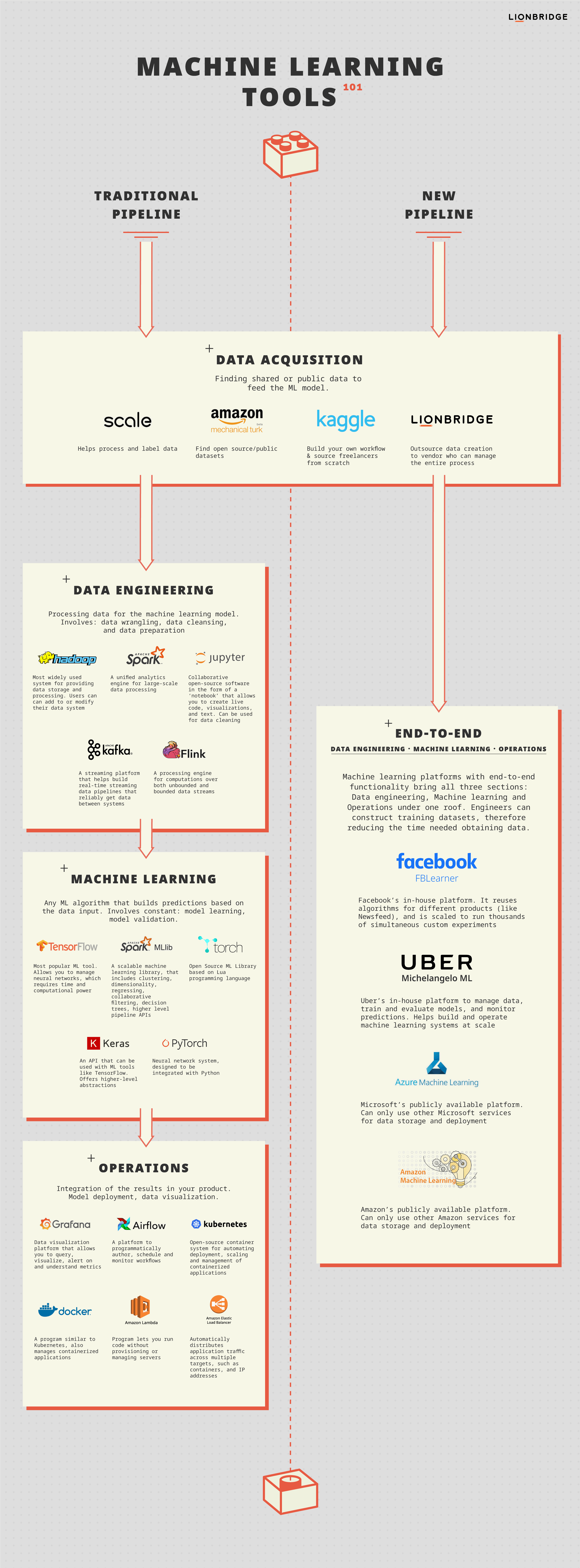
**BÁO CÁO NGHIÊN CỨU KHOA HỌC: TÌM HIỂU THƯ VIỆN “TENSORFLOW” VÀ “FLOWER”**

***Task: Tìm hiểu 2 thư viện TFF, Flower và tìm dataset tương ứng domain***

*Hình 1. Tensorflow trong vũ trụ Machine Learning*

1. ***Tensorflow***

I.1) Định nghĩa:

Nền tảng Machine Learning mã nguồn mở này hỗ trợ bạn xây dựng hiệu quả các mô hình học máy. TensorFlow hỗ trợ mọi giai đoạn của quy trình, từ chuẩn bị dữ liệu cho đến chạy các mô hình. TensorFlow cung cấp cho bạn quyền truy cập vào một số công cụ và thư viện cho cả học máy và học sâu bằng nhiều ngôn ngữ. Python thường là lựa chọn phù hợp để sử dụng TensorFlow, nhưng nó cũng hỗ trợ Java, C++, JavaScript, Go, Swift và C.

I.2) TensorFlow thường được sử dụng ở đâu?

Google đã phát triển TensorFlow để giúp việc xây dựng các ứng dụng học máy trở nên dễ tiếp cận trong nhiều ngành công nghiệp khác nhau. Đây là một công cụ hữu ích để làm việc với học máy truyền thống. Bạn cũng có thể sử dụng nó để xây dựng các mô hình học sâu và đào tạo các mạng nơ-ron phức tạp, dữ liệu lớn.

Một số ngành công nghiệp cụ thể sử dụng TensorFlow bao gồm công nghệ thông tin, an ninh mạng, thương mại điện tử, phương tiện truyền thông xã hội và chăm sóc sức khỏe. Các doanh nghiệp xây dựng các ứng dụng bằng TensorFlow để thực hiện các tác vụ như phát hiện video, nhận dạng văn bản, học chuyển giao, xử lý hình ảnh và tự động phản hồi email, với các thuật toán TensorFlow cung cấp khả năng tạo ra nhiều mô hình khác nhau.

I.3) Cách Tensorflow hoạt động:

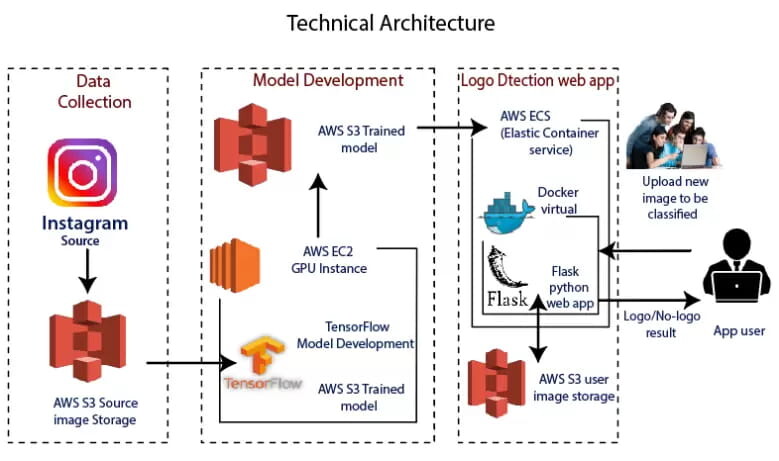
TensorFlow cho phép các lập trình viên tạo ra dataflow graph, cấu trúc mô tả làm thế nào dữ liệu có thể di chuyển qua 1 biểu đồ, hay 1 sê-ri các node đang xử lý. Mỗi node trong đồ thị đại diện 1 operation toán học, và mỗi kết nối hay edge giữa các node là 1 mảng dữ liệu đa chiều, hay còn được gọi là ‘tensor’.

TensorFlow cung cấp tất cả những điều này cho [lập trình viên theo phương thức của ngôn ngữ Python](https://topdev.vn/viec-lam-it/python-kt34). Vì Python khá dễ học và làm việc, ngoài ra còn cung cấp nhiều cách tiện lợi để ta hiểu được làm thế nào các high-level abstractions có thể kết hợp cùng nhau. Node và tensor trong TensorFlow là các đối tượng Python, và các ứng dụng TensorFlow bản thân chúng cũng là các ứng dụng Python.

Các operation toán học thực sự thì không được thi hành bằng Python. Các thư viện biến đổi có sẵn thông qua TensorFlow được viết bằng các binary C++ hiệu suất cao. Python chỉ điều hướng lưu lượng giữa các phần và cung cấp các high-level abstraction lập trình để nối chúng lại với nhau.

TensorFlow cho phép nhà phát triển tạo biểu đồ luồng dữ liệu, mô tả cách dữ liệu di chuyển qua biểu đồ hoặc các nút xử lý. Mỗi nút trong biểu đồ đại diện cho một phép toán và các kết nối giữa chúng là mảng dữ liệu đa chiều hoặc tensor.

TensorFlow có thể chạy trên nhiều nền tảng khác nhau như máy cục bộ, đám mây, iOS, Android, CPU, hoặc GPU. Nhìn chung, quy trình huấn luyện mô hình máy học của TensorFlow như sau:

* **Xây dựng đồ thị tính toán**: bằng cách mô tả biến (variables) và phép tính (operations) trong mô hình.
* **Xác định hàm mất mát**: để đo lường khoảng chênh lệch giữa kết quả dự đoán của mô hình và giá trị thực tế.
* **Tối ưu hóa mô hình**: để tìm ra giá trị tham số tối ưu nhằm giảm thiểu hàm mất mát.
* **Huấn luyện dữ liệu mô hình**: được đưa và để điều chỉnh và cập nhật giá trị tham số.
* **Đánh giá mô hình**: sử dụng dữ liệu kiểm tra để đánh giá hiệu suất tổng thể của mô hình.
* **Sử dụng mô hình**: Mô hình đã đào tạo được áp dụng để dự đoán và phân loại dữ liệu mới.

*Hình 2. Sơ đồ kiến trúc kỹ thuật của Tensorflow*

I.4) Lợi ích từ TensorFlow

Lợi ích dễ thấy nhưng quan trọng nhất mà TensorFlow cung cấp cho việc lập trình machine learning chính là abstraction. Thay vì phải đối phó với những tình huống rườm rà từ việc thực hiện triển khai các thuật toán, hay tìm ra cách hợp lý để chuyển output của 1 chức năng sang input của 1 chức năng khác, giờ đây bạn có thể tập trung vào phần logic tổng thể của 1 ứng dụng hơn. TensorFlow sẽ chăm sóc phần còn lại thay cho bạn.

Ngoài ra TensorFlow còn ung cấp các tiện ích bổ sung cho các lập trình viên cần debug cũng như giúp bạn tự suy xét các ứng dụng TensorFlow. Chế độ [eager execution](https://www.tensorflow.org/programmers_guide/eager) cho phép bạn đánh giá và sửa đổi từng operation của biểu đồ 1 cách riêng biệt và minh bạch, thay vì phải dựng toàn bộ biểu đồ dưới dạng 1 đối tượng độc lập vốn khá mơ hồ hay phải đánh giá chung tổng thể. Cuối cùng, 1 tính năng khá độc đáo của TensorFlow là [TensorBoard](https://www.tensorflow.org/programmers_guide/summaries_and_tensorboard). TensorBoard cho phép bạn quan sát 1 cách trực quan những gì TensorFlow đang làm.

TensorFlow còn có nhiều cải tiến từ sự hậu thuẫn từ các ekíp thương mại hạng A tại Google. Google không những tiếp lửa cho tiến độ nhanh chóng cho sự phát triển đằng sau dự án, mà còn tạo ra nhiều phục vụ độc đáo xung quanh TensorFlow để nó dễ dàng deploy và sử dụng: như silicon TPU mình đã nói ở trên để tăng tốc hiệu suất đám mây Google, 1 online hub cho việc chia sẻ các model được tạo với framework, sự hiện diện của [in-browser](https://js.tensorflow.org/) và gần gũi với mobile của framework, và nhiều hơn thế nữa…

Lưu ý: Trong 1 số công việc training, vài chi tiết về việc triển khai của TensorFlow làm cho nó khó có thể quyết định được hoàn toàn kết quả training model. Đôi khi 1 model được train trên 1 hệ thống này sẽ có thay đổi 1 chút so với 1 model được train trên hệ thống khác, ngay cả khi chúng được cung cấp dữ liệu như nhau. Các nguyên nhân cho điều này cũng xê xích hay 1 số hành vi khi không được xác định khi sử dụng GPU. Điều này nói rằng, các vấn đề đó có thể giải quyết được, và đôi ngũ của TensorFlow cũng đang xem xét việc kiểm soát nhiều hơn để ảnh hưởng đến tính quyết định trong quy trình làm việc.

I.5) Các Component của TensorFlow

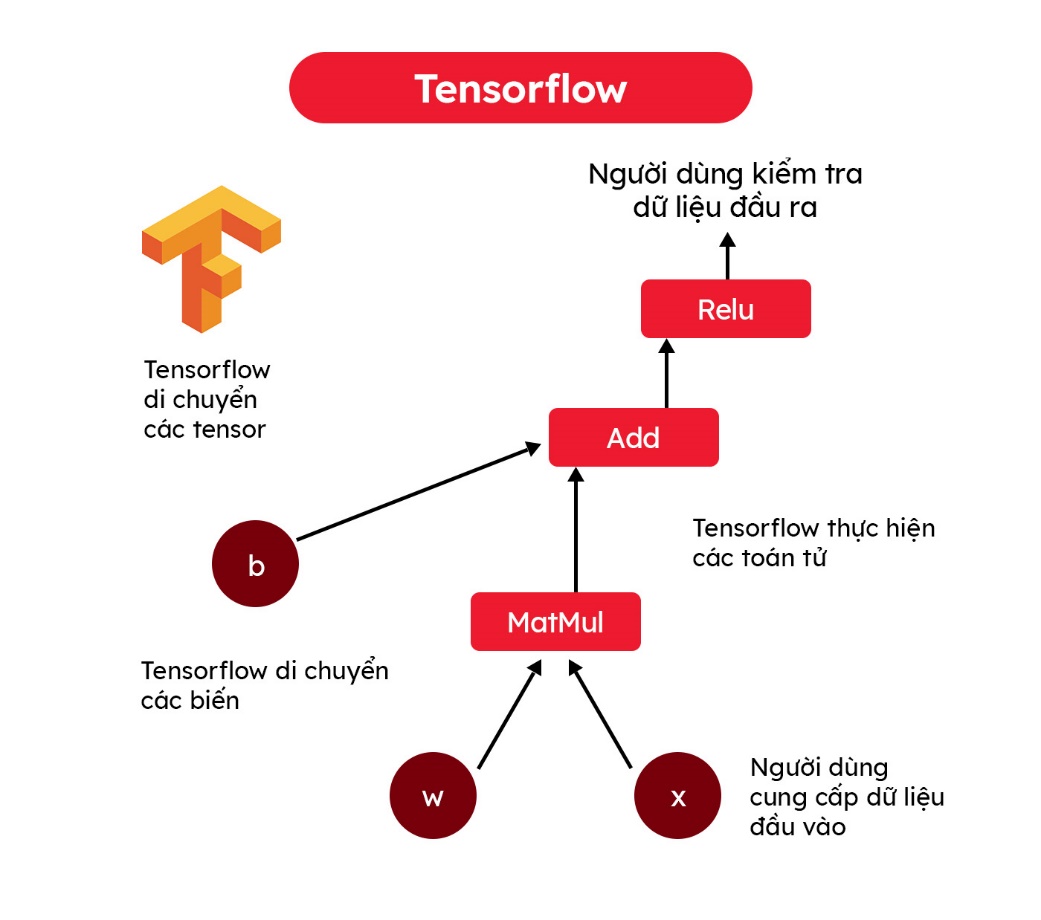
**Tensor**

A number symbols in a row

Description automatically generated with medium confidenceTên của TensorFlow được đưa ra trực tiếp là nhờ vào framework cốt lõi của nó: Tensor. Trong TensorFlow, tất cả các tính toán đều liên quan tới các tensor. 1 tensor là 1 vector hay ma trận của n-chiều không gian đại diện cho tất cả loại dữ liệu. Tất cả giá trị trong 1 tensor chứa đựng loại dữ liệu giống hệt nhau với 1 shape đã biết (hoặc đã biết 1 phần). Shape của dữ liệu chính là chiều của ma trận hay mảng.

*Hình 3.Ma trận đặc biệt trong Machine Learning*

1 tensor có thể được bắt nguồn từ dữ liệu input hay kết quả của 1 tính toán. Trong TensorFlow, tất cả các hoạt động được tiến hành bên trong 1 graph – biểu đồ. Biểu đồ là 1 tập hợp tính toán được diễn ra liên tiếp. Mỗi operation được gọi là 1 op node  (operation node) và được kết nối với nhau.

Biểu đồ phát thảo các op và kết nối giữa các node. Tuy nhiên, nó không hiển thị các giá trị. Phần edge của các node chính là tensor, 1 cách để nhập operation với dữ liệu.

Hinh Sơ đồ flow của các Tensor trong framework trong Tensorflow

**Graph**

TensorFlow sử dụng framework dạng biểu đồ. Biểu đồ tập hợp và mô tả tất cả các chuỗi tính toán được thực hiện trong quá trình training. Biểu đồ cũng mang rất nhiều lợi thế:  
– Nó được làm ra để chạy trên nhiều CPU hay GPU, ngay cả các hệ điều hành trên thiết bị điện thoại.

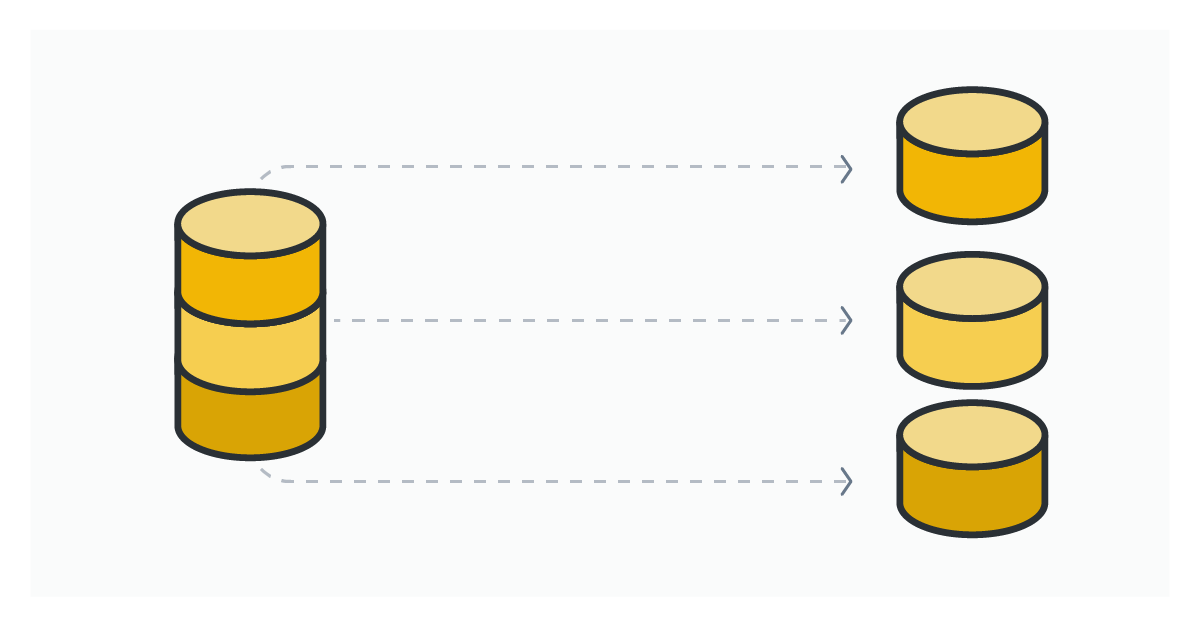
– Tính di động của biểu đồ cho phép bảo toàn các tính toán để bạn sử dụng ngay hay sau đó. Biểu đồ có thể được lưu lại để thực thi trong tương lai.

– Tất cả tính toán trong biểu đồ được thực hiện bằng cách kết nối các tensor lại với nhau. 1 tensor có 1 node và 1 edge. Node mang operation toán học và sản xuất các output ở đầu cuối. Các edge giải thích mối quan hệ input/output giữa các node.

II.1) Định nghĩa

Flower Datasets là một thư viện Python mới cho phép người dùng dễ dàng tải xuống và phân vùng các tập dữ liệu cho Federated Learning, Federated Analytics và Federated Evaluation. Điều này giúp bạn làm việc hiệu quả hơn và giải pháp của bạn rõ ràng và dễ sử dụng với tất cả các khuôn khổ ML phổ biến.

II.2) Cách sử dụng

Flower Datasets xử lý tất cả sự phức tạp này cho người dùng. Nó cho phép người dùng dễ dàng tải xuống một tập dữ liệu và phân vùng nó trong khi chỉ định tất cả các tham số. Tất cả chỉ gói gọn trong 3 dòng mã.

Nó tạo ra một giải pháp rõ ràng, dễ tái tạo và tích hợp liền mạch với các khuôn khổ khác nhau. Bất kể người dùng viết một bài nghiên cứu hay muốn thử nghiệm quy trình phát triển của mình, người dùng có thể tiết kiệm thời gian và không phải lo lắng về các chi tiết triển khai vì các lược đồ phân vùng được sử dụng có độ tin cậy cao và đã được một số nhà nghiên cứu và kỹ sư giỏi nhất trong lĩnh vực này kiểm tra và đánh giá cẩn thận.

III. Một số dataset liên quan :

<https://www.tensorflow.org/datasets/catalog/colorectal_histology>

<https://www.tensorflow.org/datasets/catalog/patch_camelyon>

<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>

<https://wiki.cancerimagingarchive.net/display/Public/Imaging+Clinical+Trials>

<https://www.cancerimagingarchive.net/collection/lidc-idri/>

Link 1: <https://www.coursera.org/articles/what-is-tensorflow?utm_source=gg&utm_medium=sem&utm_campaign=b2c_apac_career-academy_coursera_ftcof_professional-certificates_arte_aug-24_dr_geo-set-2-multi-audience_pmax_gads_lg-all&utm_content=b2c&campaignid=21573875733&adgroupid=&device=c&keyword=&matchtype=&network=x&devicemodel=&adpostion=&creativeid=&hide_mobile_promo&gad_source=1&gclid=Cj0KCQiAlsy5BhDeARIsABRc6ZuXenOvPO4VVUFYB9MEZSjzMoRU2pckDeuHymtkZ77qh417IHtrjOsaAnJ1EALw_wcB>

Link 2: <https://topdev.vn/blog/tensorflow-la-gi/?utm_source=google&utm_medium=cpc&utm_campaign=pmax-branding&utm_content=performance&gad_source=1&gclid=Cj0KCQiAlsy5BhDeARIsABRc6ZuBb8YSgh6LO_-nR6snGnmEjoB1Nz52cvN1bYKfU5ggVDDnmP-VAOoaAh5yEALw_wcB>

Link 3:<https://flower.ai/blog/2023-11-23-announcing-flower-datasets/>

Link 4: <https://gitiho.com/blog/tensorflow-la-gi.html>