

Dataset	Beneficence		Non-maleficence	
1) The datasets created by the Lake Lucern Institute 2) To test machine learning models for lesion segmentation 3) Over 250 patients were involved in the data collection process 4) The data was formatted in a way to distinguish each channel and ground truth of the segmentation, the data has been preprocessed on Zenodo 5) Some data files were corrupted 6) The data is publicly available since it was use for the ISLES22 competition	<div>□ What are the expected benefits of analyzing this data? For whom?</div> <div>1) Analyzing this data will allow to create precise machine learning that could ultimately predict correct lesion segmentation. These segmentation can be used for further therapeutic approaches.</div>		Risks	Mitigation
			<div>□ Does the dataset contain unsafe data (violence, nudity...)?</div> <div>□ What kind of impacts can errors in the data or in the analysis have?</div> <div>□ Could the data or the conclusions from the analysis be used in harmful ways?</div> <div>1) No, the data set does not contain unsafe data</div> <div>2) Errors in the data or in the analysis can lead to erroneous lesion segmentation that could lead to medical errors</div> <div>3) No, they can only be used to improve medical follow-up</div>	
	Privacy		Fairness	
	Risks	Mitigation	Risks	Mitigation

	<ul style="list-style-type: none"> □ Does the data contain personal or sensitive information? □ Can personal or sensitive information be derived or inferred from the data or from the analysis? <ol style="list-style-type: none"> 1) The data contains public information from anonymous patients 2) No personal or sensitive information can be derived from this dataset 		<ul style="list-style-type: none"> □ Is the data representative from a larger set (population)? How are subgroups represented? □ What kinds of biases may affect the data? □ Can the outcomes of the analysis be different for different groups? □ Could the data or analysis results contribute to discrimination against people or groups? <ol style="list-style-type: none"> 1) The data contains 250 patients, mainly from Germany and Switzerland, so not that many subgroups are represented unfortunately 2) The age, geographical region, known medical histories 3) Totally, these data are specific to certain groups 4) No, they cannot be used in such a way 	
	Sustainability		Empowerment	
	Risks	Mitigation	Risks	Mitigation
	<ul style="list-style-type: none"> □ What is the carbon and water footprint generated by the storage of the data and by the computation in the analysis process? □ What type of human manual labor is involved in the data (e.g. labeling)? □ Does the data or the analysis require updates? <ol style="list-style-type: none"> 1) CO2 is produced during the training process and water is used to cool down the CPU's 2) The data is originally manually labeled 		<ul style="list-style-type: none"> □ How are the people concerned involved with the data or the analysis: have they been notified, have they consented? □ Are the people concerned able to make choices (e.g. revoke consent, modify or delete data) regarding the data or the analysis? <ol style="list-style-type: none"> 1) All patients have consented to providing their data for research 2) They can revoke their consent whenever they want 	

