The model that appeared to perform the best on the test set were the ones with dropout layers added. The first of these models had a dropout layer added before the fully connected layer while the second model had a dropout layer added after each Relu activation function. The second dropout layer model edged out the first with a test accuracy of 29.62% but also had the highest training time. The model with the lowest training time was Variant 2 but also had the lowest test accuracy.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Train Accuracy | Validation Accuracy | Test Accuracy | Parameters | Time |
| Base | 0.10188 | 0.1004 | 0.1012 | 1186378 | 28min 16s |
| Variant 1 (SELU) | 0.59248 | 0.6034 | 0.115 | 1186378 | 27min 50s |
| Variant 1 (SWISH) | 0.0998 | 0.096 | 0.115 | 1186378 | 26min 23s |
| Variant 2 | 0.1 | 0.0978 | 0.1 | 1147978 | 5min 7s |
| Variant 3 (Dropout 1) | 0.28162 | 0.2996 | 0.2917 | 1186378 | 25min 15s |
| Variant 3 (Dropout 2) | 0.4232 | 0.4306 | 0.2962 | 1186378 | 30min 10s |
| Variant 4 | 0.10056 | 0.0996 | 0.0028 | 1186378 | 24min 47s |