Data Science: Artificial Intelligence (BU.920.624)

Group Assignment #2

Due on November 20th via Canvas

Group Members: _____

1. Read the Harvard Business School case "The Future in Sight: LumineticsCore and the First Autonomous AI for Diagnostics" (Product #: 626019-PDF-ENG), available through the coursepack: https://hbsp.harvard.edu/import/1357348. The case has been pre-paid by the instructor. Use the link for your own access only and do not share it outside the class. Your answers must rely on specific facts, data, and design choices from the case and its exhibits.

(a) Explain how LumineticsCore works and how it was trained. Discuss why Dr. Abramoff's design and data strategy enabled him to secure FDA authorization ahead of competitors such as Google, drawing on the case's description of ground truth standards, dataset construction, and the autonomy requirement. Based on your own professional background and career interests, and if helpful the perspectives of your team members, critique the main strengths and limitations of his algorithmic and dataset choices.

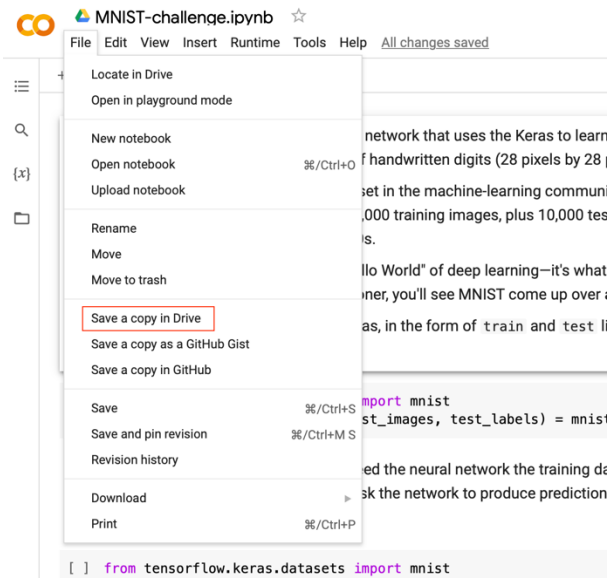
(b) Describe the pivotal trial and evaluation method used for FDA clearance, including the target condition, site selection, patient sample, performance thresholds, diagnosability rate, and the use of sensitivity and specificity as primary endpoints. Drawing from your own background and those of your teammates, explain why these metrics were chosen and how this evaluation approach aligns or conflicts with the priorities of patients, primary care physicians, ophthalmologists, the FDA, and payers, using concrete examples from the case.

(c) If you are tasked with rebuilding LumineticsCore today, outline how you would redesign its clinical role (autonomous diagnosis versus triage versus decision support), technical architecture, regulatory strategy (de novo versus 510(k), trial design), and business/reimbursement model. Feel free to incorporate your own and your team members' professional experience. Show how your redesign would address at least two adoption obstacles highlighted in the case.

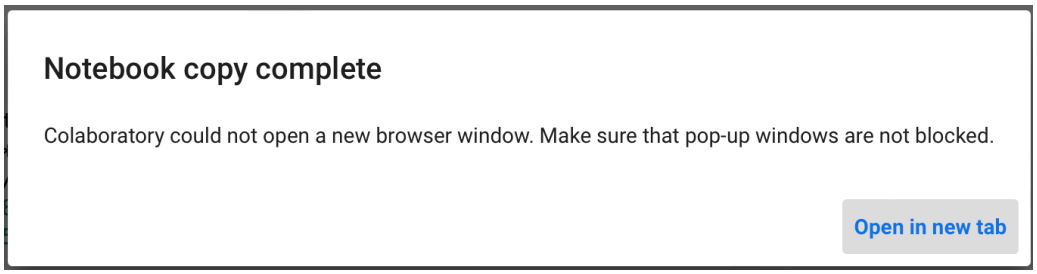
(d) Propose at least three questions you would ask Dr. Abramoff. Aim for questions that reflect the background and interests of your team. For each question, briefly explain why it matters.

2. Open the Google Colab notebook "MNIST-challenge.ipynb" by clicking on the following link: https://colab.research.google.com/drive/1An8WZ2pj7NOFChDouMkOWuk0GtjF0Dvp?usp=sharing

After you've opened the above notebook, go to the "File" menu, and select "Save a copy in Drive" (as shown in the screenshot below):



When you've done saving a copy to Drive, select "Open in new tab":



The notebook is now yours!

(a) Read and run the code in its entirety. How accurate is the model on the training data? How good is the model on the test data? Is the model prone to overfitting?

(b) In the last section of your Google Colab notebook:

```
model.fit(
    train_images,train_labels,
    epochs=5,
    batch_size=128,
    validation_data=(test_images,test_labels),
    )
```

change "`epochs=5,`" to "`epochs=20,`".

Run the code again. How accurate is the newly trained model on the training data? How accurate is the newly trained model on the test data? Does this model still have an overfitting problem? If so, is the situation better or worse?

(c) Propose and implement strategies to deal with overfitting. You can use the dropout technique, for example, by changing the Python code below "Let's build the network":

```
model=keras.Sequential([
 layers.Dense(512,  activation = "relu"),
 layers.Dense(10, activation = "softmax")
])
```

to

```
model=keras.Sequential([
  layers.Dense(512,  activation = "relu"),
  layers.Dropout(0.5),
  layers.Dense(10, activation = "softmax")
])
```

Feel free to change the dropout rate to any value you like. In addition to the dropout technique, consider increasing the size of your neural network (e.g., by changing the number of neurons from "512" to "784") or the number of hidden layers (e.g., by adding another layer using `layers.Dense(16, activation="relu"))`. You can also use the regularization technique (L1/L2).

Run your revised code and demonstrate that you have a better model in the sense that overfitting is less of a problem.

3. Use a generative AI tool, such as ChatGPT, Claude.ai, or Google Gemini, to help you build and train an AI model in Google Colab. You may choose one MedMNIST dataset (PneumoniaMNIST, BreastMNIST, DermaMNIST, or PathMNIST), or, if you prefer a non-health application, use CIFAR-10 or CIFAR-100. With guidance from your GenAI assistant, write Colab-ready code that automatically loads your chosen dataset, builds an initial model, and then iteratively improves it to increase accuracy. In your submission, explain what your model aims to accomplish and why it is useful, provide your full implementation, describe how you used the GenAI tool during development, report your final test accuracy, and offer a concise reflection on which model changes led to the greatest improvement.

4. Consider the following two-dimensional 5 × 5 input feature map:

0	1	1	1	0
0	-1	0	-1	1
0	0	-1	0	1
0	-1	0	-1	1
0	0	0	0	0

We now apply Convolution, Activation (ReLU) and MaxPooling to it.

(a) First, we apply the following two-dimensional, 2 × 2 convolutional filter

1	0
0	1

to the input feature map. Represent the output feature map.

(b) Apply the ReLU activation function to the result from step (a). Represent the output feature map.

(c) Maxpool the result from step (b) with 2 × 2 filters and stride 2. Represent the output feature map.