In my point of view, doppelganger effects are NOT unique to biomedical data, because in many fields of machine learning, they share characteristics such as "high sample similarity" and "small sample training set", which are highly correlated to doppelganger effects.

There are some other kinds of fields in which may have doppelganger data:

- Quantitative Trading
- Autopilot
- Plant image recognition
- ……

Possible solutions:

1. On the basis of the existing model, establish one new kind of model named "amplification feature model", which only uses the data processed by "feature amplification" to decrease the similarty between doppelganger data, both in training and validation. In actual use, the data should be “amplified” before being analyzed, too. (However, this may lead to new problems: how to ensure that the data after feature amplification is not distorted)

Due to deadline, the contents above is the whole content of my report. I will continue my thinking on this problem.