be suitable for student projects, although the preparation of smaller subsets may be necessary. Collections above 100 GB are more demanding in terms of download, storage and processing and are therefore generally more appropriate for advanced projects or research-oriented activities.

Taken together, the number of videos and total size show that dataset scale cannot be assessed using only one parameter. A dataset containing many short or low-resolution clips may be smaller than a collection containing fewer but longer or higher-resolution recordings. Both indicators should therefore be considered when selecting video datasets for practical AI teaching.

3.1.4 Video Formats and Technical Characteristics

The technical suitability of the collected video datasets was assessed through their main data formats, video length, resolution and frame rate. These parameters influence compatibility with commonly used AI and video-processing tools, the level of preprocessing required and the computational resources needed for practical educational activities.

Video Formats

As shown in Figure 3.5, the MP4/MPEG-4 family is the dominant format, used in 17 datasets (56.7%). AVI is represented in 4 datasets (13.3%), while 3 datasets (10.0%) use specialised formats such as ROS bag, TFRecord or CSD/HDF5. Frame- or image-based collections and mixed-

format datasets each account for 2 datasets (6.7%). WebM and MKV are each represented by one dataset (3.3%).

The predominance of MP4/MPEG-4 and AVI is favourable for educational use because these formats are widely supported by common video-processing tools and Python libraries. Specialised and multimodal formats may require additional conversion, dedicated software or more advanced technical knowledge before they can be used in practical exercises.

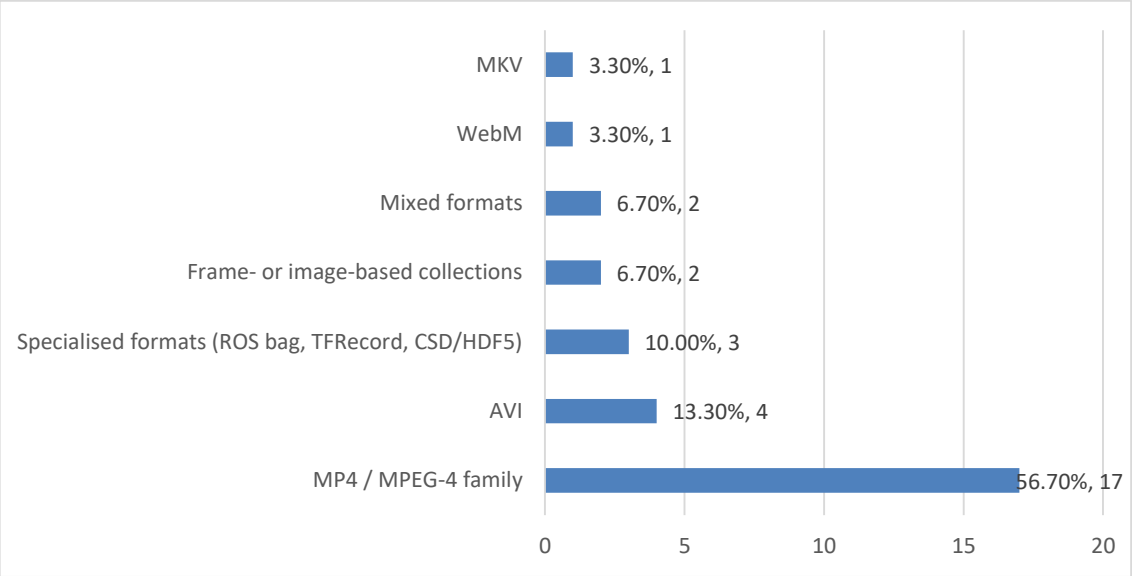


Figure 3.5. Distribution of video datasets by main data format

Video Length

The collected video datasets also differ considerably in the duration of individual recordings. Some datasets contain very short clips lasting only a few seconds, while others include longer recordings, continuous video streams or segments with variable duration. This variation reflects the different purposes for which the datasets were developed.

Short video clips are particularly suitable for tasks such as action recognition, gesture classification and event detection because they usually focus on a single activity or clearly defined event. Longer recordings provide richer temporal information and are more appropriate for tracking, behaviour analysis, surveillance, autonomous systems and multimodal video understanding. However, they also require more storage, memory and processing time.

For the purpose of the analysis, video length can be grouped into the following categories:

- below 10 seconds;
- 10–60 seconds;
- 1–10 minutes;
- above 10 minutes;

- variable, continuous or not specified.

Datasets containing long or continuous recordings may require temporal segmentation, frame sampling or the extraction of shorter clips before being used in practical educational activities. In contrast, datasets consisting of short and clearly defined clips are generally easier to integrate into introductory and intermediate AI exercises.

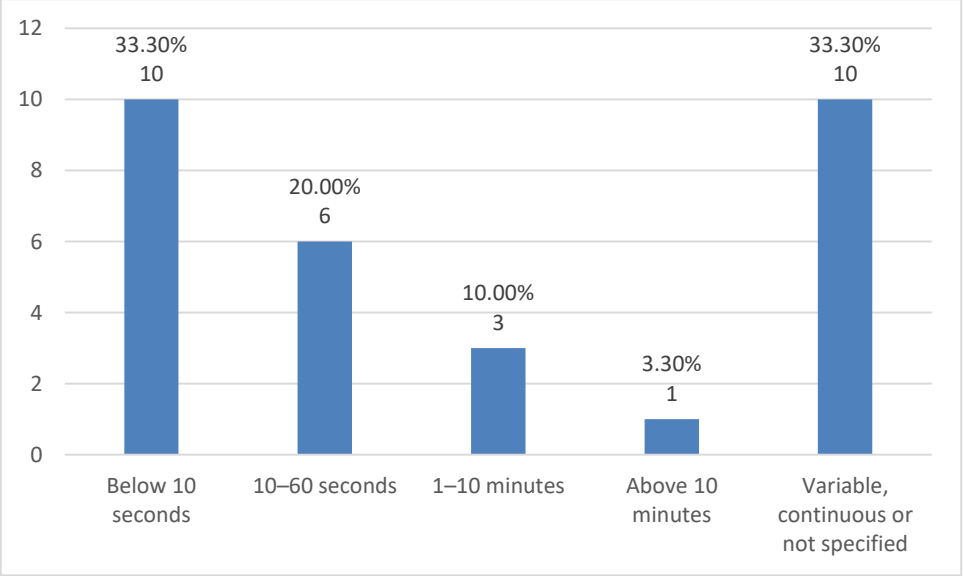


Figure 3.6. Distribution of video datasets by video length

Where average or median duration was available, it was used for classification. Datasets with wide duration ranges, continuous recordings or unavailable duration information were classified as variable, continuous or not specified.

Video Resolution

For the purpose of this analysis, the reported video resolutions were grouped into low resolution, HD or near-HD, Full HD, 4K or higher, multiple or variable resolutions, and not specified. The collected datasets show substantial variation across these categories. Low-resolution videos represent 8 datasets (26.7%), while 6 datasets (20.0%) contain HD or near-HD material. Seven datasets (23.3%) include multiple or variable resolutions, reflecting differences between recordings, cameras or sensors. Full HD is the main resolution in 3 datasets (10.0%), while 4 datasets (13.3%) provide 4K or higher-resolution recordings. Resolution information was not specified for 2 datasets (6.7%).

Lower-resolution datasets generally require fewer computing resources and may therefore be more suitable for introductory exercises. Full HD and 4K datasets provide richer visual information but require more storage, memory and processing capacity. Datasets containing multiple resolutions may also require resizing or standardisation before model training.

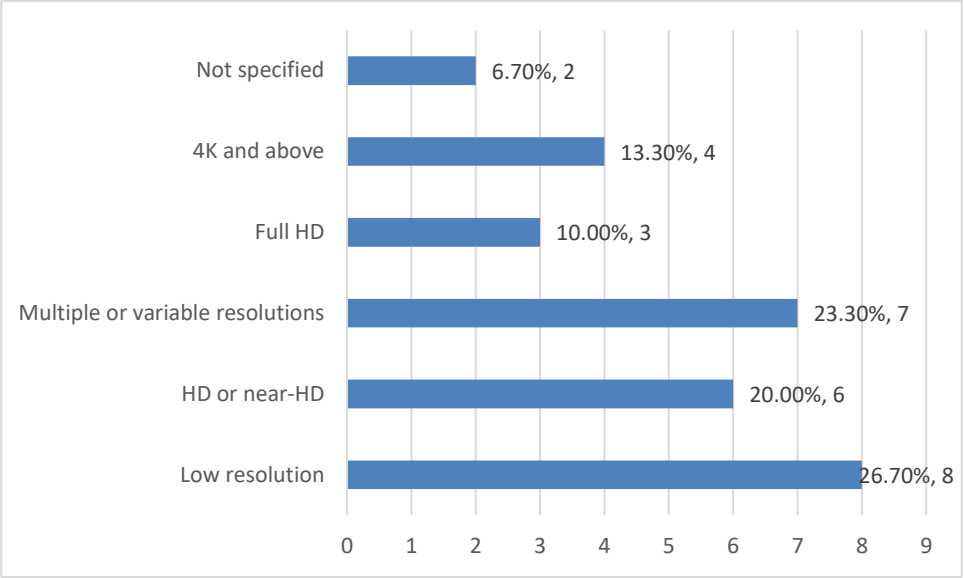


Figure 3.7. Distribution of video datasets by resolution category

Frame Rate

Half of the collected datasets, 15 datasets (50.0%), use frame rates between 20 and 30 fps. Six datasets (20.0%) have frame rates below 20 fps, while one dataset (3.3%) exceeds 30 fps. Five datasets (16.7%) contain variable or multiple frame rates, and frame-rate information was not specified for 3 datasets (10.0%).

Frame rates between 20 and 30 fps are suitable for many common tasks, including action recognition, classification, detection and tracking. Lower frame rates may reduce computational requirements but can provide less temporal detail, while higher or variable frame rates may require frame sampling or temporal standardisation before datasets are used within a common educational workflow.

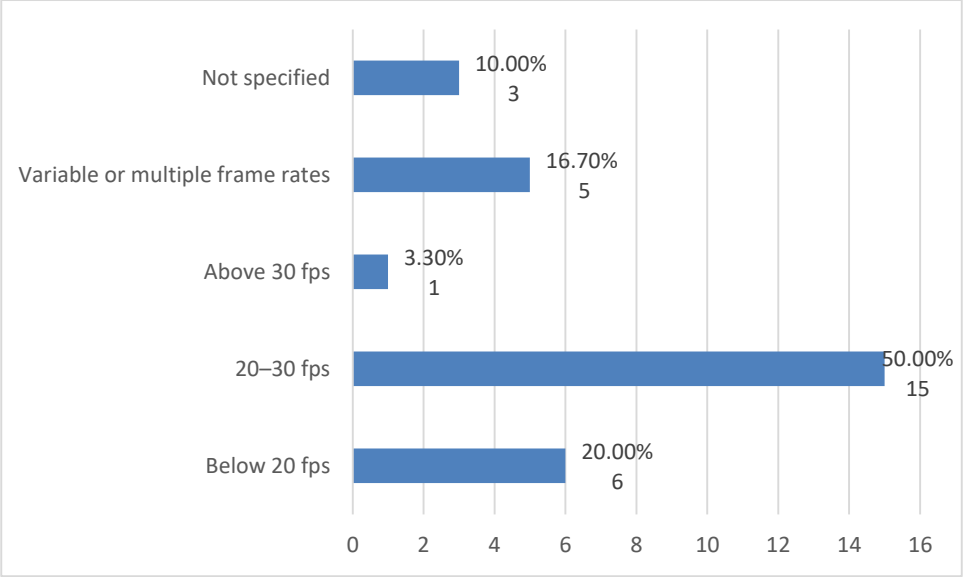


Figure 3.8. Distribution of video datasets by frame rate

The recorded codec information also shows that H.264 and related MPEG-4 codecs are widely represented, although several datasets use VP9, XVID or specialised codecs, and codec information is incomplete for some collections. Overall, most video datasets use formats and technical specifications that can be processed using standard AI and computer-vision tools, while datasets with specialised structures, very high resolutions or variable frame rates require additional preparation.

3.1.5 Educational Applicability of Video Datasets

Based on the recorded characteristics, the collected video datasets offer a wide range of opportunities for AI teaching and research, from basic video classification exercises to advanced multimodal learning and autonomous-system applications. In practice, the most appropriate dataset is not necessarily the largest or the most technically complex one. For introductory activities, clearly defined tasks, understandable labels, manageable dataset size and compatibility with commonly used tools are usually more important. As students progress, larger collections, more complex annotations and multimodal data can be introduced to support more demanding model-development and research activities.

For introductory practical exercises, smaller datasets containing short videos and clearly defined classes are particularly useful. The Highway Traffic Videos Dataset, for example, contains 254 five-second videos with manually assigned labels describing light, medium and heavy traffic. Its small size of approximately 92 MB, low resolution and clearly defined classification task make it suitable for demonstrating the complete machine-learning workflow, including video loading, frame extraction, preprocessing, model training and performance evaluation. The KTH Action

Recognition Dataset is another suitable example. It contains 600 videos representing six human actions performed by 25 participants in different scenarios. Its limited number of classes, relatively small size and low-resolution recordings allow students to focus on the basic principles of action recognition without excessive computational requirements.

At the intermediate level, students can work with datasets that include more classes, higher-resolution recordings or more demanding preprocessing requirements. The WLASL Dataset, containing 12,000 videos of 2,000 American Sign Language words, can support projects involving gesture and sign-language recognition. Although the large number of classes makes the complete dataset relatively demanding, carefully selected subsets could be used for intermediate projects focusing on a smaller vocabulary. The CBVD-5 Cow Behavior Video Dataset provides another relevant example. It contains 887 videos, accompanying annotations and Full HD recordings intended for cow-behaviour recognition, tracking and livestock monitoring. Its use would allow students to work with realistic video data while learning how to perform frame sampling, resizing and annotation processing.

Large-scale, multimodal and technically specialised datasets are more appropriate for advanced student projects and research-oriented activities. The nuScenes Dataset, which combines camera data with LiDAR and radar recordings, can support advanced work on autonomous driving, object detection, tracking and sensor fusion. However, its total size of approximately 350 GB and heterogeneous data structure require substantial storage, computational resources and data-preparation skills. Similarly, the FieldSAFE Dataset combines stereo, thermal, web and 360-degree cameras with LiDAR, radar and localisation data collected during a synchronised two-hour agricultural event. Its raw sensor streams and ROS bag structure make it particularly relevant for research involving multimodal obstacle detection and agricultural robotics, but also require specialised software and knowledge of sensor-data processing. The EGO4D Dataset, with more than 24,000 first-person videos and approximately 20 TB of data, provides valuable opportunities for advanced research on egocentric perception, although its scale generally makes the use of selected subsets necessary in an educational setting.

Table 3.1. Examples of video datasets suitable for different educational levels

Educational level	Dataset and suitable AI task	Reason for suitability	Required preparation or conditions
Introductory practical exercises	**Highway Traffic Videos Dataset** - traffic-level classification	Small collection of 254 short videos, clearly defined classes and manually assigned labels; CC0 public-domain licence	Basic frame extraction and video preprocessing

Introductory practical exercises	**KTH Action Recognition Dataset** - human action recognition	Six clearly defined action classes, 600 videos, low resolution and manageable size	Possible AVI/codec conversion, depending on the software used; non-commercial licence
Intermediate student projects	**WLASL Dataset** – sign-language and gesture recognition	Standard MP4 format, manageable total size and a clearly defined recognition task	Extraction of a smaller vocabulary subset is advisable; use is subject to the C-UDA
Intermediate student projects	**CBVD-5 Cow Behavior Video Dataset** - behaviour recognition and tracking	Annotated Full HD videos supporting realistic classification and tracking tasks; CC0 public-domain licence	Frame sampling, resizing and annotation processing may be required
Advanced or research-oriented activities	**nuScenes Dataset** - autonomous driving, detection, tracking and sensor fusion	Multimodal data combining camera recordings, LiDAR and radar	Approximately 350 GB; specialised processing and substantial computing resources required; non-commercial use
Advanced or research-oriented activities	**FieldSAFE Dataset** - multimodal obstacle detection and agricultural robotics	Synchronised data from several camera types, LiDAR, radar and localisation sensors; CC BY 4.0 licence	ROS bag and raw sensor processing, temporal synchronisation and specialised software required

The proposed classification is indicative rather than fixed. A large or technically complex dataset can also be used at an introductory or intermediate level if an appropriate subset is prepared in advance and the task is simplified. Conversely, a relatively small dataset can support advanced work when it is used to investigate more complex modelling, evaluation or preprocessing methods. In all cases, dataset selection should be aligned with the intended learning outcomes, available teaching time and computational resources. Access conditions and licence requirements should also be verified before a dataset is incorporated into teaching materials or project outputs. These recommendations are based on the characteristics documented by the dataset providers and recorded in the shared dataset collection document.

3.2 Audio and Sound Datasets

The collected audio and sound datasets cover a broad range of AI applications, including speech processing, environmental sound recognition, music analysis, bioacoustics and biomedical sound analysis. The datasets differ considerably in their application areas, scale, formats, technical characteristics and licensing conditions. The following analysis examines these characteristics and their relevance for AI education and research.

3.2.1 Main Applications of Audio and Sound Datasets

The collected audio and sound datasets cover a wide range of applications relevant to Artificial Intelligence, machine learning and digital signal processing. As shown in Figure 3.9, the largest group consists of datasets intended for music analysis and music information retrieval, accounting for 10 datasets or 33.33% of the collection. This category includes music genre and instrument recognition, automatic music tagging and transcription, audio synthesis, melody extraction, music generation and source separation.

Environmental sound and general audio event recognition represents the second largest category, with 7 datasets (23.33%). These datasets support tasks such as environmental sound classification, acoustic scene recognition, audio tagging and sound event detection. Speech recognition, speaker identification and speech synthesis are represented by 6 datasets (20.00%), covering automatic speech recognition, keyword spotting, speaker verification, text-to-speech synthesis and spoken digit classification.

Bioacoustics and animal sound recognition account for 2 datasets (6.67%), while another 2 datasets (6.67%) are intended for speech emotion recognition. Biomedical sound analysis is also represented by 2 datasets (6.67%), focusing on heartbeat and respiratory sound classification. The remaining dataset (3.33%) supports vehicle speed estimation using audio and video recordings.

This distribution demonstrates that the collection supports both widely used audio-processing tasks and more specialised educational and research applications.

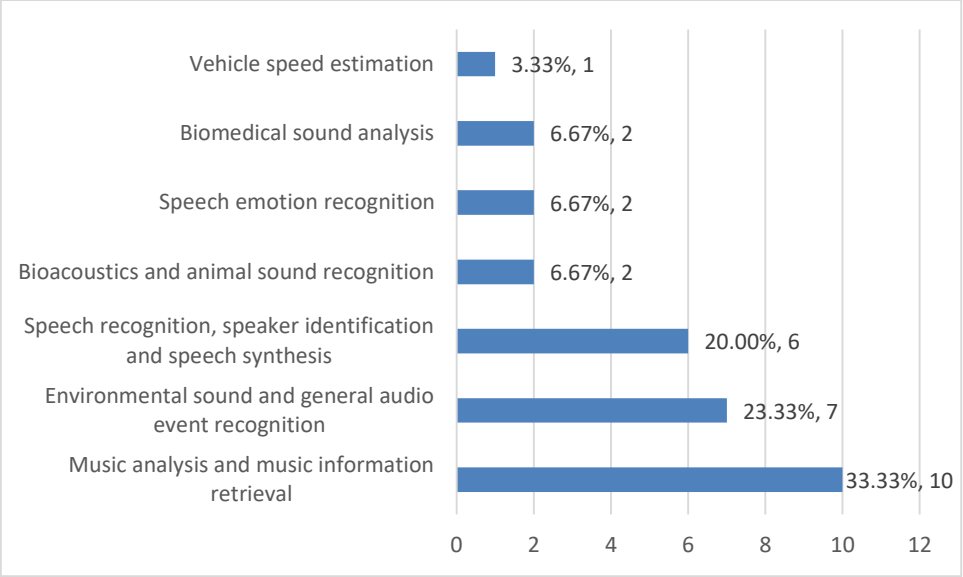


Figure 3.9. Distribution of audio and sound datasets by main application

For the purpose of the analysis, each dataset was assigned to one main application category according to its primary intended use. Although some datasets may support more than one AI task, only the dominant application was considered in order to avoid multiple counting. A detailed list of the datasets assigned to each main application category is provided in Annex 2.

3.2.2 Licence and Conditions of Use

The analysis of the licence information shows that the audio and sound datasets are generally accompanied by clearly stated conditions of use. As presented in Figure 3.10, Creative Commons licences represent the largest category, covering 22 datasets or 73.3% of the collection. This category includes attribution-based licences as well as Creative Commons licences containing non-commercial, share-alike or other specific conditions.

Two datasets, or 6.7%, are available under public-domain or other open-use conditions. A further 5 datasets, representing 16.7%, are subject to separate research-only, non-commercial, custom or otherwise restricted conditions. Licence information was not clearly specified for one dataset, corresponding to 3.3% of the collection.

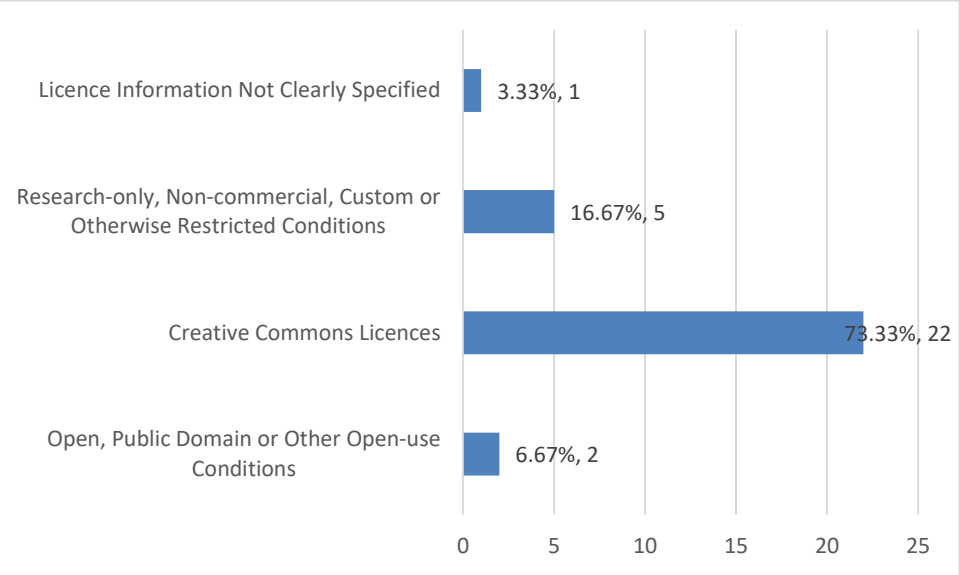


Figure 3.10. Distribution of audio and sound datasets by licence and conditions of use

These results indicate that most audio and sound datasets can support educational and research activities. However, the precise licence conditions should be checked before datasets or derived materials are redistributed or included in publicly available teaching resources.

3.2.3 Dataset Scale: Number of Audio Files and Total Size

The scale of the collected audio and sound datasets was analysed using two complementary parameters: the number of audio files or recordings and the total dataset size. The number of audio files indicates the amount of material available for model training, validation and evaluation, while the total size provides an indication of download, storage and computational requirements.

Number of Audio Files

The scale of the collected audio and sound datasets was assessed according to the reported number of audio files or equivalent units. Since the collections use different organisational structures, tracks, clips, utterances, recordings and performances were treated as the closest equivalents to individual audio files. For the purpose of this analysis, the datasets were grouped into five categories: fewer than 100 files, 100–999 files, 1,000–9,999 files, 10,000–99,999 files, and 100,000 or more files.

As shown in Figure 3.11, one dataset (3.33%) contains fewer than 100 audio files, while 5 datasets (16.67%) include between 100 and 999 files. Nine datasets (30.00%) contain between 1,000 and

9,999 files, and 7 datasets (23.33%) include between 10,000 and 99,999 files. The remaining 8 datasets (26.67%) contain 100,000 or more audio recordings or equivalent audio units.

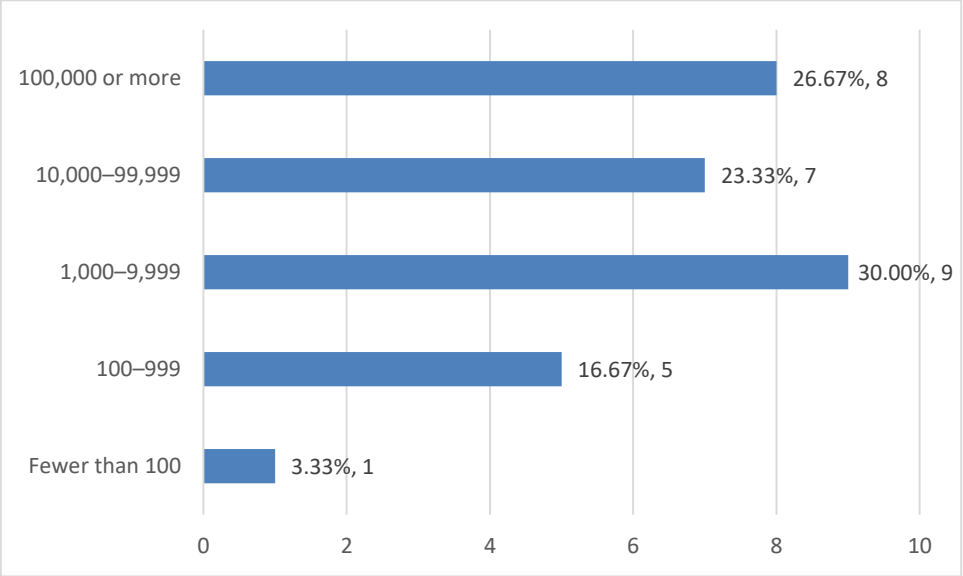


Figure 3.11. Distribution of audio/sound datasets by number of audio files

The results show that 16 datasets contain between 1,000 and 99,999 audio files or equivalent units, while 8 datasets contain at least 100,000 recordings. Smaller collections may be easier to download, inspect and process during practical classes and student projects. In contrast, datasets containing tens or hundreds of thousands of recordings provide greater potential for training and evaluating more complex models, but generally require additional preparation, storage capacity and computational resources.

The number of audio files should be considered together with the duration and total size of the dataset. A collection containing a large number of short audio clips may require less storage than a dataset consisting of fewer but substantially longer recordings. These parameters should therefore be considered jointly when selecting datasets for educational activities.

Total Dataset Size

The total size of the audio and sound datasets also varies substantially. Three datasets (10.00%) are smaller than 1 GB, while 8 datasets (26.67%) range from 1 to less than 10 GB. The largest group consists of 12 datasets (40.00%) with sizes ranging from 10 to less than 100 GB. Four datasets (13.33%) fall within the range of 100 GB to 1 TB, and 2 datasets (6.67%) exceed 1 TB. Size information was not available for one dataset (3.33%).

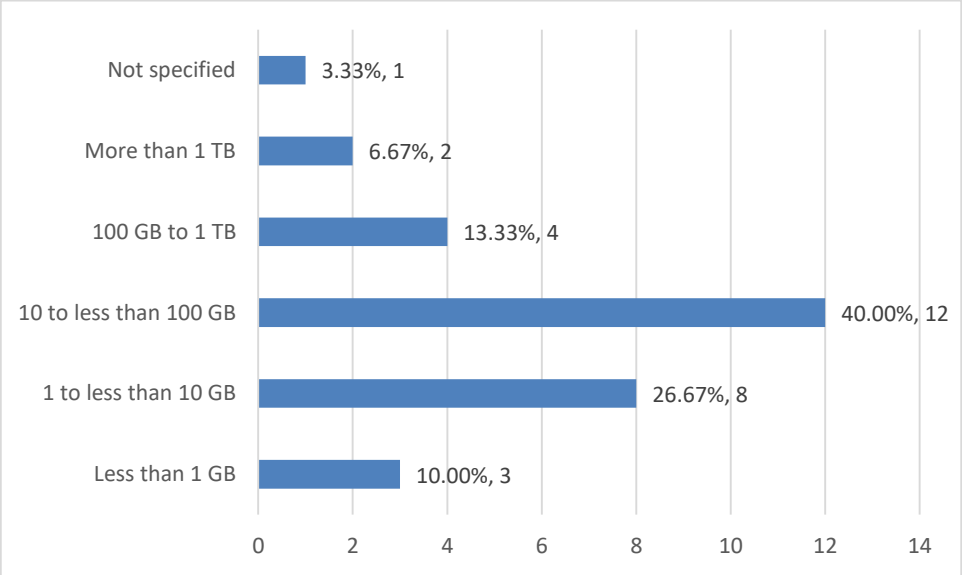


Figure 3.12. Distribution of audio/sound datasets by total size

More than one third of the collected datasets are smaller than 10 GB and can therefore be considered relatively manageable for educational use. Datasets between 10 and 100 GB may also be suitable for student projects, although downloading selected subsets or reducing the number of recordings may be necessary. Collections larger than 100 GB are more demanding in terms of storage, transfer time and processing capacity and are therefore generally more appropriate for advanced projects or research-oriented activities.

3.2.4 Audio Formats and Technical Characteristics

The technical suitability of the collected audio and sound datasets was assessed through their main audio formats, duration of recordings, sampling rates and channel configurations. These parameters influence compatibility with commonly used audio-processing tools and Python libraries, the level of preprocessing required and the computational resources needed for practical educational activities.

Audio Formats

As shown in Figure 3.13, WAV is the dominant audio format, used as the primary format in 19 datasets (63.33%). Five datasets (16.67%) contain mixed audio or multimodal formats, such as WAV combined with MP3, M4A, AIF, MIDI or STEMS files. MP3 is the primary format in 3 datasets (10.00%), while FLAC is represented in one dataset (3.33%). One dataset (3.33%) provides pre-extracted audio features in CSV and TensorFlow Record formats, while another dataset (3.33%) is distributed through compressed source formats such as MP4 and WebM.

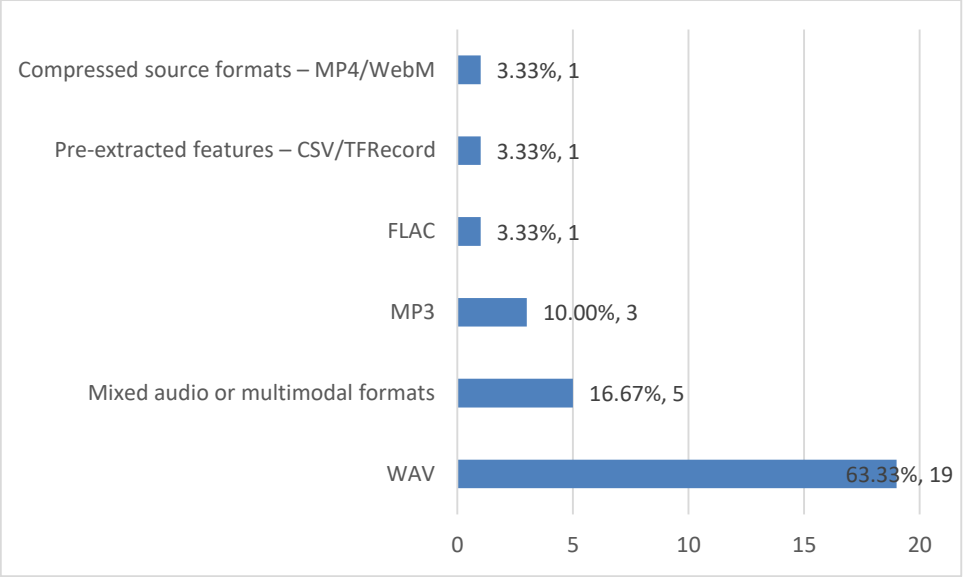


Figure 3.13. Distribution of audio and sound datasets by main data format

The predominance of WAV is favourable for educational use because it is widely supported by audio-processing tools and Python libraries. Mixed, compressed or specialised formats may require conversion or additional preprocessing before practical use.

Audio Duration

The collected datasets differ considerably in the duration of individual audio recordings. As shown in Figure 3.14, one dataset (3.33%) contains recordings shorter than one second. Eight datasets (26.67%) primarily contain recordings lasting from 1 to less than 10 seconds, while 9 datasets (30.00%) contain recordings with a typical duration from 10 to 30 seconds. Three datasets (10.00%) mainly contain recordings longer than 30 seconds, including full music tracks lasting several minutes. The remaining 9 datasets (30.00%) contain recordings with wide or variable duration ranges and were therefore classified separately.

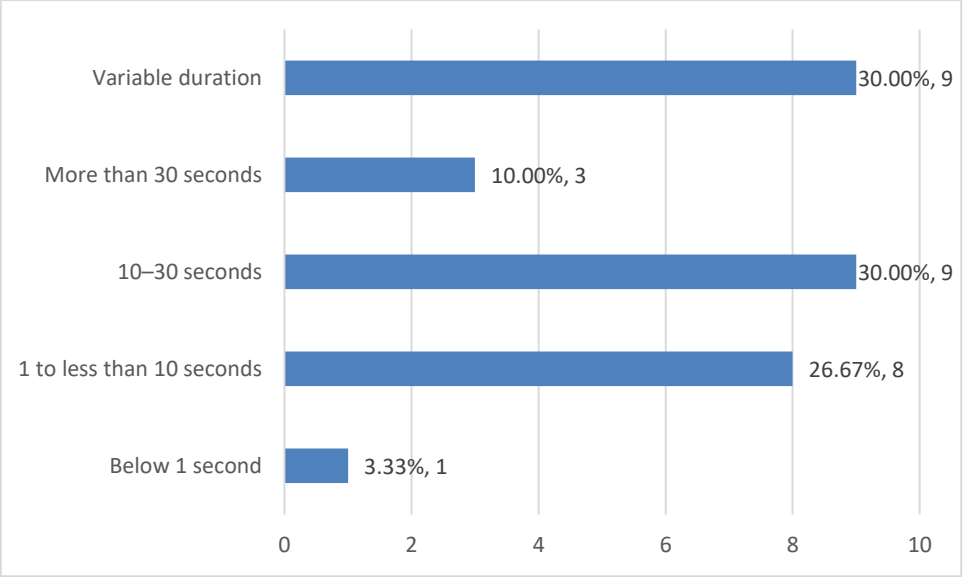


Figure 3.14. Distribution of audio and sound datasets by audio duration

Short and standardised recordings are generally easier to use in practical classes and are suitable for classification, keyword spotting and sound event detection. Longer or variable-duration recordings provide richer temporal information but may require trimming, segmentation or padding before model training.

Sampling Rate

The collected datasets use a range of sampling rates. As presented in Figure 3.15, 44.1 kHz is the most common sampling rate and is used in 11 datasets (36.67%). A sampling rate of 16 kHz is used in 7 datasets (23.33%), while 48 kHz is represented in 3 datasets (10.00%). Two datasets (6.67%) use 22.05 kHz, and one dataset (3.33%) uses 8 kHz. The remaining 6 datasets (20.00%) contain variable or multiple sampling rates.

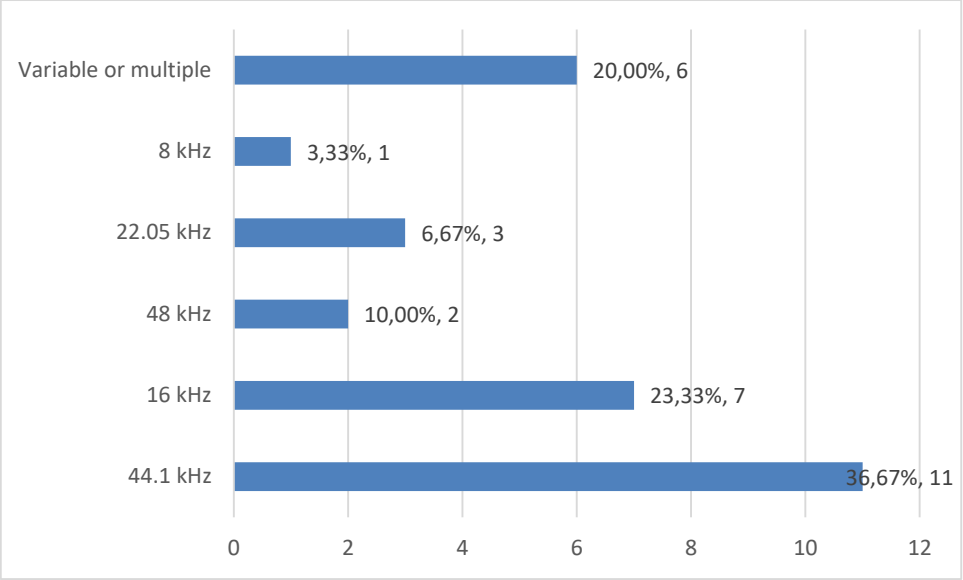


Figure 3.15. Distribution of audio and sound datasets by sampling rate

Sampling rates of 16 kHz are commonly used for speech-related tasks, while 44.1 and 48 kHz are widely used for music and environmental sound analysis. Datasets containing variable sampling rates may require resampling and standardisation before combined processing or model training.

Channel Configuration

Mono recordings are dominant in the collection and are used in 19 datasets (63.33%). Six datasets (20.00%) contain stereo recordings, while 2 datasets (6.67%) include both mono and stereo recordings. The remaining 3 datasets (10.00%) have variable channel configurations or do not provide clearly specified channel information.

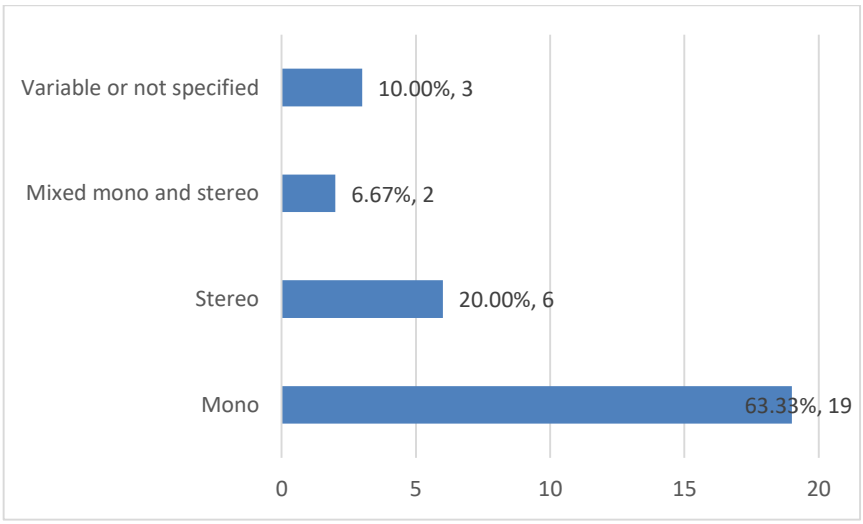


Figure 3.16. Distribution of audio and sound datasets by channel configuration

Mono recordings require less storage and are suitable for many speech- and sound-classification tasks, while stereo recordings preserve additional spatial information relevant to music and acoustic-scene analysis. Mixed channel configurations may require standardisation before model training.

3.2.5 Educational Applicability of Audio and Sound Datasets

The collected audio and sound datasets offer opportunities for teaching a wide range of AI tasks, including sound classification, speech recognition, speaker identification, emotion recognition, music analysis and bioacoustic monitoring. Their educational suitability depends not only on dataset size, but also on the duration and consistency of the recordings, sampling rate, channel configuration, availability of labels or transcripts, data format and the amount of preprocessing required. In an introductory practical class, a small and clearly labelled collection is often more useful than a technically complex benchmark containing hundreds of thousands of recordings. Larger and more heterogeneous datasets become more valuable as students develop the skills and computational resources needed to work with them.

For introductory practical exercises, datasets containing short WAV recordings, clearly defined classes and limited technical variation are particularly suitable. The Free Spoken Digit Dataset (FSDD) is a good example. It contains 3,000 recordings of spoken digits from zero to nine, occupies approximately 10 MB and provides recordings shorter than one second in a consistent mono WAV format. Its manageable size and clear digit labels make it suitable for introducing audio loading, waveform visualisation, feature extraction, spectrogram generation and basic classification. The ESC-50 Dataset provides a natural next step. It contains 2,000 five-second recordings organised into 50 balanced environmental sound classes. Because the files have a consistent sampling rate and channel configuration, students can focus on comparing audio representations and classification methods without first having to resolve extensive data-standardisation issues.

At the intermediate level, students can work with collections that introduce greater technical variation or more complex recognition tasks. UrbanSound8K contains 8,732 labelled urban sound excerpts from ten classes, with start and end times recorded for each excerpt. The dataset can support sound-event classification and the comparison of different feature-extraction approaches. At the same time, its variable sampling rates provide a realistic opportunity to teach resampling and audio standardisation before model training. The RAVDESS Dataset can be used for speech-emotion recognition through its 2,452 short audio-only recordings. Since the complete collection also includes video-only and audio-video files, it can provide a useful transition from single-modality audio analysis to more advanced multimodal emotion-recognition activities. Depending on the selected task, students may need to standardise channel configuration, extract the audio-only subset or align information from different modalities.

Advanced student projects and research-oriented activities can make use of substantially larger or more specialised datasets. LibriSpeech, which contains approximately 1,000 hours of read English speech and around 292,000 utterances with corresponding transcripts, is suitable for

automatic speech recognition and the evaluation of large speech-processing models. Its 60 GB size and large number of files make it more demanding in terms of storage, data organisation and model-training time. AudioSet represents an even more complex case, with more than two million ten-second clips covering a broad hierarchy of acoustic events. Its scale, YouTube-based source structure, variable technical characteristics and clip-level weak labels make it particularly relevant for research on large-scale audio tagging and sound-event recognition. However, working with the collection may require extensive data retrieval, validation, preprocessing and computational resources. Specialised collections such as Maestro, which combines aligned piano audio and MIDI data, can also support advanced work on music transcription, generation and multimodal music representation.

Table 3.2. Examples of audio and sound datasets suitable for different educational levels

Educational level	Dataset and suitable AI task	Reason for suitability	Required preparation or conditions
Introductory practical exercises	**Free Spoken Digit Dataset (FSDD)** – spoken-digit classification	Approximately 10 MB, 3,000 short recordings, consistent mono WAV format and clearly defined digit labels	Basic waveform normalisation and feature extraction; CC BY-SA 4.0 licence
Introductory practical exercises	**ESC-50** - environmental sound classification	2,000 five-second recordings, 50 balanced classes and consistent technical characteristics	Track-level labels are provided, but no temporal annotations; non-commercial use under CC BY-NC 3.0
Intermediate student projects	**UrbanSound8K** – urban sound-event classification	8,732 labelled excerpts from ten classes, accompanied by metadata and excerpt start and end times	Variable sampling rates may require resampling and standardisation; non-commercial use
Intermediate student projects	**RAVDESS** – speech-emotion recognition and multimodal emotion analysis	Short labelled recordings of several emotions, with audio-only and audio-video options	Channel standardisation or extraction of the audio-only subset may be required; CC BY-NC-SA 4.0 licence
Advanced or research-oriented activities	**LibriSpeech** – automatic speech recognition	Approximately 1,000 hours of speech, around 292,000 utterances and corresponding transcripts; CC BY 4.0 licence	Approximately 60 GB; large-scale file handling, FLAC processing and greater training resources may be required
Advanced or research-	**AudioSet** – audio tagging and sound-	More than two million human-labelled clips	YouTube-based sources, variable technical

oriented activities	event recognition	covering a broad range of acoustic events	characteristics, weak clip-level labels and approximately 2.4 TB of audio; access and usage conditions should be checked

The distinction between educational levels should not be understood as fixed. Many datasets can be adapted by selecting an appropriate subset or simplifying the intended task. The Free Music Archive Dataset, for example, is available as both a smaller collection of approximately 7.2 GB and a full collection of approximately 900 GB. The smaller version could support a student project in music classification, while the complete collection is more appropriate for large-scale research. Similarly, an introductory exercise may use only a limited number of classes from a larger dataset, whereas advanced students may work with the full collection and address issues such as data imbalance, heterogeneous recording conditions or model scalability.

Dataset selection should therefore be aligned with the intended learning outcomes, students' prior knowledge, available teaching time and computational resources. The licence and conditions of use should also be checked before recordings are reused, shared or incorporated into educational materials. The recommendations presented here are based on the characteristics documented by the dataset providers and recorded in the shared dataset collection document.

3.3 Numerical Datasets

The collected numerical datasets cover a broad range of applications relevant to Artificial Intelligence, machine learning, statistical analysis and data-driven decision-making. They include data related to healthcare, finance, energy, environmental monitoring, transportation, economics, engineering and other application areas. The datasets differ considerably in scale, structure, data formats, number of features, availability of time-series information, missing values and licensing conditions. The following analysis examines these characteristics and their relevance for AI education and research.

3.3.1 Main Applications of Numerical Datasets

The collected numerical datasets support a diverse range of AI and data-analysis tasks. As shown in Figure 3.17, the largest group consists of datasets primarily intended for regression and numerical prediction, accounting for 10 datasets (33.33%). These datasets support applications such as house-price prediction, building-energy assessment, materials-property prediction, market analysis and other numerical forecasting tasks.

Classification and decision-support applications represent the second largest category, with 9 datasets (30.00%). They cover medical diagnosis, customer churn prediction, defect detection, activity recognition, credit-risk assessment and other classification problems. Time-series forecasting and monitoring are represented by 7 datasets (23.33%), including applications related to energy consumption, air quality, climate, epidemiology and transportation demand.

Three datasets (10.00%) primarily support socioeconomic, macroeconomic and panel-data analysis, while one dataset (3.33%) is mainly intended for recommendation and combined predictive analysis.

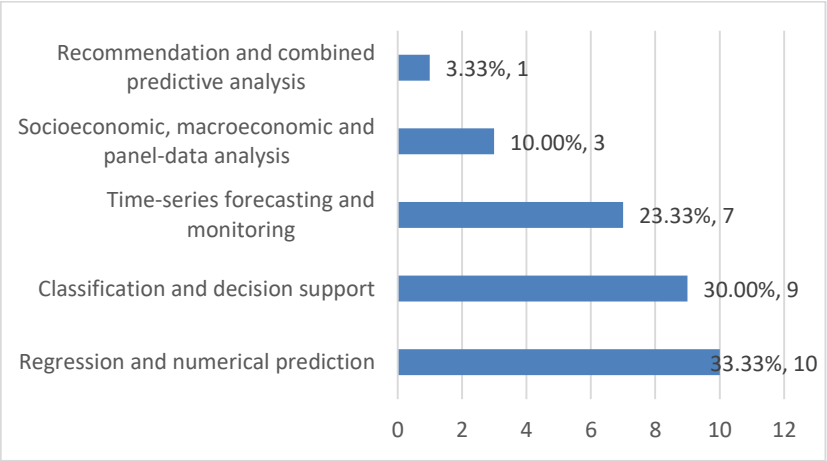


Figure 3.17. Distribution of numerical datasets by main application

This distribution demonstrates that the numerical dataset collection supports both fundamental machine-learning tasks and more specialised educational and research applications.

For the purpose of the analysis, each dataset was assigned to one main application category according to its primary intended use. Although some datasets support several tasks, only the dominant application was considered to avoid multiple counting. A detailed list of the datasets assigned to each main application category is provided in Annex 2.

3.3.2 Licence and Conditions of Use

The analysis of the licence information shows that the numerical datasets are generally available under comparatively open and clearly stated conditions of use. As presented in Figure 3.18, Creative Commons licences represent the largest category, covering 17 datasets or 56.7% of the collection. This category includes both attribution-based licences and Creative Commons licences containing additional non-commercial, share-alike or no-derivatives conditions.

A further 11 datasets, representing 36.7% of the collection, are available under public-domain or other open-use conditions, including CC0, MIT, open-data and similar reuse arrangements. The remaining 2 datasets, or 6.7%, are subject to separate non-commercial, institutional or otherwise restricted conditions. Licence information was recorded for all numerical datasets.

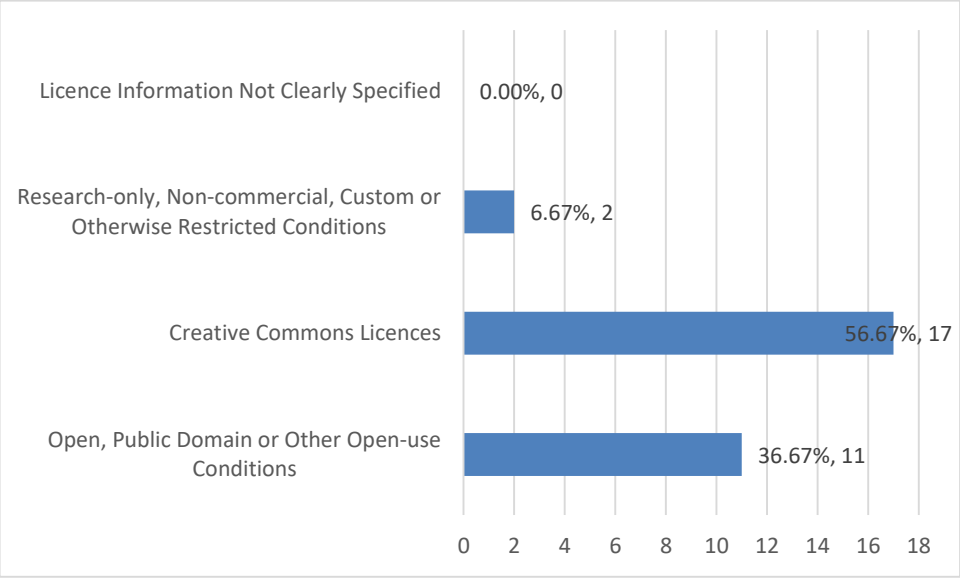


Figure 3.18. Distribution of numerical datasets by licence and conditions of use

Overall, the licensing structure of the numerical dataset collection is favourable for educational and research activities. Nevertheless, the specific conditions of use, including attribution requirements, restrictions on commercial use and institutional limitations, should be checked

before datasets or derived materials are redistributed or included in publicly available teaching resources.

3.3.3 Dataset Scale: Number of Records and Total Size

The scale of the collected numerical datasets was analysed using two complementary parameters: the number of records and the total dataset size. The number of records indicates the amount of data available for model training and evaluation, while total size provides an indication of download, storage and processing requirements.

Number of Records

As shown in Figure 3.19, 4 datasets (13.33%) contain fewer than 1,000 records, while 6 datasets (20.00%) contain between 1,000 and 9,999 records. The largest group consists of 12 datasets (40.00%) containing between 10,000 and 99,999 records. Six datasets (20.00%) contain between 100,000 and 999,999 records, while 2 datasets (6.67%) contain one million or more records.

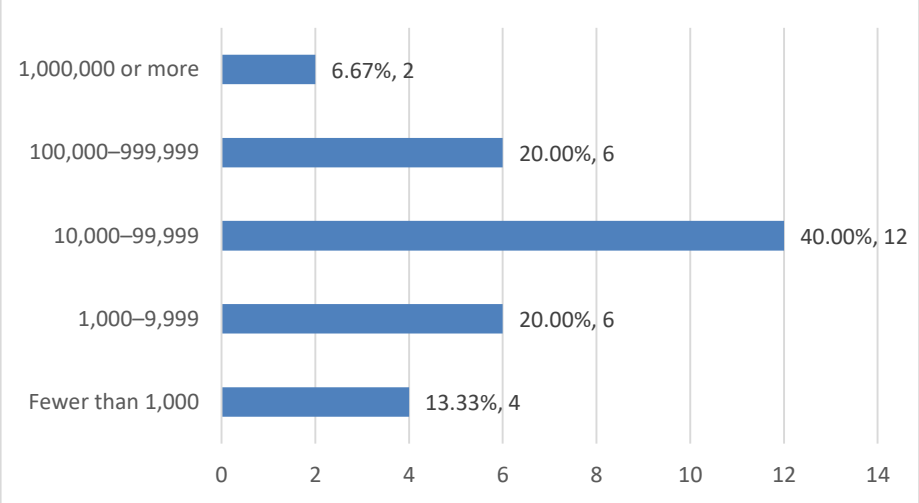


Figure 3.19. Distribution of numerical datasets by number of records

Datasets containing fewer than 10,000 records are generally easier to inspect and process during introductory practical classes. Medium-sized datasets support a broad range of classification, regression and exploratory data-analysis exercises, while datasets containing hundreds of thousands or millions of records are more appropriate for advanced projects and research-oriented activities.

Where a dataset was provided through several files, regions, cities or monthly releases, the reported number of records for the representative or complete collection was used based on the information available in the dataset collection document.

Total Dataset Size

The numerical datasets are generally smaller than the collected video and audio datasets. Nine datasets (30.00%) are smaller than 1 MB, while 8 datasets (26.67%) range from 1 to less than 10 MB. A further 8 datasets (26.67%) range from 10 to less than 100 MB. Four datasets (13.33%) have sizes between 100 MB and 1 GB, and only one dataset (3.33%) exceeds 1 GB.

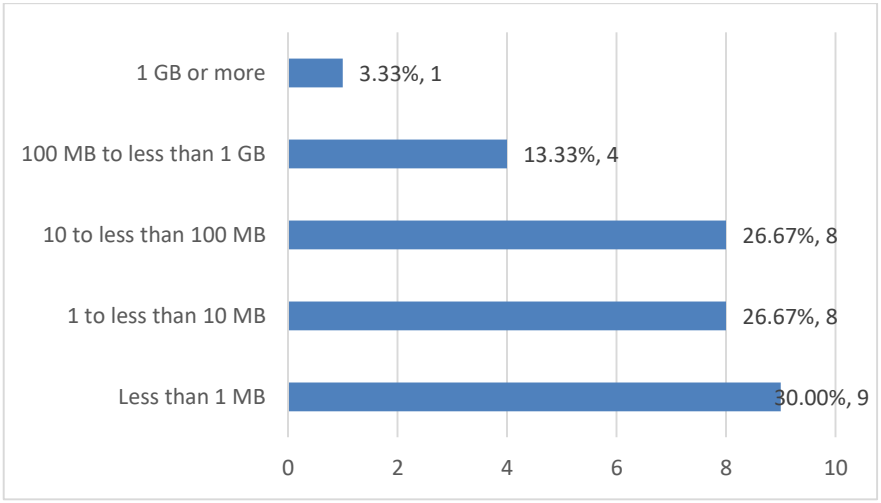


Figure 3.20. Distribution of numerical datasets by total size

More than half of the numerical datasets are smaller than 10 MB and can therefore be considered highly manageable for educational use. Even most of the larger collections can be processed using standard computers, while the dataset exceeding 1 GB may require greater memory, longer processing time or the extraction of smaller subsets.

3.3.4 Data Formats and Technical Characteristics

The technical suitability of the numerical datasets was assessed through their main data formats, number of features, feature types, presence of missing values and time-series properties. These characteristics influence compatibility with commonly used data-analysis tools and Python libraries, the level of preprocessing required and the types of educational activities that can be performed.

Data Formats

CSV is the dominant format and is used as the primary format in 21 datasets (70.00%). Five datasets (16.67%) use mixed tabular formats, such as CSV combined with TXT, TSV, XLSX or Parquet. Three datasets (10.00%) are primarily distributed in Excel formats, while one dataset (3.33%) is provided in TXT format.

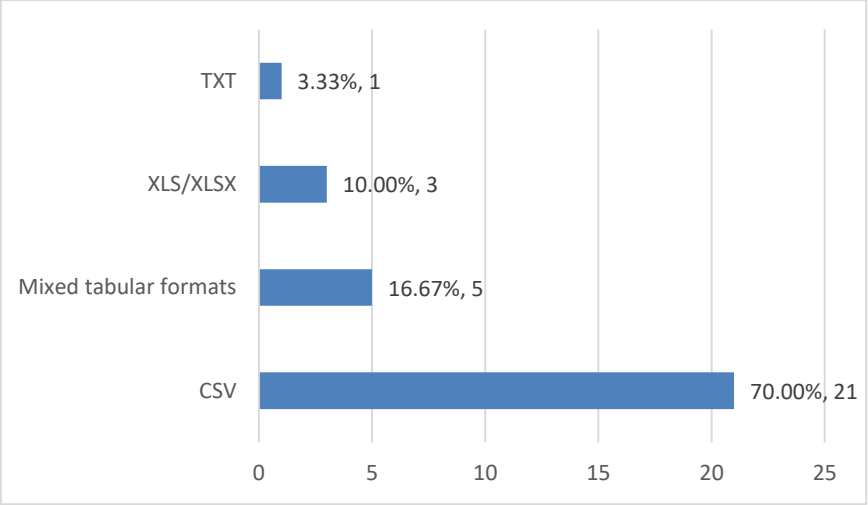


Figure 3.21. Distribution of numerical datasets by main data format

The predominance of CSV is favourable for educational use because this format is widely supported by spreadsheet software, statistical packages and Python libraries. Excel, TXT, TSV and Parquet formats are also compatible with commonly used tools, although mixed-format collections may require additional import or conversion steps.

Number of Features

The numerical datasets also differ substantially in dimensionality. Nine datasets (30.00%) contain 10 or fewer features, while 10 datasets (33.33%) contain between 11 and 20 features. Three datasets (10.00%) contain between 21 and 50 features, and 5 datasets (16.67%) contain between 51 and 100 features. The remaining 3 datasets (10.00%) contain more than 100 features.

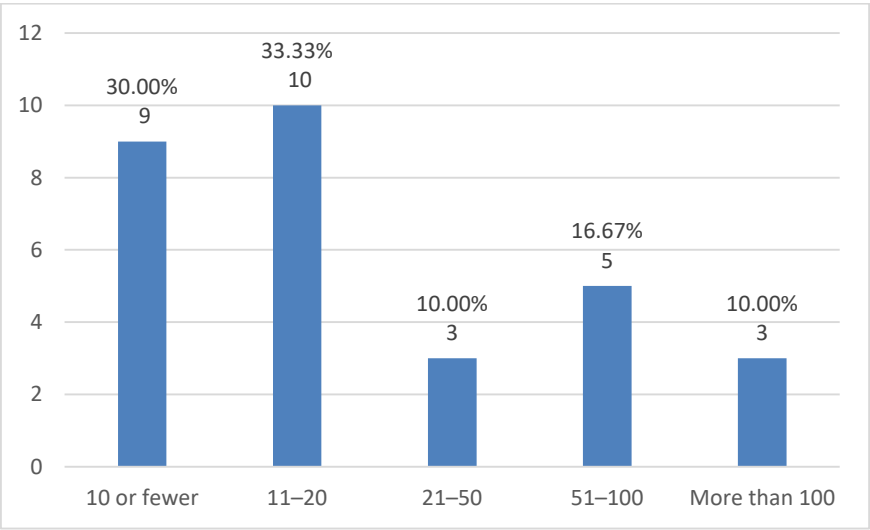


Figure 3.22. Distribution of numerical datasets by number of features

Datasets with a limited number of features are generally easier to understand, visualise and use in introductory exercises. Higher-dimensional datasets provide opportunities for teaching feature selection, dimensionality reduction and more advanced modelling techniques but may require additional preprocessing and interpretation.

Missing Values

Thirteen datasets (43.33%) contain no reported missing values. Four datasets (13.33%) contain only minimal or limited missing values, while 13 datasets (43.33%) include clearly reported missing or sparse data.

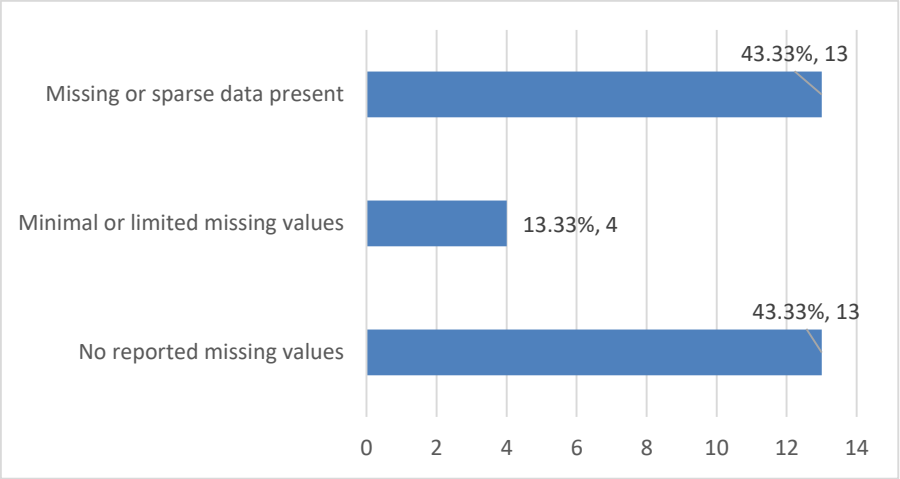


Figure 3.23. Distribution of numerical datasets by availability of missing values

Datasets without missing values are generally easier to integrate into introductory exercises. Datasets containing incomplete or sparse records provide useful opportunities for teaching data-quality assessment, missing-value visualisation, deletion strategies and imputation methods.

Time-Series Properties

Thirteen datasets (43.33%) contain time-series or temporally ordered data, while 17 datasets (56.67%) are primarily cross-sectional or non-time-series collections.

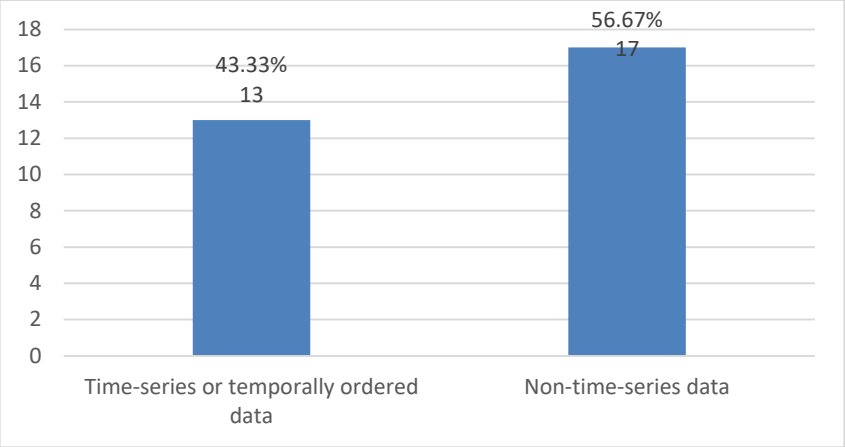


Figure 3.24. Distribution of numerical datasets by time-series properties

The presence of time-series data supports exercises involving trend analysis, seasonality, lag features and forecasting. Non-time-series datasets are suitable for classification, regression, clustering and exploratory data analysis. Datasets with different temporal frequencies may require sorting, aggregation, resampling or time-based train-test splitting.

3.3.5 Educational Applicability of Numerical Datasets

Numerical datasets provide a particularly accessible starting point for AI education because many of them are available in standard tabular formats and can be processed using spreadsheet software, statistical packages and commonly used Python libraries. However, a dataset that is easy to open is not necessarily simple to analyse. Its educational suitability also depends on the number and type of features, the presence of missing values, the clarity of the target variable, the temporal structure of the data and the amount of preparation required before modelling. Smaller and well-structured datasets are generally suitable for introducing the main stages of a machine-learning workflow, while larger, higher-dimensional and time-series collections provide opportunities for more advanced data preparation, modelling and evaluation.

For introductory practical exercises, datasets with a limited number of features, clearly defined target variables and no missing values are especially useful. The Energy Efficiency Dataset, for example, contains 768 records and eight numerical input features describing different building designs, with heating and cooling loads as target variables. Its small size, clearly defined regression task and complete data make it suitable for introducing exploratory data analysis, correlation analysis, train-test splitting, regression modelling and performance evaluation. The California Housing Dataset provides a slightly larger but still manageable example, with 20,640 records, eight numerical features and no missing values. Since the dataset is intended for house-value prediction and is available in a standard tabular format, students can compare different regression methods and learn how model performance changes when working with a larger number of observations.

At the intermediate level, datasets can introduce missing values, temporal ordering or higher dimensionality. The Air Quality Dataset contains 9,358 hourly records from a gas multisensor device, together with reference pollutant concentrations. Missing values are recorded using the value −200, which provides a practical opportunity to teach data-quality assessment, missing-value identification and cleaning before regression or sensor-calibration tasks. Because the observations are ordered in time, the dataset can also introduce time-aware data splitting and the importance of preserving temporal order during model evaluation. The Human Activity Recognition Using Smartphones Dataset contains 10,299 records and 561 normalised features derived from smartphone accelerometer and gyroscope signals. It supports classification of six human activities and is suitable for teaching feature selection, dimensionality reduction and the comparison of classification models. Its large number of features makes it more demanding than a conventional introductory tabular dataset, even though its overall size remains manageable.

Advanced student projects and research-oriented activities can use large-scale, sparse or continuously updated datasets. The NYC Taxi Trip Record Dataset contains millions of trip records per month and includes pickup and drop-off times, locations, trip distances, fares and payment information. It can support research on demand analysis, travel-time or fare prediction and temporal patterns in urban mobility. Its monthly releases and Parquet or CSV formats make it possible to work with selected periods or progressively larger subsets, depending on the available computing resources. However, the full collection may require efficient file handling, timestamp processing, aggregation and scalable data-analysis methods. The Lending Club Loan Dataset is another advanced example, containing approximately 2.26 million accepted loan records and 151 borrower, loan and repayment attributes. Its size, mixed feature types and many sparse columns provide opportunities for advanced work on credit-risk classification, feature selection and missing-data treatment, but may also require extensive preprocessing and careful interpretation of model results. The World Development Indicators Dataset, with long-term data for 217 economies and approximately 1,400 development indicators, can support research involving panel data, macroeconomic trends and time-series analysis. Its broad scope and sparse structure mean that relevant indicators, countries and periods should normally be selected before analysis.

Table 3.3. Examples of numerical datasets suitable for different educational levels

Educational level	Dataset and suitable AI task	Reason for suitability	Required preparation or conditions
Introductory practical exercises	**Energy Efficiency Dataset** – prediction of building heating and cooling loads	768 records, eight numerical features, clearly defined targets and no missing values; CC BY 4.0 licence	Basic data inspection, scaling and train-test splitting

Introductory practical exercises	**California Housing Dataset** – house-value prediction	20,640 records, eight numerical features, no missing values and a clearly defined regression task; public domain	Basic exploratory analysis, scaling and comparison of regression models
Intermediate student projects	**Air Quality Dataset** – pollutant prediction and sensor calibration	Hourly time-series data with 9,358 records, 15 features and a realistic data-quality problem; CC BY 4.0 licence	Values coded as −200 need to be identified and treated as missing; temporal order should be considered
Intermediate student projects	**Human Activity Recognition Using Smartphones Dataset** – activity classification	10,299 records, 561 normalised features and six activity classes; CC BY 4.0 licence	High dimensionality may require feature selection or dimensionality reduction; TXT files may require structured import
Advanced or research-oriented activities	**NYC Taxi Trip Record Dataset** – demand analysis, fare prediction and mobility modelling	Millions of records per month, time and location information and regular monthly releases; public-domain data	Selection of relevant periods, timestamp processing, aggregation and efficient Parquet or CSV handling may be required
Advanced or research-oriented activities	**Lending Club Loan Dataset** – credit-risk classification and regression	Approximately 2.26 million records and 151 borrower, loan and repayment attributes; CC0 public-domain licence	Approximately 1.5 GB, with mixed feature types and many sparse columns; extensive cleaning and feature selection may be required

The proposed classification is intended as practical guidance rather than a fixed division. The same dataset may support different educational levels depending on how the activity is designed. A small subset of the NYC Taxi Trip Record Dataset could be used in an intermediate project, while the complete multi-year collection would be more appropriate for advanced analysis. Similarly, a compact dataset can become conceptually demanding when it contains many feature types, missing values or data from a specialised domain. The Cleveland Heart Disease Dataset, for example, contains only 303 records and 13 features, but its medical context requires careful interpretation and makes it more suitable for demonstrating classification methods than for drawing clinical conclusions.

Dataset selection should therefore reflect the intended learning outcomes, students' previous experience, the complexity of the required preprocessing and the available computational resources. For time-series datasets, temporal ordering and appropriate train-test splitting should be considered, while datasets containing categorical or incomplete data may require encoding, cleaning or imputation. Licence conditions should also be verified before datasets are reused or incorporated into educational materials. The recommendations presented here are based on the information documented by dataset providers and recorded in the shared dataset collection document.

3.4 Image Datasets

The collected image datasets cover a broad range of applications relevant to Artificial Intelligence (AI), Machine Learning (ML) and computer vision. They include general and fine-grained image classification, object detection, semantic segmentation, facial analysis, autonomous driving, medical imaging, remote sensing and agricultural applications. The datasets differ considerably in scale, formats, resolution, colour space, annotation level and licensing conditions. The following analysis examines these characteristics and their relevance for AI education and research.

3.4.1 Main Applications of Image Datasets

The collected image datasets support a diverse range of computer vision tasks. As shown in Figure 3.25, the largest group consists of datasets primarily intended for image classification and fine-grained recognition, accounting for 11 datasets (36.67%). These datasets support tasks such as object, scene, vehicle, clothing and handwritten-digit classification.

Object detection, localisation and pose estimation represent the second largest category, with 7 datasets (23.33%). Five datasets (16.67%) primarily support semantic segmentation and scene understanding. Face analysis and generative modelling, remote sensing and agricultural analysis, and fashion retrieval and image generation are each represented by 2 datasets (6.67%). One dataset (3.33%) is primarily intended for biomedical image analysis.

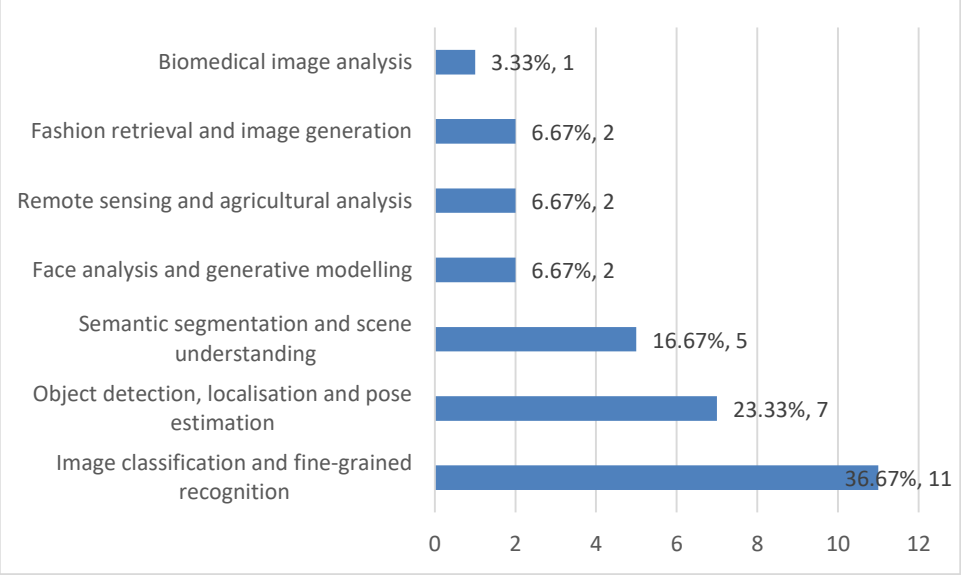


Figure 3.25. Distribution of image datasets by main application

This distribution demonstrates that the image dataset collection supports both widely used computer-vision tasks and more specialised educational and research applications.

For the analysis, each dataset was assigned to one main application category according to its primary intended use. Although some datasets support several computer-vision tasks, only the dominant application was considered to avoid multiple counting. A detailed list of the datasets assigned to each main application category is provided in Annex 2.

3.4.2 Licence and Conditions of Use

The analysis of the licence information shows that the collected image datasets differ considerably in their conditions of use. As presented in Figure 3.26, Creative Commons licences represent the largest category, covering 11 datasets or 36.7% of the collection. This category includes attribution-based licences as well as Creative Commons licences containing non-commercial or share-alike conditions.

A further 10 datasets, representing 33.3% of the collection, are subject to separate non-commercial, academic, research-only, custom or otherwise restricted conditions. Nine datasets, or 30.0%, are available under public-domain or other open-use conditions, including CC0, MIT and other open research arrangements. Licence information or stated conditions of use were recorded for all 30 image datasets.

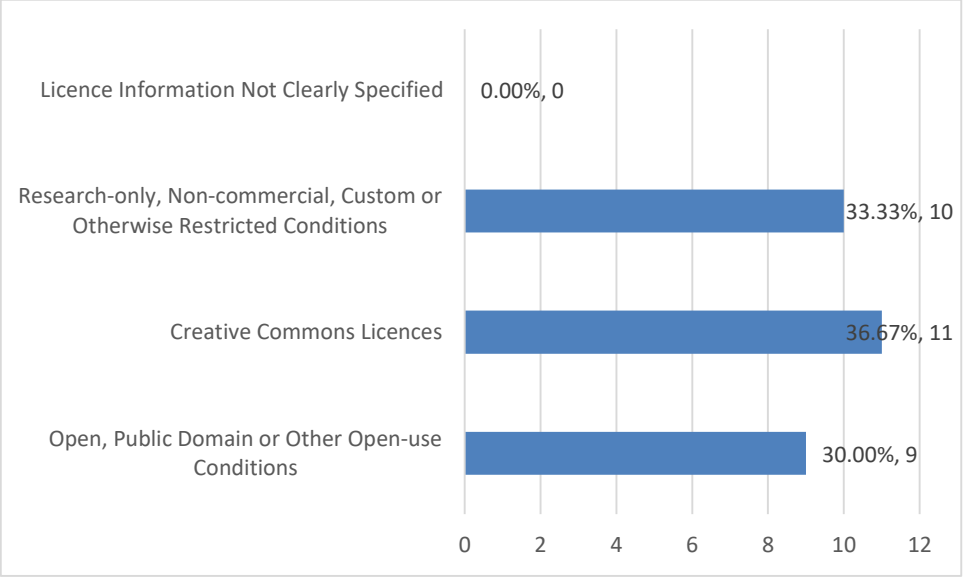


Figure 3.26. Distribution of image datasets by licence and conditions of use

Overall, the image dataset collection provides a broad range of resources for educational and research activities. However, the specific license conditions should be checked before datasets, or derived materials are redistributed or incorporated into publicly available teaching resources.

3.4.3 Dataset Scale: Number of Images and Total Size

The scale of the collected image datasets was analysed using two complementary parameters: the number of images and total dataset size. The number of images indicates the amount of material available for model training and evaluation, while total size provides an indication of download, storage and computational requirements.

Number of Images

As shown in Figure 3.27, 4 datasets (13.33%) contain fewer than 10,000 images, while the largest group, consisting of 14 datasets (46.67%), contains between 10,000 and 99,999 images. Nine datasets (30.00%) contain between 100,000 and 999,999 images, while 3 datasets (10.00%) contain one million or more images.

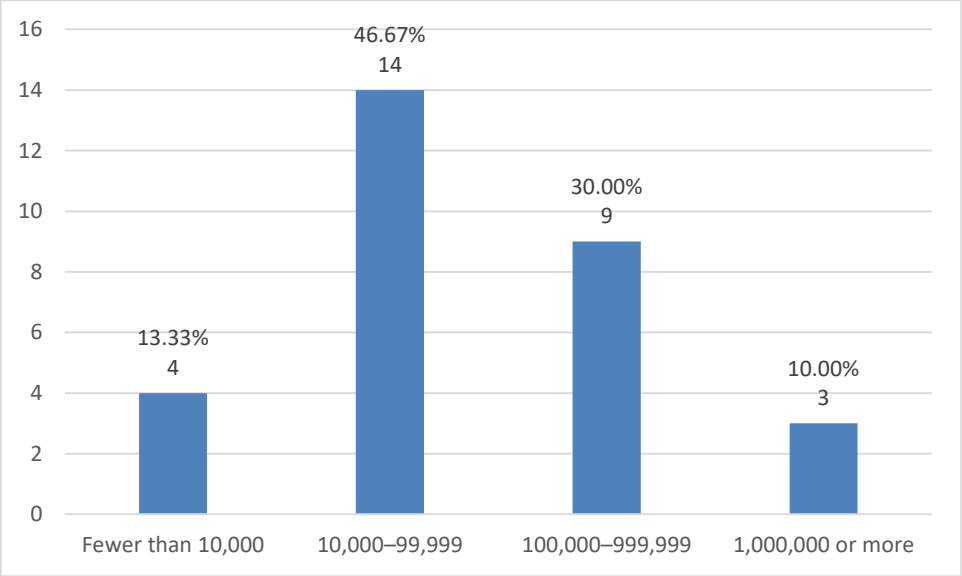


Figure 3.27. Distribution of image datasets by number of images

Smaller image datasets may be easier to manage in practical classes and student projects. Datasets containing tens or hundreds of thousands of images provide greater potential for model training and evaluation but generally require more preparation and computational resources. Collections containing millions of images are more appropriate for advanced projects or research-oriented activities.

Where datasets include frames, image patches or cropped objects rather than conventional image files, these units were treated as the closest equivalents to individual images.

Total Dataset Size

The total size of the collected image datasets also varies substantially. Eight datasets (26.67%) are smaller than 1 GB, while 9 datasets (30.00%) range from 1 to less than 10 GB. A further 9 datasets (30.00%) have sizes ranging from 10 to less than 100 GB, and 4 datasets (13.33%) range from 100 GB to 1 TB.

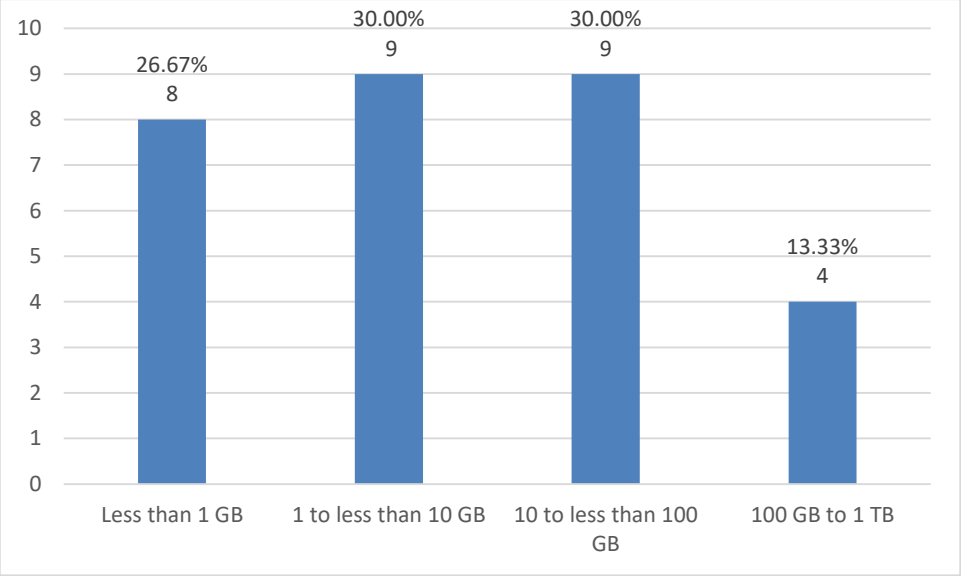


Figure 3.28. Distribution of image datasets by total size

More than half of the image datasets are smaller than 10 GB and can therefore be considered relatively manageable for educational use. Datasets between 10 and 100 GB may require the extraction of smaller subsets, while collections larger than 100 GB are generally more suitable for advanced student projects and research-oriented activities.

3.4.4 Image Formats and Technical Characteristics

The technical suitability of the collected image datasets was assessed through their main image formats, resolution, colour space and annotation level. These characteristics influence compatibility with commonly used computer-vision tools and Python libraries, the level of preprocessing required and the types of educational activities that can be performed.

Image Formats

JPEG is the dominant image format and is used as the primary format in 16 datasets (53.33%). PNG is the main format in 7 datasets (23.33%), while 3 datasets (10.00%) contain mixed standard image formats, such as JPEG combined with PNG or TIFF. Four datasets (13.33%) use proprietary, raw or specialised storage formats.

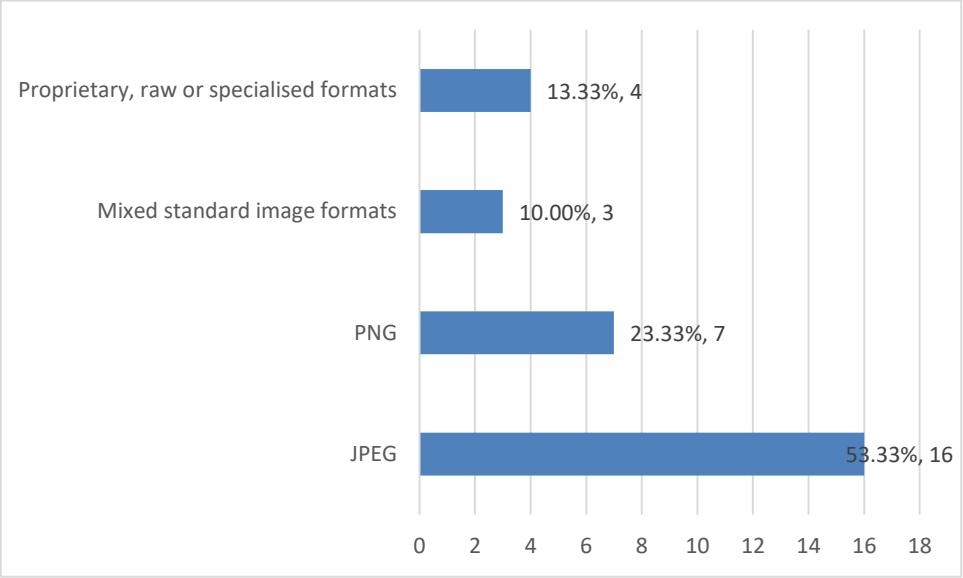


Figure 3.29. Distribution of image datasets by main data format

The predominance of JPEG and PNG is favourable for educational use because these formats are widely supported by computer-vision software and Python libraries. Proprietary, raw and specialised formats may require additional extraction, decoding or conversion before practical use.

Image Resolution

The image datasets contain a wide range of resolutions. Six datasets (20.00%) primarily contain low-resolution images, while 7 datasets (23.33%) mainly contain medium-resolution images. Five datasets (16.67%) contain high-resolution images, and 12 datasets (40.00%) include variable or multiple resolutions.

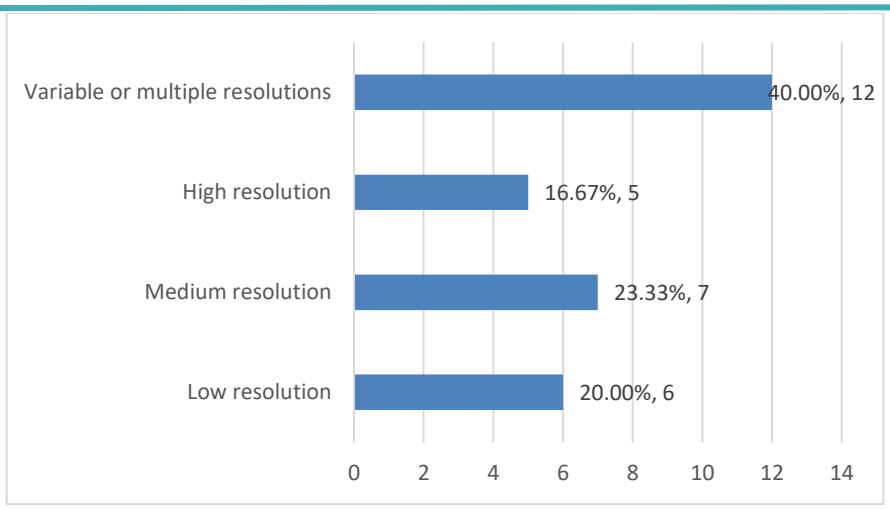


Figure 3.30. Distribution of image datasets by resolution category

Low-resolution datasets generally require fewer computational resources and may be suitable for introductory classification exercises. High-resolution and variable-resolution datasets provide richer visual information but may require resizing, cropping or other standardisation before model training.

Image Colour Space

RGB images dominate the collection and are used in 24 datasets (80.00%). Three datasets (10.00%) contain grayscale images. One dataset (3.33%) includes both RGB and grayscale images, one dataset (3.33%) contains multispectral images, and one dataset (3.33%) combines RGB images with depth information.

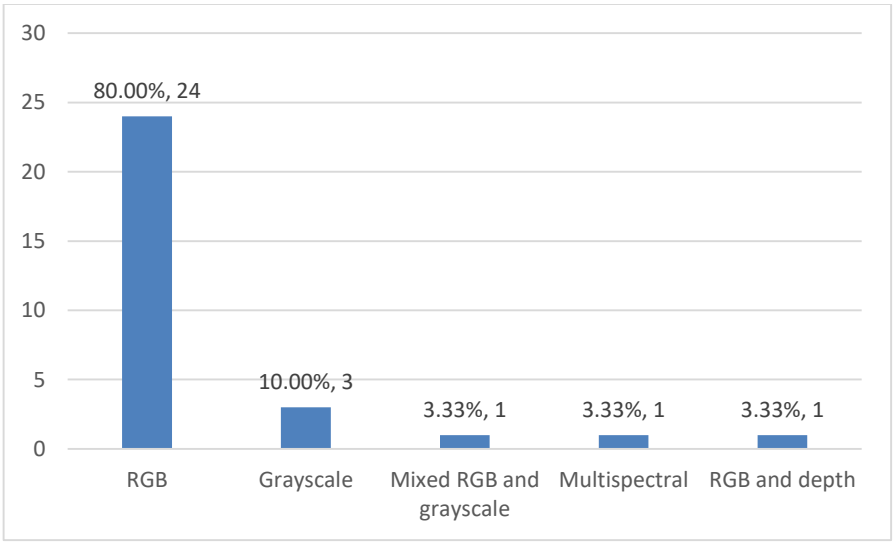


Figure 3.31. Distribution of image datasets by colour space

RGB images are compatible with most standard computer-vision models and tools. Grayscale images require fewer computational resources, although they may be converted to three-channel representations when used with models pre-trained on RGB images. Multispectral and RGB-depth datasets provide additional spectral or spatial information but generally require more specialized preprocessing and modelling approaches.

Annotation Level

The collected datasets also differ in the type and level of available annotations. Twelve datasets (40.00%) primarily provide image-level labels, such as object, scene, identity or class categories. Eleven datasets (36.67%) contain object-level annotations, including bounding boxes, landmarks, key points or pose coordinates. Seven datasets (23.33%) include pixel-level annotations for semantic, instance or panoptic segmentation.

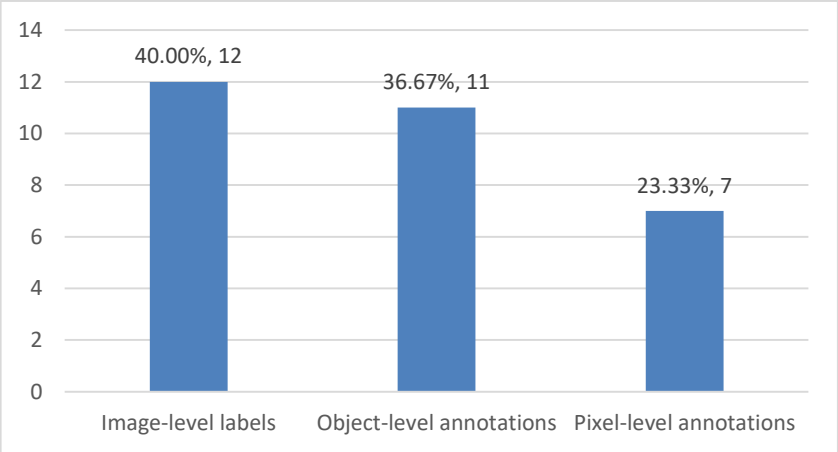


Figure 3.32. Distribution of image datasets by annotation level

Image-level labels are generally suitable for classification exercises, while bounding boxes, landmarks and key points support object detection, localisation and pose-estimation tasks. Pixel-level annotations enable semantic and instance segmentation but require more complex models and greater computational resources.

Where datasets provide several annotation types, they were classified according to the most detailed or primary annotation level used for their main intended application.

3.4.5 Educational Applicability of Image Datasets

The collected image datasets support a broad range of educational activities, from basic image classification to object detection, semantic segmentation, pose estimation and three-dimensional scene understanding. Their level of difficulty is not determined by the number of images alone. A collection containing many small, standardised images with image-level labels can be easier to use than a much smaller collection containing variable-resolution images,

bounding boxes or pixel-level segmentation masks. Educational suitability therefore depends on image resolution and format, annotation type, dataset size, technical complexity, required preprocessing and the computational resources available for model training.

For introductory practical exercises, standardised datasets with clearly defined image-level classes are particularly suitable. CIFAR-10, for example, contains 60,000 RGB images of 32 × 32 pixels organised into ten classes, with predefined training and test sets. Its small total size and uniform PNG format allow students to focus on the main stages of an image-classification workflow, including image loading, normalisation, model training and performance evaluation. The PlantVillage Dataset provides a more application-oriented example. It contains 54,306 RGB images with a standardised resolution of 256 × 256 pixels and 38 image-level labels representing crop species and disease categories. Its folder-based structure, JPEG format and public-domain licence make it suitable for introductory classification exercises and initial experiments with transfer learning.

At the intermediate level, students can work with datasets containing more detailed annotations and variable image characteristics. The Oxford-IIIT Pet Dataset includes 7,349 images of 37 cat and dog breeds, together with breed labels, head-region bounding boxes and pixel-level trimap segmentations. This allows the same dataset to support a gradual progression from fine-grained classification to object localisation and semantic segmentation. Because image resolutions vary and annotations are provided in XML and PNG formats, students may need to perform resizing and annotation processing before model training. PASCAL VOC 2012 offers a further step towards realistic computer-vision projects. It contains 11,530 images, 27,450 annotated object instances, bounding-box coordinates and pixel-level segmentation masks. The dataset can support object detection and semantic-segmentation activities, while its XML annotations and variable image resolutions provide useful experience in preparing heterogeneous image data.

Datasets with specialised image channels can also be introduced at the intermediate or advanced level. The EuroSAT Dataset, for example, contains 27,000 satellite-image patches for land-use and land-cover classification. Its RGB JPEG version can support relatively accessible classification exercises, while the 13-band multispectral TIFF version introduces more advanced issues such as handling additional spectral channels, 16-bit image data and remote-sensing-specific preprocessing. This makes EuroSAT a useful example of how the same dataset can support different educational levels depending on the selected data representation.

Advanced student projects and research-oriented activities can make use of larger collections with complex annotation structures or combined two-dimensional and three-dimensional information. The COCO Dataset contains approximately 328,000 images and supports object detection, instance segmentation, keypoint estimation and image captioning. Its variable image resolutions and extensive JSON annotations allow students to work on realistic deep-learning

tasks, but may require subset extraction, annotation parsing and substantial computational resources. The KITTI-360 Dataset is even more technically demanding, with more than 320,000 frames, high-resolution fisheye images and unified two-dimensional and three-dimensional semantic instance annotations. Its structured sequences and PLY-based 3D data make it relevant for autonomous-driving research, 3D scene understanding and novel-view synthesis, although working with the complete 150 GB collection may require dedicated processing tools and significant storage and computing capacity. Similarly, SUN RGB-D, which combines RGB images with depth information and 3D-oriented bounding boxes, can support advanced research on indoor-scene understanding and 3D object detection.

Table 3.4. Examples of image datasets suitable for different educational levels

Educational level	Dataset and suitable AI task	Reason for suitability	Required preparation or conditions
Introductory practical exercises	**CIFAR-10** – image classification	60,000 standardised 32 × 32 RGB images, ten clearly defined classes and predefined training and test sets; CC0/public-domain licence	Basic normalisation, optional data augmentation and model training
Introductory practical exercises	**PlantVillage Dataset** – crop and plant-disease classification	54,306 standardised 256 × 256 JPEG images organised into 38 image-level classes; CC0 1.0 public-domain licence	Preparation of training, validation and test subsets may be required
Intermediate student projects	**Oxford-IIIT Pet Dataset** – fine-grained classification and semantic segmentation	Manageable size with breed labels, bounding boxes and pixel-level trimap annotations; CC BY-SA 4.0 licence	Variable image resolutions, XML annotation processing, resizing and mask preparation
Intermediate student projects	**PASCAL VOC 2012** – object detection and semantic segmentation	11,530 images with bounding boxes and pixel-level segmentation masks; public-domain/open-research conditions	Parsing XML annotations, processing segmentation masks and standardising image dimensions
Advanced or research-oriented activities	**COCO Dataset** – object detection, instance segmentation, keypoint estimation and captioning	Approximately 328,000 images with several detailed annotation types; CC BY 4.0 licence	Approximately 25 GB, variable resolutions and complex JSON annotations; subset extraction and substantial training resources may be required
Advanced or	**KITTI-360** – 3D	More than 320,000 frames	Approximately 150 GB,

research-oriented activities	scene understanding and autonomous-navigation research	with high-resolution images and unified 2D/3D semantic instance annotations	structured sequences and PLY-based 3D data; dedicated processing tools and significant computing resources may be required

The proposed classification is indicative rather than fixed. CIFAR-10 contains substantially more images than the Oxford-IIIT Pet Dataset, but its uniform resolution and simple image-level labels make it easier to use. Conversely, the Oxford-IIIT Pet Dataset is smaller but becomes more demanding when its bounding boxes and segmentation masks are included. Similarly, a selected subset of COCO could support an intermediate object-detection project, while the complete collection is more appropriate for advanced model development and research.

Dataset selection should therefore reflect the intended learning outcomes, the type of annotation required for the selected AI task and the available computational resources. Image-level labels are generally sufficient for classification exercises, while bounding boxes, key points and pixel-level masks are needed for detection, pose-estimation and segmentation tasks. High-resolution, multispectral, RGB-depth and 3D datasets may require resizing, channel transformation, specialised file handling or the use of smaller subsets. Licence and access conditions should also be verified before images or annotations are reused or included in educational materials. The recommendations presented here are based on the characteristics documented by the dataset providers and recorded in the shared dataset collection document.

4. Comparative Assessment and Educational Applicability

The analysis shows that the four dataset modalities collectively offer broad potential for AI education and research, although the suitability of individual datasets differs depending on dataset scale, technical complexity, format, licensing conditions and available computational resources.

The comparative assessment therefore focuses on the main similarities and differences between video, audio/sound, numerical and image datasets, as well as the requirements associated with their use in practical exercises, student projects and research-oriented activities. Since the basic data units differ across modalities, the number of videos, audio files, numerical records and images should not be compared directly. Instead, the assessment considers overall manageability, technical requirements, preprocessing needs and conditions of use.

4.1 Comparative Findings

The complete collection comprises 120 datasets, equally distributed across four data modalities: video, audio/sound, numerical and image datasets. Although each modality is represented by 30 datasets, the numbers of individual data items cannot be directly compared because a video recording, an audio file, an image and a numerical record represent fundamentally different units of data.

The consolidated comparison therefore focuses only on characteristics that were recorded and analysed across all four modalities and that can be interpreted consistently. These include licence and access conditions, availability of information on total dataset size, the use of standard or specialised data formats, general preprocessing requirements, relative computational demands and potential educational suitability. Modality-specific parameters, such as video frame rate, audio sampling rate, the number of numerical features and image colour space, are not combined because they are not directly comparable.

Licence and Access Conditions

For the cross-modal comparison, Creative Commons attribution-based licences, including variants with non-commercial, share-alike or no-derivatives conditions, were grouped into one category. CC0 licences were included under open, public-domain or other open-use conditions. Separate research-only, academic, non-commercial, custom or otherwise restricted conditions were included in a distinct category.

Table 4.1. Consolidated overview of licence conditions across the dataset collection

Licence category	Video	Audio/ Sound	Numerical	Image	Total
Open, Public Domain or Other Open-use Conditions	4	2	11	9	26
Creative Commons Licences	12	22	17	11	62
Research-only, Non-commercial, Custom or Otherwise Restricted Conditions	9	5	2	10	26
Licence Information Not Clearly Specified	5	1	0	0	6
Total	**30**	**30**	**30**	**30**	**120**

Creative Commons licences represent the largest category, covering 62 datasets or 51.7% of the complete collection. A further 26 datasets, or 21.7%, are available under public-domain or other open-use conditions. Together, these two categories account for 88 datasets, corresponding to 73.3% of the collection.

Another 26 datasets, or 21.7%, are subject to separate research-only, non-commercial, custom or otherwise restricted conditions. Licence information was not clearly specified for 6 datasets, representing 5.0% of the collection. These cases occur mainly among the video datasets.

The results show that a substantial part of the collection can support educational and research activities. However, datasets within the same broad category do not necessarily provide identical reuse rights. In particular, Creative Commons licences may include attribution, non-commercial, share-alike or no-derivatives conditions. The specific licence terms should therefore be checked before datasets or derived materials are redistributed or incorporated into publicly available teaching resources.

Dataset Size and Format Characteristics

Information on total dataset size was available for 117 of the 120 datasets. It was recorded for 28 video datasets, 29 audio/sound datasets, and all 30 numerical and 30 image datasets. This provides a generally strong basis for assessing download, storage and processing requirements. Nevertheless, dataset sizes should not be compared using identical thresholds across modalities. A dataset considered large within the numerical group may still require substantially less storage than a medium-sized video, audio or image collection. The modality-specific size categories presented in Chapter 3 are therefore more appropriate for detailed assessment, while the

consolidated comparison is limited to the availability of size information and general resource requirements.

Standard and widely supported formats are dominant in all four groups. MP4/MPEG-4 and AVI are commonly represented among video datasets, WAV, MP3 and FLAC among audio and sound datasets, CSV and related tabular formats among numerical datasets, and JPEG and PNG among image datasets. Their prevalence facilitates integration into commonly used data-processing and Machine Learning workflows.

At the same time, the collection includes mixed, specialised and multimodal structures. Examples include ROS bag, TFRecord and CSD/HDF5 in the video group; combinations of different audio formats, MIDI, STEMS and TFRecord files in the audio/sound group; mixed tabular formats and time-series structures among numerical datasets; and raw, proprietary or annotation-specific structures among image datasets. These formats increase the diversity and research value of the collection but may require additional extraction, conversion or specialised processing.

Comparative Assessment of Practical Requirements

Table 4.2 provides a qualitative synthesis of the general patterns identified in Sections 3.1-3.4. The entries were derived from the modality-specific analyses of dataset size, formats, technical characteristics, preprocessing requirements and educational applicability. They do not represent a formal scoring of each individual dataset.

Table 4.2. Comparative overview of common dataset characteristics

Comparative characteristic	Video	Audio/Sound	Numerical	Image
Dataset size information available	28 of 30	29 of 30	30 of 30	30 of 30
Typical dataset size	Highly variable; often medium to high	Highly variable; often medium to high	Generally low	Highly variable; from low to high
Main standard formats represented	MP4/MPEG-4 and AVI	WAV, MP3 and FLAC	CSV, XLSX and TXT	JPEG, PNG and mixed standard image formats
Examples of mixed or specialised formats and structures	ROS bag, TFRecord, CSD/HDF5 and multimodal sensor data	Mixed audio formats, MIDI, STEMS and TFRecord	Mixed tabular formats and time-series structures	Raw, proprietary and annotation-specific structures

Typical preprocessing	Frame extraction and sampling, temporal segmentation, resizing, and frame-rate or format standardisation	Segmentation, resampling, channel standardisation, normalisation and feature extraction	Missing-value treatment, categorical encoding, scaling and time-aware splitting	Resizing, normalisation, colour-space conversion and annotation transformation
Relative computational requirements	Generally high	Generally moderate to high	Generally low to moderate	Generally moderate to high
Typical educational suitability	Mainly intermediate, advanced and research-oriented; subsets may support introductory use	Introductory to advanced, depending on size and technical consistency	Particularly suitable for introductory and intermediate activities	From introductory classification to advanced detection and segmentation
Main practical limitations	Large size, long processing time and multimodal complexity	Variable duration, sampling rates, channels and labelling practices	Missing values, heterogeneous variables and temporal organisation	Large collections, high resolution and complex annotations

Numerical datasets generally provide the most accessible entry point for introductory AI and Machine Learning activities. Their predominantly tabular structure, widespread use of standard formats and comparatively modest storage requirements allow students to focus on data exploration, preprocessing, classification, regression and time-series analysis.

Image and audio/sound datasets support a broad progression from introductory to advanced activities. Smaller and technically consistent collections can be used for basic image or sound classification, while larger datasets, detailed annotations and heterogeneous technical characteristics support more demanding tasks such as object detection, segmentation, speech recognition, music analysis and acoustic-event detection.

Video datasets generally require substantial preprocessing effort and computational resources because of their temporal structure, frame-level processing and, in some cases, multimodal complexity. Audio/sound datasets may also involve considerable storage requirements, particularly when they contain large numbers of long or high-quality recordings. Large image collections can similarly require substantial storage and processing capacity. Where complete

collections are too large or technically demanding, smaller representative subsets may be prepared for classroom use.

Across all modalities, educational suitability is not determined by data type alone. Dataset size, licence conditions, format, technical consistency, documentation, availability of labels or annotations, preprocessing requirements and available computational resources should be considered together. Standard-format and well-documented datasets are generally easier to integrate into introductory and intermediate teaching, while large, multimodal or technically specialised datasets are more appropriate for advanced student projects and research-oriented activities.

4.2 Datasets Suitable for Different Educational Uses

The collected datasets can support different levels and types of educational activities, depending on their scale, technical complexity, format, licence conditions and preprocessing requirements. For the purpose of this assessment, the educational uses were grouped into three broad categories: introductory practical exercises, intermediate student projects and advanced or research-oriented activities.

Smaller datasets in standard formats are generally the most suitable for introductory practical exercises. Numerical datasets in CSV or spreadsheet formats can support activities involving data exploration, visualisation, classification and regression. Smaller image datasets with standardised resolutions and image-level labels are suitable for basic image-classification exercises, while audio datasets containing short recordings in WAV format can be used for introductory sound-classification and speech-recognition tasks. Small video datasets or extracted subsets may also be used, although they generally require more preparation and computational resources.

Intermediate student projects can use datasets that introduce a moderate level of preprocessing, more complex structures, variable technical characteristics or more detailed annotations. Suitable activities include time-series analysis with numerical data, object detection and segmentation with image datasets, sound-event or emotion recognition with audio datasets, and action recognition or tracking with video datasets. These projects allow students to compare different preprocessing approaches and models and to evaluate performance using more realistic data.

Large-scale, multimodal and technically specialised datasets are most appropriate for advanced student projects and research-oriented activities. They can support deep-learning model development, multimodal learning, high-resolution computer vision, large-scale speech or sound analysis, autonomous-driving applications and other computationally demanding tasks. Their use generally requires greater storage and processing capacity, as well as more extensive preprocessing, standardisation or subset extraction.

Table 4.3. Dataset characteristics suitable for different educational uses

Educational use	Suitable dataset characteristics	Typical examples
Introductory practical exercises	Small or manageable size, standard formats, limited preprocessing and clearly defined labels or target variables	Tabular classification and regression, basic image classification, short-audio classification and small video-classification tasks
Intermediate student projects	Moderate size, more complex features or annotations and a meaningful level of data cleaning or standardisation	Time-series analysis, high-dimensional classification, object detection, semantic segmentation, sound-event recognition, emotion recognition and action recognition
Advanced or research-oriented activities	Large-scale, multimodal or specialised data, extensive preprocessing and higher computational requirements	Deep learning, multimodal analysis, autonomous systems, large-scale speech processing, high-resolution vision and scalable data analysis

To provide more concrete guidance for dataset selection, Table 4.4 presents a representative selection from the complete collection. The recommendations are based on the dataset characteristics analysed in Chapter 3, including scale, format, technical consistency, labels or annotations, preprocessing requirements and licence conditions. The table is not intended to represent an exhaustive ranking of all 120 datasets, and some datasets may be suitable for more than one educational level depending on the selected subset, task and available computational resources.

Table 4.4. Representative datasets recommended for different educational uses

Educational level	Dataset	Modality	Suitable AI task	Reason for recommendation	Important limitation or preparation requirement
Introductory practical exercises	**Highway Traffic Videos Dataset**	Video	Traffic-level video classification	Small collection of short videos with three clearly defined classes; supports a complete introductory video-classification	Requires frame extraction and basic video preprocessing; low resolution limits more detailed vision tasks

				workflow	
Introductory practical exercises	**Free Spoken Digit Dataset**	Audio/ Sound	Spoken-digit classification	Small size, short mono WAV recordings and clearly defined digit labels; suitable for waveform analysis, spectrograms and basic classification	Basic feature extraction or spectrogram generation is required
Introductory practical exercises	**Energy Efficiency Dataset**	Numerical	Regression	Small, complete and well-structured dataset with clearly defined input features and two target variables	Feature scaling and the interpretation of two related target variables may be required
Introductory practical exercises	**CIFAR-10**	Image	Image classification	Standardised low-resolution images, ten clearly defined classes and predefined training and test sets	Low image resolution limits applicability to more complex detection or segmentation tasks
Intermediate student projects	**CBVD-5 Cow Behavior Video Dataset**	Video	Cow-behaviour recognition, classification and tracking	Contains 887 annotated Full HD videos and supports realistic computer-vision applications in livestock monitoring; CC0 public-domain licence	Frame sampling, resizing and annotation processing may be required; Full HD recordings increase computational requirements
Intermediate student projects	**UrbanSound8K**	Audio/Sound	Urban sound-event classification	Clearly labelled excerpts from ten classes and sufficient scale for comparing audio features and classification models	Variable sampling rates require resampling and technical standardisation; the CC BY-NC 3.0 licence permits non-commercial use only.
Intermediate student	**Air Quality**	Numerical	Regression, sensor	Introduces missing values, temporal	Missing values coded as −200

projects	**Dataset**		calibration and time-series analysis	ordering and realistic environmental sensor data	must be identified and treated; chronological splitting is recommended
Intermediate student projects	**Oxford-IIIT Pet Dataset**	Image	Fine-grained classification, localisation and segmentation	Supports progression from breed classification to bounding-box localisation and pixel-level segmentation	Variable image sizes and XML/PNG annotations require resizing and annotation processing
Advanced or research-oriented activities	**FieldSAFE Dataset**	Video	Autonomous-system perception and sensor fusion	Combines video and other sensor streams and reflects realistic autonomous-system data-processing challenges	ROS bag structure requires specialised software, sensor synchronisation and substantial preprocessing
Advanced or research-oriented activities	**LibriSpeech**	Audio/Sound	Automatic speech recognition	Large-scale speech corpus with transcripts and approximately 1,000 hours of audio; appropriate for deep speech-processing models	High storage and computational requirements; subset selection may be necessary
Advanced or research-oriented activities	**NYC Taxi Trip Record Dataset**	Numerical	Mobility-demand analysis, travel-time prediction and large-scale temporal modelling	Millions of records and recurring releases support scalable data analysis and research on urban mobility	Requires efficient file handling, timestamp processing, aggregation and usually a selected time period
Advanced or research-oriented activities	**KITTI-360**	Image	Autonomous-driving perception, semantic understanding and 3D scene	Provides complex visual and spatial information suitable for advanced computer-vision	Large scale, multimodal structure and complex annotations require

			analysis	and autonomous-driving research	substantial storage, preprocessing and computing resources; the CC BY-NC-SA 4.0 licence includes non-commercial and share-alike conditions.

The introductory recommendations prioritise manageable size, standard formats and clearly defined tasks. They are therefore suitable for teaching the main stages of a Machine Learning workflow without allowing technical preparation to dominate the activity. For example, the Energy Efficiency Dataset and CIFAR-10 allow students to focus directly on regression and classification, while the Free Spoken Digit Dataset and Highway Traffic Videos Dataset introduce audio and video processing through relatively small collections. The individual analyses in Chapter 3 already identify these characteristics as particularly useful for introductory work.

The intermediate recommendations introduce realistic challenges such as variable sampling rates, missing values, temporal organization, frame or clip processing and more detailed annotations. They remain manageable for student projects but require students to make and justify decisions about preprocessing, model selection and evaluation. CBVD-5, UrbanSound8K, the Air Quality Dataset and the Oxford-IIIT Pet Dataset are especially useful because they introduce realistic requirements related to frame processing, annotation handling, resampling, missing-value treatment and temporal organization without requiring research-scale infrastructure.

The advanced recommendations were selected because they involve large-scale, multimodal or technically specialised data. FieldSAFE and KITTI-360 support autonomous-system and multimodal perception activities, LibriSpeech supports large-scale speech recognition, and the NYC Taxi Trip Record Dataset supports scalable numerical and temporal analysis. In each case, the educational value comes partly from addressing realistic challenges in data organisation, preprocessing, resource management and model development.

These recommendations should be treated as guidance rather than fixed assignments. A large dataset can be adapted to an introductory or intermediate activity by preparing a smaller and clearly documented subset, while a comparatively small dataset can support advanced work when the task involves complex modelling, detailed evaluation or methodological comparison. Before selecting a dataset for a specific educational activity, the intended learning outcomes, available time and computing resources, data documentation and applicable licence conditions

should be reviewed.

4.3 Datasets Selected for Subsequent Project Activities

4.3.1 Selection Process and Criteria

Following the collection and analysis of 120 datasets, each of the four partners responsible for a particular data modality selected three datasets for subsequent project activities. This resulted in a final selection of 12 datasets, comprising three video, three audio/sound, three numerical and three image datasets. The selected datasets or appropriately prepared subsets were uploaded to the project repository for further technical preparation and the development of Python-based educational materials.

The selection considered the educational relevance of each dataset, the availability of labels, annotations or target values, data format and technical compatibility, dataset size and manageability, licence and conditions of use, and the level of preprocessing required. The partners also considered whether the datasets could support clearly defined AI tasks and be adapted to practical exercises, student projects or more advanced project activities.

4.3.2 Overview of the Selected Datasets

Table 4.5 provides a consolidated overview of the 12 datasets selected for subsequent project activities, including three datasets from each data modality: video, audio/sound, numerical and image data. For each dataset, the table presents the responsible partner, dataset name, modality, version or source information, and the applicable licence or conditions of use. The information relates to the datasets made available for further technical preparation and the development of Python-based educational materials.

Table 4.5. Overview of datasets selected for subsequent project activities

Partner	Dataset	Modality	Uploaded version	Licence
UNILJ	Chinese Hand numerals	video	v1	Open Access
UNILJ	Traffic Signs Detection Europe Computer Vision Model	video	v14	CC BY 4.0
UNILJ	Popeye C artoon	video	v1	Open Access
UoM	FSDD - Free Spoken Digit Dataset	Audio/Sound	Complete version downloaded from the official GitHub repository,	Creative Commons

			including WAV recordings, documentation, metadata, and licence information	Attribution -ShareAlike 4.0 International - CC BY-SA 4.0
UoM	Sopele Sounds	Audio/Sound	Complete Sopele music dataset provided by the project partner UNIRI	Creative Commons Attribution 4.0 International - CC BY 4.0
UoM	ESC-50	Audio/Sound	Complete original dataset downloaded from the official ESC-50 GitHub repository (2,000 WAV audio files with metadata)	Creative Commons Attribution - NonCommercial 3.0 (CC BY-NC 3.0)
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	The data was obtained from Federal Hydrometeorological Institute (FBiH)	Institutional data (FHMZ FBiH) - restricted use
UNSA	High-performance concrete mix designs (cement, blast furnace slag, fly ash, water, superplasticizer, coarse & fine aggregate, age) with	numerical	Complete version downloaded from the link: https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength	Free reuse with citation retained

	measured compressive strength. Local working copy (xls/xlsx + source paper).			
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	Complete version downloaded from the link: https://www.kaggle.com/datasets/taweilo/loan-approval-classification-data	Open data (Kaggle - see dataset page)
UNBG	CarLogos and CarIDs	Image	Version 1.0 (Link: https://drive.google.com/drive/folders/1naZoEEmuz1UO0VwztfRhts__ekhCCOia?usp=drive_link). This dataset was collected by the Laboratory for Data Analysis and Applied Artificial Intelligence, University of Belgrade – School of Electrical Engineering.	CC BY 4.0
UNBG	Brand logos and icons dataset	Image	Version 1.0. This dataset available at: https://drive.google.com/drive/folders/16udIS2UjhZPKCNuARiFjl5Vbojm14aZH?usp=drive_link This dataset was collected from Kaggle (Link: https://www.kaggle.com/datasets/programmer3/brand-logos-and-icons-dataset, accessed date March 20th, 2026)	CC0: Public Domain
UNBG	CIFAR-10 Image Classification	image	Version 1.0. This dataset available at: https://www.kaggle.com/competitions/cifar10-mu/data This dataset was collected by the University of Toronto, Department of Computer Science, Canadian Institute for Advanced Research (CIFAR) (accessed date April 8th, 2026)	CC0: Public Domain

Legend:
Partner - partner responsible for selecting and uploading the dataset to the project repository.
Dataset - full name of the selected dataset.
Modality - data type: video, audio/sound, numerical or image.
Uploaded version - indicates whether the full dataset or a prepared subset was uploaded.
Licence - licence and the main conditions governing the use of the dataset.

4.3.3 Technical Characteristics and Required Preprocessing

For each data modality, the responsible partner provided information on the technical characteristics of the three selected datasets, the reasons for their selection and the preprocessing required for their subsequent use. Tables 4.6–4.9 summarise these inputs for the selected video, audio/sound, numerical and image datasets. Since the datasets represent different data modalities, the assessment applies modality-specific parameters, including file format, dataset scale and structure, recording or image characteristics, and the availability of labels, annotations or target values. Across all modalities, the main preparation activities include data organisation and validation, format standardisation, normalisation or scaling, verification of labels and annotations, appropriate division into training, validation and test subsets and, where necessary, feature extraction or data augmentation. The required level of preprocessing ranges from limited preparation for well-structured benchmark datasets to more extensive conversion, segmentation or annotation for technically heterogeneous collections.

Table 4.6. Technical characteristics of the selected video datasets

No	***Dataset***	***Intended AI task***	***Labels/targets***	***File char.***	**Reason for selection**	**Required preprocessing**
1	Chinese Hand numerals	Hand-gesture recognition and numeral classification	Hand-gesture numerals	MOV format, approximately 4 GB	Provides a relatively large video dataset for developing and evaluating models that recognise numerical hand gestures. It is suitable for studying hand localisation, motion patterns, and gesture classification	Convert MOV files into a consistent video format if required; extract or sample frames; resize frames to a standard resolution; normalise pixel values; verify or create gesture labels; split data into training, validation, and test sets; apply

					under different recording conditions	augmentation such as rotation, cropping, brightness adjustment, and horizontal flipping where appropriate
2	Traffic Signs Detection Europe	Traffic-sign object detection and classification	Traffic-sign classes	MP4 format, approximately 277 MB	Supports the development of computer-vision systems for road-scene analysis and automated traffic-sign recognition. The dataset is relevant for intelligent transport systems and includes visually challenging conditions such as varying sign sizes, viewing angles, lighting, and backgrounds	Extract representative frames; resize and normalise images; verify traffic-sign class labels and bounding boxes; standardise annotation format; address class imbalance; apply augmentation such as scaling, contrast adjustment, blur, and perspective transformation
3	Popeye Cartoon	Cartoon-character detection and classification	Character class	MP4 format, approximately 71.9 MB	Enables evaluation of character-recognition methods on animated content, where visual appearance, poses, expressions, and	Extract frames at an appropriate sampling rate; remove repeated or near-identical frames; resize and normalise images; verify or create character labels and

					backgrounds can vary	bounding boxes; segment scenes where necessary; split data by episode or scene to reduce data leakage; apply augmentation such as cropping, scaling, flipping, and brightness or saturation adjustment

Table 4.7. Technical characteristics of the selected audio/sound datasets

No	***Dataset***	*Number of files*	*Audio duration*	*Sampling rate*	**Reason for selection**	**Required preprocessing**
1	FSDD	*3000 clips*	*<1 s*	*8000 Hz*	Small, clearly labelled and easy-to-use dataset, suitable for introductory audio-classification exercises.	Minimal preprocessing: amplitude normalization, padding or truncation to equal length, and train/validation/test splitting.
2	Sopele Sounds	*>80 files*	*21-38 s per file*	*44 100 Hz*	Unique, openly licensed dataset with clear tone labels, suitable for automatic music transcription and mono/polyphonic audio classification	Audio segmentation, amplitude normalization, optional conversion to mono, and extraction of frequency-based features or spectrograms
3	ESC - 50	*2000 files*	5 seconds per file	*44 100 Hz*	Clearly labelled, manageable and widely used benchmark dataset suitable	Audio normalisation and feature extraction, such as spectrogram or Mel-spectrogram

					for environmental sound classification exercises.	generation; optional resampling and silence trimming

Table 4.8. Technical characteristics of the selected numerical datasets

No	***Dataset***	Number of items	Total size	Structure	**Reason for selection**	**Required preprocessing**
1	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.	17,257 records	1.07 MB	tabular, time-series	A locally relevant time-series dataset containing hourly meteorological variables and clearly defined PM10 and PM2.5 target values. It is suitable for regression, air-quality prediction and time-series modelling exercises.	Parse and chronologically order timestamps, identify missing or duplicate records, check measurement units and detect outliers. Depending on the intended model, numerical features may be scaled and lagged or rolling-window variables generated, with chronological training, validation and test splits.
2	High-performance concrete mix designs (cement, blast furnace slag, fly	1,030 records	66 KB (xlsx)	tabular	A compact and well-structured benchmark dataset with clearly defined mixture	Standardise variable names and units, check for missing or duplicate

	ash, water, superplasticizer, coarse & fine aggregate, age) with measured compressive strength. Local working copy (xls/xlsx + source paper).				components and measured compressive strength as a continuous target. It is suitable for teaching regression, feature importance and model comparison.	records and convert the working file into a consistent tabular format where necessary. Numerical scaling and division into training, validation and test sets may be applied depending on the selected algorithm.
3	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.	45,000 records	3.6 MB	tabular	A sufficiently large dataset containing both numerical and categorical applicant attributes and a clearly defined binary loan-approval label. It supports classification, feature-selection, model-interpretability and fairness-related exercises.	Check missing values, duplicates and class balance, encode categorical variables and scale numerical features where required. Potential data leakage and bias should be examined before creating stratified training, validation and test subsets.

Table 4.9. Technical characteristics of the selected image datasets

No	***Dataset***	Number of items	Total size (GB)	Structure	**Reason for selection**	**Required preprocessing**
1	The CarLogos and CarIDs dataset contains vehicle images collected across multiple cities in Serbia under diverse real-world conditions. It includes car logos from more than 40 brands and vehicle registration plates from over 30 cities and municipalities. Images were captured from different angles and distances, during both daytime and nighttime. The dataset is suitable for vehicle and logo recognition, Serbian licence plate recognition and object detection in	700 images	2.02 GB	JPG files (Uncompressed JPEG)	The reason for selecting this specific dataset lies in its high degree of environmental and technical diversity, which is crucial for building resilient computer vision models. Unlike synthetic or highly curated datasets, this collection reflects the unpredictable nature of real-world urban traffic, incorporating a wide variety of camera angles, distances, and fluctuating lighting conditions ranging from bright daylight to low-light night scenarios. By capturing over 40 distinct brand logos and more than 30 regional license plate identifiers, the dataset provides the necessary variability to	Regarding the required preprocessing, raw images from this dataset will likely need a standardized pipeline to maximize model performance. Initial steps should include image resizing to a consistent resolution suitable for the target neural network architecture, along with rigorous normalization of pixel intensity values to mitigate the impact of varying lighting conditions. Furthermore, data augmentation techniques, such as rotation, blurring, and

	complex urban environments.				prevent model overfitting, ensuring that the resulting AI systems can generalize effectively across different vehicle models and geographical markers while maintaining high accuracy in challenging, non-laboratory environments.	slight color adjustment, may be applied to expand the dataset's variance, while manual or semi-automated annotation of bounding boxes for logos and license plates is essential to establish ground truth for object detection tasks. Ensuring that the dataset is balanced across different logo classes and city identifiers through oversampling or strategic data selection will be a vital step in maintaining fairness and predictive stability during the training phase.
2	Brand logos and icons dataset is a curated	1955 images	0.03 GB	1810 PNG images and 145 images	The reason for selecting this dataset typically centers on its	The primary step involves unifying the image

	collection comprising 1,955 total image files, organized into two distinct subfolders designated for brands and icons. The "brands" subfolder features high-resolution 400×400 pixel images showcasing recognizable corporate trademarks such as FlyDubai, Nestle, Abercrombie, Mazda, McDonald's, etc, while the "icons" subfolder contains smaller 120×120 pixel graphics representing popular user interface symbols, emojis, and tech icons like Apple. This structured organization provides a clean,			in JPG formats	dual-resolution structure and specific focus on commercial branding and UI icons, making it ideal for specialized computer vision applications like automatic logo detection in web scraping, or user interface element recognition. Unlike natural photography datasets, these images feature clean backgrounds, sharp edges, and distinct color profiles, which allow researchers and developers to test how well classification models handle graphic art, transparency in PNG files, and varying icon styles. Its manageable size also enables rapid prototyping and efficient model training without the heavy computational overhead required by massive,	dimensions to match the input requirements of the target neural network architecture through resizing or padding. Additional preprocessing must handle the the alpha channel (transparency) present in many of the PNG files, typically by converting RGBA images to RGB format using a solid background colour (such as white or black) to avoid channel mismatch errors during training. Furthermore, standard pixel normalization (scaling to $[0,1]$) and data augmentation techniques, such as slight rotations, scaling, and colour

	categorized repository specifically tailored for tasks involving graphical asset recognition, brand identification, and digital icon classification.				uncurated web image collections.	jittering, are recommended to improve the model's robustness against real-world variations when detecting logos and icons from diverse digital sources.
3	The CIFAR-10 dataset is a widely recognized benchmark collection in the field of computer vision and machine learning, consisting of 60,000 colour images divided equally across 10 distinct classes such as airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Each image is a small, low-resolution thumbnail measuring 32×32 pixels	60 000 images	0.137 GB	PNG images	The reason for selecting CIFAR-10 often stems from its ideal balance between complexity and manageability, making it the definitive proving ground for new neural network architectures and training strategies before scaling up to massive datasets like ImageNet. Because the images are small and low-resolution, experiments can be conducted rapidly on standard hardware, allowing researchers to quickly iterate on hyperparameters,	CIFAR-10 images generally require minimal geometric manipulation since they are already uniformly sized at 32×32 pixels, but they demand careful normalization and augmentation to optimize model training. Standard preprocessing typically involves scaling the pixel values from the integer range $[0,255]$ to a floating-point

	in RGB format, with 5,000 training images and 1,000 test images per class. Developed by researchers at the Canadian Institute for Advanced Research (CIFAR), this dataset has served as a foundational standard for evaluating and comparing the performance of various image classification algorithms, ranging from traditional machine learning models to deep convolutional neural networks.				test regularization techniques, and validate architectural innovations like residual connections or attention mechanisms. Furthermore, the multi-class nature and natural variations in background, pose, and color within the classes provide a robust challenge that helps assess a model's ability to generalize without requiring excessive computational resources.	range of $[0,1]$ or $[-1,1]$, often utilizing the channel-wise mean and standard deviation of the dataset for normalization. To prevent overfitting and enhance model robustness against variations, data augmentation techniques such as random cropping (e.g., padding by 4 pixels and cropping back to 32×32) and horizontal random flipping are routinely applied during the training phase.

4.3.4 Readiness for Subsequent Project Activities

The readiness assessment presented in this section builds on the dataset overview provided in Section 4.3.2 and the technical characteristics and preprocessing requirements summarized in Section 4.3.3. Rather than considering readiness solely in terms of dataset availability, the assessment also takes into account the completeness and organization of the uploaded files, the clarity of labels or target values, compatibility with commonly used Python tools and machine-

learning frameworks, license conditions, and the level of preparation required before the datasets can be used in practical project activities.

UNILJ selected three video datasets that support different computer-vision tasks and provide a suitable basis for subsequent project activities.

The Chinese Hand Numerals dataset is intended for hand-gesture recognition and numeral classification. Its relatively large size makes it suitable for training and evaluating models that must identify hand shapes and movements. Before use, the videos will require frame extraction, standardisation of resolution and format, verification of labels, and division into training, validation, and test subsets.

The Traffic Signs Detection Europe Computer Vision Model dataset is relevant for traffic-sign detection and classification in road scenes. It can support the development and testing of models for intelligent transport and automated visual recognition. The main preparation activities will include verification of class labels and bounding boxes, removal of duplicate or low-quality frames, and conversion of annotations into a consistent format.

The Popeye Cartoon dataset was selected for cartoon-character detection and classification. It enables the evaluation of models on animated content, where visual features differ from real-world images and videos. Required preparation will include frame extraction, removal of repeated frames, verification or creation of character annotations, and organisation of the data into separate training, validation, and test sets.

Overall, the selected datasets are considered suitable for the next project phase. Some preprocessing and annotation validation will be required before model development.

UoM selected three audio and sound datasets that support different audio-processing and classification tasks and provide a suitable basis for subsequent project activities.

The Free Spoken Digit Dataset (FSDD) is intended for spoken-digit recognition and introductory audio-classification exercises. Its relatively small size, clear labels and short recordings make it suitable for demonstrating basic preprocessing and model-training procedures. Before use, the audio clips will require amplitude normalisation, padding or truncation to a consistent length, and division into training, validation and test subsets.

The Sopele Sounds dataset was selected for automatic music transcription and mono- and polyphonic audio classification. It contains clearly labelled recordings of characteristic musical tones and offers a specialised resource for analysing frequency patterns and musical structures. Required preparation will include audio segmentation, amplitude normalisation, optional conversion to mono, and extraction of frequency-based features or spectrograms.

The ESC-50 dataset is relevant for environmental sound classification and is a widely used benchmark for evaluating audio-classification models. Its clearly defined 50 sound classes, consistent recording duration and manageable size make it suitable for practical exercises and student projects. Preprocessing will include audio normalisation and feature extraction, such as spectrogram or Mel-spectrogram generation, with optional resampling and silence trimming where required.

Overall, the selected datasets are considered suitable for the next project phase. Some preprocessing, feature extraction and verification of dataset splits and labels will be required before model development.

UNSA selected three numerical datasets that represent time-series regression, conventional regression and binary classification tasks and provide a suitable basis for subsequent project activities.

The Sarajevo meteorological and air-quality dataset contains hourly observations and PM10 and PM2.5 concentrations as target variables. It is suitable for air-quality prediction and time-series modelling, although timestamp processing, data-quality checks, feature preparation and chronological dataset splitting will be required before model development.

The concrete compressive-strength dataset provides clearly defined numerical input variables and a continuous target and can therefore support regression exercises and model-comparison activities. Only limited preparation is expected, primarily involving format standardisation, data inspection, optional feature scaling and division into training, validation and test subsets.

The loan-approval dataset contains 45,000 records, mixed numerical and categorical attributes and a binary approval label. It is suitable for classification and model-interpretability activities, but requires categorical encoding, missing-value and class-balance assessment, and verification of potential data leakage or bias.

Overall, the selected numerical datasets are considered suitable for the next project phase and are compatible with commonly used Python data-analysis and machine-learning libraries. Their relatively manageable sizes and standard tabular structures allow them to be used after limited to moderate preprocessing, subject to verification of the applicable access and licence conditions.

UNBG selected three image datasets that represent object detection, multi-class image classification and graphical asset recognition tasks, providing a suitable basis for subsequent project activities.

The CarLogos and CarIDs dataset contains vehicle images captured across multiple cities in Serbia, featuring brand emblems spanning different historical periods and license plates from over 30 cities with daytime and nighttime conditions. It is suitable for object detection, vehicle

recognition, and license plate reading, although image resizing, normalization, bounding box annotation for logos and plates, and data augmentation will be required before model development.

The CIFAR-10 dataset provides 60,000 low-resolution color images (32×32 pixels) divided equally across 10 distinct classes such as airplanes, animals, and vehicles. It can therefore support standard image classification exercises and neural network benchmarking activities. Only limited preparation is expected, primarily involving pixel value scaling, dataset splitting into training and test sets, and standard augmentation techniques like random cropping and flipping.

The Brand Logos and Icons dataset contains 1,955 image files divided into two subfolders for corporate brand logos (400×400 pixels) and UI icons and emojis (120×120 pixels). It is suitable for graphical asset recognition and brand classification, but requires resolution unification, handling of RGBA PNG transparency channels, pixel normalization, and verification of class balance.

Overall, the selected image datasets are considered suitable for the next project phase and are compatible with commonly used Python computer vision and deep learning libraries. Their manageable sizes and structured formats allow them to be used after limited to moderate preprocessing, subject to verification of the applicable access and licence conditions.

Taken together, the 12 selected datasets provide a balanced and technically diverse basis for subsequent WP3 activities. They are considered ready for further technical preparation and the development of Python-based educational materials, subject to the completion of modality-specific preprocessing, verification of labels, annotations and data quality, and confirmation of the applicable licence conditions. Their complementary characteristics support a range of AI tasks and educational activities across all four data modalities.

5. Conclusion

This report presented a structured analysis of 120 datasets collected within WP3-A1, comprising 30 video, 30 audio/sound, 30 numerical and 30 image datasets. The analysis examined their main application areas, conditions of use, scale, formats, modality-specific technical characteristics and potential applicability in Artificial Intelligence education and research.

The findings show that the collected datasets cover a broad range of AI and Machine Learning tasks, including classification, regression, object detection, segmentation, speech and sound recognition, time-series analysis, action recognition, autonomous driving and multimodal learning. Standard and widely supported formats are well represented across all four modalities, enabling the use of commonly available data-processing tools, Python libraries and Machine Learning frameworks. However, large-scale, multimodal and technically specialised datasets may require conversion, standardisation, additional preprocessing or the preparation of smaller subsets before practical use.

The comparative assessment confirms that no single data modality is equally suitable for all educational purposes. Numerical datasets generally provide an accessible starting point for introductory data-analysis, classification and regression exercises. Smaller and technically consistent image and audio/sound datasets can support introductory and intermediate activities, while video datasets, high-resolution image collections and more complex or heterogeneous datasets are generally more appropriate for advanced student projects and research-oriented work. Dataset selection should therefore take into account the intended learning outcomes, student level, available time, computational resources and required level of preparation.

Based on the analysis, the WP3 partners selected 12 datasets for subsequent project activities, including three datasets from each modality. The selected datasets support a diverse range of tasks, from spoken-digit, environmental-sound and image classification to time-series prediction, regression, object detection and vehicle, gesture and traffic-sign recognition. Their technical characteristics, required preprocessing and readiness for further use were assessed by the responsible partners. The selected collection provides a balanced basis for further technical preparation and the development of Python-based educational materials, although modality-specific preprocessing, verification of labels and annotations, data-quality checks and appropriate dataset splitting will be required.

Licence and access conditions remain an important consideration. Although many datasets are available under open or clearly defined licences, others are subject to non-commercial, research, institutional or other specific conditions. The applicable terms should therefore be verified

before datasets or derived materials are redistributed, incorporated into teaching resources or included in publicly available project outputs.
The preparation of this report was coordinated by UoM and based on contributions from the partners responsible for the four data modalities: UNILJ for video, UoM for audio/sound, UNBG for image and UNSA for numerical datasets. UNIRI supported the process by reviewing the collected datasets and advising on the structure and formats of the collection.

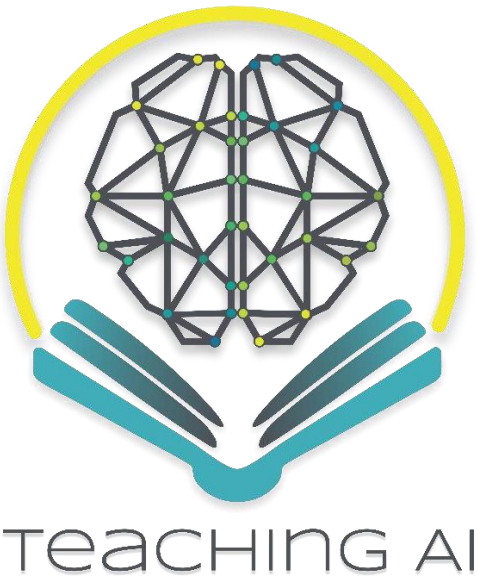
Overall, the analysis confirms that the collected and selected datasets provide a diverse, relevant and technically suitable basis for the next WP3 activities. They support the development of practical exercises, student projects and research-oriented learning activities and provide the necessary foundation for the preparation of Python-based educational materials across different areas of Artificial Intelligence.

6. Annexes

Annex 1: Shared Dataset Collection Document

Annex 1 contains the shared dataset collection document used as the primary data source for the analysis presented in this report. It provides detailed descriptive, legal, structural and technical information on all 120 datasets covering video, audio/sound, numerical and image modalities.

Annex 2: Classification of Datasets by Main Application


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

Annex 1: Shared Dataset Collection Document

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
Commente/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	220847		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	13320	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	2900	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/atalaydenknalbant/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	635000	~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	11732	~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, ■. (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: 2048x933, 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 of bees infested and not infested by Varroa Destructor	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 Bornø 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/fire-smoke-detection	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 CarlDs	image	https://drive.google.com/drive/folders/1naZoEEmuz1UO0VwztlRhts__ekhCCOia?usp=drive_link	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.kaggle.com/datasets/programmer3/brand-logos-and-icons-dataset	/	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.kaggle.com/competitions/cifar10-mu/data	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			

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*						
4	UNBG	ImageNet (ILSVRC)	image	https://www.image-net.org/	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://cocodataset.org/	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.kaggle.com/datasets/hojjatk/mnist-dataset	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.box.com/v/ChestXray-NIHCC	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.com/zalandoresearch/fashion-mnist	Zalando Research	MIT	Xiao, Han, Kashif Rasul, and Roland Vollgraf. "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms." arXiv preprint arXiv:1708.07747 (2017). Link: https://doi.org/10.48550/arXiv.1708.07747	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.cuhk.edu.hk/projects/CelebA.html	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.cvlibs.net/datasets/kitti/	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.ox.ac.uk/pascal/VOC/voc2012/	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.stanford.edu/housenumbers/	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.robots.ox.ac.uk/~vgg/data/pets/	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.edu/~acoates/stl10/	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.csail.mit.edu/	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.kaggle.com/datasets/jessicali9530/lfw-dataset	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.kaggle.com/datasets/imbikramsaha/caltech-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.org/records/7711810#.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	2	Directory structure separated by land type 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.cityscapes-dataset.com/	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	12	Split hierarchy by city, partitioned into leftImg8bit and gtFine	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.princeton.edu/	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	35	Frame-indexed folders holding image directories and MATLAB orientation matrices	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.cuhk.edu.hk/datasets/comp_cars/	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	15	Folders grouped by structural design year or source viewpoints	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.com/spMohanty/PlantVillage-Dataset	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	0.83	Categorized class subfolders	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.cuhk.edu.hk/projects/DeepFashion.html	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	35	Segmented zip archives mapping multi-task attributes via indices	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://shuoyang1213.me/WIDERFACE/	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)	3.4	Target event category directories (e.g., Parade, Meeting) + text label maps	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-inf.mpg.de/departments/computer-vision-and-machine-learning/software-and-datasets/mpii-human-pose-dataset	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)	12.5	Unpacked image directories paired with centralized MATLAB annotations	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.cvlibs.net/datasets/kitti-360/	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	150	Structured log-sequence folders split by camera stream arrays	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.com/fyu/lsun	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	400	Category-specific LMDB binary files	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.com/nightrome/cocostuff	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	25	Flat mask asset files matching base COCO filename index tracks	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.csail.mit.edu/	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	3.8	Folders partitioned into index mappings split by training/validation paths	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_dataset | https://www.kaggle.com/datasets/eduardo4jesus/stanford-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	2	Split train/test folders paired with a central metadata file (car_devkit)	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.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data	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 (.wav)	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.edu/web/mtg/irmas?utm_source=chatgpt.com	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	9579	~11GB	files	WAV audio files (.wav)	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://urbansounddataset.weebly.com/urbansound8k.html	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 (.wav)	Uncompressed	8732	up to 4 s	Variable - 8,000 Hz to 192,000 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.ucg.ac.me/science/vs13/	Slobodan Djukanović, Nikola Bulatović, Ivana Šavor, Faculty of Electrical Engineering, University of Montenegro (@inproceedings{djukanovic2022dataset, author = "Slobodan Djukanović\{c} and Nikola Bulatović\{c} and Ivana \v{S}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. Šavor, "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 specified	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.sciencedirect.com/science/article/pii/S2352340919311953?via%3Dihub	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., Ljubić, 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 tone	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		

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*						
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.org/	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.com/mdeff/fma	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.com/karolpiczak/ESC-50	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.openslr.org/12	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.tensorflow.org/datasets/nsynth	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.robots.ox.ac.uk/~vgg/data/voxceleb/	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.weebly.com/	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.google.com/audioset/ontology/index.html	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://fsannotator.upf.edu/fsd/release/FSD50K/	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 (.wav)		
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.ac.uk/codeapps/the-magnatagatune-dataset	City University London	Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported	Law, L., West, K., Mandel, M., Bay, M. & Downie, J.S., Evaluation of algorithms using games: the case of music annotation. In Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR), 2009	Benchmarking tool in Music Information Retrieval (MIR) and Audio Machine Learning	25863	3 GB	file	MP3 audio file	Compressed	25863	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 Technology Research (CSTR), University of Edinburgh	CC BY 4.0	Yamagishi, J., Veaux, C., and MacDonald, K. (2019). CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit (version 0.92). University of Edinburgh, Centre for Speech Technology Research. https://doi.org/10.7488/ds/2645	Multi-speaker text-to-speech synthesis, speaker adaptation, speaker identification, accent recognition and speech recognition	44,455 utterances - mic1 configuration	10.94 GB compressed download	Audio files organised by speaker, with corresponding text transcriptions and speaker metadata	WAV audio files and TXT transcriptions	ZIP archive; uncompressed PCM WAV audio	44,455 - mic1 configuration	Approximately 44 hours	48,000 Hz	mono	16 bit	WAV	Text transcriptions, speaker identity, accent and gender labels	
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., & Serra, X., The MTG-Jamendo Dataset for Automatic Music Tagging. Machine Learning for Music Discovery Workshop, ICML, 2009	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.tensorflow.org/datasets/catalog/speech_commands	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/CheyneyComputerScience/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.google.com/audioset/	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.com/LJ-Speech-Dataset/	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.com/Jakobovski/free-spoken-digit-dataset	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.tensorflow.org/datasets/maestro	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.github.io/datasets/musdb.html	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://bhichallenge.med.auth.gr/ICBHI_2017_Challenge	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	Project partner	Dataset description	Modality	Link	Source / Owner	License	Reference	Applications	Number of items	Total size (GB)	Structure	Data format	Compression	No. of rows (tabular)	No. of features	Feature types	Missing data (%)	Time-series	Sampling frequency (tabular)	Units
Comment/example			video, image, audio, numerical, multimodal								files, sequences, tabular, time-series, ...	mp3, mp4, ...								
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)	

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*						
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 hidrometeorolosko zavod FBiH (FHMZ)	Institutional data (FHMZ FBiH) - restricted use	Federalni hidrometeorolosko zavod FBiH. Meteorološki 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)	

Annex 2: Classification of Datasets by Main Application

Table A2.1. Classification of video datasets by main application

Main application category	Datasets included	Number of datasets
Human action and behaviour recognition	WLASL - Word-Level American Sign Language; Something-Something V2; UCF101; Kinetics-700; FallVision; Workout/Exercise Recognition Dataset; KTH Human Action Dataset; Word-Level Arabic Sign Language Dataset	8
Agriculture and animal monitoring	Durum Wheat Kernels and Impurities Dataset; UAV Videos for Mango Yield Estimation; EV2 Bee Varroa Dataset; TeaWeed-Action; CBVD-5 Cow Behavior Video Dataset; Raw Oil Palm Fruit Video Dataset	6
Traffic, surveillance and safety	PEViD-UHD; Highway Traffic Video Dataset; Annotated Traffic Lights Video Dataset; Fire and Smoke Detection Video Database; HWID12 Highway Incidents Detection Dataset	5
Object detection and tracking	UAV Vineyard Botrytis Dataset; GrapeMOTS; Underwater Tracking Benchmark Dataset	3
Autonomous driving and navigation	nuScenes; FieldSAFE; UAV Monocular RGB Video Dataset for SLAM, SfM and 3D Reconstruction	3
Multimodal understanding and affect analysis	Qualcomm Interactive Video Dataset - QIVD; Ego4D; CMU-MOSEI	3
General video classification and recognition	Car Damage Recognition Video Dataset; YouTube-8M Segments Dataset	2

Note: Each dataset was assigned to one main application category according to its primary intended use. Although some datasets support several AI tasks, only the dominant application was considered in order to avoid multiple counting.

Table A2.2. Classification of audio and sound datasets by main application

Main application category	Datasets included	Number of datasets
Music analysis and music information retrieval	GTZAN; IRMAS; Sopele Sounds; FMA - Free Music Archive; NSynth; MedleyDB; MagnaTagATune; MTG-Jamendo; MAESTRO; MUSDB18	10
Environmental sound and general audio event recognition	UrbanSound8K; ESC-50; AudioSet Ontology; FSD50K; MUSAN; TAU Urban Acoustic Scenes 2020 Mobile; AudioSet	7
Speech recognition, speaker identification and speech synthesis	LibriSpeech; VoxCeleb - Speaker Recognition and Verification; CSTR VCTK Corpus; Google Speech Commands; LJ Speech Dataset; Free Spoken Digit Dataset - FSDD	6
Bioacoustics and animal sound recognition	Xeno-canto; BirdVox-DCASE-20k	2
Speech emotion recognition	RAVDESS; CREMA-D	2
Biomedical sound analysis	Heartbeat Sounds; ICBHI 2017 Respiratory Sound Database	2
Vehicle speed estimation	VS13 Audio-Video Vehicle Speed Dataset	1

Note: Each dataset was assigned to one main application category according to its primary intended use. Although some datasets support several audio-related AI tasks, only the dominant application was considered in order to avoid multiple counting.

Table A2.3. Classification of numerical datasets by main application

Main application category	Datasets included	Number of datasets
Regression and numerical prediction	California Housing Dataset; Energy Efficiency Dataset; Air Quality Dataset; Combined Cycle Power Plant Dataset; Diamonds Dataset; Auto MPG Dataset; Airbnb Listings Dataset; FIFA Players Dataset; Concrete Compressive Strength Dataset; Housing Price Dataset	10
Classification and decision support	Cleveland Heart Disease Dataset; Bank Marketing Dataset; Steel Plates Faults Dataset; Human Activity Recognition Using Smartphones Dataset; Telco Customer Churn Dataset; Kepler Objects of Interest Dataset; Wine Reviews Dataset; Lending Club Loan Dataset; Loan Approval Classification Dataset	9
Time-series forecasting and monitoring	Beijing PM2.5 Dataset; NYC Yellow Taxi Trip Records; COVID-19 Dataset; Global Terrorism Database; Electricity Transformer Temperature Dataset; PJM Hourly Energy Consumption Dataset; Sarajevo Air Quality Dataset	7
Socioeconomic, macroeconomic and panel-data analysis	World Development Indicators; CO_2 and Greenhouse Gas Emissions Dataset; Eurostat National Accounts Dataset	3
Recommendation and combined predictive analysis	Spotify Tracks Dataset	1

Note: Each dataset was assigned to one main application category according to its primary intended use. Although some datasets support several analytical and Machine Learning tasks, only the dominant application was considered in order to avoid multiple counting.

Table A2.4. Classification of image datasets by main application

Main application category	Datasets included	Number of datasets
Image classification and fine-grained recognition	Brand Logos and Icons Dataset; CIFAR-10; ImageNet - ILSVRC; MNIST; Fashion-MNIST; SVHN - Street View House Numbers; STL-10; Places365; Caltech 101; CompCars; Stanford Cars Dataset	11
Object detection, localisation and pose estimation	CarLogos and CarIDs; COCO - Common Objects in Context; KITTI; PASCAL VOC 2012; SUN RGB-D; WIDER FACE; MPII Human Pose Dataset	7
Semantic segmentation and scene understanding	Oxford-IIIT Pet Dataset; Cityscapes; KITTI-360; COCO-Stuff; ADE20K	5
Face analysis and generative modelling	CelebA - CelebFaces Attributes; LFW - Labeled Faces in the Wild	2
Remote sensing and agricultural analysis	EuroSAT; PlantVillage Dataset	2
Fashion retrieval and image generation	DeepFashion; LSUN - Large-Scale Scene Understanding	2
Biomedical image analysis	ChestX-ray14	1

Note: Each dataset was assigned to one main application category according to its primary intended use. Although some datasets support several computer-vision tasks, only the dominant application was considered in order to avoid multiple counting.