German Traffic Sign Recognition Benchmark Dataset (GTSRB)

Dataset:https://sid.erda.dk/public/archives/daaeac0d7ce1152aea9b61d9f1e19370/published-archive.html
Data Card Authors: Dhivya S

The **German Traffic Sign Recognition Benchmark** (**GTSRB**) contains 43 classes of traffic signs, split into 39,209 training images and 12,630 test images. The images have varying light conditions and rich backgrounds.

Authorship

Publishers

PUBLISHING ORGANIZATION	INDUSTRY SECTOR	PUBLISHER CONTACT
Write the names of the institution or organization responsible for publishing the dataset.	*Bold to select all applicable.* ✋ *Do not delete any unselected choices.*	*Provide publisher contact details. For dataset owners, see next row.*
Organization Name **Institut für Neuroinformatik**	Corporate **Academic** Not-for-profit Individual Others (please Specify)	• Johannes Stallkamp, Institut f¨ur Neuroinformatik Ruhr-Universit¨at Bochum 44780 Bochum, Germany, johannes.stallkamp@ini.rub.de • Marc Schlipsing, Institut f¨ur Neuroinformatik Ruhr-Universit¨at Bochum 44780 Bochum, Germany , marc.schlipsing@ini.rub.de • Jan Salmen, Institut f¨ur Neuroinformatik Ruhr-Universit¨at Bochum 44780 Bochum, Germany , jan.salmen@ini.rub.de • Christian Igel , Department of Computer Science University of Copenhagen 2100 Copenhagen, Denmark, igel@di.ku.dk Group Email: tsr-benchmark@ini.rub.de Website:https://www.ini.rub.de/index.html.en

Dataset Owners

DATASET TEAM(S)	DATASET CONTACT	DATASET AUTHORS
Write the names of the groups or team(s) that own the dataset.	*How can dataset owners be contacted for questions about the model? See previous row for publishing institution.*	*Write the names of all authors associated with the dataset. Provide the affiliation and year if different from publishing institutions or multiple affiliations:*
Real-time Computer Vision Research group, Ruhr-Universit¨at Bochum, Institut für Neuroinformatik	• Group Email: tsr-benchmark@ini.rub.de • Website:https://benchmark.ini.rub.de/gtsrb_news.html	• Johannes Stallkamp, PhD, Institut f¨ur Neuroinformatik Ruhr-Universit¨at Bochum 44780 Bochum, Germany,2011 • Marc Schlipsing, PhD, Institut f¨ur Neuroinformatik Ruhr-Universit¨at Bochum 44780 Bochum, Germany ,2011 • Jan Salmen, PhD, Institut f¨ur

		Neuroinformatik Ruhr-Universit¨at Bochum 44780 Bochum, Germany ,2011 • Christian Igel , PhD, Department of Computer Science University of Copenhagen 2100 Copenhagen, Denmark ,2011

Funding Sources

FUNDING INSTITUTION(S)	FUNDING DETAILS
Write the names of the funding institutions.	*Provide a short summary of funding sources and other support, including details such as programs or projects that may have funded the creation, collection, or curation of the dataset. Include links to relevant documents where applicable.*
• Federal Ministry of Education and Research • NISYS GmbH	• Federal Ministry of Education and Research sponsored this dataset for the the Real-Time Computer Vision group which published **The German Traffic Sign Recognition Benchmark** at the International Joint Conference on Neural Networks (IJCNN) in 2011. • NISYS GmbH supplyed the data collection and annotation software for the dataset

Dataset Overview

DATASET SUBJECT	DATASET SNAPSHOT	DESCRIPTION OF CONTENT
Bold to select all applicable. *Do not delete any unselected choices.*	*Fill out details as indicated, adding rows as needed. Include links to additional table(s) with more detailed breakdowns in the caption.*	*Provide a short summary of the dataset content. Include links where applicable.*
Sensitive Data about people Non-Sensitive Data about people Data about natural phenomena **Data about places and objects** Synthetically generated data Data about systems or products and their behaviors Unknown Others* (*please specify)	Size of dataset 347.9 MB Number of Instances 50000+ Number of Fields 8 Labeled Classes 40+ Number of Labels 40+	• The images contain one traffic sign each • Images contain a border of 10 % around the actual traffic sign (at least 5 pixels) to allow for edge-based approaches • The bounding box of the traffic sign is part of the annotations .

DESCRIPTIVE STATISTICS

Add basic statistics for each field here, as relevant. If there is insufficient space, focus on the most important or critical fields for this dataset. E.g., some statistics will be relevant for numeric data, but not for strings.

Statistic	width	Height	ROI.x1	ROI.y1	ROI.x2	ROI.y2	ClassID	Path
Count	39.2k	39.2k	39.2k	39.2k	39.2k	39.2k	39.2k	39.2k
Mean	50.8	50.3	6	5.96	45.2	44.7	15.8	NA
Std	24.3	23.1	1.48	1.39	23.1	22	12	NA
Min	25	25	0	5	20	20	0	NA
25%	35	35	5	5	29	30	5	NA
50%	43	43	6	6	38	38	12	NA
75%	58	58	6	6	53	52	25	NA
Max	243	225	20	20	223	205	42	NA

Descriptive statistics for train.csv

Statistic	width	Height	ROI.x1	ROI.y1	ROI.x2	ROI.y2	ClassID	Path
Count	12.6k	12.6	12.6k	12.6k	12.6k	12.6k	12.6k	12.6k
Mean	50.5	50.4	6	5.98	44.9	44.8	15.6	NA
Std	25.1	23.7	1.54	1.43	23.8	22.5	11.9	NA
Min	25	25	1	5	20	20	0	NA
25%	34	35	5	5	29	29	5	NA
50%	43	43	6	6	38	38	12	NA
75%	58	57	6	6	53	52	25	NA
Max	266	232	23	19	244	212	42	NA

Descriptive statistics for test.csv

Statistic	ColorID	ShapeID	ClassID	Path
Count	43	43	43	43
Mean	0.51	0.79	21	NA
Std	0.92	0.79	12.4	NA
Min	0	0	0	NA
25%	0	0	10	NA
50%	0	1	21	NA
75%	1	1	32	NA
max	3	4	42	NA

Descriptive statistics for meta.csv

Sensitivity of Data

SENSITIVE DATA	FIELDS WITH SENSITIVE DATA	SECURITY AND PRIVACY HANDLING
Bold to select all applicable. *Do not delete any unselected choices.*	*Please indicate which features or fields might contain sensitive or personally identifiable information, and if or not collection was intentional using the format below:*	*Provide a short summary of measures or steps to handle sensitive data in this dataset. Include links and metrics where applicable.*
User Content User Metadata User Activity Data **Identifiable Data** Sensitive Data Business Data Employee Data Pseudonymous Data Anonymous Data Health Data Children's Data None Others* (*please specify)	Intentionally Collected Sensitive Data None Unintentionally Collected Sensitive Data None	NA
	RELEVANT LINKS	RISKS AND MITIGATIONS
	Provide link(s) to documents that describe any S/PII where available:	*Provide a short summary of how risks from PII or sensitive information have been mitigated in the dataset. Include links and metrics where applicable.*
	NA	NA

Dataset Version and Maintenance

VERSION STATUS	DATASET VERSION	MAINTENANCE PLAN
Bold to select ONE. *Do not delete any unselected choices.*	*Provide details about this version of the dataset.*	*Provide a short summary of how the dataset is maintained, including information about refreshes, versioning criteria, errors, feedback and/or recourse. Include links and metrics where applicable.*
Regularly Updated New versions of the dataset have been or will continue to be made available. Actively Maintained **No new versions will be made available, but this dataset will be**	Current Version 1.0 Last Updated 05/2019 Release Date 07/2011	• The datasets, software packages, and results still available for download but the dataset will not be updated or maintained.

actively maintained, including but not limited to updates to the data. Limited Maintenance The data will not be updated, but any technical issues will be addressed. Deprecated This dataset is obsolete or is no longer being maintained.		
	NEXT PLANNED UPDATE	EXPECTED UPDATES OR CRITERIA
⚠ *Fill this if this dataset is* *(a) Regularly updated* *(b) Actively maintained and another version is planned*	*Provide details about the next planned update.*	*Provide a short summary for readers to understand updates to the dataset and/or data. Include links, charts, and visualizations as appropriate.*
	Version affected NA Next data update NA Next Version NA Next Version update NA	NA

Motivations & Use

Motivations

Bold to select ONE (primary modality). ✋ *Do not delete any unselected choices.*	*Link to multiple data points or exploratory demos. If access is restricted, consider adding a fake example that provides a realistic description of data points in the dataset.*	*Provide a list of fields in data points, including a description and notes on how to interpret fields in an example of data in this dataset.*
Image Data Text Data Tabular Data Audio Data Video Data Time Series Graph Data Geospatial Data Multimodal (Please specify) Others (please specify) Unknown	• https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign?select=Train.csv	• Width: Width of image • Height: Height of image • Roi.X1: Upper left X coordinate of sign on image • Roi.Y1: Upper left Y coordinate of sign on image • Roi.X2: Lower right X coordinate of sign on image • Roi.Y2: Lower right Y coordinate of sign on image • ClassId: Class of provided image • Path: Path to provided image
	EXAMPLE: TYPICAL DATA POINT	EXAMPLE: OUTLIER DATA POINT
	.	
	E.g. of Data Point: **27,26,5,5,22,20,Train/20/00020_00000_00000.png**	E.g. of Data Point: **20**

DATASET PURPOSE(S)	KEY DOMAINS AND APPLICATION(S)	PRIMARY MOTIVATION(S)
Bold to select ONE. *Do not delete any unselected choices.*	*Use comma-separated tags to indicate the key domains for this dataset.*	*List the primary motivations for creating or curating this dataset:*
Monitoring **Research** Production Others (please specify)	Domains **Machine Learning, Object Recognition, Computer Vision** Problem Space Gender accuracy in text translations that describe occupations.	E.g. • Recognition of traffic signs is a challenging real-world problem of high industrial relevance. • Traffic sign recognition is a multi-class classification problem with unbalanced class frequencies. • Although commercial systems have reached the market and several studies on this topic have been published, systematic unbiased comparisons of different approaches are missing and comprehensive benchmark datasets are not freely available.

Intended Use

DATASET USAGE	INTENDED AND/OR SUITABLE USE CASE(S)	UNSUITABLE USE CASE(S)
Bold to select ONE. *Do not delete any unselected choices.*	*Summarize the intended and known use cases of this dataset:*	*Summarize any known problematic use cases of this dataset:*
Safe for production use **Safe for research use** Conditional use- some unsafe applications Only approved use Others (please specify)	• Research applications in computer vision • Traffic sign detection in self-driving cars	• Traffic sign detection for signs apart from the listed 43 classes
	PROBLEM SPACE AND RESEARCH QUESTIONS(S)	PUBLICATION GUIDELINES
	Describe the specific problem space that this dataset intends to address. Include any specific research questions.	*Include any guidelines and steps for citing this dataset in research and/or production work.*

	Assessing the performance of state-of-the-art machine learning algorithms on a publicly available traffic sign dataset	To reference this dataset in your paper, please follow the following guidelines: • Reference any prior publications that have referenced the dataset,

Access, Retention, & Wipeout

Access

ACCESS TYPE	DOCUMENTATION LINKS	ACCESS PREREQUISITES
Bold to select ONE. *Do not delete any unselected choices.*	*Provide links that describe documentation to access this dataset:*	*Please describe any required training or prerequisites to access: this dataset.*
Unrestricted Conditional **Open Access** Others (please specify)	• Website:https://benchmark.ini.rub.de/gtsrb_news.html • Colab Example: http://benchmark.ini.rub.de/Dataset/GTSRB_Python_code.zip	No prerequisites
	DIRECT LINKS TO DATASET	ACCESS POLICY
	Provide links to access this dataset:	*Summarize the access policy associated with this dataset. Use this space to include any other information or links that might be relevant to accessing the dataset.*
	• Direct download link: https://sid.erda.dk/public/archives/daaeac0d7ce1152aea9b61d9f1e19370/published-archive.html • Other repository: https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign/download Code to download data `# …`	The dataset is available for public access.

Retention

	RETENTION DURATION	RETENTION POLICY
	Specify the duration for which this dataset can be retained:	*Summarize the retention policy for this dataset. Use this space to include any other information or links that might be relevant.*
	No specified time	No specified policy
	RETENTION STEPS	EXCEPTIONS AND EXEMPTIONS
	Summarize any additional requirements and related steps to retain the dataset.	*Summarize any additional exceptions and related steps to retain the dataset:*

	-	-

Wipeout and Deletion

	WIPEOUT DURATION	DELETION EVENT
	Specify the duration after which this dataset should be deleted or wiped out:	*Summarize the sequence of events and allowable processing for data deletion:*
	No specified time	-
	ACCEPTABLE MEANS OF DELETION	POST-DELETION OBLIGATIONS
	List the acceptable means of deletion:	*Summarize the sequence of obligations after a deletion event:*
	-	-
	OPERATIONAL REQUIREMENTS	EXCEPTIONS AND EXEMPTIONS
	List any wipeout integration operational requirements:	*Summarize any additional exceptions and related steps to a deletion event:*
	-	-

Dataset Provenance

Data Collection & Sources

DATA COLLECTION METHODS	DATA SOURCES	DESCRIPTION OF DATA SOURCE(S)
Bold to select all applicable. ✋ *Do not delete any unselected choices.*	*Describe the source for each collection method. Add rows as meaningful. Refer to guidance on Duplicate for each collection method as necessary.*	*Provide a brief description of each Data Source by type. Include appropriate breakdowns if data sources contain data from other sources. Include links to more information, metrics, visualizations, etc.*
API Artificially Generated **Crowdsourced - Paid** Crowdsourced - Volunteer Vendor Collection Efforts Scraped or Crawled Survey, forms or polls Taken from other existing datasets Unknown	Crowdsourced - The dataset was created from approx. 10 h of video that were recorded while driving on different road types in Germany during daytime. A Prosilica GC 1380CH camera was used with automatic exposure control and a frame rate of 25 fps. The camera images, from which the traffic sign images are extracted, have a resolution of 1360 × 1024 pixels. The video sequences are stored in a raw Bayer-pattern format (Bayer, 1975). Data collection, annotation and image extraction was performed using the NISYS Advanced Development and Analysis Framework	Videos: 10 h of video that were recorded while driving on different road types in Germany during daytime.

To be determined Others (please specify)	(ADAF), an easily extensible, module-based software system Date of Collection: March 2010- October 2010	
DATASET TYPE	COLLECTED DATA	DATA PROCESSING
Bold to select all applicable. *Do not delete any unselected choices.*	*List or describe any fields or data that were collected for this dataset, and indicate if they were included in the dataset or excluded from the dataset. Include links, descriptive statistics, and visualizations where relevant.* *Duplicate for each collection method as necessary.*	*If multiple methods were used to collect data, how was the data aggregated, processed, or connected? Include relevant descriptions, statistics, metrics or visualizations, links and libraries in your response. Break down by source type.*
Static **Data was collected once from single or multiple sources.** Streamed Data is continuously acquired from single or multiple sources. Dynamic Data is updated regularly from single or multiple sources. Others* (*please specify)	Crowdsourcing videos Collected and included • Color images • HOG features • HAAR like features • Color histograms Collected and excluded • Discard tracks with less than 30 images. • Discard classes with less than 9 tracks. 3. • For the remaining tracks: If the track contains more than 30 images, equidistantly sample 30 images.	.Data processing, annotation and image extraction was performed using the NISYS Advanced Development and Analysis Framework (ADAF), an easily extensible, module-based software system.

Criteria

SELECTION CRITERIA	INCLUSION CRITERIA	EXCLUSION CRITERIA
Please describe the data selection criteria. Break down by method as applicable. Include links, descriptive statistics, and visualizations where relevant.	*Please describe the data inclusion criteria. Break down by method as applicable. Include links, descriptive statistics, and visualizations where relevant..*	*Please describe the data exclusion criteria. Break down by method as applicable. Include links, descriptive statistics, and visualizations where relevant.*
Crowdsourced - videos Only images containing traffic signs describing known categories were used.	Crowdsourced- videos 50k randomly sampled images containing identifiable traffic sign associated with it's class.	Crowdsourced- videos Low resolution: Traffic signs at high distance result in low resolution while closer ones are prone to motion blur. The illumination may change, and the motion of the car affects the perspective with respect to occlusions and background Images taken at low velocity: The car passes different traffic sign instances with different velocities, depending on sign position and the overall traffic situation. In the recording, this leads to different numbers of traffic sign images per track (approximately 5–250 images per track). Consecutive images of a traffic sign that was passed with low velocity are very similar to each other. They do not contribute to the diversity of the dataset.

Relationship to Source

USE	BENEFITS AND VALUE	LIMITATIONS AND TRADE-OFFS
If at all, how is the resulting dataset aligned with the purposes, motivations, or intended use of the upstream source(s)? Break down by source type.	*What are the benefits of the resulting dataset to its consumers, compared to the upstream source(s)? Break down by source type.*	*What are the limitations of the resulting dataset to its consumers, compared to the upstream source(s)? Break down by source type.*
Crowdsourced - videos Traffic sign recognition is a multi-category classification problem with unbalanced class frequencies. It is a challenging realworld computer vision problem of high practical relevance, which has been a research topic for several decades.	Crowdsourced - videos The model can be used in technologies for driver assistance systems, which represent an important and challenging field of application. These intelligent systems analyze the vehicle´s environment via different types of sensors, for instance video and radar, thus, increasing safety and comfort for the driver. Can also be used for research applications in computer vision.	Crowdsourced - videos Dataset cannot be used to predict any other class of traffic sign other the trained 43 classes

Updates to Dataset

⚠ *Fill this next row if: this is not the first version of the dataset, and there is no data card available for the first version.*

	FIRST VERSION	NOTES ON FIRST VERSION
	Provide a basic description of the first version of this dataset.	*Optional.* *Provide a short summary describing caveats or nuances of the first version of this dataset. Include links, charts, and visualizations as appropriate.*
Not applicable	Release date NA Link to dataset NA Status NA [Actively Maintained/Limited Maintenance/ Deprecated] Size of Dataset NA Number of Instances NA	
DATASET UPDATE FREQUENCY	DATASET UPDATE SCHEDULE	CHANGES ON UPDATE
Bold to select ONE ✋ *Do not delete any unselected choices.*	*Please describe the update schedule*	*What happens when the dataset is refreshed? Break down by sources as necessary. Include any applicable policies and changes to the dataset that occur during a refresh.*

Yearly Quarterly Monthly Biweekly Weekly Daily Hourly Static **Not updated**	Date of last update NA DD/MM/YYYY NA Frequency of Updates NA[Yearly / Quarterly / Monthly/ Biweekly / Weekly / Daily / Hourly / Static / Others (please specify) Data points affected NA Data points updated NA Data points added NA Data points removed NA Date of next update NA	NA

Human and Other Sensitive Attributes

SENSITIVE HUMAN ATTRIBUTES	INTENTIONALITY OF COLLECTIONS	RATIONALE FOR COLLECTING HUMAN ATTRIBUTES
Bold to select ALL ATTRIBUTES that are present in the dataset. *Do not delete any unselected choices.*	*For each human attribute indicated, specify if this information was collected intentionally or unintentionally:*	*Briefly describe the motivation, rationale, considerations or approaches that caused this dataset to include the indicated human attributes. Summarize why or how this might affect the use of the dataset.*
Race Gender Ethnicity Socio-economic status Geography Language Sexual Orientation Religion Age Culture Disability Experience or Seniority **No sensitive attributes** Others (please specify)	Intentionally Collected Attributes (human attributes that were labeled or collected as a part of the dataset creation process) NA Unintentionally Collected Attributes (human attributes that were not explicitly collected as a part of the dataset creation process but can be inferred using additional methods) NA	NA
	SOURCE(S) OF HUMAN ATTRIBUTES	COLLECTION METHODS
	Indicate the source of the sensitive attributes using the format provided.	*Describe the methods used to collect human attributes in the dataset. Break down by human attribute as necessary. Include information related to the tasks, platforms, visualizations, links to additional documentation as applicable.*
	NA	NA

DISTRIBUTION OF HUMAN ATTRIBUTES

Duplicate and populate the following row for each human attribute previously selected. Include the key takeaways in the caption.

NA

[Human attribute]	Label or Class	Label or Class	Label or Class	Label or Class

	KNOWN CORRELATIONS	RISK, TRADE-OFFS AND CAVEATS
	List or describe any known correlations with the indicated sensitive attributes in this dataset. Summarize why or how this might affect the use of the dataset. Include visualizations, metrics, or links where necessary.	*Provide a statement, list or summarize any expectations, systemic or residual risks, trade-offs and caveats due to human attributes in this dataset. Break down by human attribute if necessary.*
	Not Applicable	Not Applicable

Extended Use

Use with Other Data

SAFETY OF USE WITH OTHER DATA	KNOWN SAFE DATASETS OR DATA TYPES	BEST PRACTICES FOR JOINING OR AGGREGATING WITH DATASET
Bold to select ONE. *Do not delete any unselected choices.*	*Which known datasets or data can this dataset be safely joined or aggregated with? Describe any relevant transformation types.*	*Summarize best practices for using this dataset in conjunction with other datasets or data type. Links to demonstrative examples where available.*
Safe to use with other data Conditionally safe to use with other data Should not be used with other data Unknown Others* (Please specify)	German Traffic Sign *Detection* Benchmark (GTSDB).	Not available
	KNOWN UNSAFE DATASETS OR DATA TYPES	KNOWN LIMITATIONS AND RECOMMENDATIONS
Fill out this row if you selected "Conditionally safe to use with other datasets" or "Should not be used with other datasets":	*Which known datasets or data should this dataset not be joined or aggregated with? List and describe any relevant transformation types.*	*Describe limitations of the dataset that might introduce foreseeable risks to intended use when the dataset is conjoined with other datasets. Include any suggested recommendations.*
	Not applicable	Not applicable

Forking & Sampling

SAFETY OF FORKING / SAMPLING	ACCEPTABLE SAMPLING METHODS	BEST PRACTICES FOR FORKING AND SAMPLING
Bold to select ONE. *Do not delete any unselected choices.*	*Bold to select all applicable.* *Do not delete any unselected choices*	*Summarize best practices for forking or sampling this dataset. Links to demonstrative examples where available.*
Safe to fork and/or sample Conditionally safe to fork and/or sample Should not be forked and/or sampled Unknown Others* (*Please specify)	Cluster Sampling Haphazard Sampling Multi-stage Sampling **Random Sampling** Retrospective Sampling **Stratified Sampling** **Systematic Sampling** Weighted Sampling Unknown Unsampled Others* (*Please Specify)	• Sampling should be done by taking into account class and track membership. This makes sure that the overall class distribution is preserved for each individual set and that • all images of one traffic sign instance are assigned to the same set, as otherwise the datasets could not be considered stochastically independent.
	KNOWN RISKS TO SAMPLING	KNOWN LIMITATIONS AND RECOMMENDATIONS

⚠ *Fill out this row if you selected "Conditionally safe to fork and/or sample" or "Should not be forked and/or sample".*	*What known or residual risks are associated with forking and sampling methods when applied to the dataset? List and describe.*	*Describe limitations of the dataset that might introduce foreseeable risks to intended use when the dataset is forked or sampled. Include any suggested recommendations.*
Not Applicable	Not Applicable	Not Applicable

Use in Machine Learning or AI Systems

DATASET USE(S)	DATASET SPLITS	USAGE GUIDELINES OR POLICIES
Bold to select all applicable. ✋ *Do not delete any unselected choices.*	*Describe and name the splits in the dataset (if more than one), and include any criteria for splitting the data.*	*Describe any usage guidelines or policies that users of the dataset should be aware of. Summarize documents and link to them as relevant.*
Training **Testing** Validation Dev Others* (* Please Specify)	Train 39.2k Test 12.6k	None
	FEATURE DISTRIBUTIONS	KNOWN CORRELATIONS
	Describe any notable feature distributions in the dataset. Include links to servers where readers can explore the data on their own.	*List or describe any known correlations with the indicated features in this dataset. Summarize why or how this might affect the use of the dataset. Include links where necessary.*
	HOG features 6052 Haar like features 11584 Hue histograms 256	No known correlations
	SPLIT STATISTICS	
	Provide the sizes of each split. As appropriate, provide any descriptive statistics for features.	

Statistic	Train	Test	Valid
Count	50%	25%	25%

Dataset split

Dataset Transformations

⚠ *Fill this section if any transformations were applied in the creation of your dataset.*

TRANSFORMATIONS APPLIED	FIELDS TRANSFORMED	LIBRARIES AND METHODS USED
Bold to select all applicable *Do not delete any unselected choices.*	*What were the data types that fields were transformed to? Break down by transformations applied*	*List any relevant libraries used to process the data, as applicable.*
Anomaly Detection Cleaning Mismatched Values Cleaning Missing Values Converting Data Types Data Aggregation Dimensionality Reduction Joining Input Sources Redaction or Anonymization **No transformations** Others* (*Please specify)	Not applicable	Not applicable

Breakdown of Transformations

Fill out relevant rows.

CLEANING MISSING VALUES	METHODS USED	COMPARATIVE SUMMARY
Which fields in the data were missing values? How many?	*How were missing values cleaned? What other choices were considered?*	*Why were missing values cleaned using this method (over others)? Provide comparative charts showing before and after missing values were cleaned.*
Not applicable	Not applicable	Not applicable

CLEANING MISMATCHED VALUES	METHODS USED	COMPARATIVE SUMMARY
Which fields in the data were corrected for mismatched values?	*How were incorrect or mismatched values cleaned? What other choices were considered?*	*Why were incorrect or mismatched values cleaned using this method (over others)? Provide a comparative analysis demonstrating before and after values were cleaned.*
Not applicable	Not applicable	Not applicable

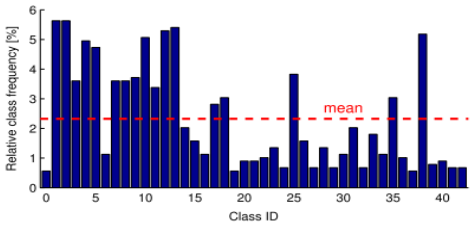
ANOMALY DETECTION	METHODS USED	OUTLIERS HANDLING
How many anomalies or outliers were detected?	*What methods were used to detect anomalies or outliers?*	*If at all, how were anomalies or outliers handled? Why or why not?*
Not applicable	Not applicable	Not applicable

DATA AGGREGATION	METHODS USED	COMPARATIVE SUMMARY
Which fields in the dataset were aggregated?	*What methods were used to aggregate the data? Include the aggregating operator. What other choices were considered?*	*Why was the data aggregated using this method (over others)? Provide comparative charts that demonstrate the choices of aggregators.*
Not applicable	Not applicable	Not applicable
DIMENSIONALITY REDUCTION	METHODS USED	COMPARATIVE SUMMARY
How many original features were collected and how many dimensions were reduced?	*What methods were used to reduce the dimensionality of the data? What other choices were considered?*	*Why were features reduced using this method (over others)? Provide comparative charts showing before and after dimensionality reduction processes.*
Not applicable	Not applicable	Not applicable
JOINING INPUT SOURCES	METHODS USED	RESIDUAL RISKS AND APPROVALS
What were the distinct input sources that were joined?	*What are the shared columns of fields used to join these sources?*	*What are the differential privacy or other residual risks from this join? Include links to relevant approvals and documentation.*
Not applicable	Not applicable	Not applicable
REDACTION OR ANONYMIZATION	METHODS USED	RESIDUAL RISKS AND APPROVALS
Which features were redacted or anonymized?	*What methods were used to redact or anonymize data?*	*What are the differential privacy or reidentification risks to redacted data or anonymization? Include links to relevant approvals and documentation.*
Not applicable	Not applicable	Not applicable
OTHERS (PLEASE SPECIFY)	METHODS USED	RESIDUAL RISKS & COMPARATIVE SUMMARY
What was done? Which features or fields were affected?	*What methods were used?*	*What are the residual risks associated with this transformation? Include links to relevant approvals and documentation. Why were features reduced using this method (over others)? Provide comparative charts showing before and after this transformation.*
Not applicable	Not applicable	Not applicable

Annotations

⚠ *Fill this section if any human or algorithmic annotation tasks were performed in the creation of your dataset.*

ANNOTATION WORKFORCE TYPE	ANNOTATION CHARACTERISTICS	ANNOTATION DESCRIPTION
Bold to select ALL APPLICABLE ✋ *Do not delete any unselected choices.*	*Describe relevant characteristics as indicated. For quality metrics, consider including accuracy, consensus accuracy, IRR, XRR at the appropriate granularity (e.g. across dataset, by annotator, by annotation, etc.). Duplicate for each annotation type if multiple methods were used.*	*Briefly describe the annotations applied to the dataset, including but not limited to: Creation of data, authoring of data, labeling, annotation, rating, etc. Include links, and indicate platforms, tools or libraries used wherever possible. Break down by annotation type as applicable.*
Annotation Target in Data **Machine-generated Annotations** Human Annotations - Expert Human Annotations - Non-expert Human Annotations - Employees Human Annotations - Contractors Human Annotations - Crowdsourcing Human Annotations - Outsourced / Managed Teams Unlabeled Others* (*Please specify)	**Machine-generated Annotations** Total number of annotation 51,840	Annotations are provided in CSV files. Fields are separated by ";" (semicolon). Annotations contain the following information: • **Filename**: Filename of corresponding image • **Width**: Width of the image • **Height**: Height of the image • **ROI.x1**: X-coordinate of top-left corner of traffic sign bounding box • **ROI.y1**: Y-coordinate of top-left corner of traffic sign bounding box • **ROI.x2**: X-coordinate of bottom-right corner of traffic sign bounding box • **ROI.y2**: Y-coordinate of bottom-right corner of traffic sign bounding box The training data annotations will additionally contain • **ClassId**: Assigned class label
	ANNOTATION DISTRIBUTION(S)	ANNOTATION TASK AND INSTRUCTIONS
	Provide a distribution of annotations for each annotation or class of annotations using the format below. Duplicate for each annotation type if multiple methods were used.	*Briefly summarize the annotation task and instructions provided to annotators or methods employed for machine annotations. Include the inter-annotation adjudication policy, and any golden questions if applicable. Add links wherever possible. Break down by annotation type as applicable.*

		The images are stored in PPM format alongside the corresponding annotations in a text file.

Description of Human Annotators

⚠ *Fill this section if human annotators were used.*

	ANNOTATOR BREAKDOWN	ANNOTATOR DESCRIPTION
	Provide a description of the annotators. Add more rows as meaningful. For inapplicable rows, refer to guidance on slide 38 of go/recommended-by. Duplicate for each annotation type if multiple methods were used.	*Provide a brief description of the annotator pool(s). Elaborate on the annotator type, training provided, selection criteria, and anything else that might affect the quality of annotations. Break down by annotation type.*
	No human annotations	Not Applicable

LANGUAGE(S) OF ANNOTATORS	LOCATION(S) OF ANNOTATORS	GENDER(S) OF ANNOTATORS
Provide distributions as available. Duplicate for each annotation type if multiple methods were used.	*Provide distributions as available. Duplicate for each annotation type if multiple methods were used.*	*Provide distributions as available. Duplicate for each annotation type if multiple methods were used.*
Not Applicable	Not Applicable	Not Applicable

Validation Methods

⚠ *Fill this section if the data in dataset was validated during or after the creation of your dataset.*

VALIDATION METHOD(S)	VALIDATION BREAKDOWN	DESCRIPTION OF VALIDATION
Bold to select ALL APPLICABLE ✋ *Do not delete any unselected choices.*	*Describe the fields and data points that were validated. Duplicate for each validation type if multiple methods were used.*	*Briefly describe the methods used to validate the dataset. Include tools, frameworks, libraries, platforms used. Indicate results, outcomes, actions and visualizations. Include links wherever possible. Break down by validation type if multiple methods were used.*
Data Type Validation Range and Constraint Validation Code/cross-reference Validation Structured Validation Consistency Validation **Not Validated** Others* (*Please specify)	Not validated # of data points validated Not applicable Fields Validated: Not applicable Field Not applicable Field Not applicable Field Not applicable	Not applicable

Description of Human Validators

⚠ *Fill this section if the dataset was validated using human validators*

	VALIDATORS CHARACTERISTIC(S)	VALIDATORS DESCRIPTION(S)
	Describe the following about the validators. Add more rows as meaningful. Duplicate for each validation type if multiple methods as necessary.	*Provide a brief description of each validator pool. Elaborate on the annotator type, training provided, selection criteria, and anything else that might affect the quality of annotations. Break down by validation type as necessary.*
	Not applicable Unique validators Not applicable # of examples per validator Not applicable Average cost/task/ validator Not applicable Training provided Not applicable Expertise required Not applicable	Not applicable

LANGUAGE(S) OF VALIDATORS	LOCATION(S) OF VALIDATORS	GENDER(S) OF VALIDATORS
Provide the following distribution as available. Duplicate for each validation type as necessary.	*Provide the following distribution as available. Duplicate for each validation type as necessary.*	*Provide the following distribution as available. Duplicate for each validation type as necessary.*
Not applicable	Not applicable	Not applicable

Sampling Methods

⚠ *Fill out the following block if your dataset employed any sampling methods.*

SAMPLING METHOD(S)	SAMPLING CHARACTERISTIC(S)	SAMPLING CRITERIA
Bold to select ALL APPLICABLE ✋ *Do not delete any unselected choices.*	*Provide the following for each sampling method used. Add additional sampling statistics as relevant. Duplicate for each sampling type, if multiple methods were used.*	*Describe any criteria used to sample the data. Break down by sampling methods as relevant. Include links and metrics where necessary.*
Cluster Sampling Haphazard Sampling Multi-stage Sampling Random Sampling Retrospective Sampling **Stratified Sampling** Systematic Sampling Weighted Sampling Unknown Unsampled Others* (*Please specify)	Stratified Sampling Upstream Source: 50k images Total data sampled: 50k Sample size: 50k Threshold applied: random Sampling Rate: random Sample Mean: Not specified Sample Std. Dev.: Not specified Sampling Distribution: Not specified Sampling Variation: Not specified Sample Statistic: Not specified	• The split was performed at random, but taking into account class and track membership. • This makes sure that (a) the overall class distribution is preserved for each individual set and that (b) all images of one traffic sign instance are assigned to the same set, as otherwise the datasets could not be considered stochastically independent.

Known Applications & Benchmarks

⚠ *Fill out the following section if your dataset was primarily created for use in AI or ML system(s)*

ML APPLICATION(S)	EVALUATION - RESULTS	EVALUATION – PROCESS
Write tags separated by commas. Focus on key tasks performed by the model	*Enumerate the models on which this dataset was used and corresponding performance metrics. Link to model cards or model documentation. Duplicate for each model.*	*Describe any notable factors in your process for evaluating your model's overall performance or assessing how the dataset contributes to the model's performance.* *Break down for each model. Include links, metrics, charts, and visualizations.*
Object detection, Computer vision, Machine learning, Deep learning	Accuracy 95.5 Loss 20.8 Precision - Recall - Performance metric -	• The proposed work is evaluated using different parameters. The training data of the German Traffic Sign Recognition Benchmark dataset dataset has 34799 image samples and the testing data has 12630 image samples. • The loss, train accuracy and valid accuracy for the CNN model was calculated. The training set gave an accuracy of 95% at the end of the 15th epoch.
	MODEL DESCRIPTION(S) AND STATISTICS	EXPECTED PERFORMANCE AND KNOWN CAVEATS
Bold to select ONE *Do not delete any unselected choices.*	*Briefly describe the model(s) and tasks that this dataset was used in. Include links where necessary. Duplicate for each model.*	*Expected performance: Briefly summarize the application and expected performance when using this dataset.* *Known Caveats: Describe the known caveats, trade-offs and consequences when using this dataset. Duplicate for each model. Include links wherever possible.*
Duplicate this row as necessary for each model type	https://drive.google.com/file/d/13EyZdegJor6c7XRfOR1nH_U1X71nrCoX/view?usp=sharing The model analyzed in this card uses a **convolutional neural network** on the **German Traffic Sign Recognition Benchmark dataset** for traffic sign detection. Model Card [Link] Model Size 823KB Model Weights - Model Layers 6 Latency 15 MINS	https://drive.google.com/file/d/13EyZdegJor6c7XRfOR1nH_U1X71nrCoX/view?usp=sharing Expected performance: Deep learning models perform well. Known Caveats: • Dataset cannot be used to predict any other class of traffic sign other the trained 43 classes

Terms of Art

Concepts and Definitions referenced in this Data Card

Use this space to include the expansions and definitions of any acronyms, concepts, or terms of art used across the Data Card. Use standard definitions where possible. Include the source of the definition where indicated. If you are using an interpretation, adaptation, or modification of the standard definition for the purposes of your data card or dataset, include your interpretation as well.

GTSRB
Definition: German Traffic Sign Recognition Benchmark

Source: https://benchmark.ini.rub.de/

Reflections on Data

Use this space to include any additional information about the dataset that has not been captured by the Data Card.

offensive, insulting, threatening data	No, the datacard does not contain any offensive, insulting, threatening data