TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
Tập 20, Số 1 (2023): 165-179
Website: https://journal.hcmue.edu.vn
ISSN: 2734-9918

HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
Vol. 20, No. 1 (2023): 165-179
https://doi.org/10.54607/hcmue.js.20.1.3652(2023)

Bài báo nghiên cứu

MỘT PHƯƠNG PHÁP TIẾP CẬN NHẬN DẠNG KHUÔN MẶT NGƯỜI BẰNG HUẤN LUYỆN HỌC MÁY

***Đàm Minh Lịnh*, Nguyễn Hoàng Thành**

Học viện Công nghệ Bưu chính Viễn thông Cơ sở tại Thành phố Hồ Chí Minh, Việt Nam

**Tác giả liên hệ: Đàm Minh Lịnh – Email: linhdm.tg@ptithcm.edu.vn*

Ngày nhận bài: 24-10-2022; ngày nhận bài sửa: 29-11-2022; ngày duyệt đăng: 09-12-2022

TÓM TẮT

Nhận dạng khuôn mặt là một kĩ thuật công nghệ sinh trắc học ánh xạ các đặc điểm khuôn mặt người. Tập dữ liệu ảnh Facial Expression Recognition 2013 (FER-2013) gồm có bảy loại biểu cảm khác nhau của khuôn mặt người, được tác giả dùng làm bộ dữ liệu huấn luyện trong nghiên cứu này. Hiện tại, tầm quan trọng của bảo mật hệ thống là hết sức cấp thiết, vì vậy triển khai ứng dụng xác thực nhận dạng khuôn mặt người để đăng nhập vào hệ thống, xác thực trên điện thoại thông minh, chấm công, đeo khẩu trang. Chúng tôi đề xuất mô hình học máy, học sâu với nhiều phương pháp huấn luyện khác nhau kết hợp với tập dữ liệu FER-2013, được mở rộng các định dạng ảnh kích thước (32x32, 48x48, 64x64, 72x72) nhằm mở rộng mục tiêu hướng nghiên cứu và tiến hành thực nghiệm với các mô hình LDA, NB, KNN, DT, SVM. Sau đó, đánh giá sự hiệu quả của từng mô hình các tiêu chí Accuracy, Precision và F1-Score. Kết quả thực nghiệm của chúng tôi đã đóng góp được ba vấn đề chính: một là, mở rộng định dạng bộ dataset với kích thước đa dạng hơn để làm nền tảng cho kết quả nghiên cứu; hai là mô phỏng các mô hình thuật toán khác nhau trong quá trình huấn luyện nhằm đánh giá và so sánh về các tiêu chí ở trên; ba là đề xuất mô hình học sâu CNN được đánh giá hiệu quả.

Từ khóa: thị giác máy tính; Học sâu; nhận diện khuôn mặt; FER-2013; mạng nơ ron

1. Giới thiệu

Nhận dạng khuôn mặt người là hình thức phát hiện dùng các thiết bị máy móc liên quan đến việc thu thập thông tin dữ liệu ảnh, sau đó xử lí ảnh được thông qua các mô hình học máy, học sâu so sánh với chiết xuất đặc trưng từ bộ dữ liệu ảnh đã huấn luyện, từ đó sẽ đưa ra kết quả nhận dạng và phát hiện đối tượng ảnh.

Woodrow W. Bledsoe, Helen Chan và Charles Bisson (Bledsoe, 1964; Bledsoe, 1966) đồng tác giả nghiên cứu nhận dạng khuôn mặt người từ năm 1964 đến 1966, đã nghiên cứu lập trình máy tính nhận dạng khuôn mặt người với bộ cơ sở dữ liệu lớn, vấn đề đặt ra là làm sao để so sánh sự trùng khớp giữa một ảnh và bộ dữ liệu lớn. Tuy nhiên, một ứng dụng nhận

Cite this article as: Dam Minh Linh, & Nguyen Hoang Thanh (2023). An approach to human face recognition by machine learning training. *Ho Chi Minh City University of Education Journal of Science, 20*(1), 165-179.

dạng khuôn mặt đầy đủ chức năng đã được Kanade thực hiện vào năm 1977. Các nghiên cứu về nhận dạng khuôn mặt hai chiều (2D) đã được nghiên cứu chuyên sâu. Các nghiên cứu về khuôn mặt ba chiều (3D) bắt đầu được thực hiện sau những năm 2000.

Nhiều mô hình học sâu Convolutional Neural Network (CNN) (Chauhan, Kumar, & Joshi, 2018; Bhairnallykar, Prajapati, Rajbhar, & Mujawar, 2020) đã được thực nghiệm ở nhiều bộ dữ liệu khác nhau như MNIST, CIFAR-10, cho kết quả chính xác. MNIST là tập dữ liệu gồm 70.000 hình ảnh, trong đó 60.000 hình ảnh dành cho việc huấn luyện và 10.000 dành cho thử nghiệm. Kích thước của mỗi hình ảnh là 28x28 pixel và có 10 nhãn phân lớp từ 0-9.

Bài viết này gồm có 4 phần, các phần còn lại được trình bày như sau: Phần 2, trình bày về đối tượng và phương pháp nghiên cứu: Mô tả tập dữ liệu chuẩn; phương pháp đánh giá; đề xuất mô hình học sâu CNN. Tiếp theo, phần 3 là kết quả và thảo luận của nghiên cứu. Cuối cùng là phần 4 kết luận cho nghiên cứu này.

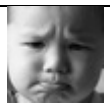
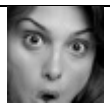
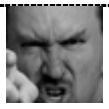
2. Đối tượng và phương pháp nghiên cứu

2.1. Đối tượng nghiên cứu

2.1.1. Mô tả tập dữ liệu chuẩn

Để đánh giá được nhận dạng khuôn mặt với độ chính xác cao, cần có bộ dữ liệu chuẩn và được nhiều công trình nghiên cứu sử dụng. Chất lượng bộ dữ liệu rất quan trọng sẽ làm ảnh hưởng kết quả cho quá trình thực nghiệm và đánh giá các phương pháp nhận dạng khuôn mặt.

Bộ dữ liệu nhận dạng biểu cảm khuôn mặt năm 2013 (SAMBARE , 2020): FER-2013 là tập dữ liệu được giới thiệu tại hội nghị quốc tế về học máy (ICML) vào năm 2013 do tác giả I. J. Goodfellow và D. Erhan và các tác giả đồng nghiên cứu khác giới thiệu (Goodellow, et al., 2015). Trong tập dữ liệu này, mỗi khuôn mặt đã được phân loại dựa trên 7 loại cảm xúc (Vui mừng, tức giận, thất vọng, sợ hãi, ghê tởm, ngạc nhiên, bình thường) khác nhau, mỗi hình ảnh có kích thước 48x48 pixel, các cảm xúc được mô tả trong Hình 1.

Vui mừng	tức giận	thất vọng	sợ hãi	ghê tởm	ngạc nhiên	bình thường
						
						

Hình 1. *Dữ liệu ảnh mô tả 7 loại cảm xúc lấy từ bộ dữ liệu FER-2013*

Tập dữ liệu FER-2013 gồm có 35.887 ảnh, trong đó: Tập dữ liệu dùng để train là 28.709 ảnh và dung cho việc test là 7178 ảnh, cho 7 loại biểu cảm khác nhau của khuôn mặt người, mô tả ở Bảng 1.

Bảng 1. *Bộ dữ liệu Facial Expression Recognition 2013 (FER-2013)*

Hình thái/ biểu cảm khuôn mặt	Tập dữ liệu gồm có 2 phần	
	Test	**Train**
Angry - tức giận	958	3995
Disgust - ghê tởm	111	436
Fear – nỗi sợ	1024	4097
Happy - vui mừng	1774	7215
Neutral - bình thường	1233	4965
Sad - thất vọng	1247	4830
Surprise - ngạc nhiên	831	3171
Tổng số	**7178**	**28709**

2.1.2. Phương pháp đánh giá

Trong nghiên cứu này, để đánh giá hiệu suất các mô hình thực nghiệm thì cần dùng công thức như là độ chính xác (Accuracy), Precision và F1-score. Ở Bảng 2, sử dụng ma trận nhầm lẫn có các thuộc tính dương tính thật (TP), âm tính thật (TN), dương tính giả (FP) và âm tính giả (FN) (Huynh & Nguyen, 2021).

Bảng 2. *Ma trận nhầm lẫn*

	Dự đoán – nhận dạng đúng	**Dự đoán - Bình thường**
Thực tế	TP	FN
Bình thường	FP	TN

Độ chính xác (Accuracy) – là mức độ gần của các phép đo với một giá trị cụ thể, số lượng dữ liệu được phân loại chính xác trên tổng số dự đoán. Độ chính xác có thể không phải là thước đo tốt nếu tập dữ liệu không được cân bằng (cả hai lớp âm và dương có số lượng dữ liệu khác nhau). Công thức tính độ chính xác được định nghĩa trong công thức (1).

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Tỉ lệ cảnh báo giả (False Alarm Rate - FAR) – còn được gọi là False Positive Rate (tỉ lệ dương tính giả). Thước đo này được tính theo công thức (2). Tỉ lệ lí tưởng cho thước đo này càng thấp càng tốt, tức là số phân loại nhầm một phân loại bình thường sang dự đoán nhận dạng đúng (FP) càng thấp càng tốt.

$$\text{FAR}=\frac{FP}{FP+TN} \quad (2)$$

Độ chính xác phép đo (Precision) – là mức độ gần của các phép đo, có giá trị gần với 1 khi kết quả là một tập phân loại tốt. Precision là 1 chỉ khi tử số và mẫu số bằng nhau (TP = TP + FP), điều này cũng có nghĩa là FP bằng 0. Khi FP tăng giá trị dẫn đến mẫu số lớn hơn tử số và giá trị chính xác giảm. Công thức tính Precision được định nghĩa trong

công thức (3).

$$\text{Precision}=\frac{TP}{TP+FP} \quad (3)$$

Tỉ lệ phát hiện (Detection Rate – DR hay Recall) – Giá trị DR càng gần với 1 sẽ cho một phân loại tốt. DR là 1 chỉ khi tử số và mẫu số bằng nhau (TP = TP + FN), điều này cũng có nghĩa là FN bằng 0. Khi FN tăng giá trị dẫn đến mẫu số lớn hơn tử số và giá trị DR giảm. Chỉ số này nhằm đánh giá mức độ tổng quát hóa mô hình tìm được và được xác định theo công thức (4).

$$\text{Detection Rate}=\frac{TP}{TP+FN} \quad (4)$$

Nếu 2 tiêu chí Precision và DR đều tốt, nghĩa là một trong hai giá trị FP và FN phải gần bằng 0 càng tốt. Cần có một tham số đo có tính đến cả Precision và DR, đó chính là F1-Score, công thức (5).

F1-Score được gọi là một trung bình điều hòa của các tiêu chí Precision và DR. Nó có xu hướng lấy giá trị gần với giá trị nào nhỏ hơn giữa 2 giá trị Precision và DR và đồng thời nó có giá trị lớn nếu cả 2 giá trị Precision và DR đều lớn. So với độ chính xác (Accuracy), F1-Score phù hợp hơn để đánh giá hiệu suất nhận dạng của các mẫu dữ liệu không cân bằng.

$$\text{F1-Score}=\frac{2(\text{Precision}\times DR)}{\text{Precision}+DR}=\frac{2TP}{2TP+FP+FN} \quad (5)$$

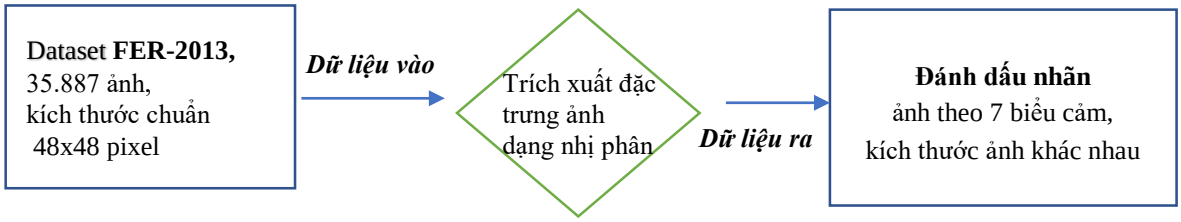
2.2. Phương pháp nghiên cứu

Trong phần này, trình bày liên quan đến ba vấn đề đóng góp của tác giả được nêu trong phần tóm tắc: chuyển đổi bộ dữ liệu gốc FER-2013; mô phỏng các thuật toán được hỗ trợ thư viện chính là sklearn, sau đó so sánh kết quả; đề xuất mô hình học sâu CNN được đánh giá hiệu quả, cải thiện hơn đối với các nghiên cứu liên quan khác.

2.2.1. Chuyển đổi dữ liệu gốc FER-2013

Từ bộ dữ liệu chuẩn **FER-2013,** có 35.887 ảnh với kích thước chuẩn 48x48 pixel. Sau đó, thực hiện chuyển đổi thành kích thước khác nhau 32x32, 64x64, 72x72 pixel mô tả qua trong hình 2 và sơ đồ bên dưới, sử dụng thư viện cv2, OS được kết hợp chiết xuất đặc trưng dạng ảnh nhị phân lưu vào ma trận 3 chiều, sau đó gán nhãn cho từng loại theo 7 loại biểu cảm khác nhau của khuôn mặt, cuối cùng sử dụng hàm xử lí trong thư viện cv2 để tăng hoặc giảm ảnh theo kích thước đề xuất.

Sơ đồ tổng quát cho việc mở rộng các định dạng ảnh khác nhau



Chiết xuất đặc trưng bộ dữ liệu ảnh FER-2013 với các kích thước (pixel) khác nhau (32x32, 64x64, 72x72) từ bộ dữ liệu gốc.

```
test_imgs = []
test_imgs_lbp = []
test_lables = []
for i in range(len(emotion_list)) :
    emotion_path = os.path.join(test_path, emotion_list[i])
    for img_name in os.listdir(emotion_path):
        img = plt.imread(os.path.join(emotion_path, img_name))
        img_lbp = skimage.feature.local_binary_pattern(img, 8, 1.0, method='var')
        if np.isnan(img_lbp).sum() == 0:
            #img1 = cv2.imread(str(img))
            resized_img = cv2.resize(img,(img_size,img_size))
            test_imgs.append(resized_img)
            test_imgs_lbp.append(img_lbp)
            test_lables.append(i)
```

Hình 2. *Mô hình hóa cho chiết xuất đặc trưng bộ dữ liệu ảnh FER-2013 với các kích thước khác nhau*

2.2.2. Mô hình Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis là một phương pháp kĩ thuật ứng dụng dạng bài toán giảm kích thước tiền xử lí dữ liệu cho các ứng dụng máy học và phân loại (Tharwat et al., 2017).

Thuật toán LDA

Bước 1. Đọc ảnh để huấn luyện $X\{x_1, x_2, \ldots, x_N\}$, trong đó $x_i(r \times g)$: i^{th} là ảnh thứ i được huấn luyện, r và g là kích thước của ảnh lần lượt tương ứng cho chiều cao, chiều rộng, N là tổng số ảnh được sử dụng để huấn luyện.

Bước 2. Chuyển đổi tấc cả ảnh dạng vectơ K $\{k_1, k_2, \ldots, k_M\}$, trong đó K là M x 1, $M = r \times g$

Bước 3. Tính giá trị trung bình của mỗi lớp μ_i, tổng giá trị trung bình của tất cả dữ liệu μ, *between-class variance* S_B (M × M) và *within-class variance* S_W (M × M) của X.

Bước 4. Tìm *k* vectơ riêng của S_B >0, U $\{u_1, u_2, \ldots, u_k\}$.

Bước 5. Tìm giá trị của $U^T S_w U$, loại bỏ vectơ có giá trị cao (sắp xếp), vectơ được chọn kí hiệu là V.

Bước 6. Ma trận(Ψ) LDA xác định S_B trong khoảng [>0...17] và $S_W = \emptyset$, $\Psi = UV$.

Bước 7. Dữ liệu gốc LDA là $Y = X\Psi = XUV$.

2.3. Mô hình Naïve Bayes (NB)

Naïve Bayes là một thuật toán phân loại dữ liệu. Thuật toán phân loại hiển thị vùng tốt nhất dưới giá trị đường cong (AUC). Kết quả các nghiên cứu thuật toán có độ chính xác hơn so với thuật toán Lazy-IBK, Zero-R và cây quyết định-J48 (Wibawa et al., 2019).

Ưu điểm của thuật toán Naive Bayes (Rajeswari, Juliet, & Aradhana, 2017): Dữ liệu đào tạo nhỏ, tính toán đơn giản, dễ để thực hiện, hiệu quả về thời gian.

Định lí Bayes

$$P(Q|X)= \frac{P(X|Q).P(Q)}{P(X)} \quad (6)$$

Trong đó: X là dữ liệu lớp không xác định, Q là dữ liệu lớp cụ thể, $P(Q|X)$ là xác suất Q đề cập tới X, $P(Q)$ là xác suất của giả thuyết Q, $P(X|Q)$ là xác suất X đề cập tới Q, $P(X)$ là xác suất X.

$$=P(Q)\prod_{i=1}^{n} P(X_n|Q) \quad (7)$$

Phương trình trên là một mô hình từ Naïve Bayes sẽ được sử dụng trong quá trình phân loại. Để phân loại với dữ liệu số có thể được xử lí bằng cách sử dụng hàm mật độ xác suất tiêu chuẩn.

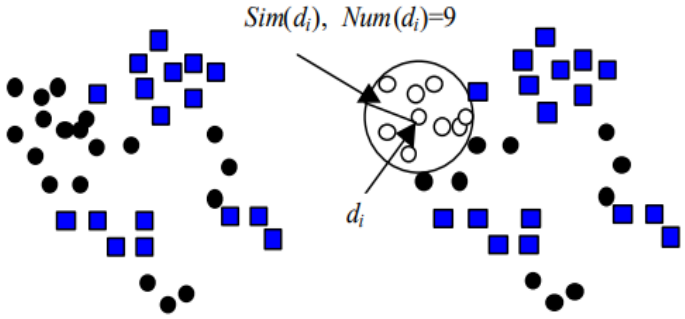
Công thức từ mật độ Gauss:

$$P\left(X_{i=}\, x_i \middle| Q=\, q_j\right) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(X_j-\mu_{ij})^2}{2\sigma^2_{ij}}} \quad (8)$$

Trong đó: P là Cơ hội, X_i là thuộc tính i, x_i là giá trị thuộc tính i, Q là lớp liên quan, q_j là lớp con của Q, μ là giá trị trung bình của tất cả các thuộc tính, σ là độ lệch chuẩn và phương sai của tất cả các thuộc tính.

2.4. Mô hình k-Nearest-Neighbours (KNN)

k-Nearest-Neighbours là thuật toán học có giám sát, áp dụng cho bài toán phân loại dữ liệu và hồi quy (Guo, Wang, Bell, Bi, & Greer, 2004) (Rajeswari, Juliet, & Aradhana, 2017).



Hình 3. *Phân loại dữ liệu trong k-Nearest-Neighbours*

Thuật toán phân loại được mô tả: Gọi M là mô hình đại diện. Trong đó <Cls (d_i), Sim (d_i), Num (d_i), Rep (d_i)> lần lượt đại diện cho nhãn lớp của d_i, độ tương đồng thấp nhất với d_i trong số các bộ dữ liệu được N_i bao phủ; nếu có nhiều hơn một vùng lân cận có cùng số lượng láng giềng tối đa, sẽ chọn một với giá trị tối thiểu của Sim (d_i).

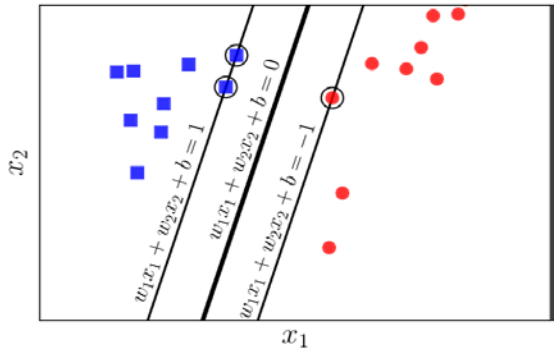
2.5. Mô hình Decision Tree (DT)

Decision trees thường được sử dụng nhiều lĩnh vực khác nhau, chẳng hạn như xử lí hình ảnh và xác định các mẫu (Taha & Mohsin, 2021). Các nút và các nhánh được cấu tạo

từ mỗi cây. Mỗi nút đại diện cho các tính năng trong một danh mục được phân loại và mỗi tập hợp con xác định một giá trị có thể được nhận bởi nút. Các loại thuật toán cây quyết định như là: Iterative Dichotomies, Classification and Regression Tree (CART), CHi-squared Automatic Interaction Detector (CHAID), Multivariate Adaptive Regression Splines (MARS), Conditional Inference Trees (CTREE).

2.6. Mô hình Support Vector Machine (SVM)

Support Vector Machine là một thuật toán học giám sát, ứng dụng cho bài toán thuộc phân loại dữ liệu, đệ quy (Srivastava & Bhambhu, 2010). Phát biểu bài toán: Áp dụng bài toán phân loại dữ liệu như Hình 4, tìm giải pháp tối ưu phải dựa vào tiêu chuẩn nào? Cặp dữ liệu của *training set* là $(x_1,y_1),(x_2,y_2),\ldots,(x_N,y_N)$ với vector $X_i \in R^d$ thể hiện *đầu vào* của một điểm dữ liệu và y_i là *nhãn* của điểm dữ liệu đó.



Hình 4. *Sự phân loại dữ liệu trong Support Vector Machine*

Phương trình phân chia giữa hai classes Hình 4

$$w^Tx+b= w_1x_1+w_2x_2+b=0 \quad (9)$$

Bài toán tối ưu trong SVM là tìm w và b sao cho *margin* này đạt giá trị lớn nhất

$$(w,b)= \arg\max_{w,b} \left\{ \min_n \frac{y_n(w^Tx+b)}{\|w\|_2} \right\} \quad (10)$$

Các mô hình huấn luyện và các tiêu chí đánh giá.

```
models.append(("DT",DecisionTreeClassifier()))
models.append(("SVM",SVC()))
```

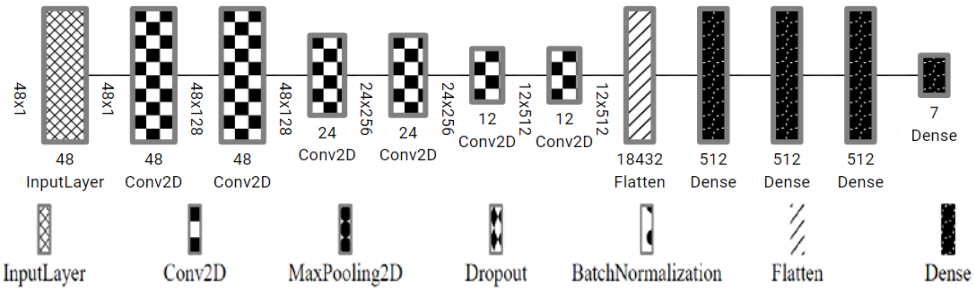
```
def f1_score(y_true, y_pred):
    true_positives = K.sum(K.round(K.clip(y_true * y_pred, 0, 1)))
    possible_positives = K.sum(K.round(K.clip(y_true, 0, 1)))
    predicted_positives = K.sum(K.round(K.clip(y_pred, 0, 1)))
    precision = true_positives / (predicted_positives + K.epsilon())
    recall = true_positives / (possible_positives + K.epsilon())
    f1_val = 2*(precision*recall)/(precision+recall+K.epsilon())
    return f1_val
```

Hình 5. *Mô hình hóa cho các phương pháp huấn luyện dữ liệu và đánh giá các tiêu chí*

2.7. Đề xuất mô hình mạng nơ ron

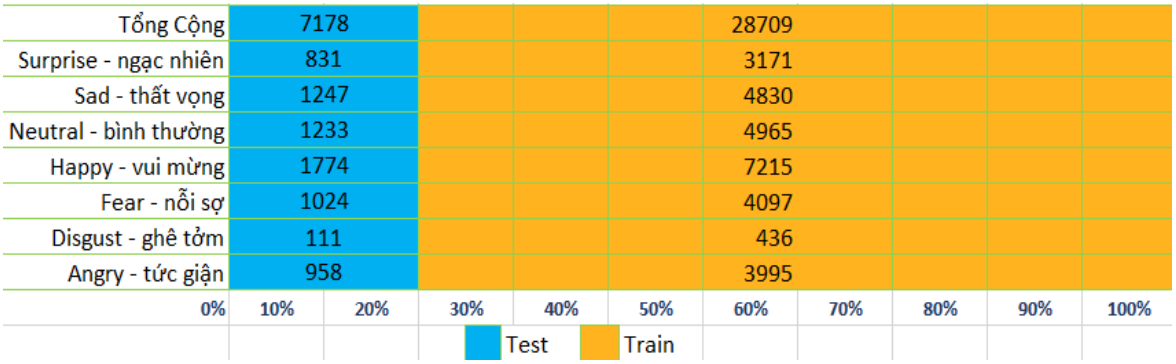
Mô hình mạng học sâu sử dụng đầu vào hình ảnh và biến đổi thông qua bộ lọc, để trích xuất các đặc trưng. Phương pháp huấn luyện khác kết hợp với tập dữ liệu FER-2013 với kích thước ảnh chuẩn 48x48 pixel, được mở rộng các định dạng ảnh kích thước 32x32, 64x64, 72x72 pixel để làm dữ liệu đầu vào của mô hình huấn luyện, như Hình 6.

Trong mô hình này, thực nghiệm trên bộ dữ liệu chuẩn, kết quả so sánh dựa vào các tiêu chí trong phần giới thiệu đánh giá các mô hình LDA, NB, KNN, DT, SVM, CNN.



Hình 6. *Mô tả đề xuất mô hình mạng nơ ron CNN*

Số lượng ảnh được phân bổ dùng để huấn luyện và kiểm tra, bao gồm có 7 loại biểu cảm khác nhau: Vui mừng, tức giận, thất vọng, sợ hãi, ghê tởm, ngạc nhiên, bình thường. như Hình 7. Trong mỗi loại biểu cảm, sẽ đưa ra 2 bộ dữ liệu dùng để train và test riêng.



Hình 7. *Mô tả số lượng dữ liệu dùng để huấn luyện và kiểm tra*

Mô hình đề xuất học sâu được sử dụng train và test với các tham số đầu vào ảnh thuộc tính **conv2d** = kích thước ảnh 72x72 pixel, đầu ra mô hình này dense_1 = 7 loại biểu cảm khác nhau, tổng số tham số được train là 75.941.095, số vòng train Epoch =100, thời gian cho mỗi epoch = 69s. Đồng thời, các tham số này được sử dụng để huấn luyện cho conv2d = kích thước ảnh 32x32 và 64x64 pixel.

```
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d (Conv2D)              (None, 72, 72, 32)        160

 conv2d_1 (Conv2D)            (None, 72, 72, 64)        8256

 batch_normalization (BatchN  (None, 72, 72, 64)        256
 ormalization)

 max_pooling2d (MaxPooling2D  (None, 36, 36, 64)        0
 )

 dropout (Dropout)            (None, 36, 36, 64)        0

 conv2d_2 (Conv2D)            (None, 36, 36, 128)       32896

 conv2d_3 (Conv2D)            (None, 35, 35, 256)       131328

 batch_normalization_1 (Batc  (None, 35, 35, 256)       1024
 hNormalization)

 max_pooling2d_1 (MaxPooling  (None, 17, 17, 256)       0
 2D)

 dropout_1 (Dropout)          (None, 17, 17, 256)       0

 flatten (Flatten)            (None, 73984)             0

 dense (Dense)                (None, 1024)              75760640

 dropout_2 (Dropout)          (None, 1024)              0

 dense_1 (Dense)              (None, 7)                 7175

=================================================================
Total params: 75,941,735
Trainable params: 75,941,095
Non-trainable params: 640
_________________________________________________________________
```

Hình 8. *Mô hình đề xuất học sâu, kích thước ảnh 72x72 pixel*

Tương tự, mô hình đề xuất học sâu được sử dụng train và test với các tham số đại diện đầu vào ảnh thuộc tính **conv2d** = kích thước ảnh 48x48 pixel, đầu ra mô hình này dense_1 = 7 loại biểu cảm khác nhau, tổng số tham số được train là 31.900.903, số vòng train Epoch =200, thời gian cho mỗi epoch = 34s. Tuy nhiên, khi train thì thời gian thực cao hơn rất nhiều, sử dụng card đồ họa NVIDIA® GeForce RTX™ 1650 **GPU** 4GB để thực hiện.

```
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d_4 (Conv2D)            (None, 48, 48, 32)        160
 dense_3 (Dense)              (None, 7)                 7175

=================================================================
Total params: 31,901,543
Trainable params: 31,900,903
Non-trainable params: 640
```

Hình 9. *Mô hình đề xuất học sâu, kích thước ảnh 48x48 pixel*

3. Kết quả và thảo luận

Trong bài thực nghiệm này, được đánh giá trên máy tính laptop Asus Rog Strix Gaming G513IH với Windows 10 Pro 20H2 cấu hình: CPU AMD Ryzen™ 7 (8 nhân, 16 luồng) up 4.2GHz - 8MB Cache, RAM 16 Gb DDR4-3200Mhz, M.2 NVMe™ PCIe® 3.0 *SSD, NVIDIA® GeForce RTX™ 1650* ***GPU*** *4GB.*

3.1. Kết quả thực nghiệm

Trong quá trình mô phỏng, đã sử dụng các thư viện như OpenCV hỗ trợ nhiều về thị giác máy tính bao gồm nhận dạng và phát hiện khuôn mặt, tìm kiếm ảnh có đặc trưng từ bộ dữ liệu lớn. Scikit-learning một thư viện hỗ trợ vấn đề bài toán phân loại, phân cụm, hồi quy,

tuyến tính và quan trọng đó là mô hình đề xuất lựa chọn. Keras hỗ trợ liên quan cho mạng nơ ron học sâu, xử lí dạng bài toán về dạng hình ảnh và văn bản. TensorFlow là một thư viện phần mềm mã nguồn mở cho hiệu suất cao tính toán linh hoạt kiến trúc nhiều nền tảng (CPU, **GPU**, TPU), hỗ trợ mạnh về vấn đề học máy và học sâu.

Thực hiện cho mô hình, tham số đề xuất ở trên với kích ảnh 72x72 pixel, kết quả thu được các mô hình LDA, NB, KNN, DT, SVM kết hợp với phương pháp đánh giá các tiêu chí: Accuracy, precision, recall, F1_score, như Hình 10 và 11.

```
Epoch 99/100
898/898 [==============================] - 69s 77ms/step - loss: 0.8520 - accuracy: 0.6879 - f1_score: 0.6686 - val_loss: 1.2
167 - val_accuracy: 0.5926 - val_f1_score: 0.5706
Epoch 100/100
898/898 [==============================] - 69s 77ms/step - loss: 0.8644 - accuracy: 0.6816 - f1_score: 0.6603 - val_loss: 1.1
862 - val_accuracy: 0.5889 - val_f1_score: 0.5705
```

Hình 10. *Kết quả dữ liệu train và test với Epoch = 100, tiêu tốn thời gian thực = 69s/ mỗi epoch với kích thước ảnh 72x72 pixel*

Sau khi thực hiện train với mô hình đề xuất bên trên và được kết hợp thư viện chính của mô hình học sâu, kết quả thu được các mô hình dựa vào tiêu chí đánh giá Accuracy 42%, F1_score = 40% của SVM là cao nhất, Hình 11.

```
========== LDA RESULT
Accuracy score:0.34
Precision score:0.31
Recall score:0.34
F1_score score:0.31
========== NB RESULT
Accuracy score:0.29
Precision score:0.28
Recall score:0.29
F1_score score:0.26
```

```
========== KNN RESULT
Accuracy score:0.30
Precision score:0.32
Recall score:0.30
F1_score score:0.30
========== DT RESULT
Accuracy score:0.29
Precision score:0.29
Recall score:0.29
F1_score score:0.29
```

```
========== SVM RESULT
Accuracy score:0.42
Precision score:0.44
Recall score:0.42
F1_score score:0.40
```

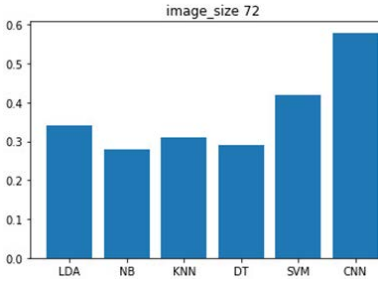
Hình 11. *Kết quả thu được các mô hình LDA, NB, KNN, DT, SVM*

Kết quả Bảng 3, mô tả quá trình thực nghiệm dữ liệu đầu vào với kích thước ảnh IMAGE-SIZE = 64x64, 32x32 pixel, mô hình đề xuất mạng nơ ron được đánh giá tỉ lệ độ chính xác cao nhất lần lượt là: **58.90**%, **62.98**%.

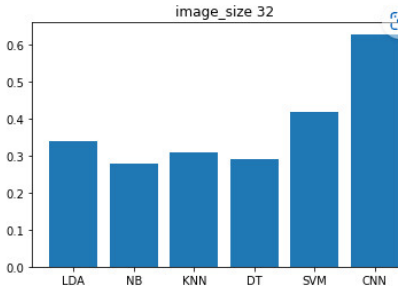
Bảng 3. *Bảng thống kê kết quả so sánh quá trình huấn luyện và kiểm tra các mô hình với các kích thước ảnh còn lại*

Kích thước ảnh	Mô hình thuật toán	Accuracy %	Precision %	F1-Score %
IMAGE-SIZE = 64x64	*LDA*	0.33	0.30	0.30
	NB	0.28	0.28	0.26
	KNN	0.31	0.33	0.30
	DT	0.29	0.29	0.29
	SVM	**0.43**	0.44	0.40
	CNN	**0.5890**	0.6181	0.5946
IMAGE-SIZE = 32x32	*LDA*	**0.34**	0.30	0.34
	NB	0.28	0.28	0.28
	KNN	0.31	0.33	0.31
	DT	0.29	0.29	0.29
	SVM	**0.42**	0.44	0.40
	CNN	**0.6298**	0.6475	0.6354

Kết quả đánh giá độ chính xác được so sánh dưới dạng biểu đồ hình cột cho 2 IMAGE-SIZE = 72x72, 32x32 pixel, được mô tả ở Hình 12, Hình 13, cao nhất là mô hình đề xuất học sâu CNN, sau đó là SVM, thấp nhất là mô hình NB và DT xấp xỉ gần bằng nhau.

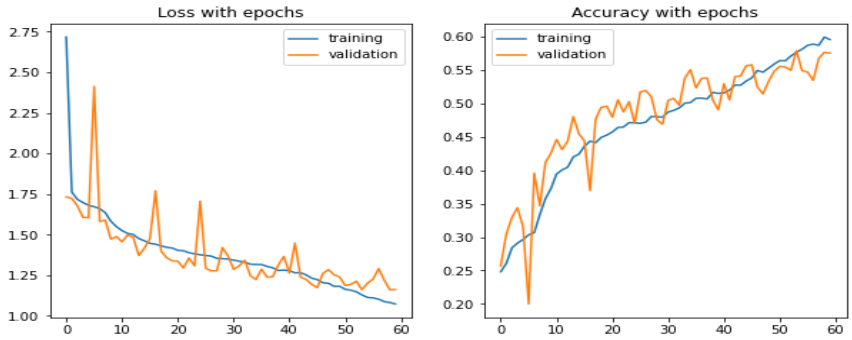


Hình 12. *Biểu đồ so sánh kết quả các thuật toán, kích thước ảnh 72x72*



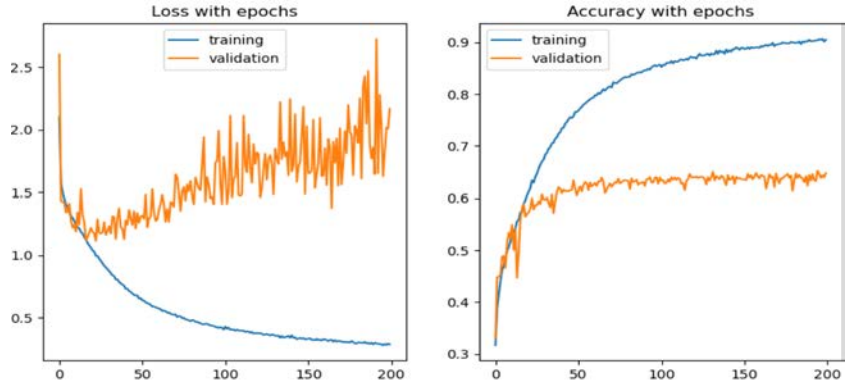
Hình 13. *Biểu đồ so sánh kết quả các thuật toán, kích thước ảnh 32x32*

Kết quả Hình 12 và Hình 13, cho 2 kích thước ảnh IMAGE-SIZE = 64x64, 48x48 pixel, đánh giá số liệu qua mỗi vòng *(epochs = 100)* với tỉ lệ *Loss và Accuracy.*



Hình 14. *Biểu đồ kết quả tỉ lệ học sâu chính xác qua mỗi vòng, kích thước ảnh 64x64*

Với mô hình đề xuất học sâu mạng nơ ron được sử dụng train và test với các tham số đầu vào ảnh thuộc tính conv2d = kích thước ảnh 48x48 pixel, đầu ra mô hình này dense_1 = 7 loại biểu cảm khác nhau, tổng số tham số được train là 31.900.903, số vòng train Epochs =200, thời gian cho mỗi epoch = 34s. Hình 15 và 16, cho kết quả tỉ lệ chính xác (Accuracy) của hai dữ liệu huấn luyện và được kiểm tra.



Hình 15. *Biểu đồ kết quả tỉ lệ học sâu chính xác qua mỗi vòng, kích thước ảnh 48x48*

```
Epoch 199/200
898/898 [==============================] - 34s 38ms/step - loss: 0.2919 - accuracy: 0.9014 - f1_score: 0.9018 - val_loss: 2.0
073 - val_accuracy: 0.6396 - val_f1_score: 0.6413
Epoch 200/200
898/898 [==============================] - 34s 38ms/step - loss: 0.2873 - accuracy: 0.9051 - f1_score: 0.9059 - val_loss: 2.1
659 - val_accuracy: 0.6478 - val_f1_score: 0.6486
```

Hình 16. *Kết quả dữ liệu train và test với Epoch = 200/200, tiêu tốn thời gian thực = 34s/ mỗi epoch với kích thước ảnh 48x48*

Kết quả khi được **test** với Epoch =200 vòng, thì tỉ lệ học chính xác (accuracy) = **64.77**%, Hình 17.

```
score = model.evaluate(test_set, steps=test_set.n//test_set.batch_size)
```

```
224/224 [==============================] - 2s 10ms/step - loss: 2.1675 - accuracy: 0.6477 - f1_score:
0.6493
```

```
print('Test loss: ', score[0])
print('Test accuracy: ', score[1])
```

```
Test loss:  2.167459726333618
Test accuracy:  0.6477399468421936
```

Hình 17. *Kết quả độ chính xác dữ liệu test với Epoch = 200, tiêu tốn thời gian thực = 34s với kích thước ảnh 48x48*

Kết quả khi được **train** với Epochs =200 vòng, thì tỉ lệ học chính xác (accuracy) = **97.57 %**, Hình 18.

```
train_loss, train_accu = model.evaluate(train_set)
test_loss, test_accu = model.evaluate(test_set)
```

```
898/898 [==============================] - 16s 18ms/step - loss: 0.0748 - accuracy: 0.9757 - f1_score: 0.9754
```

Hình 18. *Kết quả độ chính xác dữ liệu train với Epoch = 200, tiêu tốn thời gian thực = 34s với kích thước ảnh 48x48*

Bảng 3. *Bảng so sánh kết quả công trình nghiên cứu, model CNN*

Tác giả	Accuracy %
(Nishime, Endo, Yamada, Toma, & Akamine, 2016)	58.0
CNN (Raghuvanshi & Choksi, 2016)	48.0
CNN (Samsani & Gottala, 2020)	61.4
Mô hình đề xuất	**64.77**

Qua khảo sát công trình nghiên cứu, tác giả các bài báo sử dụng model CNN để thực nghiệm đánh giá, theo kết quả cho thấy kết quả dữ liệu học kiểm thử của (Nishime, Endo, Yamada, Toma, & Akamine, 2016), CNN (Raghuvanshi & Choksi, 2016), CNN (Samsani & Gottala, 2020), lần lượt đạt tỉ lệ chính xác là: 58%, 48%, 61.4% và mô hình đề xuất CNN là **64.77**%. Theo kết quả trên Bảng 3 thì mô hình đề xuất của chúng tôi đạt hiệu quả tốt hơn. Ngoài ra, kết quả khi được **train** với Epochs =200 vòng, thì tỉ lệ học đạt chính xác (accuracy) = **97.57%**, Hình 18, kết quả này được đánh giá rất tốt.

3.2. Thảo luận

Qua quá trình thực nghiệm của bài nghiên cứu này, chúng tôi đã đưa ra ba vấn đề chính: một là, mở rộng bộ bộ dữ liệu chuẩn FER-2013 với 35.887 ảnh, trong đó: để train là 28.709 ảnh và test là 7178 ảnh, kích thước ảnh gốc 48x48 pixel, được sử dụng bộ chiết xuất đặc trưng ảnh dạng nhị phân, được lưu trữ dạng matrix 3 chiều, gán nhãn theo 7 loại biểu cảm; hai là, thực hiện kiểm tra mô phỏng các mô hình LDA, NB, KNN, DT, SVM được đánh giá theo tiêu chí chính xác (Accuracy) làm kết quả so sánh, cho thấy kết quả mô hình SVM tương ứng kích thước 72x72, 64x64, 32x32 có giá trị tốt nhất tương ứng: 42%, 43%, 42%, tỉ lệ học thấp nhất là mô hình NB = 29%, còn các mô hình còn lại nằm trong khoảng giữa hai mô hình trên. Ba là, mô hình đề xuất mạng nơ ron, khi huấn luyện Epochs = 100, dense = 7, kích thước tương ứng: 64x64, 32x32 pixel là **58.90%, 62.98%**. Riêng kích thước ảnh gốc của tập dữ liệu chuẩn FER-2013, khi sử dụng số vòng train Epochs = 200 thì kết quả thu được trên tập dữ liệu train có độ chính xác (accuracy) = **97.57%**, test = **64.77** %.

Trong nghiên cứu này, đã so sánh với các bài nghiên cứu của của (Nishime, Endo, Yamada, Toma, & Akamine, 2016), CNN (Raghuvanshi & Choksi, 2016), CNN (Samsani & Gottala, 2020) với kết quả mô hình đề xuất ở bảng 3 lần lượt là: 58%, 48%, 61.4%, **64.77%**, kết quả mô hình đề xuất tỉ lệ học chính xác, hiệu quả cao nhất.

Đánh giá chung về mô hình đề xuất còn hạn chế, tiếp tục khắc phục: Cần phải kết hợp các mô hình học sâu tăng cường, mô hình học sâu dạng kết hợp thì tỉ lệ độ chính xác tốt hơn cho việc nhận dạng ảnh.

4. Kết luận

Như vậy kết quả thực nghiệm này, chúng tôi đã nêu ra ba vấn đề chính: một là, mở rộng định dạng bộ dataset Facial Expression Recognition 2013 với kích cỡ ảnh gốc 48x48 pixel, sau đó được mở rộng nhiều kích thước ảnh khác như là 72x72, 64x64, 32x32 pixel để làm nền tảng cho đánh giá kết quả nghiên cứu thêm đa dạng; hai là, mô phỏng các mô hình thuật toán sử dụng từ các thư viện python trong quá trình huấn luyện nhằm đánh giá và so sánh về các tiêu chí Accuracy, Precision, F1-Score; ba là, đề xuất mô hình học sâu CNN được so sánh với các công trình nghiên cứu khác là đáng tin cậy và thiết thực. Điều đó, đã được minh chứng rất rõ và chi tiết ở phần 3.

Từ kết quả thực nghiệm với tập dữ liệu FER-2013 gồm có 35.887 ảnh, trong đó: tập dữ liệu dùng để train là 28.709 ảnh và dùng cho việc test là 7178 ảnh. Trong phần đánh giá kết quả mô hình đề xuất mạng nơ-ron CNN về xác nhận dạng khuôn mặt người được so sánh với các công trình nghiên cứu liên quan khác, cho thấy mô hình đề xuất đạt hiệu quả độ chính xác cao hơn. Qua nghiên cứu này, chúng tôi sẽ khắc phục những hạn chế của mô hình đề xuất học sâu này như đã đánh giá ở mục 3.2 thảo luận.

❖ ***Tuyên bố về quyền lợi:*** *Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.*

TÀI LIỆU THAM KHẢO

Bhairnallykar, S., Prajapati, A., Rajbhar, A., & Mujawar, S. (2020). Convolutional Neural Network (CNN) for Image Detection. *International Research Journal of Engineering and Technology (IRJET)*.

Bledsoe, W. W. (1964). The Model Method in Facial Recognition. *Technical Report PRI 15*.

Bledsoe, W. W. (1966). Some Results on Multicategory Pattern Recognition. *Journal of the ACM*, 304-316.

Chauhan, R., Kumar , K. G., & Joshi, R. (2018). Convolutional Neural Network (CNN) for Image Detection and Recognition. *First International Conference on Secure Cyber Computing and Communication (ICSCCC).*

Goodellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., . . . H, D. (2015). Challenges in representation learning: A report on three machines learning contests. *International Conference on Neural Information Processing*, 59-63.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2004). *KNN Model-Based Approach in Classification.* Northern Ireland, UK.

Huynh, T. T., & Nguyen, T. H. (2021). On the performance of intrusion detection systems with hidden multilayer neural network using DSD training. *International Journal of Computer Networks & Communications (IJCNC)*, 117-137.

Krizhevsky, A., Nair, V., & Hinton, G. (2006). *CIFAR-10 dataset*. Retrieved from https://www.cs.toronto.edu/~kriz/cifar.html

Nishime, T., Endo, S., Yamada, K., Toma, N., & Akamine, Y. (2016). Feature Acquisition From Facial Expression Image Using Convolutional Neural Networks. *Journal of Robotics, Networking and Artificial Life*, 9-12.

Raghuvanshi , A., & Choksi, V. (2016). Facial Expression Recognition with Convolutional Neural Networks. *CS231n Course Projects Winter*.

Rajeswari, R. P., Juliet, K., & Aradhana. (2017). Text Classification for Student Data Set using Naive Bayes Classifier and KNN Classifier. *International Journal of Computer Trends and Technology (IJCTT)*, 8-12.

Sambare , M. (2020). *FER-2013*. Retrieved from: https://www.kaggle.com/datasets/msambare/fer2013

Samsani, S., & Gottala, V. A. (2020). A real-time automatic human facial Expression recognitiom system using deep neural networks. *Information and Communication Technology for Sustainable Development, Singapore*, 431-441.

Srivastava, D., & Bhambhu, L. (2010). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, 1-7.

Taha , B. J., & Mohsin, A. A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends (JASTT)*, 20-28.

Tharwat, A., Gaber, T., Tharwat, A., Ibrahim, Hassanien, & A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, 169-190.

Wibawa, A. P., Kurniawan, A. C., Murti, D. M., Adiperkasa, R. P., Putra, S. M., Kurniawan, S. A., & & Nugraha, Y. R. (2019). Naïve Bayes Classifier for Journal Quartile Classification. *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, 91-98.

AN APPROACH TO HUMAN FACE RECOGNITION BY MACHINE LEARNING TRAINING

Dam Minh Linh*, Nguyen Hoang Thanh

Posts and Telecommunications Institute of Technology in Ho Chi Minh City, Vietnam

**Corresponding author: Dam Minh Linh – Email: linhdm.tg@ptithcm.edu.vn*

Received: October 24, 2022; Revised: November 29, 2022; Accepted: December 09, 2022

ABSTRACT

Facial recognition is a biometric technology technique that faces human facial features. The Facial Expression Recognition 2013(FER-2013) image dataset, including seven different types of human facial expressions, was used as the training dataset in this study. The importance of system security is urgent to deploy a human face recognition authentication application to log in to the system, authenticate on smartphones, and time attendance. We propose a machine learning, deep learning model with many different training methods combined with the FER-2013 dataset, which is expanded with image sizes (32x32, 48x48, 64x64, 72x72) to conduct experiments with LDA, NB, KNN, DT, and SVM models. Then, evaluating the effectiveness of each model in terms of accuracy, precision, and F1-Score was conducted. The experimental results have three main contributions: (a)expanding the dataset format with more diverse sizes; (b) simulating different algorithmic models during training to evaluate and compare the above criteria, and (c) showing the effectiveness of the proposed CNN deep learning model.

Keywords: computer vision; deep learning; face recognition; FER-2013; neural networks