

Effect of the SVM algorithm on deep classification networks

Samar Abd AlGanai Al-Halabi

Fida Emad Khalil

Faculty of Engineering || Al-Wataniya Private University || Syria

Abstract: The study aimed at applying several convolutional neural networks on two data samples containing a large group of human images in order to identify them through transfer learning technology, and study the effect of applying support vector machine (SVM) on these networks.

Human facial recognition technology is an important problem; this technology is applied broadly in security (such as criminal identification), education (such as attendance systems), protection systems (such as secure electronic banking), etc. Traditional algorithms didn't give optimal results in the classification field, so we used the newest (CNNs) in its. Additionally, we replaced classification layer in each of the networks studied with (svm), to study its effect on the performance of these networks in accuracy and time.

At the end, we got good results that achieved accuracy about 99% and reducing training time and classification error rate in some cases.

Keywords: Transfer learning, Machine Learning, Image Classification, Convolutional Neural Networks, Support Vector Machine (SVM).

تأثير خوارزمية SVM على شبكات التصنيف العميقة

سمر عبد الغني الحلبي

فداء عماد خليل

كلية الهندسة || الجامعة الوطنية الخاصة || سوريا

المستخلص: هدفت الدراسة إلى تطبيق عدة شبكات عصبونية تلافيفية على عيني بيانات تضمنان مجموعة كبيرة من الصور البشرية بهدف التعرف عليها من خلال تقنية التعلم بالنقل، ودراسة مدى تأثير تطبيق خوارزمية متجه دعم الآلة على هذه الشبكات. تعد تقنية التعرف على الوجوه البشرية من المسائل المهمة، فقد تم تطبيقها على نطاق واسع في العديد من المجالات منها الأمنية كالاعتراف على المجرمين، والتعليمية كأنظمة الحضور، وأنظمة الحماية كالخدمات المصرفية الالكترونية الآمنة، وغيرها. لا تعطي الخوارزميات التقليدية نتائج مثلى في هذا المجال لذلك قمنا بالعمل على عدة شبكات عصبونية تلافيفية اعتمدت في مجال التصنيف من الأقدم إلى الأحدث، بالإضافة إلى ذلك استبدلنا طبقة التصنيف في كل شبكة من الشبكات المدروسة بخوارزمية متجه دعم الآلة للاستفادة من مزاياها في فصل البيانات غير الخطية ودراسة مدى تأثيرها على أداء هذه الشبكات من ناحية الدقة والزمن. حصلنا في نهاية البحث على مجموعة من النتائج الجيدة، والتي حققت دقة عالية بلغت تقريباً 99% وانخفاض في زمن التدريب ونسبة خطأ التصنيف في بعض الحالات المدروسة.

الكلمات المفتاحية: التعلم بالنقل، تعلم الآلة، التصنيف، شبكة عصبونية تلافيفية، خوارزمية متجه دعم الآلة.

1. المقدمة.

تعتبر تقنية التعرف على الوجوه من أكثر طرائق المقاييس الحيوية استخداماً، وواحدة من أكثر المهام الأساسية تحدياً في مجال تمييز الأنماط [Amirthalingam 2017]، فقد تم استخدامها في العديد من المجالات من أهمها: المجال الأمني والتحقيقات الجنائية للتعرف على المجرمين، مجال أنظمة الحماية كالخدمات المصرفية الالكترونية الآمنة وبطاقات الائتمان، المجال التعليمي كأنظمة الحضور والغياب، ومجال الإعلانات الموجهة وذلك لتوجيه الإعلان إلى المستخدمين المحتملين لشراء منتج أو طلب خدمة ما.

تطبق تقنية التعرف على الوجوه من خلال ثلاث خطوات رئيسية، بحيث في الخطوة الأولى يتم اكتشاف الوجه، وفي الخطوة الثانية يتم تتبع الوجه واستخراج الميزات، أما الخطوة الأخيرة فيتم فيها التعرف على الوجه وتصنيفه.

تقسم خوارزميات التعرف على الوجوه بشكل عام إلى الخوارزميات التقليدية والخوارزميات المعتمدة على مفهوم التعليم العميق [Wang 2010][Rosero 2020]، ولكن بسبب تأثر مسألة التعرف على الوجوه بالعديد من العوامل منها الدقة المنخفضة للصور، الإضاءة، الوضعيات المختلفة وعوامل أخرى، فإن الخوارزميات التقليدية لا تعمل بشكل دقيق [Guo 2019].

بينما تعتمد معظم الخوارزميات الحديثة المستخدمة للتعرف على الوجوه على الشبكات الاصطناعية العميقة والتي أعطت نتائج ماهرة في الوقت الحالي، وتعتبر الشبكات العصبونية التلافيفية (CNN) من أهمها، والتي تعتمد مبدأ استخراج ميزات الوجوه من قبل طبقات الالتفاف طبقة تلو الأخرى، ومن ثم استخدام هذه الميزات كعينات لتدريب الشبكة. وبالتالي كلما استخدمت طبقات أكثر يمكن تعلم ميزات أكثر، وفي عام 1998 تم تشكيل أول شبكة عصبونية تلافيفية تدعى بـ LUNET، ومع التطوير الملحوظ لهذا النوع من الشبكات تم استخدامها لحل العديد من مسائل الرؤية الحاسوبية. لذلك اعتمدنا في بحثنا على خمس شبكات عصبونية تلافيفية مدربة مسبقاً وذلك بهدف تطبيق تقنية التعلم بالنقل لاختصار زمن التدريب، وكانت جميع هذه الشبكات تعتمد في الطبقة الأخيرة على تابع softmax لعملية التصنيف، ولذلك قمنا في بحثنا بتطبيق إحدى خوارزميات التصنيف والتي تدعى متجه دعم الآلة (Support Vector Machine (SVM)) لتوضيح مدى تأثيرها على دقة تصنيف هذه الشبكات، وتم اختبار نتائج الشبكات على عيني بيانات هما LFW و Casia-webface للتأكد من صحة النتائج.

2. مشكلة الدراسة:

تعتبر عملية التعرف على الوجوه البشرية من القضايا المهمة في العديد من المجالات الحياتية كمجال المراقبة (مراقبة مداخل ومخارج مؤسسة ما) ومجال الامن (عملية التعرف على زبائن مصرف عند التعامل مع حساباتهم)، وفي مجالات عديدة أخرى.

بما أن عمل الخوارزميات التقليدية في هذا المجال لا يزال ذو دقة منخفضة نوعاً ما خاصة في شروط الإضاءة السيئة أو في حال كان الوجه في وضعيات غير واضحة، لذلك فقد تم التوجه إلى الخوارزميات المعتمدة على التعلم العميق. بالرغم من أن الشبكات العصبونية التلافيفية (CNN) قد استخدمت على نطاق واسع في مجال التصنيف [Dai, 2020] [Khan, 2020] [Levi 2016]، إلا أنه لا يزال هناك بعض المشاكل العملية، بالإضافة إلى مشكلة محدودية حجم بيانات التدريب والتي تعتبر مشكلة عامة في مجال التعليم العميق.

التعلم بالنقل الذي تم تقديمه لتبني النماذج المدربة في العديد من المجالات، تبين من خلال الدراسات أنه طريقة فعالة أيضاً لحل مسألة التصنيف [Seeri, 2020][Pan, 2010][Alipour, 2021].

بما أن مسائل التصنيف عامة ذات بيانات متداخلة نوعاً ما فقد أظهرت خوارزمية متجه دعم الآلة كفاءة عالية في فصل هذه النوعية من البيانات واتخاذ القرار بالاعتماد على جزء من البيانات [Arikabi, 2021]. لهذا قمنا بالعمل على دراسة العديد من النماذج المستخدمة في مجال التصنيف والاعتماد على تصنيف خوارزمية svm ضمن مسألة التعرف على الوجوه البشرية وتطبيق ذلك على أكثر من عينة بيانات ضمن بحثنا.

3. فرضيات الدراسة: تفترض الدراسة:

- 1- معرفة بتقنيات تعلم الآلة والتعلم العميق وأساسياتهما.
- 2- معرفة بمفاهيم خوارزميات التصنيف وألية عملهم.
- 3- اطلاع على عينات البيانات المدروسة من خلال الروابط التالية:

Casia-webface:

https://drive.google.com/file/d/1Of_EVz-yHV7QVWQGihYfvtny9Ne8qXVz/view

Labeled Faces in the Wild (LFW): <http://vis-www.cs.umass.edu/lfw>

4. أهداف الدراسة:

1. نظراً لأهمية تقنية التعرف على الوجوه في العديد من مجالات الحياة فقد كان هدف بحثنا: تطبيق تقنية التعلم بالنقل على مسألة تصنيف الوجوه البشرية وذلك لاختصار زمن التدريب وحل مشكلة البيانات القليلة.
2. تحقيق مسألة تصنيف الوجوه البشرية بأكثر من شبكة من شبكات التعلم العميق.
3. دراسة مدى تأثير خوارزمية متجه دعم الآلة (svm) على أداء الشبكات العميقة المدروسة من حيث الزمن والدقة.

5. أهمية الدراسة:

تعتبر الأهمية العلمية للدراسة كون تصنيف الوجوه البشرية يعتبر من المجالات المهمة لتطبيق نماذج تعلم الآلة عموماً على عينات من هذا المجال، والتعلم العميق خصوصاً نظراً للنتائج المرضية التي نتجت عن تطبيقها في أغلب مجالات الحياة العملية بمختلف جوانبها [Deffo 2018]. سعينا في هذه الدراسة لتطبيق الشبكات العصبونية التلافيفية الأكثر استخداماً في مجال التصنيف عموماً (MobileNet_V2 - Inception_V3 - ResNet50 - VGG19 - VGG16) على عيني بيانات كبيرة تحويان صوراً ملونة (Casia-webface - LFW (Labeled Faces in the Wild)) لدراسة دقة تصنيفها للوجوه البشرية، من ثم طبقنا خوارزمية متجه دعم الآلة ((Support Vector Machine (SVM)) على كل شبكة من الشبكات سابقة الذكر لدراسة تأثيرها على الدقة، زمن التدريب ونسبة خطأ التصنيف، كما تناولت الدراسة تفاصيل بنية كل شبكة من الشبكات المدروسة، وألية عمل خوارزمية SVM، إضافةً إلى النتائج والملاحظات التي حصلنا عليها بعد تنفيذ هذه الدراسة.

6. هيكلية الدراسة:

تم تقسيم هذه الدراسة إلى ثلاثة أقسام، يتناول القسم الأول منها الإطار النظري والدارسات السابقة للدراسة، بينما يتطرق القسم الثاني إلى دراسة الحالة التجريبية وتطبيق الشبكات التي تم اختيارها لدراسة نتائجها

في تصنيف الوجوه البشرية ومن ثم تطبيق الخوارزمية SVM، ويعرض القسم الثالث نتائج الدراسة والدقة التي حصلنا عليها في نهاية عملية التدريب للنماذج المقترحة.

القسم الأول-الإطار النظري والدراسات السابقة.

أولاً-الإطار النظري:

1. تقنية التعلم بالنقل (Transfer Learning)

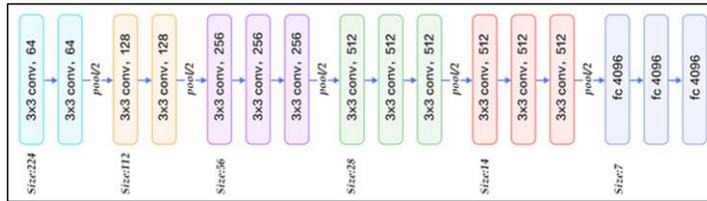
تستغرق الشبكات العصبونية التلافيفية زمناً طويلاً للتدريب على مجموعة بيانات ضخمة. لذلك كان التعلم بالنقل إحدى الطرق التي تختصر هذا الزمن من خلال إعادة استخدام أوزان نموذج من النماذج المدربة مسبقاً على مجموعة بيانات معيارية (مثل ImageNet).

يشير مصطلح التعلم بالنقل إلى تأثير المعرفة المكتسبة من قبل الشبكة العصبونية المدربة مسبقاً على مجموعة بيانات معينة (كبيرة عادة) متاحة لحل المشكلات الجديدة التي تتوفر لها أمثلة تدريب قليلة، ودمج المعرفة الحالية مع المعرفة الجديدة المستفادة من الأمثلة القليلة من مجموعة البيانات الخاصة بالمسألة [Guo 2019] [Sedrick 2020]. ويتم ذلك بتحديد النموذج المصدر المناسب للمسألة المطروحة، ثم إعادة استخدامه كنقطة انطلاق لنموذج في مهمة ثانية، ومن ثم ضبط النموذج بتحديد بيانات زوج الإدخال والإخراج المتاحة للمهمة الهدف.

2. الشبكات المستخدمة في البحث:

• شبكة مجموعة الهندسة المرئية ((Visual Geometry Group(VGG):

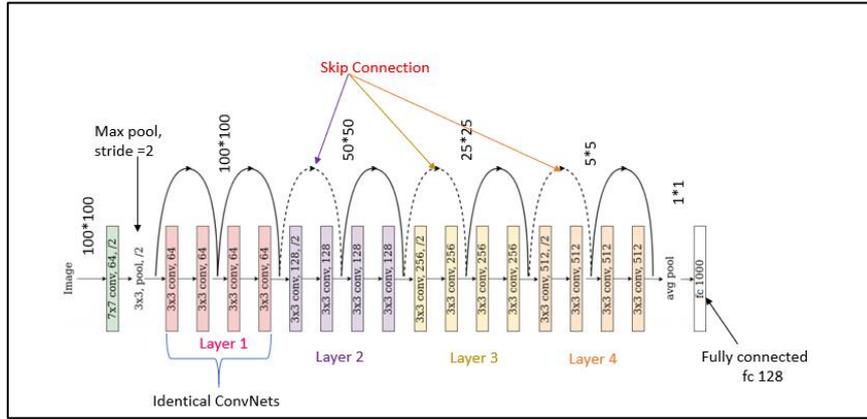
في عام 2014، تم اقتراح نموذج شبكة VGG، بحيث تتكون من طبقات الالتفاف (Convolutions layers) 3*3، طبقات التجميع القصوى (2 * 2 Max pooling layers)، وطبقات متصلة بالكامل (Fully connected)، من أشهر نسخها VGG16 و VGG19، ويشير الرقم 16 أو 19 إلى عدد طبقات الشبكة. نظراً لعمقها والعدد الكبير من العقد المتصلة بالكامل فيها فإن مساحة تخزين VGG16 تبلغ 533 ميغابايت ومساحة تخزين VGG19 تبلغ 574 ميغابايت [Simonyan 2015].



الشكل (1) يوضح بنية الشبكة VGG16

• شبكة الرواسب ((Residual Network (ResNET):

في عام 2015، ظهرت شبكة ResNet50 وكان هدفها الرئيسي حل مشكلة تلاشي التدرج اللوني الذي قد يؤدي إلى إيقاف الأوزان في الشبكة عن التحديث أو التغيير وذلك من خلال استخدام الطبقات المتبقية وتخطي الاتصالات. على الرغم من أن ResNet50 أعمق من VGG16 و VGG19 إلا أن حجم الشبكة 98 ميغابايت [Zhang, 2015].



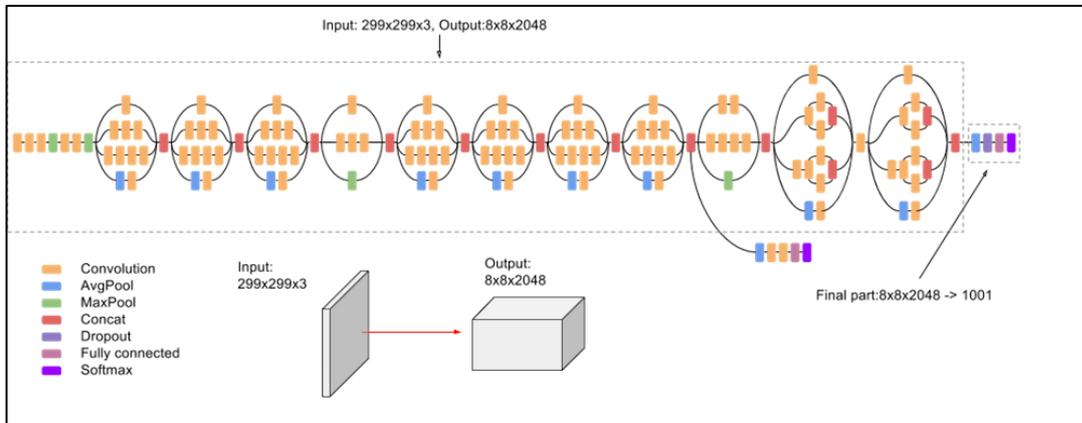
الشكل (2) يوضح بنية الشبكة ResNet

• شبكة التأسيس (Inception):

تمّ وصف واستخدام وحدة بناء جوجل Inception VN في نموذج شبكة جوجل GoogLeNet عام 2015 من قبل الباحثة كريستين زيدي ورفاقها في ورقة بحثية [Szegedy, 2014], بعنوان "الاتجاه نحو العمق مع عمليات الطّي".

الغرض من وحدة البناء أن تكون كتلة من طبقات طّي متوازية ذات أحجام مرشحات مختلفة (مثلا 1*1، 3*3، 5*5) وطبقة تجميع وفق القيمة الأكبر بعدد 3*3، من ثم جمع نتائج جميع الطبقات. تُعتبر هذه البنية وحدة بناء بسيطة وفعالة تسمح للنموذج بتعلم مرشحات متوازية وبأحجام مختلفة أو متساوية مما يسمح بالتعلّم على مستويات متعددة.

يشير الرقم N في شبكة Inception VN إلى رقم الإصدار الذي حدته Google، وسوف نعتمد على الإصدار الثالث في بحثنا ذو الحجم 96 ميجابايت [Vanhoucke, 2015].



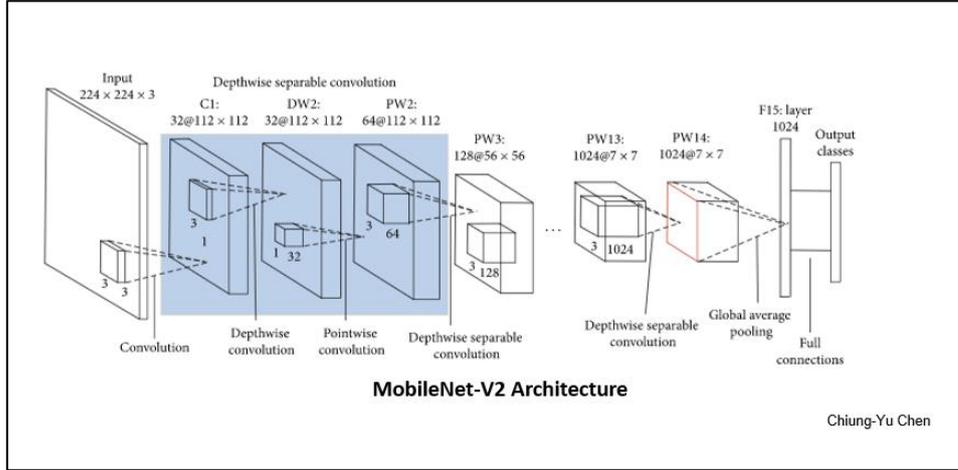
الشكل (3) يوضح بنية الشبكة Inception

• شبكة MobileNet:

هي شبكة عصبونية تلافيفية عميقة خفيفة الوزن اقترحتها Google للأجهزة المحمولة في عام 2017، وهدفها الأساسي تقليل معلمات الشبكة من خلال التحليل الذكي لنواة الالتفاف.

تقسم عملية الالتفاف إلى مرحلتين منفصلتين الأولى هي الالتفاف بالعمق (Depthwise convolution) والتي تقوم بترشيح قنوات الدخل ومعالجة كل منها بشكل منفصل، الثانية هي الالتفاف النقطي (Pointwise).

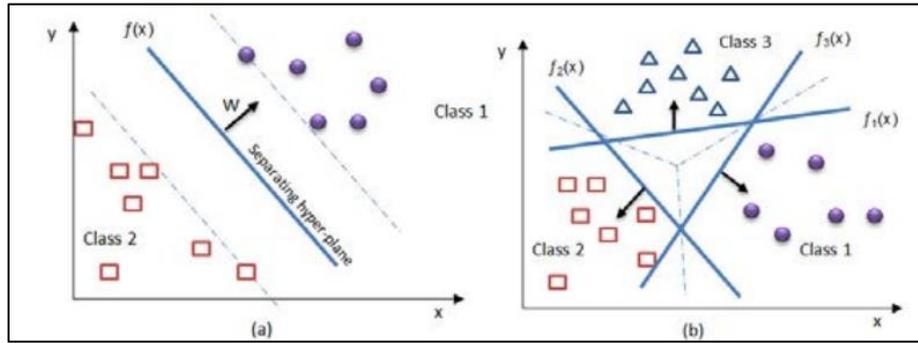
(convolution) والتي تقوم بتجميع القيم المرشحة لإنشاء مزايا جديدة. حققت هذه العملية كلفة حساب أقل بنحو 9 مرات تقريباً مع تقليل الذاكرة المستهلكة بالتالي سرعة أكبر مع الحصول على الدقة ذاتها. في عام 2019، تم اقتراح شبكة MobileNetV2 التي سننتمدها في بحثنا، وقد اختلف فيما عمل طبقة الالتفاف النقطي عن الإصدار الأول لتصبح مهمتها إنقاص عدد القنوات من خلال إسقاط البيانات التي لها عدد كبير من الأبعاد في مصفوفة ذات أبعاد أقل بكثير، ولذلك دعيت بطبقة الإسقاط أو بعنق الزجاجة (bottleneck) باعتبارها تقلل كمية البيانات التي تتدفق عبر الشبكة [Sandler, 2019].



الشكل (4) يوضح بنية الشبكة MobileNetV2

3. خوارزمية متجه دعم الآلة (Support Vector Machine (SVM):

تعتبر من أفضل خوارزميات تعلم الآلة لحل مشكلة التصنيف الخطي وغير الخطي، طورت من قبل الباحث Vapnik عام 1998، هدفها البحث عن أفضل مستوى فاصل لتصنيف البيانات، وذلك يجعل المسافة بين بيانات التدريب والمستوى الفاصل أكبر ما يمكن [Cortes,1995] [Vapnik,1998] [Burges, 1998].



الشكل (5) يوضح آلية عمل SVM

(a) للفصل بين صنفين (b) للفصل بين أكثر من صنفين

تقسم خوارزمية SVM إلى نوعين خطي وغير خطي. يتم تمثيل المستوى الفاصل في خوارزمية SVM الخطية

بالشكل التالي:

$$f(x) = W^T \cdot X = 0$$

$$\text{حيث } W, X \in R^n$$

تُعتبر W عن متجه الأوزان، X بيانات التدريب والتي يمكن تمثيلها رياضياً بالشكل التالي:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in R^n$$

حيث x_n متجه الميزات، y_n العنوان الخاص بـ x_n فعلى سبيل المثال إذا كان $y_n = 1$ هذا يعني أن البيانات تنتمي إلى الصنف الأول.

بينما في حال كانت البيانات غير قابلة للفصل خطياً. يتم تطبيق طريقة النواة (kernel) لتحويل مجموعة البيانات من فضاءها الحالي (الفضاء ثنائي البعد) إلى فضاء أعلى، وذلك بإضافة بعد جديد ومن ثم اختبار قابلية فصل البيانات، ففي حال التحقق تبحث الخوارزمية عن المستوى الفاصل المناسب لعملية الفصل، وإلا تستمر الخوارزمية بإضافة أبعاد ومن ثم اختبار البيانات.

من أهم التوابع المستخدمة في طريقة النواة (kernel) هي:

Gaussian (RBF) kernel

$$k(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$$

Polynomial kernel

$$k(x_i, x_j) = (1 + x_i^t x_j)^d$$

Linear kernel

$$k(x_i, x_j) = x_i \cdot x_j$$

تُعبّر x_i عن بيانات التدريب و x_j عن قيم احتمالات ظهور x_i .

بما أن بيانات مسألتنا (تصنيف الوجوه) متداخلة نوعاً ما فنحن بحاجة إلى طريقة لفصل بيانات غير خطية، لذلك اعتمدنا في بحثنا على خوارزمية SVM، بالإضافة إلى كفاءة الخوارزمية العالية في اتخاذ القرار بالاعتماد على جزء من بيانات التدريب وليس على كاملها، بينما استخدمنا التابع الخطي (Linear kernel) في طريقة النواة لاعتماده على فضاء ميزات التدريب الأساسية دون أي تغيير كما في التوابع الأخرى مما يجعله الأسرع وهذا يناسب مسألتنا.

قمنا باستبدال طبقة التصنيف في كل شبكة من الشبكات العميقة المدروسة بالمصنف SVM، وذلك لدراسة مدى تأثيره عليها.

ثانياً-الدراسات السابقة:

تشهد الأبحاث في مجال التصنيف عموماً والتعرف على الوجوه خصوصاً في السنوات الأخيرة ارتفاعاً ملحوظاً لاسيما مع الشبكات العصبونية التلافيفية (CNNs). نذكر منها:

- دراسة [Gruber 2017] اقترح خوارزمية للتعرف على الوجوه من خلال مرحلتين الأولى استخدمت شبكة (CNN) بسيطة مؤلفة من 32 طبقة التفاضلية للتدريب على 925 صنف من قاعدة Casia-webface وتوصلت إلى دقة 71%، أما الثانية فقد استخدمت شبكة ResNet50 مع 60k تكرار وتوصلت إلى دقة 87% ومع مضاعفة التكرارات تحسنت الدقة إلى 90%.

- دراسة [Sinha, 2019] اقترح نموذج مؤلف من ثلاث بنى هجينة لشبكة MobileNet مع استبدال لتابع التنشيط غير الخطي ReLU بـ Drop Activation وإلغاء لبعض الطبقات غير الضرورية طبقت على قاعدة CIFAR-10 وقد توصلت لدقة 85%. [Sushma 2020] درس مسألة تصنيف الصور باستخدام الشبكات (vgg16 – vgg19) – (vgg16 – vgg19) وكان متوسط الدقة لجميع أصناف الصور في الشبكات الثلاث 87% - 83% - 85% على التوالي. [Gwyn 2021] درس مسألة تصنيف الوجوه على قاعدة LFW مستخدماً الشبكات AlexNet – Xception – inceptionv3 – inceptionv2 – Resnet101 – Resnet50 – Vgg16 – Vgg19 ، وقارن دقتهم بثلاث نسب لتقسيم البيانات (20/80 ، 30/70 ، 40/60)، وحصل على نتائج متفاوتة فبعض الشبكات تحسن أدائها والبعض الآخر تراجع، وتراوحت دقة الشبكات بالتقسيم 20/80 بين 60% إلى 84%. [Alrikabi2021] اقترح طريقة للتعرف على الوجوه باستخدام خوارزمية svm مع شبكتي inception_v1 , Vgg16، وطبقها على قواعد

بيانات للوجوه البشرية بعضها بصور ملونة والبعض الآخر بصور رمادية، وحصل على دقة 95% لشبكة inception_v1 و 94% لشبكة vgg16 من أجل قاعدة LFW. بناءً على هذه الدراسات انطلقنا في بحثنا بهدف دراسة أداء عدة شبكات عصبونية التفاضلية (vgg16, vgg19, resnet50, inception_v3, mobilenet_v2) مستخدمة في مجال التصنيف على مسألة تصنيف الوجوه البشرية، ومن ثم دراسة تأثير خوارزمية svm على أداء هذه الشبكات من ناحية الدقة والزمن المستغرق في عملية التدريب ونسبة خطأ التصنيف.

القسم الثاني-دراسة الحالة التجريبية والتعرف على الوجوه البشرية باستخدام الشبكات المدروسة.

أولاً-دراسة الحالة التجريبية:

لعبت مجموعات البيانات الكبيرة والعالية الجودة دوراً مهماً في التقدم في مجالات التعليم العميق، تتكون المهمة المختارة لتجربة التعلم بالنقل من تصنيف صور لوجوه بشرية وقد اخترنا خمسة أصناف. يرجع اختيار هذه المهمة بشكل أساسي إلى توفر هذه المجموعة من البيانات، بالإضافة إلى مجال المشكلة، وهو عام بما يكفي ليكون مناسباً للتطبيق الفعال للتعلم بالنقل مع الشبكات العصبونية المدربة مسبقاً على مجموعة بيانات ImageNet الشهيرة.

عينة البيانات المدروسة:

تعاملنا في دراستنا مع عيني البيانات Casia-webface و LFW ، حيث تعد عينة البيانات Casia-webface ثاني أكبر قاعدة بيانات للصور البشرية الملونة والمستخدم في عملية التصنيف تحوي حوالي 450000 صورة موزعة إلى أكثر من 10000 صنف [Shlizerman, 2016].



الشكل (6) يوضح عينة من بيانات Casia-webface

بينما قاعدة البيانات LFW مصممة لدراسة مشكلة التعرف على الوجوه تحوي أكثر من 13000 صورة لوجوه تم جمعها من الويب حيث فيها 1680 من الأشخاص لهم صورتان أو أكثر [Huang, 2007].



الشكل (7) يوضح عينة من بيانات LFW

تحقق كلا العينتين وجود صور مراعية لشروط الإضاءة والوضعية المختلفة للوجوه. تم اختيار الأصناف التي تحوي أكبر عدد ممكن من الصور في كلا العينتين والتي تراعي تعدد الوضعيات وشروط الإضاءة المختلفة ولكلا

الجنسين وبأعمار مختلفة أيضاً. وعلى الرغم من شهرة قاعدة LFW إلا أن غالبية أصنافها تحوي صور أو صورتين وهذا العدد لا يناسب الشبكات العميقة مما اضطرنا إلى تحديد عدد الأصناف. تم تقسيم كل عينة إلى لمجموعتي تدريب واختبار بنسبة 20/80.

الجدول (1) معلومات البيانات المستخدمة

القاعدة المستخدمة	عدد صور كل صنف	عدد الأصناف المدروسة	حجم الصور	نوع الصور
Casia-webface	490	5	250*250	Jpg
LFW	300	5	250*250	Jpg

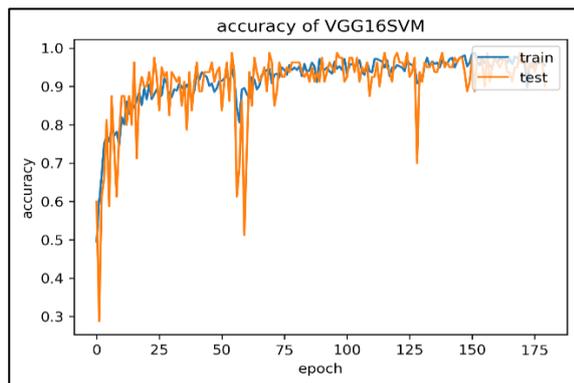
ثانياً-تطبيق الشبكات المدروسة في مسألة التعرف على الوجوه البشرية ومناقشة النتائج:
قمنا بتطبيق شبكات التعلم العميق المذكورة سابقاً من خلال تدريبها ثم مقارنة نتائجها قبل وبعد إضافة خوارزمية SVM من حيث دقة المقياس Top-5 بالإضافة إلى زمن التدريب ونسبة الخطأ في التصنيف.
قمنا بمعالجة الصور في عينات الدراسة لتأخذ نفس الحجم و من ثم زيادتها عن طريق Data Augmentation بتطبيق بعض التحويلات المتاحة في Image DataGenerator كالتالي:
Rotation =15 , zoom_range =0.2 , shear_range = 0.2 horizontal_flip=true, rescale=1./255
كما استخدمنا batch_size =8 لكل الشبكات المدروسة، أما دخل الشبكات فقد تم اختياره اعتماداً على أبعاد الصور في عيني البيانات المدروسة وقد كانت ابعاد الصور في العينتين 250*250 لذلك اعتمدنا دخل لجميع الشبكات 224 * 224 .
تمت تهيئة الشبكات المدروسة بأوزان تم تدريبها مسبقاً على ImageNet، ومن ثم اعتمدنا على Adam Optimizer ومعدل التعلم 0.0001 وعدد تكرارات 180 .
تم تطبيق الشبكات وتدريبها واختبارها على حاسب محمول بمعالج Intel فئة Core i7 10th Gen وسرعة معالجة 2.2 غيغاهرتز، وذاكرة 16 غيغابايت، ووحدة معالجة الرسومات GeForce GTX 1660 Ti بذاكرة 6 غيغابايت. تم تسجيل النتائج لكل الحالات المدروسة وعرضها في القسم التالي.

القسم الثالث- نتائج الدراسة ومناقشتها.

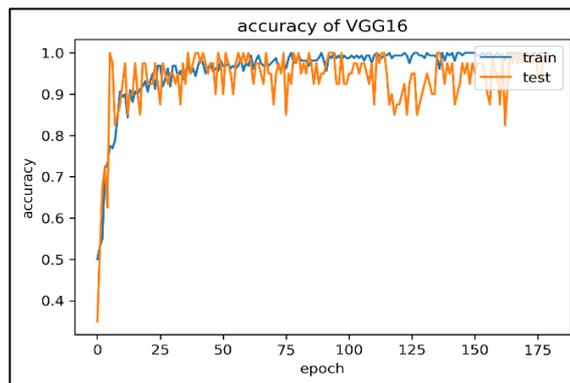
يحتوي هذا القسم نتائج التدريب على عيني الدراسة ومن ثم تقييم الشبكات من حيث المقاييس التالية:

- الدراسة من حيث الدقة:

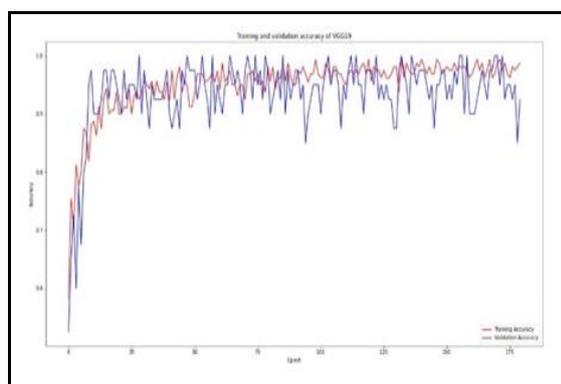
سنعرض مخططات التدريب على عينة البيانات Casia-webface.



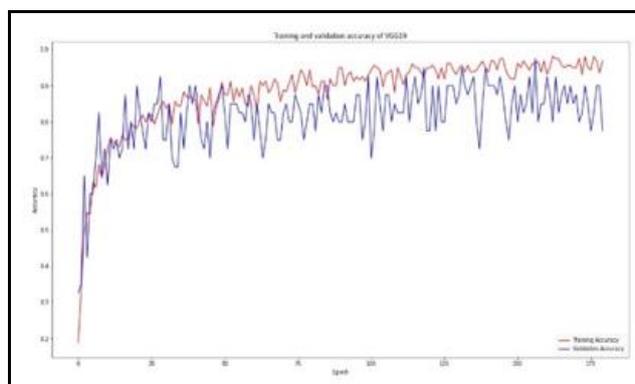
Vgg16+svm



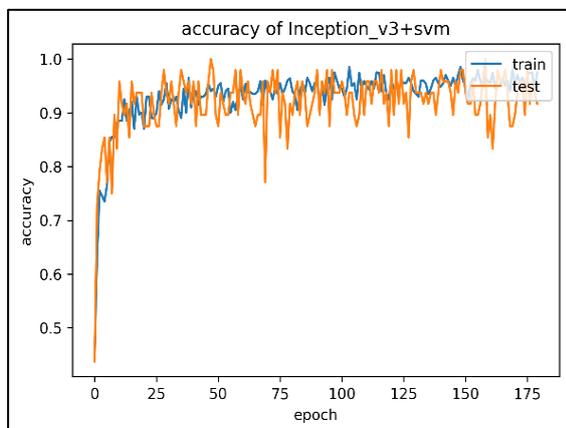
Vgg16



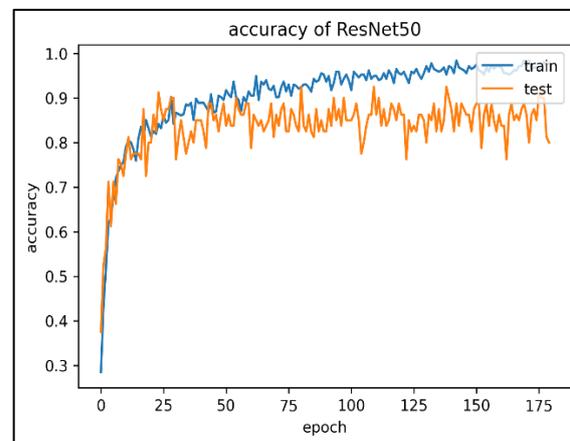
Vgg19+svm



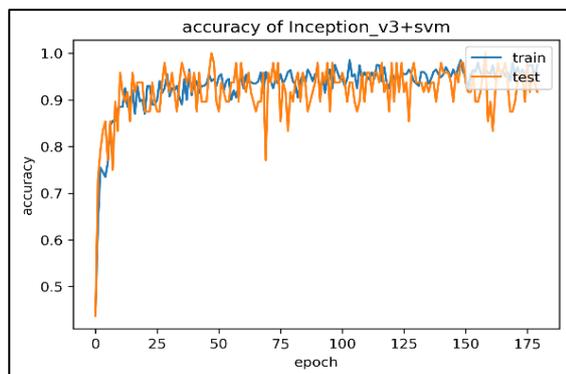
Vgg19



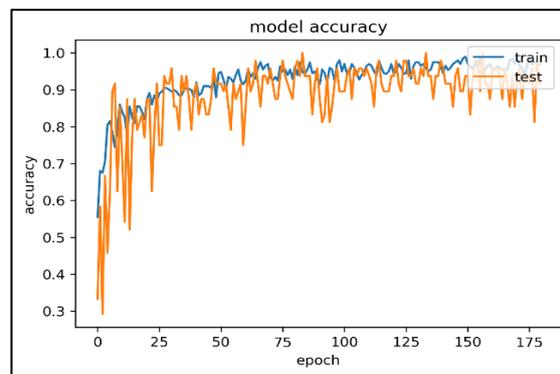
ResNet50+svm



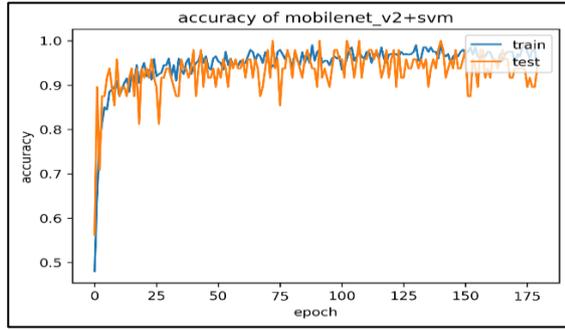
ResNet50



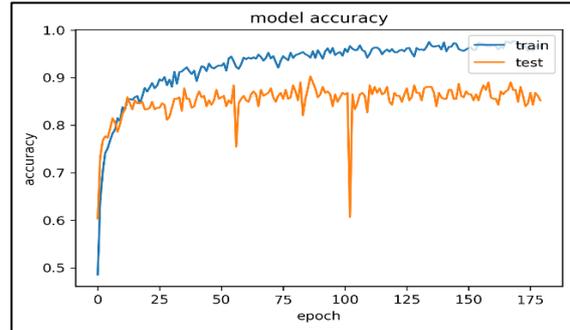
Inception_v3+svm



Inception_v3



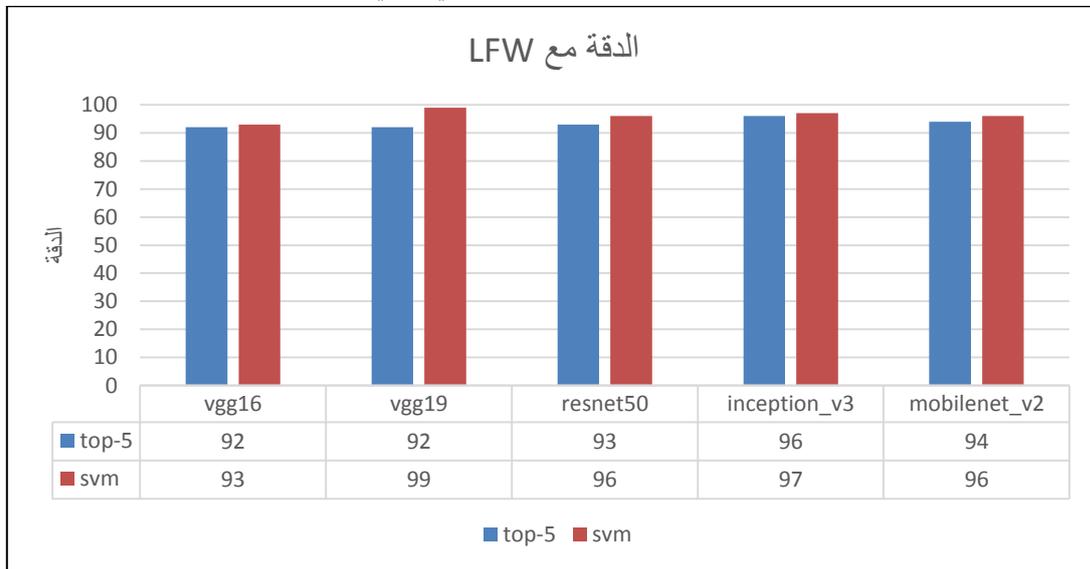
Mobilnet_v2+svm



Mobilnet_v2

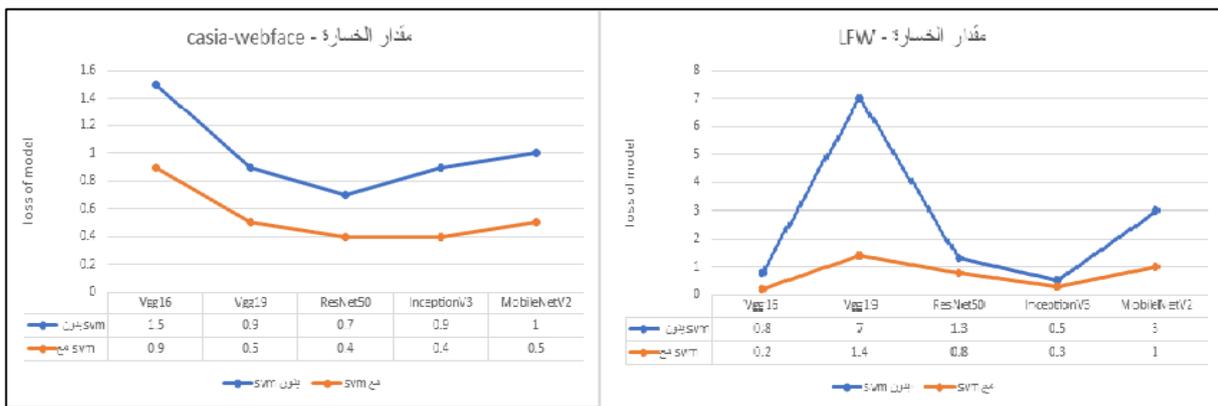
الشكل (8) نتائج تدريب الشبكات على عينة البيانات Casia-webface

من المخططات السابقة نلاحظ ارتفاع دقة جميع الشبكات المدروسة عند استخدام خوارزمية SVM فقد حصلنا على دقة أعلى من 91 في جميع الشبكات، وكانت دقة شبكة inception_v3 بدون SVM ومعه هي الأفضل. بينما عينة البيانات LFW سنورد نتائجها ضمن المخطط البياني التالي:



الشكل (9) نتائج تدريب النماذج على عينة البيانات LFW

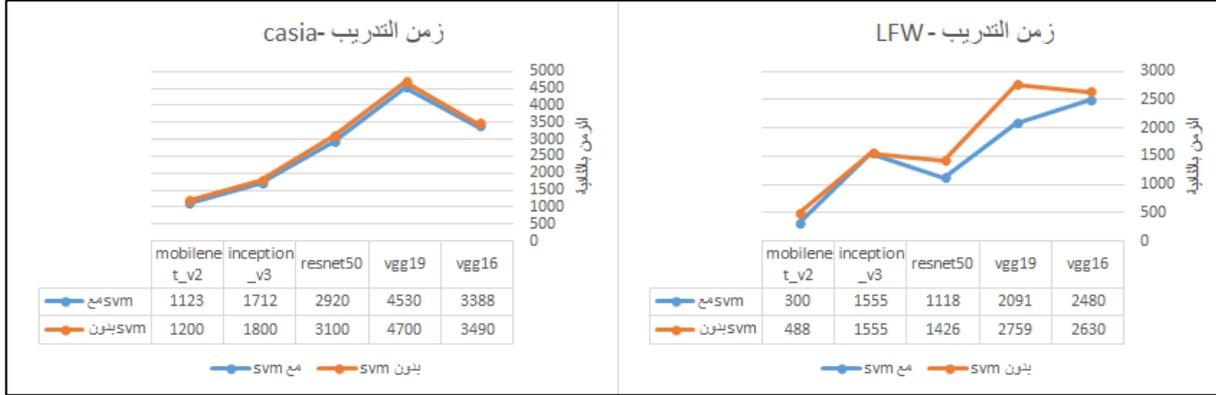
من المخطط السابق نلاحظ أيضاً ارتفاع دقة جميع الشبكات المدروسة عند استخدام خوارزمية SVM. ترافق ارتفاع دقة الشبكات المدروسة مع انخفاض مقدار الخطأ فيها، لكن عند محاولة تقليلها أكثر لم ترتفع دقة الشبكة أكثر مما استدعى التوقف عند هذا المقدار.



الشكل (10) مقدار خطأ الشبكات في عيني البيانات

المخططات السابقة توضح مقدار خطأ الشبكات الحاصل عند استخدامها خوارزمية SVM وعدم استخدامها، وقد لوحظ مع ارتفاع الدقة عند استخدام خوارزمية SVM انخفاض بخطأ الاختبار لدى كل الشبكات المدروسة.

• الدراسة من حيث زمن التدريب:



الشكل (11) زمن تدريب النماذج على عيني البيانات

تحقق حسب الشكل السابق انخفاض زمن التدريب عند استخدام خوارزمية SVM، حيث كان أقل زمن لشبكة mobilenet_v2 وذلك في عيني الدراسة، مع العلم أن الزمن يتعلق بحجم الشبكة المستخدمة ومعاملاتها ونوعية البيانات وحجمها.

• الدراسة من حيث نسبة خطأ التصنيف:

تم الاعتماد على اختبار أول 20 صورة لكل صنف بعينة Casia-webface وأول 16 صورة لكل صنف بعينة LFW، وقد تم اختيار عدد الصور بما يتناسب مع حجم كل عينة، حيث LFW تحوي صور أقل بكثير من العينة Casia-webface، بالإضافة إلى أنه بعد الاختبار تبين أن تغيرات النتائج أصبحت ضئيلة بعد هذا العدد فتم الاكتفاء به. الجداول التالية توضح نسب خطأ التصنيف في العينتين، والتي تم حسابها بالاعتماد على عينة الاختبار لكل صنف والبالغة 20% من صورته وذلك باختبار هذه الصور صورة تلو الأخرى وتسجيل عدد الصور المصنفة بشكل خاطئ ومن ثم حساب النسبة المئوية للخطأ الحاصل بكل صنف من عيني البيانات المدروسة.

الجدول (2) نتائج التصنيف بعينة Casia-webface

MobileNetV2		InceptionV3		ResNet50		Vgg19		Vgg16		الشبكة
مع svm	بدون svm	مع svm	بدون svm	مع svm	بدون svm	مع svm	بدون svm	مع svm	بدون svm	نسبة الخطأ
1	1	0.8	0.8	0.05	0.05	0.05	0.05	0.05	0.05	الصنف 1
1	1	0.4	1	0.1	0.3	0.5	0.5	0.05	0.2	الصنف 2
0.05	0.8	0.6	0.8	0.2	0.3	0.05	0.5	0.1	0.3	الصنف 3
0.8	0.9	0.8	1	0.2	0.2	0.8	0.9	0.4	0.9	الصنف 4
0.2	0.5	0.05	0.3	0.1	0.2	0.4	0.5	0.1	0.1	الصنف 5

الجدول (3) نتائج التصنيف بعينة LFW

MobileNetV2		InceptionV3		ResNet50		Vgg19		Vgg16		الشبكة
مع svm	بدون svm	مع svm	بدون svm	مع svm	بدون svm	مع svm	بدون svm	مع svm	بدون svm	نسبة الخطأ
1	1	0.8	0.9	0.05	0.3	0.1	0.05	0.05	0.05	الصنف 1
0.05	0.05	0.05	0.05	0.06	0.06	0.5	1	0.4	0.5	الصنف 2
0.6	0.8	0.7	0.7	0.5	0.6	0.1	1	0.02	0.03	الصنف 3
0.8	1	0.8	0.9	0.3	0.2	0.1	1	0.8	0.6	الصنف 4
0.6	0.5	0.4	0.4	0.2	0.4	0.4	0.4	0.2	0.2	الصنف 5

نلاحظ انخفاض نسبة خطأ التصنيف في أغلب الشبكات المدروسة على العينة Casia-webface عند استخدام SVM، وقد حصلنا على نتائج مقاربة أيضاً عند دراسة العينة LFW، وأقل نسبة خطأ تصنيف كانت في شبكة ResNet50.

مناقشة النتائج:

عموماً النتائج السابقة تؤكد على ارتفاع الدقة عند استخدام خوارزمية SVM مع كل شبكة من الشبكات المدروسة وذلك في عيني الدراسة بشكل متفاوت لكل منها، ويعود ذلك إلى اختلاف عدد طبقات الالتفاف والتجميع المسؤولين عن استخراج أكبر عدد من الميزات وانتقاء الأكثر أهمية منها في كل شبكة، بالإضافة إلى طبقات dropout والتي تلعب دوراً في التأثير على دقة الشبكة من أجل تجنب مشكلة overfitting، وبالتالي المصنف SVM يتلقى عدد من الميزات مختلف من كل شبكة مما أدى إلى اختلاف التحسين.

لوحظ من النتائج السابقة كانت شبكة inception_v3 الأعلى دقة قبل استخدام خوارزمية SVM لاعتمادها على طبقات الطي في زيادة عمق الشبكة وبالتالي زيادة عدد الميزات المستخرجة من الصور ومن ثم تصنيفها بطريقة الاتصال الكامل إلى عدة أصناف بالاعتماد على دالة softmax، وبعد استبدالها بخوارزمية svm زادت الدقة أكثر نظراً لكفاءة SVM في فصل البيانات غير الخطية. بينما كانت شبكة vgg19 الأكثر تحسناً عند استخدام خوارزمية svm لاعتماد بنيتها على أكثر من طبقة اتصال كامل لتصنيف الميزات المستخرجة من الصور إلى عدة أصناف، ومع استبدال الطبقة الأخيرة بخوارزمية svm أدى إلى تكديس المصنفات وبالتالي إلى ارتفاع الدقة وانخفاض مقدار خطأ الاختبار بشكل واضح.

كما لوحظ انخفاض زمن التدريب في جميع الشبكات المدروسة وذلك لسرعة تابع النواة الذي تم اختياره لعملية التصنيف في خوارزمية SVM، وكانت شبكة mobilenet_v2 الأقل زمناً قبل استخدام خوارزمية SVM لاعتمادها في بنيتها على طريقة الالتفاف النقطي لتقليل كمية البيانات المتدفقة عبر الشبكة بالإضافة إلى سرعة تابع النواة المستخدم في svm أدى ذلك إلى زمن تدريب أقل.

وبنتيجة التجريب لاختبار الشبكات المدروسة بعد استخدام المصنف SVM لوحظ انخفاض نسبة خطأ التصنيف في جميع الشبكات المدروسة.

خلاصة البحث:

تمت دراسة مجموعة من الشبكات العصبونية التلافيفية المستخدمة في مجال التصنيف عموماً وطبقناها على مسألة تصنيف الوجوه البشرية وكانت هذه الشبكات هي (Inception_V3 - ResNet50 - VGG19 - VGG16) - MobileNet_V2)، من ثم استبدال طبقة التصنيف في كل شبكة بمتجه دعم الآلة (SVM) لدراسة مدى تأثيره على كل شبكة من الشبكات سابقة الذكر من حيث الدقة وزمن التدريب ونسبة خطأ التصنيف، وقد تمت الدراسة على عينتي بيانات كبيرة تحويان صوراً ملونة (Casia-webface - LFW (Labeled Faces in the Wild)) ، وكانت النتائج بأننا توصلنا إلى زيادة دقة جميع الشبكات المدروسة بحيث جميعها تجاوزت 90، وإلى انخفاض في زمن تدريب هذه الشبكات بمقدار تراوح ما بين 100 إلى 200، بالإضافة إلى انخفاض عام في نسبة خطأ تصنيف جميع الشبكات المدروسة.

التوصيات والمقترحات.

بناءً على النتائج التي تم التوصل إليها توصي الباحثان وتقترحان ما يلي:

1. كان الهدف من الدراسة هو دراسة الشبكات التلافيفية الأكثر استخداماً في مجال التصنيف وتطبيقها على مسألة تصنيف الوجوه البشرية من ثم دراسة تأثير تطبيق خوارزمية SVM على الشبكات المدروسة من ناحية الدقة والزمن ونسبة الخطأ في التصنيف، وتم الوصول إلى نتائج جيدة من حيث زيادة الدقة وتخفيض الزمن وتخفيض نسبة الخطأ.
2. نظراً لأهمية خوارزمية SVM في مجال التصنيف غير الخطي وتوسعها في إيجاد توابع النواة المناسبة لفصل البيانات غير الخطية، فيمكن الاستفادة من هذه الخوارزمية في تطبيقها على أنواع أخرى من الشبكات العميقة والاستفادة منها في مجالات تصنيف أخرى.
3. يمكن تطبيق هذه الدراسة على مسائل تصنيف في مجالات عديدة مهمة منها الطبية مثلاً، كما يمكن توسيع هذه الدراسة أيضاً لتشمل تطبيق أنواع أخرى من خوارزميات التصنيف على عينة البيانات ومقارنة نتائجها مع نتائج الدراسة الحالية.

قائمة المراجع.

- C. Burges (1998) - A tutorial on support vector machines for pattern recognition, Data Mining Knowledge Discovery, vol. 2, No.2, pp. 121-167.
- C. Cortes, V. Vapnik (1995) - Support vector networks, Machine Learning, vol. 20, pp. 273-297.
- C. Sedrick, Z. Michele (2020) - Transfer Learning in Image Classification: how much training data do we really need?
- C.Szegedy, W.Liu,Y.Jia, P.Sermanet, S.Reed, D.Anguelov, D.Erhan, V.Vanhoucke, A.Rabinovich (2014)- Going deeper with convolutions, CoRR,vol. abs/1409.4842.
- D. Sinha, M. El-Sharkawy (2019) - Thin MobileNet: An Enhanced MobileNet Architecture, IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON).
- G. Amirthalingam, S. Subramaniam (2017)- Multi Modal Biometric System: A Review on Recognition Method, International Journal of Engineering Research & Technology (IJERT), Vol. 6 ,No. 5.

- G. Guo, N. Zhang (2019) - A survey on deep learning based face recognition, Computer Vision and Image Understanding.
- G. Levi, T. Hassner (2016) - Age and Gender Classification using Convolutional Neural Networks.
- Gruber, M. Hlaváč, M. Zelezný, A. Karpov (2017) - Facing Face Recognition with ResNet: Round One, Springer International Publishing, pp. 67–74.
- Huang, G.B. Ramesh, M. Berg, T. Learned-Miller (2007) - Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Available online: <https://hal.inria.fr/inria-00321923> (accessed on 1/9/2021).
- J. M. Alrikabi , K.H. Alibraheemi (2021)- Feature Level Combination for Face Recognition Based on Convolutional Neural Networks, Journal of Al-Qadisiyah for Computer Science and Mathematics, Vol. 13, No.3, pp. 99–113.
- K. Shlizerman, I. Seitz, S.M. Miller, D. Brossard (2016) - The Mega Face benchmark: 1 million faces for recognition at scale.
- K. Simonyan, A. Zisserman (2015)-Very deep convolutional networks for large-scale image recognition, in International Conference on Learning Representations.
- L. L. S. Deffo, E.T. Fute, E. Tonye (2018) - CNNsFR: A Convolutional Neural Network System for Face Detection and Recognition, International Journal of Advanced Computer Science and Applications, Vol. 9, No. 12.
- L. Sushma, K.P. Lakshmi (2020)- An Analysis of Convolution Neural Network for Image Classification using Different Models, International Journal of Engineering Research & Technology (IJERT), Vol. 9, No. 10.
- M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen (2019)-MobileNetV2: Inverted Residuals and Linear Bottlenecks, Mobile networks for classification, detection and segmentation," CoRR, vol. abs/1801.04381.
- N. Alipour, O. Tarkhaneh, M. Awrangjeb, H.Tian (2021)- Flower Image Classification Using Deep Convolutional Neural Network, Journal of Korea Multimedia Society Vol. 23, No. 4.
- R.H. Khan, K.W. Kang, S.J. Lim, S.D. Youn, Oh-J. Kwon, S.H. Lee, K.R. Kwon (2020) - Animal Face Classification using Dual Deep Convolutional Neural Network.
- S. Rosero Vasquez (2020) - Facial Recognition: Traditional Methods vs Methods Based on Deep Learning, Springer Nature Switzerland.
- S. Seeri, P.S. Hiremath, Jayalaxmi G.N (2020) - Flower Classification using Deep Learning models, IEEE.
- S.J. Pan, and Q. Yang (2010) - A survey on transfer learning, IEEE.
- T. Gwyn, K. Roy, M. Atay (2021) - Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study, Future Internet, <https://doi.org/10.3390/fi13070164>.

- V. Vanhoucke, C. Szegedy, S. Ioffe, J. Shlens, Z. Wojna (2015)- Rethinking the Inception Architecture for Computer Vision, CoRR, vol. abs/1512.00567.
- V. Vapnik (1998) - Statistical Learn Theory, John Wiley, New York.
- W. Dai, Y. Dai, K. Hirota, Z. Jia (2020)- A Flower Classification Approach with MobileNetV2 and Transfer Learning, The 9th International Symposium on Computational Intelligence and Industrial Applications, vol.3.
- X. Zhang, K. He, S. Ren, and J. Sun (2015)-Deep residual learning for image recognition, CoRR, vol. abs/1512.03385.
- Z. Wang, X. Xie (2010) - An Efficient Face Recognition Algorithm Based on Robust Principal Component Analysis.