# BÁO CÁO ĐỢT 2

1. Viết báo cáo chương 1 + 2: tổng quan đề tài và cơ sở lý thuyết + related work
2. Vấn đề imbalanced data:

* Dữ liệu cho người không bị bệnh chênh lệch khá nhiều so với dữ liệu của người bị bệnh ở mỗi giai đoạn nhỏ. Nếu lấy quá nhiều dữ liệu trong tập training ra làm dữ liệu validation, phần dữ liệu còn lại của tập training là không đủ để xây dựng mô hình. Lúc này, tập validation phải thật nhỏ để giữ được lượng dữ liệu cho training đủ lớn. Tuy nhiên, một vấn đề khác nảy sinh: tập validation quá nhỏ, hiện tượng overfitting có thể xảy ra, model sẽ cố fit với dữ liệu người không bị bệnh.
  + Solution: cross-validation
  + *Cross validation* làCross validation là một kỹ thuật lấy mẫu để đánh giá mô hình học máy trong trường hợp dữ liệu không được dồi dào cho lắm. 3 phương pháp cross-validation phổ biến:

1. Hold-out: phương pháp đơn giản nhất. Dữ liệu được chia một cách ngẫu nhiên thành một tập dữ liệu học và một tập dữ liệu kiểm tra. Dùng tập đầu tiên để dạy máy rồi dùng ngay tập còn lại để kiểm tra.
2. K-fold: đây là phương pháp nâng cấp của hold-out. Toàn bộ dữ liệu được chia thành K tập con. Quá trình học của máy có K lần. Trong mỗi lần, một tập con được dùng để kiểm tra và K-1 tập còn lại dùng để dạy.
3. Leave-one-out: Tương tự như K-Fold nhưng tối đa hóa số tập con (K = số dữ liệu).
   * Một cách thường được sử dụng là chia tập training ra k tập con không có phần tử chung, có kích thước gần bằng nhau. Tại mỗi lần kiểm thử , được gọi là *run*, một trong số k tập con được lấy ra làm *validata set*. Mô hình sẽ được xây dựng dựa vào hợp của k -1 tập con còn lại. Mô hình cuối được xác định dựa trên trung bình của các *train error* và *validation error*. Cách làm này còn có tên gọi là **k-fold cross validation**.

* Threshold: đặt ngưỡng cho final output layer, thay vì số ouput của layer cuối là 5, thì thay bằng 1 và đặt các ngưỡng để mở rộng phạm vi chính xác.
* EfficientNet: mô hình được phát triển dựa trên CNN, đạt độ chính xác và hiệu năng khá cao.
* Augmentation (Tăng cường dữ liệu)