**SỞ GIÁO DỤC VÀ ĐẠO TẠO THÀNH PHỐ ĐÀ NẴNG**

**TRƯỜNG THPT CHUYÊN LÊ QUÝ ĐÔN**

A picture containing text, clipart, aircraft

Description automatically generated

**CUỘC THI KHOA HỌC KĨ THUẬT CẤP TRƯỜNG**

**NĂM HỌC: 2021 – 2022**

**Tên đề tài:**

**VACIRIS – CÔNG NGHỆ NHẬN DIỆN KHUÔN MẶT NHỮNG NGƯỜI ĐÃ TIÊM VẮC-XIN HỖ TRỢ PHÒNG CHỐNG DỊCH BỆNH**

**Mã số đề tài: 16**

**Lĩnh vực dự thi: PHẦN MỀM HỆ THỐNG**

**MỤC LỤC**

**Đà Nẵng, ngày 16 tháng 11 năm 2021**

**Tóm tắt nội dung dự án3**

**1. Giới thiệu và tổng quan về vấn đề nghiên cứu4**

1.1. Vấn đề nghiên cứu4

1.2. Mục đích nghiên cứu4

1.3. Mục tiêu nghiên cứu4

**2. Giả thuyết khoa học và câu hỏi nghiên cứu5**

**3. Phương pháp nghiên cứu6**

3.1. Đối tượng nghiên cứu6

3.2. Quá trình thu thập thông tin6

3.3. Kết quả dự kiến6

**4. Kết quả nghiên cứu7-8-9**

**5. Phân tích kết quả và thảo luận giải pháp10**

5.1. So sánh giữa các mô hình nhận diện khuôn mặt10

5.1.1. Bộ lọc nhận diện mặt10

5.1.1. Mô hình nhận diện và ứng dụng các phép đo11-12

5.2. Bàn luận giải pháp12

**6. Kết luận13**

**7. Tài liệu tham khảo13**

**TÓM TẮT NỘI DUNG DỰ ÁN:**

Hiện nay, COVID-19 vẫn đang có những diễn biến rất phức tạp và khó lường trên toàn thế giới nói chung và Việt Nam nói riêng. Với số ca nhiễm tăng lên liên tục theo từng ngày, virus có thể dần tiến hóa tạo các biến thể mới, tạo ra nhiều mối an nguy cho đời sống người dân. Chính vì vậy, chúng tôi đã đề ra dự án nhằm tập trung vào việc thu thập thông tin, nghiên cứu và tạo ra một ứng dụng có khả năng giảm thiểu sự bùng dịch tại nơi sinh sống thông qua nhận diện sinh trắc học.

Chúng tôi được sự trợ giúp từ nhiều đối tượng trên tinh thần tự nguyện nhằm thu thập và lưu trữ thông tin trên hệ thống, phục vụ quá trình nghiên cứu và thực nghiệm. Trên tinh thần đó chúng tôi sẽ thử nghiệm và truy xuất các thông tin thông qua phương pháp nhận diện sinh trắc học.

**Đóng góp mới của dự án mới này chính là tận dụng công nghệ nhận diện khuôn mặt để truy xuất thông tin về vắc-xin người đó đã tiêm, hoặc là các F0 đã khỏi bệnh, nhằm hỗ trợ các lực lượng chức năng dễ dàng kiểm soát việc ra vào vùng dịch cũng như số lượng người dân chưa tiêm đủ vắc xin, đồng thời giúp truy vết ra các F1, F2 trong cộng đồng.**

**1. GIỚI THIỆU VÀ TỔNG QUAN VỀ VẤN ĐỀ NGHIÊN CỨU**

***1.1.* Vấn đề nghiên cứu**

COVID-19 có lẽ là đại dịch không quá xa lạ trong thời điểm hiện tại, bởi đến bây giờ hậu quả của nó vẫn đang tiếp tục diễn ra từng giờ, từng phút, từng giây. Số ca nhiễm hàng ngày ở thế giới cũng như ở Việt Nam liên tục tăng lên với những mức kỉ lục đáng kinh ngạc. Theo những nghiên cứu từ các trang báo, COVID-19 đã lập rất nhiều con số kỉ lục đáng lo ngại nhất trong thời điểm dịch bùng phát. Ngay tại Việt Nam vào ngày 15/8/2021 đã phá kỉ lục về số ca nhiễm hàng ngày khi ghi nhận thêm 9.716 ca nhiễm mới và sắp vượt con số 6.000 ca tử vong 1. Tính đến thời điểm ngày 10/11/2021, thế giới đã ghi nhận tổng cộng 251.400.157 ca mắc COVID-19 và trong đó đã có 5.076.462 ca tử vong 2.

Chính vì lẽ đó, một trong những giải pháp hàng đầu chính là việc triển khai rộng rãi các chương trình tiêm chủng vắc-xin trên cả thế giới, nhằm tạo ra miễn dịch cộng đồng.

Dựa trên giải pháp đó, chúng tôi đã quyết định nghiên cứu và tạo ra một ứng dụng có thể quản lí được thông tin vắc-xin người dân để thuận tiện cho việc kiểm soát ra vào vùng dịch và truy xuất thông tin các F0, F1, từ đó có khả năng giảm sự lây lan trong cộng đồng.

***1.2. Mục đích nghiên cứu***

Như đã đề cập ở trên, sau khi phân tích tổng quan tình hình nghiên cứu trong nước, chúng tôi muốn tạo ra một ứng dụng có thể giúp quản lí thông tin vắc-xin của người dân hoặc là các F0 đã khỏi bệnh để từ đó giúp các lực lượng chức năng dễ dàng hơn trong việc kiểm soát truy vết các F1, F2 trong cộng đồng.

***1.3.*** ***Mục tiêu nghiên* cứu**

- Hoàn thành phần mềm, tối ưu giao diện người dùng và thuật toán nhận diện khuôn mặt.

- Thử nghiệm thực tế với 2 giai đoạn:

+ Đạt tính chính xác tối thiểu 93% với số lượng người tham gia ít (cụ thể là 20 bạn học sinh lớp 10A5).

+ Đạt tính chính xác tổi thiểu là 90% đối với số lượng người tham gia nhiều.**2. GIẢ THUYẾT KHOA HỌC VÀ MỤC ĐÍCH NGHIÊN CỨU**

Hiện nay, đại dịch COVID-19 vẫn đang có những diễn biến rất phức tạp và khó lường trên thế giới nói chung và Việt Nam nói riêng. Với số ca nhiễm tăng lên liên tục theo từng ngày, virus sẽ dần tiến hoá nên các biến thể mới, tạo ra nhiều mối nguy cho đời sống người dân 3.

Sách giáo khoa sinh học sau đại dịch COVID-19 có thể đưa virus corona trở thành ví dụ điển hình nhất minh họa cho thuyết tiến hóa của Darwin. Từ chủng virus ban đầu ở Vũ Hán cho tới biến thể Alpha, Beta và giờ là Delta, sự biến đổi của chủng virus này có thể giải thích phần lớn những gì đã xảy ra và vẫn đang tiếp diễn 3.

Thông qua quá trình chọn lọc tự nhiên, các virus mới, hay biến thể của chúng sẽ xuất hiện. Các đột biến nào thuận lợi cho sự tồn tại của chúng sẽ được giữ lại và di truyền sang cho thế hệ virus tiếp theo 3.

Chúng ta biết vắc xin COVID-19 được thiết kế để chống lại chủng SARS-CoV-2 ban đầu ở Vũ Hán. Nó có tác dụng chống lại biến thể Delta nhưng ít hiệu quả hơn. Thuyết tiến hóa đã dự đoán điều này, các biến thể của virus có thể tránh được vắc xin. Và nó sẽ truyền lại khả năng này cho con cháu của mình vì đây là một khả năng có lợi thế về mặt tiến hóa 3.

Nói một cách khác, khi vắc xin mất hiệu lực, virus sẽ gây bệnh nặng hơn với những người chưa tiêm. Họ có thể dễ tử vong hơn 3.

Chính vì lẽ đó, một trong những giải pháp hàng đầu chính là việc triển khai rộng rãi các chương trình tiêm chủng vắc xin trên cả thế giới, nhằm tạo ra miễn dịch cộng đồng.

Mục đích của dự án lần này là để quản lý số người đã tiêm vắc-xin hoặc đã đạt được cơ chế miễn dịch hoàn toàn thông qua các chốt kiểm soát. Giả sử một người dân đang trong vùng dịch, nếu người dân chưa có thông tin về số mũi tiêm trên hệ thống, hệ thống sẽ không cho phép người đó ra khỏi khu vực và ngược lại. Điều đó tạo cho người dân tâm lý không chủ quan với dịch bệnh và thúc đẩy việc tự giác đi tiêm vắc-xin nhằm đạt được mong muốn tự do đi lại trong vùng dịch.

**3. PHƯƠNG PHÁP NGHIÊN CỨU**

***3.1.* Chọn đối tượng nghiên cứu**

- Tiêu chí: Người dân, học sinh, sinh viên (những người đã tiêm 1 mũi vắc xin trở lên hoặc các F0 đã khỏi bệnh) có nguyện vọng, mong muốn đóng góp, giúp đỡ cho dự án.

- Đạo đức nghiên cứu: Dựa theo sự tự nguyện của đối tượng, chúng tôi tôn trọng tính riêng tư gắn với thông tin cá nhân của người dân và đảm bảo thông tin cá nhân của người dân được xử lí một cách bảo mật, an toàn và chỉ được sử dụng cho mục đích nghiên cứu.

- Phạm vi: Các trường học, khu dân cư trong khu vực TP.Đà Nẵng.

- Mục tiêu về số lượng: Khoảng 1000 người.

- Mục tiêu về chất lượng: Trong quá trình truy xuất thông tin thông qua nhận diện sinh trắc học, thông tin dữ liệu người dùng được cung cấp phải đảm bảo tính trung thực, đúng đắn.

***3.2.* Quá trình thu thập dữ liệu :**

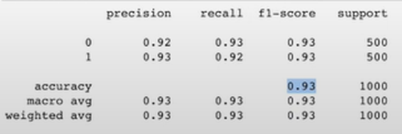
Sẽ bắt đầu thông qua các cơ sở tiêm chủng, các trường học và các khu dân cư có số dân ít trên địa bàn TP. Đà Nẵng và lưu trữ thông tin người tiêm lên hệ thống quản trị xác nhận đã tiêm ngừa vắc xin COVID-19.

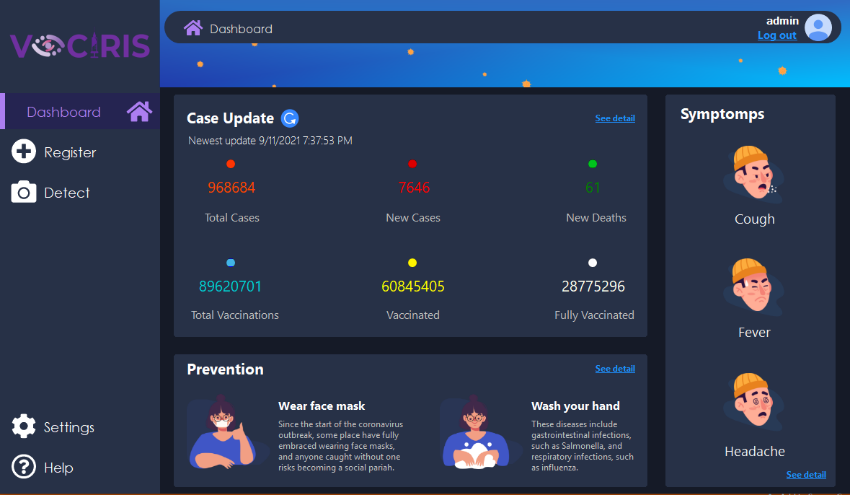
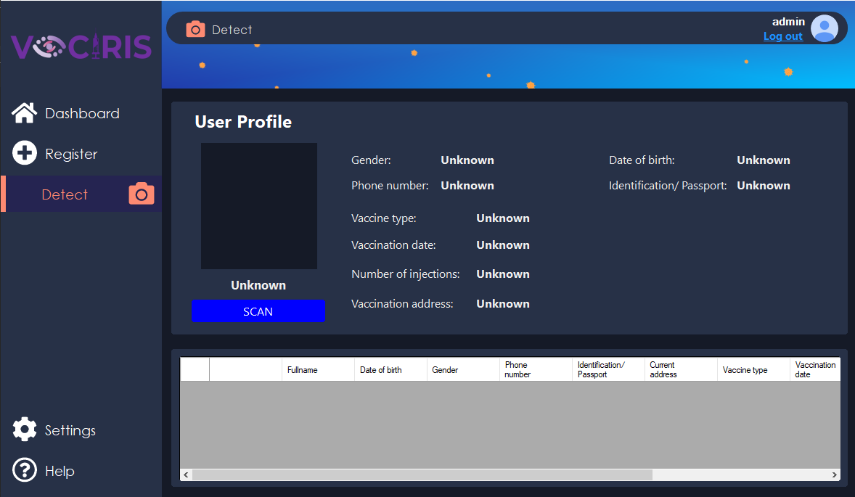
***3.3.* Kết quả dự kiến:**

Qua những những dữ liệu nghiên cứu đó, chúng tôi sử dụng công nghệ nhận diện khuôn mặt (AI) kiểm tra xem người được quét đã được tiêm chủng chưa, từ đó thông báo đến các cơ quan chức năng có thẩm quyền.

4. KẾT QUẢ NGHIÊN CỨU

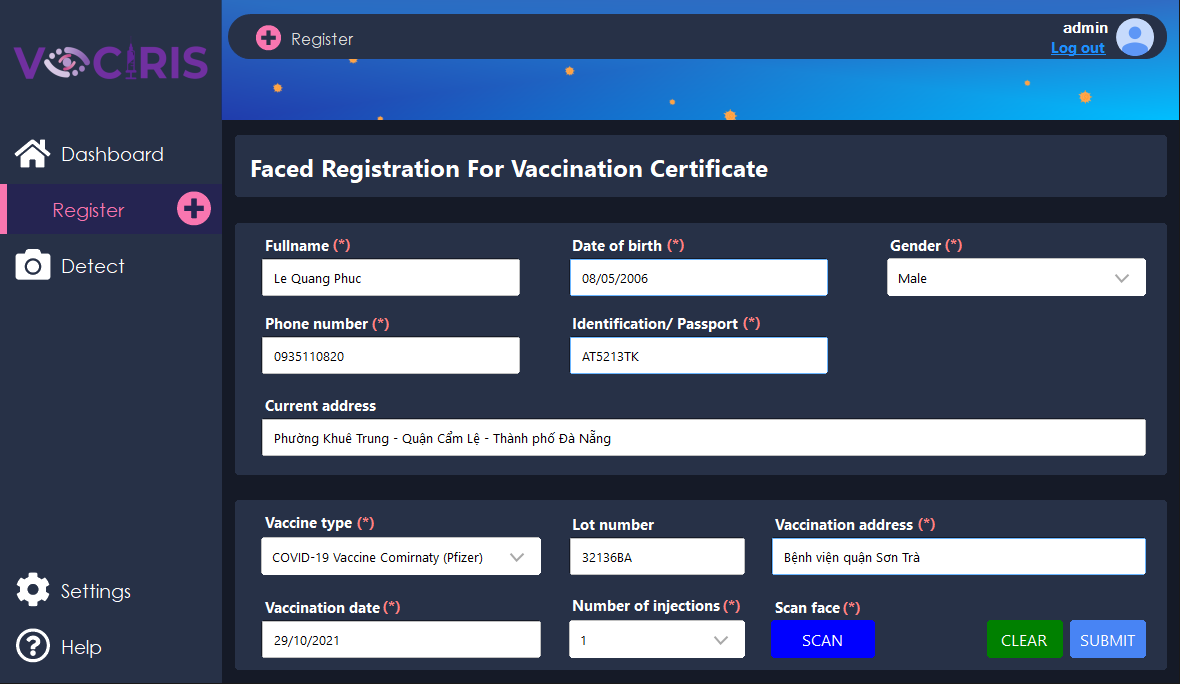
- Sau quá trình nghiên cứu và thực nghiệm trên giai đoạn 1, các kết quả ứng dụng trả về với độ chính xác lên đến 99,41% trong môi trường đầy đủ ánh sáng và đạt tối thiểu 93% trong môi trường thiếu sáng 4.



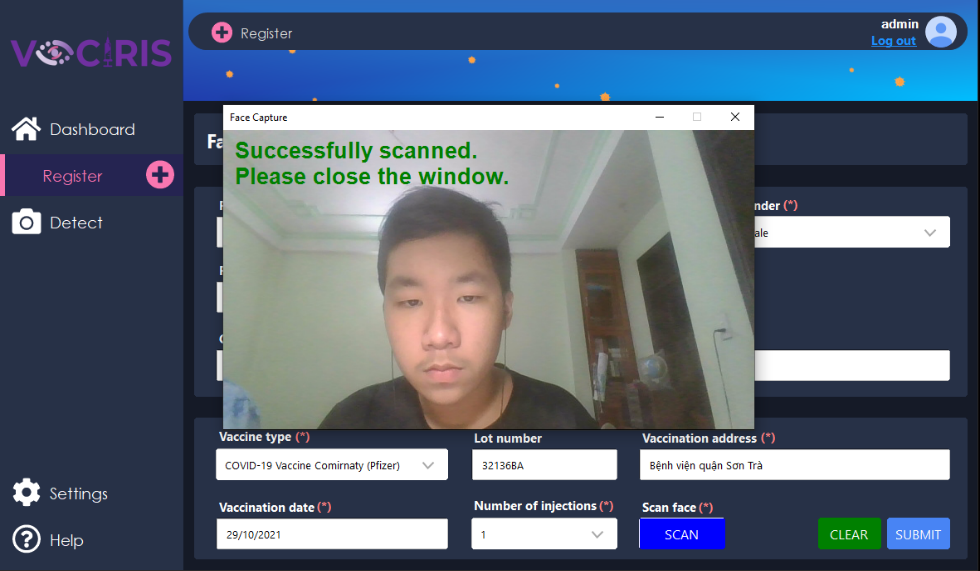
- Một số hình ảnh về giao diện chính của VACIRIS:

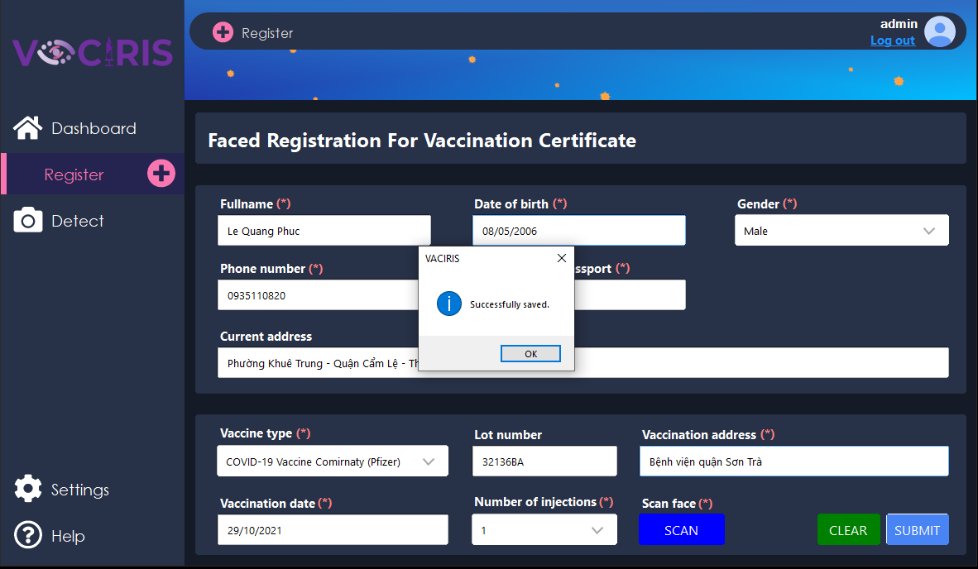
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  | |  |  |
|  |  |  |  |  |
|  |  |  | | |
|  |  |  |  |  |
|  |  | | |  |

 - Quy trình hoạt động của VACIRIS:

+ Đăng ký khuôn mặt:

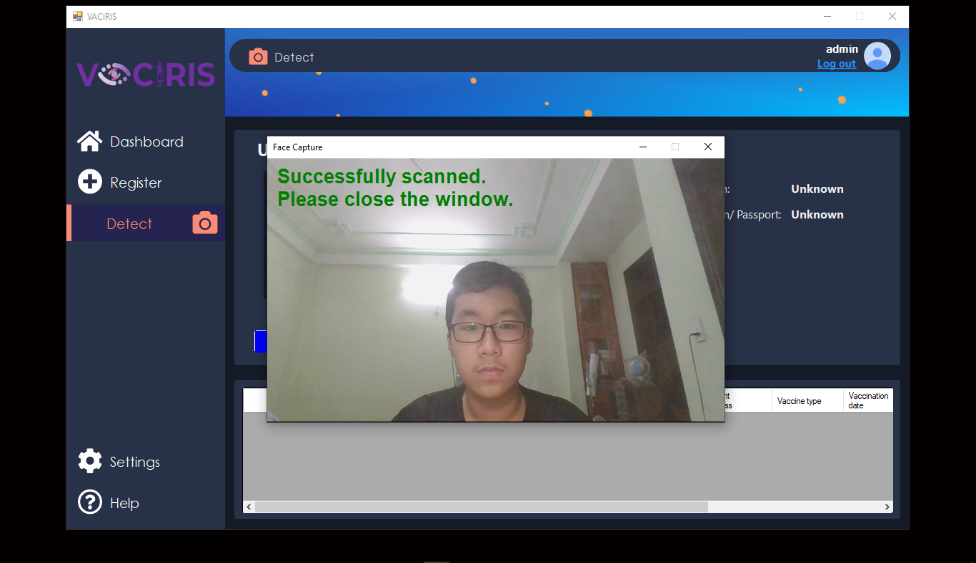
**B1:** Điền thông tin cá nhân của người dùng

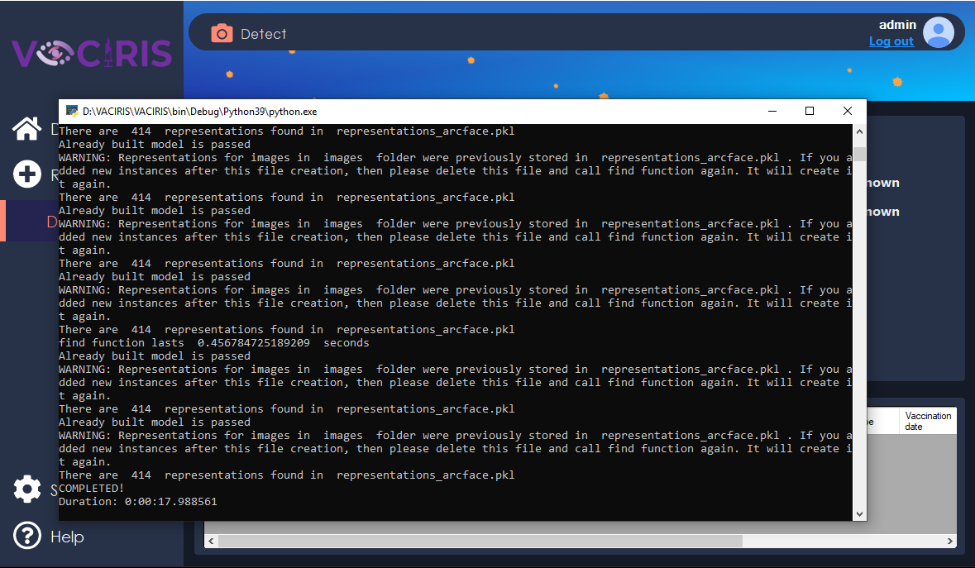
**B2:** Quét khuôn mặt làm cơ sở dữ liệu



**B3:** Lưu thông tin vào hệ thống

+ Nhận diện khuôn mặt:

**B1:** Quét khuôn mặt

**B2:** Phần mềm thực hiện tính toán



**B3:** Phần mềm trả về kết quả thông tin người dùng

**5. PHÂN TÍCH KẾT QUẢ VÀ THẢO LUẬN GIẢI PHÁP**

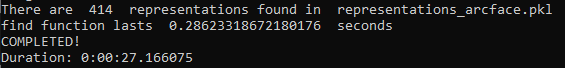
***5.1.* So sánh giữa các mô hình nhận diện khuôn mặt**

***5.1.1.* Bộ lọc nhận diện**



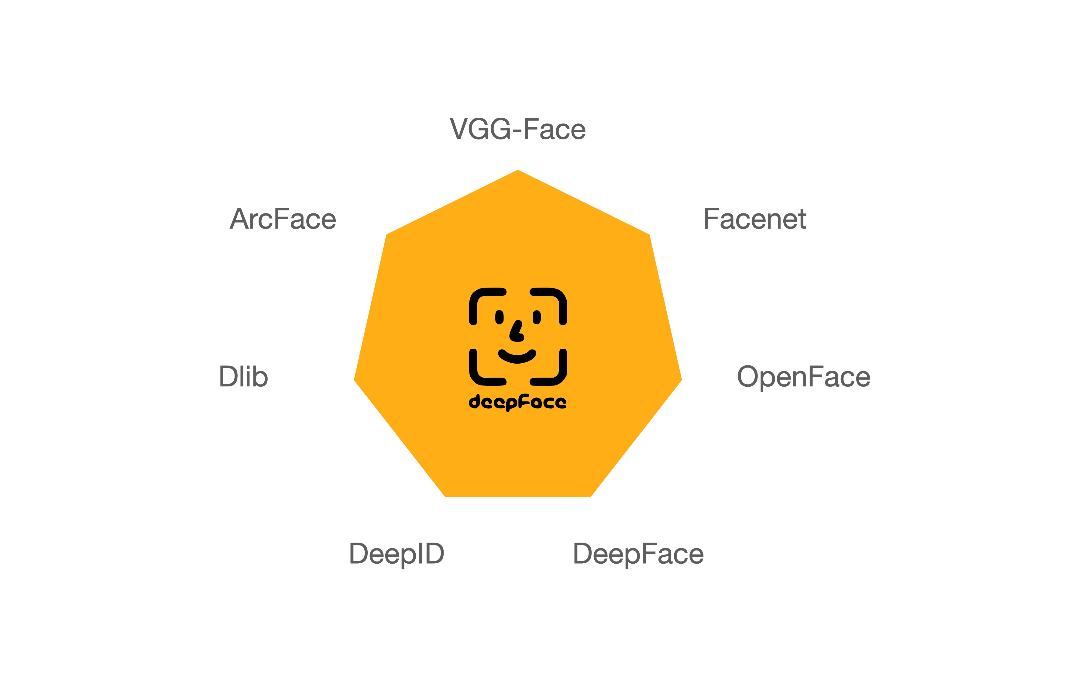
**So sánh tính chính xác của các bộ lọc** 5

Có một số bộ lọc nhận diện khuôn mặt phổ biến như [Haar Cascade](https://viblo.asia/p/haar-cascade-la-gi-luan-ve-mot-ky-thuat-chuyen-dung-de-nhan-biet-cac-khuon-mat-trong-anh-E375zamdlGW) và [SSD](https://medium.com/featurepreneur/object-detection-using-single-shot-multibox-detection-ssd-and-opencvs-deep-neural-network-dnn-d983e9d52652) của [OpenCV](https://en.wikipedia.org/wiki/OpenCV), [HOG và MMOD của Dlib](https://viblo.asia/p/xay-dung-he-thong-kiem-soat-nhan-dang-khuon-mat-voi-opencv-dlib-va-deep-learning-4P8566ma5Y3) và [MTCNN](https://towardsdatascience.com/face-detection-using-mtcnn-a-guide-for-face-extraction-with-a-focus-on-speed-c6d59f82d49) 4. Theo kết quả quá trình thử nghiệm thực tế cho thấy, Haar Cascade có khả năng nhận diện còn tương đối thấp, còn MTCNN cho đầu ra với độ chính xác cao. Chính vì vậy, chúng tôi sử dụng Haar Cascade cho quá trình chụp ảnh làm cơ sở dữ liệu vì tốc độ xử lý nhanh chóng (~50 bức/2 giây) và đòi hỏi người dùng phải đặt khuôn mặt chuẩn để nhận diện, giúp tăng độ chính xác của dữ liệu. Ngoài ra, chúng tôi còn sử dụng MTCNN cho quá trình training nhằm tạo ra tính chính xác cao với lượng dữ liệu lớn. Tuy nhiên, quá trình training vẫn còn tương đối chậm (~30 giây/50 bức/người).

****

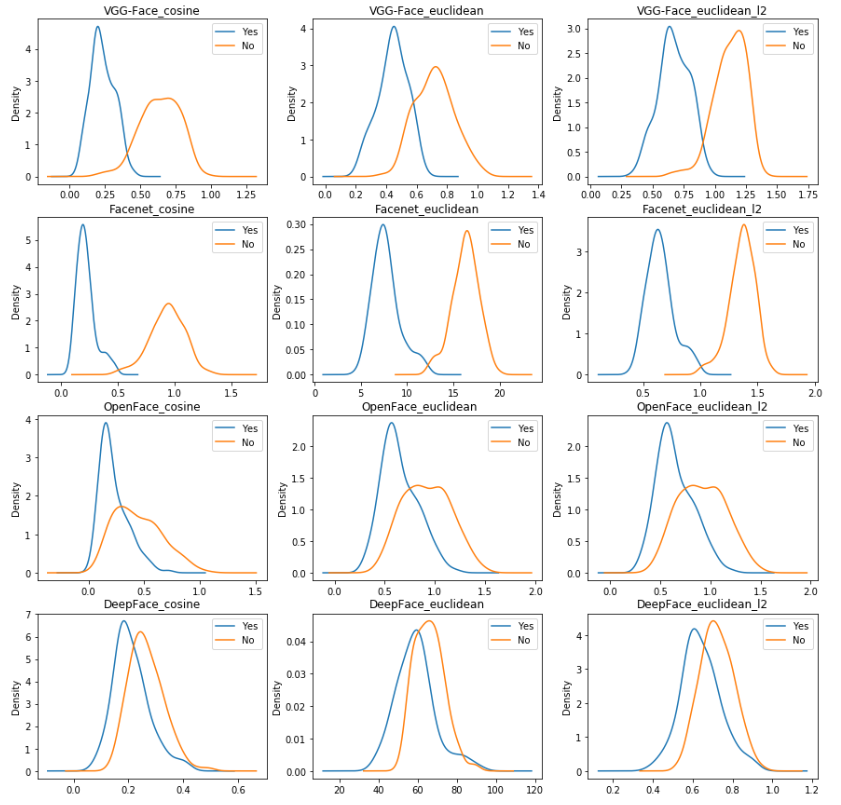
**Thời gian chạy của chương trình (50 bức ảnh)**

***5.1.2.* Mô hình nhận diện và ứng dụng các phép đo**



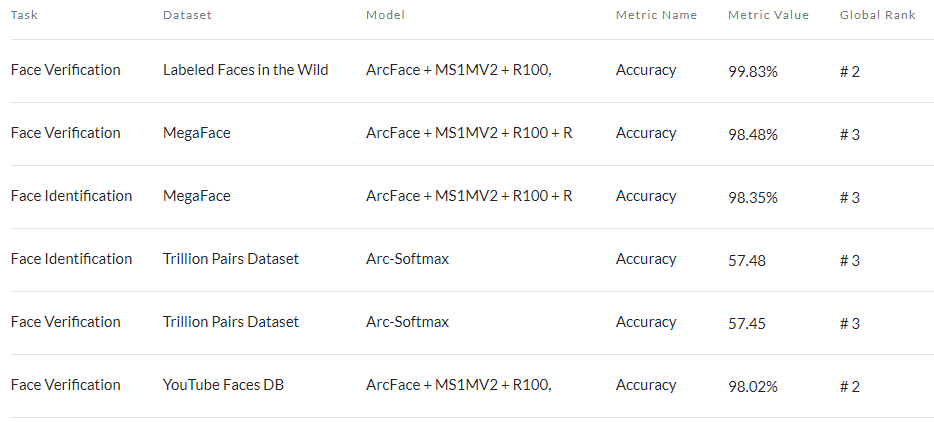
Thông qua quá trình thực nghiệm, chúng tôi nhận thấy rằng [FaceNet](https://www.section.io/engineering-education/facenet-unified-embedding-face-recognition-clustering/#:~:text=FaceNet%20is%20a%20face%20recognition,vector%20representation%20of%20any%20object.), [VGG-Face](https://exposing.ai/vgg_face/), [ArcFace](https://paperswithcode.com/method/arcface) và [Dlib](https://viblo.asia/p/xay-dung-he-thong-kiem-soat-nhan-dang-khuon-mat-voi-opencv-dlib-va-deep-learning-4P8566ma5Y3) hoạt động tốt hơn [OpenFace](https://cmusatyalab.github.io/openface/), [DeepFace](https://github.com/serengil/deepface) và [DeepID](https://sefiks.com/2020/06/16/face-recognition-with-deepid-in-keras/) 4. Chi tiết hơn, FaceNet-512d đạt 99,65%; FaceNet-128d đạt 99,2%; ArcFace đạt 99,41%; Dlib đạt 99,38%; VGG-Face đạt 98,78%; DeepID đạt được 97,05%; OpenFace đạt 93,80% trên [tập dữ liệu LFW](http://vis-www.cs.umass.edu/lfw/) 4.

Các mô hình nhận dạng khuôn mặt là mạng nơ-ron học sâu và chúng có nhiệm vụ biểu diễn khuôn mặt dưới dạng vector 4. Hiểu đơn giản rằng, ảnh 2 khuôn mặt gần nhau thì 2 vector tương ứng cũng phải có khoảng cách gần nhau, ngược lại ảnh 2 khuôn mặt khác nhau thì 2 vector tương ứng cũng phải cách xa nhau. Có nhiều phép đo cho phép chúng ta làm việc đó, và điển hình nhất chính là [tính tương tự Cosin](https://sefiks.com/2018/08/13/cosine-similarity-in-machine-learning/), [khoảng cách Euclid và dạng L2](https://vi.wikipedia.org/wiki/Kho%E1%BA%A3ng_c%C3%A1ch_Euclid) 4. Thực nghiệm cho thấy rằng dạng Euclid L2 ổn định hơn so với Cosin và dạng Euclid thông thường 4.



**Bảng so sánh tính chính xác giữa các mô hình và phép đo** 6

Vì yêu cầu độ chính xác cao cùng thời gian chạy chương trình thấp trên lượng dữ liệu lớn, chúng tôi đã quyết định sử dụng mô hình ArcFace cùng dạng Euclid L2 và cho ra kết quả khả quan như đã đề cập bên trên và số liệu bên dưới.



**Kết quả từ các dataset khác nhau** 7

***5.2.* Bàn luận giải pháp**

Dự kiến trong tương lai gần, chúng tôi sẽ thay đổi bộ lọc và mô hình mới, đồng thời phối hợp nhiều mô hình khác nhau để cho ra kết quả tối ưu hơn về mặt thời gian lẫn tính chính xác.

**6. KẾT LUẬN**

VACIRIS là sản phẩm với giao diện đơn giản, tối ưu dùng để nhận diện khuôn mặt từ đó truy xuất thông tin của người dùng lấy từ cơ sở dữ liệu có sẵn với số lần tiêm vắc-xin hay các F0 đã khỏi bệnh. Bước đầu phần mềm đã thành công trong việc thiết kế giao diện và nhận diện khuôn mặt để truy xuất thông tin. Phần mềm sẽ tiếp tục được hoàn thành với dự kiến mang đến sự thuận tiện trong việc kiểm soát vùng dịch và hỗ trợ truy vết các F1, F2 cho các lực lượng chức năng. Bên cạnh đó, chúng tôi cũng mong muốn giảm đi tỉ lệ những ca lây nhiễm cộng đồng và tạo ra sự yên tâm trong lòng người dân trước đại dịch COVID-19.

**7. TÀI LIỆU THAM KHẢO:**

1. Trọng Nghĩa. Covid-19 : Việt Nam phá kỷ lục về số ca nhiễm hàng ngày, số tử vong sắp vượt 6.000 ca. https://www.rfi.fr/vi/việt-nam/20210815-covid-19-việt-nam-phá-kỷ-lục-về-số-ca-nhiễm-hàng-ngày-trong-lúc-số-tử-vong-sắp-vượt-6-000-ca (2021).

2. Worldometers. https://www.worldometers.info/coronavirus/.

3. THANH LONG. Biến thể Delta chưa phải là ‘trùm cuối’ của đại dịch, đây là cách mà nó còn có thể tiến hóa. https://genk.vn/bien-the-delta-chua-phai-la-trum-cuoi-cua-dai-dich-day-la-cach-ma-no-con-co-the-tien-hoa-20210824054351551.chn (2021).

4. Sefik Ilkin Serengil. DeepFace. https://github.com/serengil/deepface (2019).

5. Sefik Ilkin Serengil. Real-Time Face Detection Battle: OpenCV Haar Cascade, SSD, Dlib HoG and MTCNN. https://www.youtube.com/watch?v=GZ2p2hj2H5k (2020).

6. Sefik Ilkin Serengil. Mastering Face Recognition with Ensemble Learning. https://sefiks.com/2020/06/03/mastering-face-recognition-with-ensemble-learning/ (2020).

7. Deng, J. *et al.* ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* (2021) doi:10.1109/TPAMI.2021.3087709.

**Cấu hình máy tính được sử dụng trong hệ thống: HP Pavilion 15-cc048TX**

* Intel Core i7-7500U (2.7Ghz Upto 3.5Ghz, 4MB Cache)
* RAM 8GB DDR4
* 1TB HDD SATA 5400rpm
* Nvidia Geforce GT 940MX 2GB GDDR3
* 15.6" HD (1366 x 768)
* HP Wide Vision HD Camera