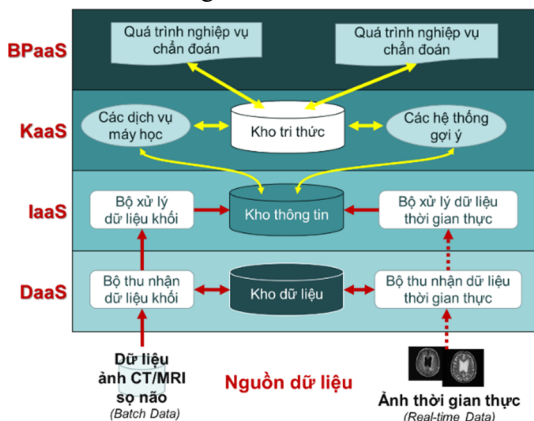


#### 4 ĐỀ XUẤT KIẾN TRÚC CHO HỆ THỐNG QUẢN LÝ TRI THỨC HƯỚNG DẪN LIỆU LỚN TRÊN ẢNH Y KHOA TRONG CHẨN ĐOÁN XUẤT HUYẾT NÃO

##### a. Kiến trúc tổng quan

Từ những vấn đề đã trình bày, tri thức bệnh xuất huyết não cần được phát triển một cách có hệ thống bao gồm cả khai phá (explore) và khai thác (exploit) tri thức. Dựa trên phương pháp nghiên cứu khoa học thiết kế (Hevner *et al.*, 2004), chúng tôi đề xuất một kiến trúc mới hướng dịch vụ cho các hệ thống quản lý tri thức hướng dữ liệu lớn trên ảnh y khoa trong chẩn đoán xuất huyết não. Kiến trúc sẽ bao gồm các thành phần: tập các cấu trúc (set of constructs) là các kiểu quan niệm khác nhau liên quan đến đối tượng tri thức bệnh xuất huyết não; một mô hình (model) diễn tả mối quan hệ giữa các quan niệm tri thức; một phương thức (method) là tập các hoạt động hỗ trợ quá trình quản lý tri thức trong KMS và cuối cùng là một minh họa cụ thể (instantiation) cho sự vận hành của hệ thống.

Theo kiến trúc này, dữ liệu cho các KMS hướng dữ liệu lớn dựa trên ảnh y khoa ứng dụng trong chẩn đoán và giáo dục y khoa (lĩnh vực xuất huyết não) sẽ bao gồm cả dữ liệu ảnh truyền thống và dữ liệu ảnh lớn thời gian thực. Tri thức có cấu tạo từ các đối tượng tri thức được phân lớp dựa trên mức độ phát triển của chúng, có thể là dữ liệu, thông tin, tri thức và trí tuệ. Khi đó, hệ thống được xây dựng cần có các khả năng: song song hóa các hoạt động xử lý, rút trích và lựa chọn đặc trưng quan trọng trong ảnh y khoa xuất huyết não, nhận dạng; hỗ trợ truy vấn ảnh dựa trên nội dung; xếp hạng các kết quả thu được; phân loại các dạng xuất huyết não chủ yếu; hỗ trợ ra quyết định giúp chẩn đoán tự động xuất huyết não từ ảnh y khoa dữ liệu lớn. Đặc biệt, hệ thống cũng sẽ có khả năng cung cấp tri thức ở các cấp độ khác nhau dưới dạng các dịch vụ.



Hình 4: Kiến trúc tổng quan được đề xuất

Để có thể hỗ trợ cho toàn bộ quá trình phát triển tri thức bệnh xuất huyết não, kiến trúc cho hệ thống quản lý tri thức hướng dữ liệu lớn trên ảnh y khoa trong chẩn đoán xuất huyết não (Hình 4) được đề xuất dựa trên bốn hoạt động chính của quá trình quản lý tri thức: thu nhận (capture), tổ chức (organization), chuyển đổi (transfer) và ứng dụng (application) tri thức.

– **Nguồn dữ liệu** bao gồm dữ liệu khối và dữ liệu thời gian thực liên quan đến bệnh xuất huyết não, đặc biệt là các ảnh CT/MRI sọ não.

– **Data-as-a-Service (DaaS)** là tầng thấp nhất có các thành phần thực hiện thu thập, lưu trữ hình ảnh xuất huyết não và các thông tin liên quan dưới dạng dữ liệu khối (batch data) và dữ liệu thời gian thực (real-time streaming data). DaaS có các dịch vụ cung cấp dữ liệu (hình ảnh thô).

– **Information-as-a-Service (IaaS)** là nơi dữ liệu khối và dữ liệu thời gian thực được đưa qua các bộ xử lý tương ứng, biến đổi thành thông tin bệnh xuất huyết não. Những thông tin này sẽ được cung cấp đến người dùng thông qua Dịch vụ thông tin (information services).

– **Knowledge-as-a-Service (KaaS)** là nơi sẽ sinh ra các tri thức bệnh xuất huyết não từ kho thông tin của tầng IaaS thông qua các hoạt động máy học hay hệ thống gợi ý. Tầng này sẽ gồm cơ sở tri thức phân tán, cấu trúc tri thức, chủ thích, các chức năng được đóng gói và công bố dưới dạng các dịch vụ tri thức (knowledge services).

– **Business Process-as-a-Service (BPaaS)** là tầng cao nhất, cho phép người dùng hình dung và phân tích kết quả truy vấn của họ. BPaaS cung cấp các dịch vụ đến ứng dụng bằng cách kết hợp chúng với các quá trình nghiệp vụ chẩn đoán xuất huyết não.

##### b. Giải pháp triển khai cụ thể các tầng

**Tầng nguồn dữ liệu:** Dữ liệu khối sẽ được cung cấp dưới dạng các tập tin hoặc cơ sở dữ liệu ảnh CT/MRI sọ não được lưu trữ tại các bệnh viện cùng với các meta-data đính kèm (thông tin bệnh nhân và các báo cáo y tế có liên quan). Dữ liệu thời gian thực cũng là các ảnh CT/MRI sọ não nhưng được gửi đến hệ thống liên tục trong quá trình khám và điều trị.

**Tầng DaaS:** HDFS (Hadoop Distributed File System) được biết đến là một hệ thống tập tin phân tán có khả năng chịu lỗi cao của Apache Hadoop. Đây cũng là một framework chiếm ưu thế trong lĩnh vực xử lý dữ liệu lớn với cơ sở hạ tầng lớn đang được triển khai và sử dụng trong nhiều lĩnh vực ứng dụng khác nhau. Thiết kế của HDFS đáng tin cậy cho việc lưu trữ các tập dữ liệu lớn cũng như truyền chúng đến các ứng dụng người dùng với băng thông