**BÀI TẬP CÁ NHÂN TUẦN 2**

Học phần: Nghiên cứu khoa học giáo dục nghề nghiệp

Mã học phần: SP1416

Sinh viên thực hiện: Nguyễn Hoàng Đỉnh

Mã sinh viên: 19004032

**Câu 1: Xác định đề tài nghiên cứu từ đó đặt tên đề tài nghiên cứu dựa trên đề tài đã chọn; xác định tính chất của đề tài nghiên cứu;**

U não được coi là một trong những căn bệnh nguy hiểm ở cả trẻ tôi và người lớn. Các khối u não chiếm từ 85 đến 90 phần trăm của tất cả các khối u nguyên phát của hệ thống thần kinh trung ương. Mỗi năm, có khoảng 11.700 người được chẩn đoán mắc bệnh u não. Tỷ lệ sống sót sau 5 năm đối với những người bị ung thư não hoặc khối u thần kinh trung ương là khoảng 34% đối với nam giới và 36% đối với nữ giới. U não được phân loại thành: U lành tính, u ác tính, u tuyến yên,… Cần thực hiện điều trị đúng cách, có kế hoạch và chẩn đoán chính xác để nâng cao tuổi thọ cho người bệnh.

Kỹ thuật tốt nhất để phát hiện u não là chụp cộng hưởng từ. Một lượng lớn dữ liệu hình ảnh được tạo ra thông qua các lần quét. Những hình ảnh này được kiểm tra bởi bác sĩ X quang. Việc áp dụng các kỹ thuật phân loại tự động bằng máy học và trí tuệ nhân tạo luôn cho thấy độ chính xác cao hơn so với phân loại thủ công. Do đó, việc đề xuất một hệ thống thực hiện phát hiện và phân loại bằng cách sử dụng thuật toán học sâu sử dụng mạng thần kinh chuyển đổi, mạng thần kinh nhân tạo và học chuyển giao sẽ hữu ích cho các bác sĩ trên toàn thế giới.

Apache Spark cho phép xây dựng các mô hình dự đoán nhanh chóng với việc tính toán được thực hiện trên một nhóm các máy tính, có thể tính toán cùng lúc trên toàn bộ tập dữ liệu mà không cần phải trích xuất mẫu tính toán thử nghiệm. Tốc độ xử lý của Spark có được do việc tính toán được thực hiện cùng lúc trên nhiều máy khác nhau. Đồng thời việc tính toán được thực hiện ở bộ nhớ trong (in-mtôiories) hay thực hiện hoàn toàn trên RAM… Nên tốc độ xử lý và đưa ra kết quả gần như tức thì. Việc này giúp cho các bác sĩ nhật xét nhanh chóng về tình trạng của bệnh nhân để có thể đưa ra giải pháp chữa bệnh kịp thời.

- **Tên đề tài nghiên cứu**: Phân tích và xử lý dữ liệu song song trên nền tảng Apache Spark.

- **Tính chất của đề tài**: Nghiên cứu lý luận.

**Câu 2: Viết “lịch sử vấn đề” dựa trên đề tài vừa chọn.**

Một bất trắc xảy ra trong việc phân đoạn khối u não là dữ liệu vô cùng lớn khiến cho quá trình huấn luyện hệ thống chuẩn đoán hoặc kiểm tra dữ liệu bị chậm lại. Điều này gây ra khó khăn lớn khi ứng dụng vào thực tế. Bởi vì các bác sĩ phải đưa ra chuẩn đoán sớm và chuẩn xác để đưa ra giải pháp thích hợp và điều trị cho bệnh nhân. Nền tảng Apche Spark là một trong những nền tảng phân tích và xử lý dữ liệu tốc độ cao và thường được người dùng gọi là nền tảng “Nhanh như chớp”.

Apache Spark hoạt động với hệ thống tệp để phân phối cụm dữ liệu và xử lý dữ liệu đó song song. Nó bao gồm nhiều khối lượng công việc như ứng dụng hàng loạt, thuật toán lặp, truy vấn tương tác, phân tích phức tạp và phát trực tuyến. Điều này khiến cho việc xử lý dữ liệu lớn diễn ra nhanh chóng hơn nhiều lần bình thường.

Spark giúp đơn giản hóa nhiệm vụ đầy thách thức và tính toán chuyên sâu để xử lý khối lượng lớn dữ liệu lưu trữ hoặc thời gian thực, cả có cấu trúc và không có cấu trúc, tích hợp liền mạch các khả năng phức tạp có liên quan như học máy và thuật toán đồ thị. Spark mang quá trình xử lý dữ liệu lớn đến với công chúng.

Mặc khác, nền tảng Apache Spark đang được sử dụng rộng rãi trong việc phân tích và xử lý dữ liệu. Nắm bắt xu hướng và sự phát triển nhanh chóng của nền tảng Apache Spark nên tôi đã quyết định chọn đề tài “Phân tích và xử lý dữ liệu song song trên nền tảng Apache Spark” làm đề tài nghiên cứu của tôi.