Steps to do this assignment:

1. Clone the original model
2. We then redefine the model to output after pool_3
3. Calculate the last convolutional layers weights and biases
4. For pruning one channel at a time, we set channel ‘chIdx’ of last convolutional layers weights and biases to 0
5. Then, we update the weights of B_clone
6. We then evaluate the model B_clone on valid dataset and if the drop in clean_accuracy_valid is greater or equal than the thresholds(2, 4, 10) we save B_clone as B_prime

Then we reconstruct the model repaired net and evaluate the repaired net for clean classification accuracy and attack success rate.

X	B_prime	B	Repaired_net
2	401/401 [==============================] - 1s 3ms/step Clean Classification accuracy for B_prime: 95.90023382696803 401/401 [==============================] - 1s 3ms/step Attack Success Rate for B_prime: 100.0	401/401 [==============================] - 1s 2ms/step Clean Classification accuracy for B: 98.62042088854248 401/401 [==============================] - 1s 2ms/step Attack Success Rate for B: 100.0	Clean Classification accuracy for repaired net: 95.74434918160561 Attack Success Rate for repaired net: 100.0
4	401/401 [==============================] - 1s 3ms/step Clean Classification accuracy for B_prime: 92.29150428682775 401/401 [==============================] - 1s 3ms/step Attack Success Rate for B_prime: 99.98441153546376	401/401 [==============================] - 1s 3ms/step Clean Classification accuracy for B: 98.62042088854248 401/401 [==============================] - 1s 3ms/step Attack Success Rate for B: 100.0	Clean Classification accuracy for repaired net: 92.1278254091972 Attack Success Rate for repaired net: 99.98441153546376
10	401/401	401/401	Clean Classification

	[==============================] - 1s 3ms/step Clean Classification accuracy for B_prime: 84.54403741231489 401/401 [==============================] - 1s 3ms/step Attack Success Rate for B_prime: 77.20966484801247	[==============================] - 1s 3ms/step Clean Classification accuracy for B: 98.62042088854248 401/401 [==============================] - 1s 2ms/step Attack Success Rate for B: 100.0	accuracy for repaired net: 84.3335931410756 Attack Success Rate for repaired net: 77.20966484801247