I – Introduce:

1. Machine Learning ?
2. Lí do chọn đề tài

* Thực tế:
  + Lí do chọn ngành y khoa
    - Nhu cầu hỗ trợ bác sĩ trong y khoa càng lúc càng tăng theo thời gian, và số lượng bệnh nhân càng lúc càng tăng và các loại bệnh cơ bản lặp đi lặp lại nhiều nên cần có công cụ hỗ trợ các bác sĩ chuẩn đoán nhanh các bệnh cơ bản đó.
  + Lí do chọn bệnh phổi
    - Vì quá trình công nghiệp hóa hiện đại hóa đang diễn ra với tốc độ nhanh chóng, từ đó kéo theo rất vấn đề như: ô nhiễm môi trường đặc biệt là không khí ,… Dẫn đến các bệnh liên quan về đường hô hấp gia tăng và ngày càng phổ biến. Nên có 1 lượng dữ liệu lớn về bệnh phổi - sau khi chuẩn đoán (bằng phương pháp chụp X - ray).
    - Số liệu thực tế: (theo: Bản đồ ung thư thế giới mới nhất: Việt Nam đang đứng thứ bao nhiêu?)
      * Theo số liệu của Viện nghiên cứu và Phòng chống ung thư, số ca mắc ung thư phổi ở nữ giới năm 2000 chỉ khoảng 2.001 ca, tỷ lệ mắc là 6,5 người/100 nghìn dân thì sau 10 năm, con số này đã lên tới 5.709 ca và tỷ lệ mắc là 13,9/100 nghìn người dân. Ước tính đến năm 2020 số ca mắc ung thư phổi ở nữ giới sẽ là 11.656 ca.
      * Số ca mắc ung thư phổi ở nam giới năm 2000 chỉ là 6.905 ca với 29,3 người/100 nghìn dân, thì đến năm 2010 con số này đã tăng lên 14.652 ca và tỷ lệ mắc là 35,1 ca/100 nghìn dân. Theo ước tính, đến năm 2020 tỉ lệ nam giới mắc bệnh ung thư phổi ở nước ta sẽ là 22.938 ca.
* Tương lai
  + Machine Learning nói riêng và AI nói chung trong tương lai sẽ phát triển rất nhanh chóng và hỗ trợ, phục vụ con người trong nhiều lĩnh vực như: kinh tế, công nghiệp, đời sống, giáo dục,… và đặt biệt là y khoa.
  + Bệnh càng ngày càng nhiều và trở nên phổ biến kéo theo lượng dữ liệu sẽ tăng lên rất nhiều lần so với hiện tại nên cần các công cụ hỗ trợ mang tính bước ngoặt trong lĩnh vực y tế.

1. Ý nghĩa đề tài

* Áp dụng Machine Learning & AI vào trong y khoa sẽ giúp các bác sĩ dễ dàng và nhanh chóng hơn trong việc chuẩn đoán bệnh để từ đó kê đơn thuốc đúng và cho từng loại bệnh của các bệnh nhân.
* Giảm thiểu sự sai sót trong chuẩn đoán bệnh.
* Hỗ trợ các bác sĩ từ các quốc gia có nên y khoa chưa phát triển từ đó giúp giảm thiểu rủi ro tử vong của bệnh nhân vì chẩn đoán không chính xác.

VD: Mỗi năm, riêng ở Mỹ, ước tính có khoảng 40.000 đến 80.000 người chết do chẩn đoán sai. (theo BCC News: Lỗi chẩn đoán gây tử vong cho bệnh nhân).

II – Goal:

* Hoàn thành dự án đúng mong đợi.
* Từ dữ liệu có sẵn, training ra được bộ model cho chẩn đoán sau này.
* Với hình ảnh test sẽ dự đoán ra được có bệnh hay không có bệnh.
* Dự đoán bệnh với độ chính xác cao (trên 80%).
* Có thể tận dụng dự án vào trong thực tế.

III – Objective

Các bước thực hiện dự án:

1. Đặt đúng vấn đề:
   * Tìm hiểu đúng thực trạng và nhu cầu sử dụng
   * Phân tích vấn đề thực tế
   * Hội ý và đưa ra giải pháp
2. Chuẩn bị dữ liệu:
   * Tìm nguồn cung cấp dữ liệu thực tế
   * Phân loại dữ liệu
   * Phân tích dữ liệu
   * Chuyển đổi dữ liệu dạng hình, nhãn thành dạng dữ liệu matrix theo vector
3. Chọn thuật toán
   * Nghiên cứu và tìm hiểu các thuật toán cơ bản và phổ biến trong Machine Learing
   * Xác định các thuật toán phù hợp với dự án
   * Áp dụng các thuật toán đã chọn vào dự án
4. Huấn luyện
   * Từ dữ liệu đã xử lí trước , thông qua việc áp dụng các thuật toán để tạo ra 1 bộ model để từ đó so sánh với dữ liệu test và cho ra kết quả theo mong đợi
5. Kiểm thử
   * Input hình cần chẩn đoán sau khi dự đoán sẽ cho ra kết quả với số liệu về độ chính xác (theo %)

IV – Theory:

V – Model: