

## INTRO/ABSTRACT

In an age of big data, being able to “do more with less” is important. We show that by following a minimalist paradigm, or “parsimonics”, high image classification accuracy can be achieved even with massively reduced-size datasets. This lets us “do nearly as much with way less”. This idea was tested against a dataset of 87,000 American Sign Language (ASL) alphabet images, yielding high accuracy results of greater than 99% over multiple iterations. These results could have implications for how big data is handled because the data needed to meet objectives may not need to be so big after all.

## METHODS

The images were converted to grayscale and downsampled to 1/16 their original size (in this way the entire dataset was reduced to 6.9% of its original size), then fed to a pretrained model, Alexnet, with high accuracy results of more than 99%. Next a simple convolutional neural network (CNN) was built and yielded similar results of greater than 99%. The training lasted 6 epochs and just over 10 hours for one iteration.

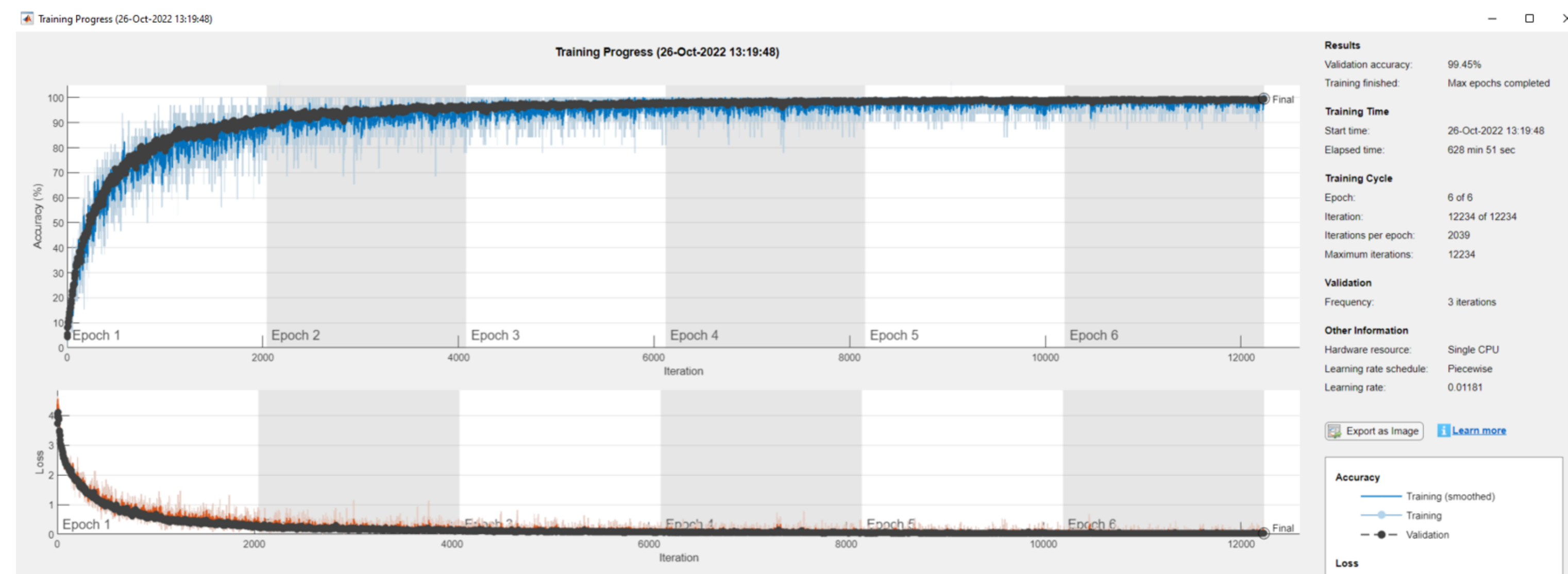


Fig.1 Results of one machine learning training iteration with a final validation accuracy of 99.45%.

## RESULTS

The high accuracy results show that for this dataset and likely others, much extraneous data can be removed upfront in order to speed up downstream processes and reduce computational, storage, and network transmission costs.

Links:  
<https://www.kaggle.com/datasets/grassknotted/asl-alphabet>  
<https://github.com/jko113/parsimonics>

# Parsimonics gives little data the power of big data, while adding the benefit of simplicity.

# By reducing a dataset's size to a minimum without compromising its utility, costs decrease and possibilities increase.



QR code to original dataset



QR code to GitHub repository