Team name

- Agile

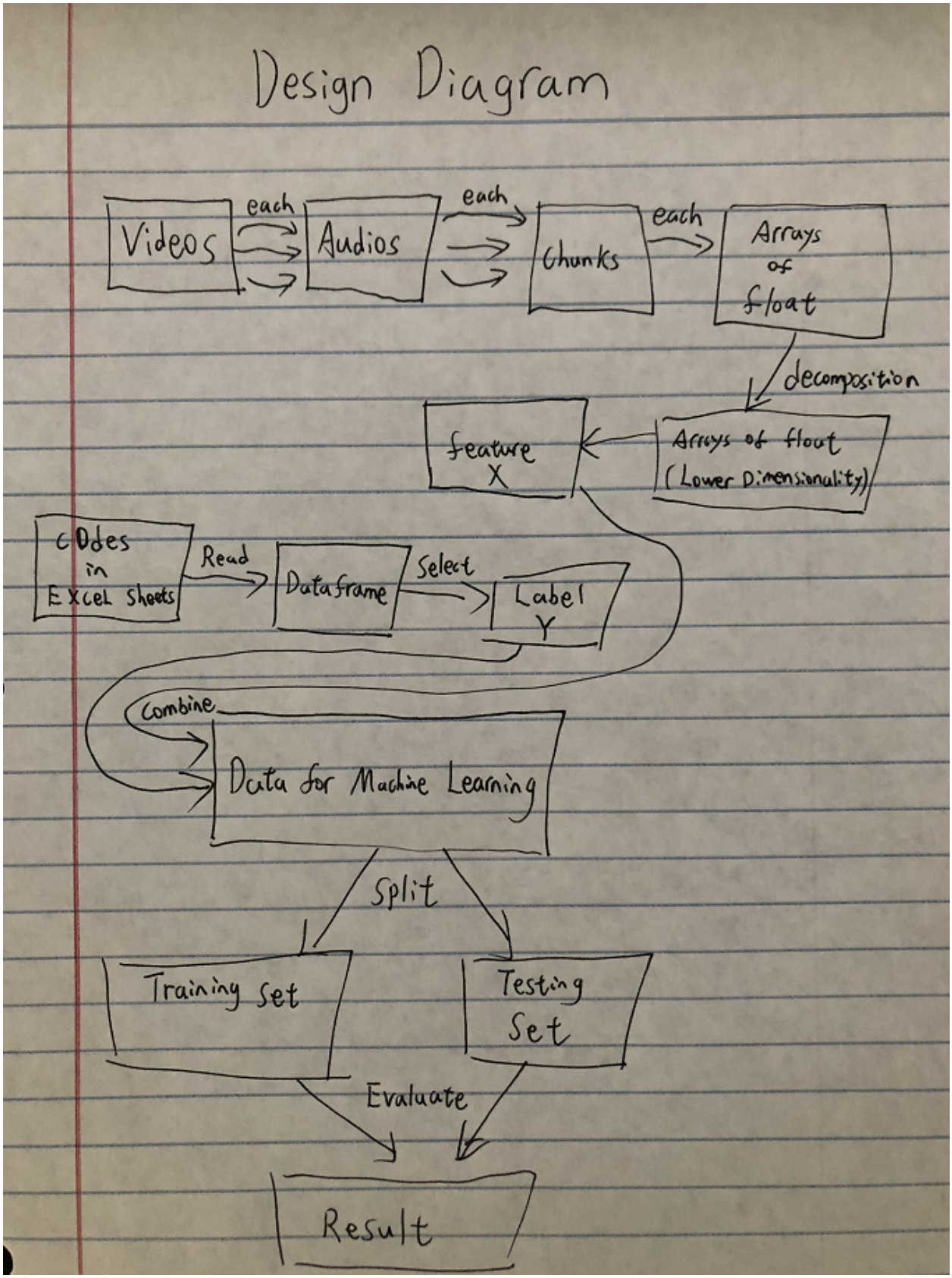
Team roles

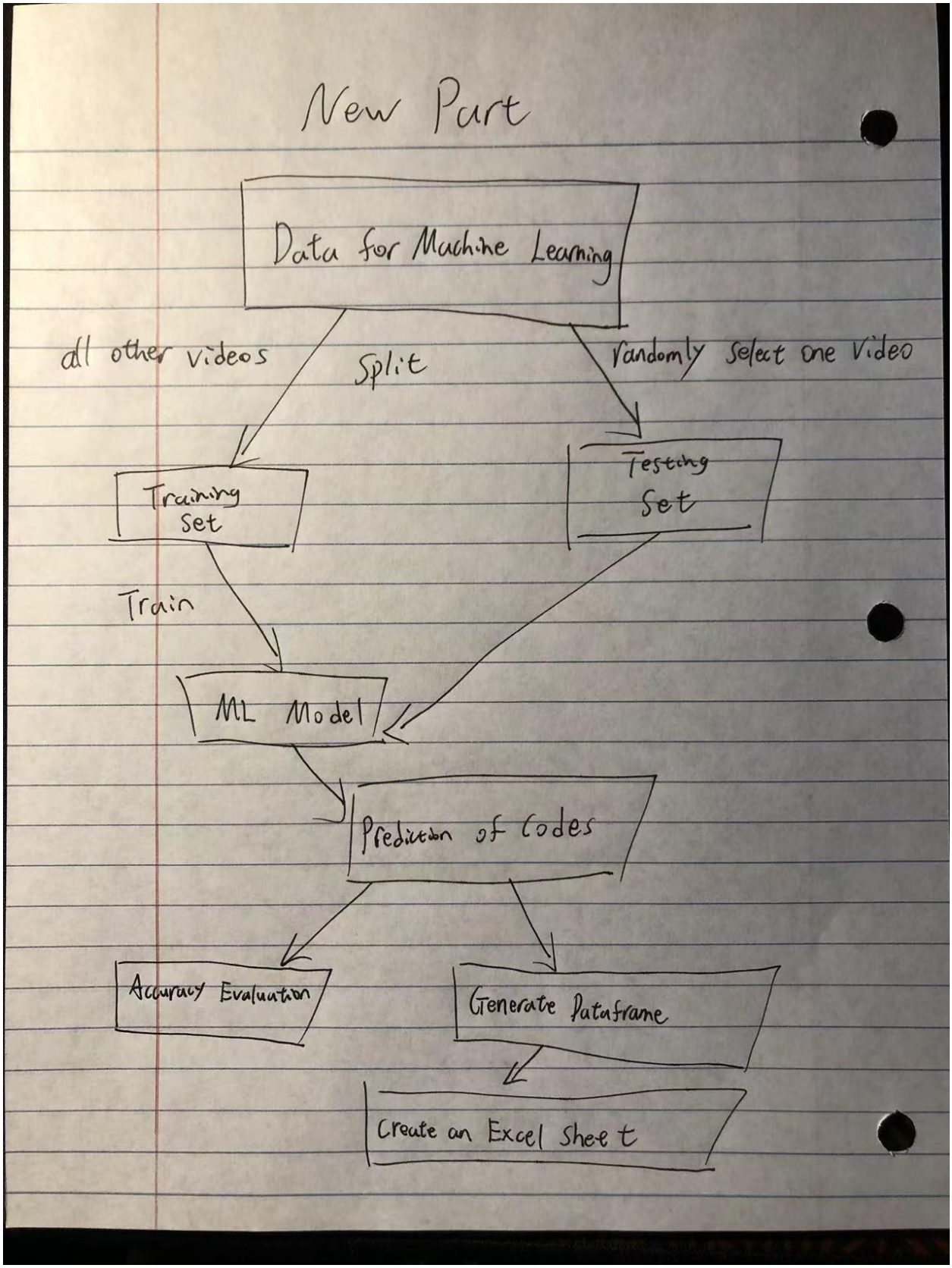
- Pei Chen: Developer
- Jin Huang: Developer
- Yuanyuan Lei: Product Owner
- Dongqing Yang: Developer
- Han Zhang: Developer
- Rongruo Zhou: Scrum Master

Customer meeting date/time/place

- First Meeting: 10/23/2021, Saturday, 10:00 am, Zoom
- Weekly regular meeting: every Wednesday (Start on Oct.27), 11 am – 12 pm, Zoom with Cindy, Guerrero, and her group members.
- Weekly regular meeting: every Wednesday (Start on Nov.3), 5 pm - 6 pm, Zoom with Dr. Irby for updates and clarifications
- Second Meeting: 10/21/2021, Wednesday, 11:00 am, Zoom
- Third Meeting: 11/10/2021, Wednesday, 11:00 am, Zoom
- Fourth Meeting: 11/17/2021, Wednesday, 11:00 am, Zoom

Design Diagram


Design Diagram
Videos
each
Audios
each
Chunks
each
Arrays of float
decomposition
Arrays of float (Lower Dimensionality)
feature X
codes in Excel sheets
Read
Dataframe
Select
Label Y
Combine
Data for Machine Learning
Split
Training set
Testing Set
Evaluate
Result


New Part
Data for Machine Learning
all other videos
Split
randomly select one video
Training Set
Testing Set
Train
ML Model
Prediction of codes
Accuracy Evaluation
Generate Dataframe
Create an Excel Sheet

Customer Feedback

- As in Wednesday's meeting this week, two of our customers listened to my report and affirmed my achievement so far. They suggested providing me with more data to help improve the accuracy of predictions. The data are uploaded today and I'm starting to work on them.
- One of the customers, Dr.Tang proposed two new to-do events to me. One is to generate the Excel sheets based on the predictions, which has been done now. The other one is to write a report to her by the end of Friday, I'm now working on it.

Note

- The code from last year is able to make predictions toward one of the eight features called "Lang of Teacher." And it's restricted to three categories in that feature.
- But the current code by me is able to make predictions toward all of the eight features. And there are no restrictions to the number of categories.
- But the accuracy for different features differs. That's because of various reasons. One point is that the number of categories of each feature differs. Some features have lots of categories which lead to relatively lower accuracy.

User Stories Finished

- Modify the method to split the training and testing data. Now I use only one video as the testing data and all other videos as training data. That's due to our limited video data. I decide to use as many videos as I can to train the model.

- Add all the features into training. And also, make predictions toward all the features to generate a whole Excel sheet based on the video.
- Select the model with the highest accuracy to generate the predictions. And combine the predictions of all the features into one Dataframe.
- Generate an Excel sheet based on the final Dataframe with a similar format as the Codes sheets provided by the customers. So the customers can use the predicted Codes in production.

GitHub

- https://github.com/oniremilia/ILOI

Slack

- https://join.slack.com/t/csce606iloi/shared_invite/zt-xglog0ed-F8eRaqVaKMo0KheFO8DD5A

Custom Grading Request

- Since the ILOI project is a pure machine learning, or natural language processing project. We do not have any requirements from the customers to do the web stuff. The customers are looking for locally running machine learning modules that can generate result forms as desired.
- Hence, I can not deploy to Heroku. But the code can run as described before locally, and each of the user stories can be tested using data similar to the data I’m using. Also, tests like RSpec or Cucumber are not doable.
- The current code completes most of our customer requirements already. I will discuss with our customers to see their requirements for the accuracy of

predictions. And improving the accuracy should be the main and only remaining goal for us.