Neural Network Activation Fonctions Comparison

Activation	Optimizer	Batch Size	Test Accuracy
relu	adam	32	0.9901
relu	adam	64	0.9886
relu	sgd	32	0.9878
relu	sgd	64	0.9848
selu	adam	32	0.9851
selu	adam	64	0.9877
selu	sgd	32	0.986
selu	sgd	64	0.9853
<function leaky_relu at 0x78f64a05ae60>	adam	32	0.9887
<function leaky_relu at 0x78f64a05ae60>	adam	64	0.989
<function leaky_relu at 0x78f64a05ae60>	sgd	32	0.9859
<function leaky_relu at 0x78f64a05ae60>	sgd	64	0.9861
<function swish at 0x78f64a05b0a0>	adam	32	0.9908
<function swish at 0x78f64a05b0a0>	adam	64	0.9891
<function swish at 0x78f64a05b0a0>	sgd	32	0.9874
<function swish at 0x78f64a05b0a0>	sgd	64	0.9851
mish	adam	32	0.9873
mish	adam	64	0.9908
mish	sgd	32	0.9876
mish	sgd	64	0.985
tanh	adam	32	0.9885
tanh	adam	64	0.989
tanh	sgd	32	0.9873
tanh	sgd	64	0.9859
gelu	adam	32	0.9901
gelu	adam	64	0.9913
gelu	sgd	32	0.9897
gelu	sgd	64	0.9873

This table presents the performance of a neural network with LeNet-5 architecture across various activation functions. The model is trained for 20 iterations on the MNIST dataset, exploring different combinations of optimizers (adam, sgd) and batch sizes (32, 64).
Activation functions under consideration include:

Relu , selu , leaky_relu , swish ,mish ,tanh ,gelu .

Results showcase test accuracy for each configuration, providing insights into the impact of activation functions on model learning. Notable performers include '**gelu**' with '**adam**' optimizer and batch size '**64**', and 'swish' with 'adam' optimizer.