Understanding COVID-19 Genome DATASET Using MLCERT/ Tensor Flow/Coq Theorem Prover[CTP] - A Simple Idea on Next Generation Bio-informatics R&D.

by Dr.Nirmal - Informatics R&D - ante Inst UTD Dallas TX USA.
Contact_info - hmfg2014@gmail.com

"We present MLCERT, a novel system for doing practical mechanized proof of the generalization of learning procedures,bounding expected error in terms of training or test error. MLCERT is mechanized in that we prove generalization bounds inside the theorem prover Coq; thus the bounds are machine checked by Coq's proof checker. MLCERT is practical in that we extract learning procedures defined in Coq to executable code; thus procedures with proved generalization bounds can be trained and deployed in real systems. MLCERT is well documented and open source; thus we expect it to be usable even by those without Coq expertise. To validate MLCERT,which is compatible with external tools such as TensorFlow,we use it to prove generalization bounds on neural networks trained using TensorFlow on the extended MNIST data set." -> Please see the reference [b] below.

[a] http://ace.cs.ohio.edu/~gstewart/mlcert.html

[b] http://ace.cs.ohio.edu/~gstewart/papers/aaai19-bagnall.pdf - **Consider Figure I & try to modify it as per your requirements.**

[c] https://www.scienpress.com/Upload/JAMB/Vol%202_2_6.pdf

[d] https://github.com/tejdnk-2019-ShortNotes/2021-Nir-Informatics

[e] http://mlcert.org/ && [f] https://coq.inria.fr/ && [g] https://ocaml.org/

[h] https://www.tensorflow.org/ && https://registry.opendata.aws/ncbi-covid-19/ - COVID-19 Genome Sequence Dataset.

Testing in Progress @ the time of submission.
Probing Bio-informatics w.r.t Covid-19 Genome datasets + Machine Learning is very interesting.

Non-Profit R&D - Inspire Others Always - Thanks to all WHO made this happen in my LIFE.

[THE END]

[01-August-2021]