# **Research Proposal: Dự đoán đợt lũ lụt**

## **1. Tên đề tài**

**Dự đoán thời gian xảy ra đợt lũ lụt ở Việt Nam dựa trên dữ liệu khí tượng và thủy văn**

## **2. Giới thiệu đề tài**

Lũ lụt là một trong những thảm họa thiên nhiên nghiêm trọng và phổ biến ở Việt Nam, gây ra thiệt hại lớn về người và tài sản hàng năm. Đặc biệt, các tỉnh miền Trung và Tây Nguyên thường xuyên hứng chịu những đợt lũ kéo dài do mưa lớn, gió mùa và bão nhiệt đới. Việc **dự đoán chính xác khoảng thời gian bắt đầu và kết thúc** của các đợt lũ có thể giúp các cơ quan chức năng và người dân chủ động hơn trong phòng chống và giảm nhẹ thiệt hại.

## **3. Mục tiêu nghiên cứu**

* Xây dựng mô hình AI nhằm **dự đoán thời điểm bắt đầu và kết thúc** của một đợt lũ lụt tại một khu vực cụ thể ở Việt Nam.
* Phân tích các yếu tố thời tiết, thủy văn và địa hình có ảnh hưởng mạnh đến sự xuất hiện của các đợt lũ.
* Tạo ra một hệ thống có thể áp dụng mở rộng cho nhiều địa phương khác nhau.

## **4. Câu hỏi nghiên cứu**

* Những yếu tố nào (địa hình, lượng mưa, độ ẩm, gió, áp suất, độ che phủ mây, v.v.) có ảnh hưởng lớn đến việc hình thành và kết thúc của một đợt lũ?
* Có thể sử dụng dữ liệu khí tượng – thủy văn và địa hìnhình lịch sử để huấn luyện mô hình dự đoán khoảng thời gian của lũ không?
* Mô hình AI nào là phù hợp nhất cho bài toán này?

## **5. Phương pháp nghiên cứu**

### **5.1. Thu thập dữ liệu**

* Dữ liệu lịch sử về **các đợt lũ ở Việt Nam** từ năm trong lịch đến để làm target cho mô hình dự đoán
* Dùng dữ liệu trên để làm mốc cho việc kéo API về để lấy data trước mốc xảy ra lũ lụt
* Dữ liệu thời tiết gồm: lượng mưa, nhiệt độ, độ ẩm, tốc độ và hướng gió, áp suất, độ che phủ mây, mực nước sông,...
* Dữ liệu sẽ được xử lý và gán nhãn theo từng đợt lũ (có start\_date, end\_date).

### **5.2. Tiền xử lý dữ liệu**

* Lựa chọn 1 biến basecase liên quan nhất đến nguyên nhân gây lũ lụt để EDA và tìm tương quan các biến còn lại với biến này. Sau đó chọn ra các biến tương quan nhất.
* Lưu dữ liệu dưới dạng .npy
* Scalling dữ liệu và transform thành tensor để đưa vào mô hình

### **5.3. Xây dựng mô hình**

* Thử nghiệm các mô hình:  
  + **Base model**: VGG-net, Resnet,....
  + **Sequence Model (LSTM/Transformer)**: Dự đoán chuỗi khả năng xảy ra lũ và xác định điểm bắt đầu – kết thúc.
  + **Hybrid Model**: Classification + Regression: Dự đoán có lũ (Yes/No), sau đó dự đoán khoảng thời gian.

### **5.4. Đánh giá mô hình**

* **MAE (Mean Absolute Error)**: chênh lệch số ngày giữa dự đoán và thực tế.

## **6. Dữ liệu đầu ra mong đợi**

* Mô hình đầu vào là dữ liệu thời tiết trong khoảng thời gian nhất định.
* Đầu ra: **ngày bắt đầu và ngày kết thúc của một đợt lũ (start\_date, end\_date)**.
* Dashboard trực quan để hiển thị khoảng thời gian dự đoán và sai số.

## **7. Ứng dụng thực tiễn**

* Hỗ trợ Trung tâm Dự báo KTTV và các Ban chỉ đạo phòng chống thiên tai trong việc ra quyết định.
* Cảnh báo sớm và chính xác hơn cho người dân tại các khu vực nguy cơ cao.
* Tích hợp với hệ thống bản đồ GIS để cảnh báo lũ theo thời gian thực.