



Universiteit  
Leiden  
The Netherlands

## Learning from small samples

Kocaman, V.

### Citation

Kocaman, V. (2024, February 20). *Learning from small samples*. Retrieved from <https://hdl.handle.net/1887/3719613>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3719613>

**Note:** To cite this publication please use the final published version (if applicable).

## Stellingen

Behorende bij het proefschrift

### **Learning from Small Samples**

1. In the context of small datasets for machine learning, the quality and relevance of data often outweigh the benefits of large quantities of data, particularly in domains like healthcare and finance where precision is crucial. [This thesis, chapter-3]
2. Implementing advanced regularization techniques, such as tailored batch normalization, significantly improves the learning efficacy and generalization capabilities of models trained on small and imbalanced datasets. [This thesis, chapter-3 and 4]
3. Utilizing salient image segmentation as a data augmentation technique within self-supervised learning frameworks significantly enhances the efficiency and performance of models, particularly in overcoming the challenges posed by small dataset sizes in image segmentation tasks. [This thesis, chapter-5]
4. Transfer learning emerges as a critical tool for overcoming the challenges of small datasets in machine learning, enabling models to leverage knowledge from related domains effectively. [This thesis, chapter-3]
5. In machine learning, the imbalance in data classes significantly contributes to model bias, particularly affecting minority classes in fields such as fraud detection and rare disease identification.
6. Integrating contextual and prior knowledge into machine learning models enhances their ability to learn effectively from small datasets, thereby improving their applicability in specialized fields like biology and medicine.
7. The development and application of advanced synthetic data generation techniques have the potential to revolutionize the way small datasets are utilized in machine learning, expanding the horizon of model training and testing.
8. Efficient anomaly detection in small datasets remains a significant challenge in machine learning, necessitating the development of specialized algorithms that can detect outliers with high accuracy.
9. The ethical use of machine learning in scenarios with limited data availability, particularly concerning minority populations, poses significant challenges when these models are employed in real-world settings. Limited data for these groups can lead to increased biases and fairness issues, underscoring the need for careful consideration and proactive strategies to ensure equitable and responsible AI applications.

Veysel Kocaman

Leiden, 20 februari 2024