

## Comparison The Performance of The Recurrent Neural Network in Reducing Training Parameters for Convolution Neural Network

Mohamad Ahmad Mounir Batikh

Mohamad Ayman Nael

Amer Boushi

Faculty of Electric and Electronic Engineering || Aleppo University || Syria

**Abstract:** The study aims to reduce the number of parameters in the Convolution Neural Network (CNN), which is one of the best techniques used to extract and categorize behavioral features in video files. These networks have a very big size and a large number of parameters which distributed in the deep layers, especially in the last layers that responsible for classification, the values of the parameters are modified at each stage of network training, which consume memory too much, and its need for a very large memory space. In this research we work to reduce the number of parameters through using lightweight Convolution Neural Network (LWCNN), we choose Alex Net network in our research, but we made some modification on it, we decrease the number of filters in convolution layer, and we replace the last layers in the network with one of the most important types of Recurrent Neural Network (RNN). We use each of Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU). The work was tested during the research period on a dataset containing 960(videos) for normal children and children with autism spectrum, which were taken in Center for psychosocial support for people with special needs. And the experimental results proved the significant decrease in the number of parameters in the system with lightweight networks after linking them with recurrent networks with 84%, as well as the recurrent network with long reliability (LSTM) gave better results than the Gated recurrent unit (GRU) in accuracy and the loss value.

**Keywords:** Recurrent Neural network, Convolution Neural network, LWCNN, LSTM, GRU.

## مقارنة أداء الشبكات العصبونية التكرارية في تخفيض بارامترات التدريب لشبكات التلافيف العصبية

محمد أحمد منير بطيخ

محمد أيمن نعال

عامر بوشي

كلية الهندسة الكهربائية والإلكترونية || جامعة حلب || سورية

المستخلص: عمل هذا البحث إلى تخفيض عدد البارامترات الموجودة في شبكات الطي العصبية (Convolution Neural Network CNN) التي تعد من أفضل التقنيات المستخدمة لاستخلاص السمات السلوكية وتصنيفها في الملفات المرئية (Videos)، تمتلك هذه الشبكات احجام كبيرة جداً وعدد كبير جداً من البارامترات الموزعة في الطبقات العميقة وخاصة في الطبقات الأخيرة المسؤولة عن التصنيف، والتي يتم تعديل قيمها في كل مرحلة من مراحل تدريب الشبكة وهو ما يسبب استهلاك كبير للذاكرة والحاجة إلى مساحة ذاكرية كبيرة جداً. تم العمل من خلال هذا البحث إلى تخفيض عدد البارامترات وذلك من خلال استخدام شبكات الطي العصبية ذات الاوزان

الخفيفة (Light Weight Convolution Neural Network LWCNN) حيث تم استخدام شبكة Alex Net وذلك بعد إجراء مجموعة من التعديلات عليها، من تخفيض عدد المرشحات المستخدمة في طبقات الطي، وكذلك استبدال الطبقات الاخيرة من الشبكة المستخدمة التي تكون مسؤولة عن التصنيف بأحد أهم أنواع الشبكات التكرارية (Recurrent Neural Network RNN) حيث تم استخدام كلاً من الشبكة التكرارية ذات الاعتمادية الطويلة (Long Short Term Memory LSTM) والشبكة التكرارية المبوبة (Gated Recurrent Unit GRU). تم اختبار العمل خلال فترة البحث على مجموعة بيانات تضم 960 فيديو لأطفال طبيعيين وأطفال مصابين بطيف التوحد، تم تسجيلها في أحد مراكز الدعم النفسي لذوي الاحتياجات الخاصة. وقد أثبتت النتائج التجريبية الانخفاض الكبير في عدد البارامترات في الشبكات العصبونية ذات الأوزان الخفيفة بعد ربطها مع الشبكات التكرارية بمقدار 84%، كذلك فإن الشبكة التكرارية ذات الاعتمادية الطويلة LSTM أعطت نتائج أفضل من الشبكة التكرارية المبوبة GRU من حيث الدقة والخطأ، حيث زادت الدقة بمقدار 3% وانخفض الخطأ بمقدار 9%.

الكلمات المفتاحية: الشبكات العصبونية التكرارية، شبكات الطي العصبونية، LWCNN، LSTM، GRU.

## 1- المقدمة.

يعد استخلاص السمات السلوكية وتصنيفها من الملفات المرئية أحد أهم مجالات الرؤية الحاسوبية (Computer Vision). يوجد العديد من الطرق التقليدية السابقة (Traditional Handcrafted Method) التي تم استخدامها لاستخلاص السمات السلوكية وتصنيفها، لكن هذه الطرق لم تستطع أن تعطي الدقة المطلوبة في تصنيف السلوك البشري الذي يتضمن الكثير من الحركات معقدة التي يصعب اكتشافها بسهولة. لكن مع ظهور الشبكات العصبونية العميقة Deep Neural Network واستخدامها في استخلاص السمات السلوكية البشرية، استطاعت أن تحقق نجاحاً كبيراً وزيادة واضحة في دقة التصنيف. تعد شبكات الطي العصبونية CNN أحد أهم أنواع الشبكات العصبونية العميقة ذات التغذية الأمامية Feed Forward Neural Network، حيث أنها تعمل بنفس المبدأ الذي يتم العمل به في الفص البصري في دماغ الإنسان [10]، وقد كانت حلاً للعديد من المشاكل في مجال الرؤية الحاسوبية من معالجة الصور وملفات الفيديو، وذلك لما تتميز به عن غيرها من الشبكات العصبونية الأخرى من الخصائص في بنيتها التصميمية. تتكون شبكات الطي العصبونية من عدة طبقات كل منها يؤدي وظيفة معينة خلال عملية استخلاص السمات السلوكية وتصنيفها. لكن المشكلة الأساسية في هذا النوع من الشبكات هو المساحة الذاكرة الكبيرة التي تتطلبها، بالإضافة إلى الكم الكبير من عمليات المعالجة الضخمة وخصوصاً خلال تدريب الشبكة العصبونية، بالإضافة إلى العدد الضخم من البارامترات الموزعة في طبقات الشبكة وما تسببه من استهلاك كبير للذاكرة لأن هذه البارامترات تتغير قيمها بشكل دوري في كل مرحلة من مراحل تدريب الشبكة. فعلى سبيل المثال شبكة ALEX NET شبكة طي عصبية حققت نجاحاً كبيراً في زيادة دقة التعرف والتصنيف مقارنة بالطرق التقليدية، لكن المشكلة الأساسية تكمن في عدد البارامترات الموجودة في الشبكة والتي تبلغ 62, 378, 344 والتي يوجد أكثر من 90% منها في الطبقات الأخيرة المسؤولة عن التصنيف. كذلك شبكة VGG- 19 تحوي حوالي 138 مليون بارامتر، كذلك شبكة Google Net التي تمتلك 7 مليون بارامتر، وشبكة ResNet التي تحوي 25 مليون بارامتر [1].

## 2- مشكلة الدراسة:

يعد تخفيض عدد البارامترات في الشبكة العصبونية ذو أهمية كبيرة جداً في تسريع عملية التدريب والتخفيض من استهلاك الذاكرة، لذلك يمكن صياغة مشكلة الدراسة في التساؤلات التالية:

1. هل يمكن استبدال الشبكة المطوية العصبونية بشبكة تؤدي نفس المهمة وتعطي نفس الدقة ولكن بعدد أقل من البارامترات.

2. هل يمكن استبدال الطبقات ذات البارامترات الكثيرة في الشبكة دون أن تتأثر النتائج التجريبية.

3- فرضيات الدراسة:

تفترض الدراسة:

1. استخدام شبكات الطي العصبية ذات الأوزان الخفيفة LWCNN.

2. استبدال الطبقات الأخيرة في LWCNN بأنواع معينة من الشبكات العصبية التكرارية RNN.

4- هدف البحث:

إن الهدف الأساسي من هذا البحث هو تطوير نظام يقوم باستخراج السمات السلوكية من الملفات المرئية باستخدام شبكات الطي العصبية ذات الأوزان الخفيفة مع استخدام أقل عدد ممكن من البارامترات فيها. وذلك من خلال الاعتماد على حذف الطبقات الأخيرة المسؤولة عن التصنيف التي تتركز البارامترات فيها بشكل كبير، واستبدالها بأحد أنواع الشبكات العصبونية التكرارية. وبالتالي بناء شبكة طي عصبية خفيفة تقوم باستخلاص السمات وتصنيفها وذلك بعد تخفيض عدد البارامترات، مع المحافظة على الدقة وبأقل خطأ ممكن.

5- أهمية الدراسة:

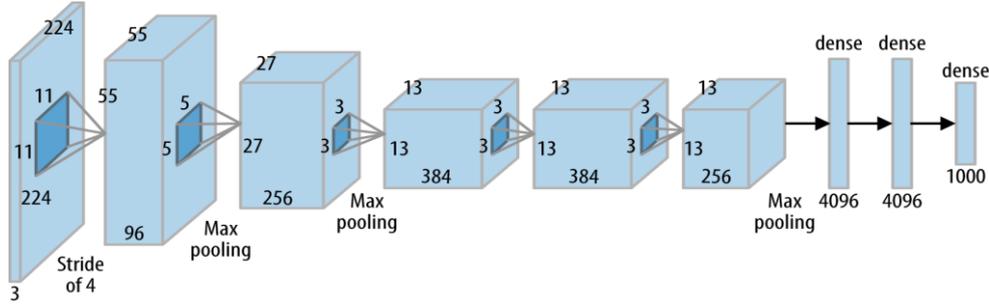
تنبع الأهمية العلمية للدراسة من أن تخفيض عدد البارامترات في شبكة الطي العصبية سيساعد كثيراً في تخفيض التعقيد الحسابي وكذلك استهلاك الذاكرة. بالإضافة إلى تسريع عملية التدريب في الشبكة العصبونية. وهنا سيتم الاستفادة أيضاً من خصائص الشبكات التكرارية في استخدام المعلومات السابقة في المعالجة الحالية، وهو ما يعطي دقة كبيرة في عملية التصنيف.

6- المبحث الأول- الإطار النظري والدراسات السابقة

1-6 الإطار النظري:

6-1-1 مفهوم شبكات الطي العصبية ذات الأوزان الخفيفة LWCNN:

نشأت فكرة بناء شبكات طي عصبية ذات أوزان خفيفة (Light Weight Convolution Neural Network) LWCNN من أجل إنشاء تطبيقات تعمل على الأجهزة المحمولة (Mobile Devices) وكذلك على أجهزة الحاسب محدودة الموارد مثل الحواسيب التي تستخدم CPU بدلاً من استخدام وحدة معالجة الرسومات (Graphic Processing Unit GPU). وأهم ما يميز استخدام شبكات LWCNN هو المحافظة على الدقة العالية في الأداء من خلال التخفيض الكبير في عدد الطبقات المستخدمة في الشبكة، وما يرافقه من إنقاص في عدد البارامترات (الأوزان)، مما يؤدي إلى زيادة في مساحة التخزين، والزيادة في السرعة وإنقاص الحسابات المعقدة الموجودة في شبكات CNN [3] فمن أجل زيادة الدقة في شبكات CNN يجب العمل على زيادة عدد طبقات الشبكة وخاصةً الطبقات المطوية (Convolutional Layer) [14]، ففي شبكة Alex Net تم استخدام 8 طبقات في الشبكة، ثم تم زيادة الدقة في شبكة VGG من خلال استخدام 16 طبقة في تصميم الشبكة، وفي شبكة Google Net تم استخدام 22 طبقة، أما في شبكة Res Net التي حققت نجاحاً كبيراً في زيادة الدقة، فقد تم استخدام 152 طبقة في بنية الشبكة. والشكل (1) التالي يوضح بنية شبكة Alex Net إحدى شبكات الطي العصبية مع الطبقات المكونة لها، من طبقات الطي الخمسة، وما يتبعها من طبقات التجميع (Max Pooling) وطبقات الارتباط الكلي الثلاث (Dense Layer) [9].



الشكل (1) طبقات شبكة Alex Net

## 2-1-6 مفهوم الشبكات العصبية التكرارية RNN:

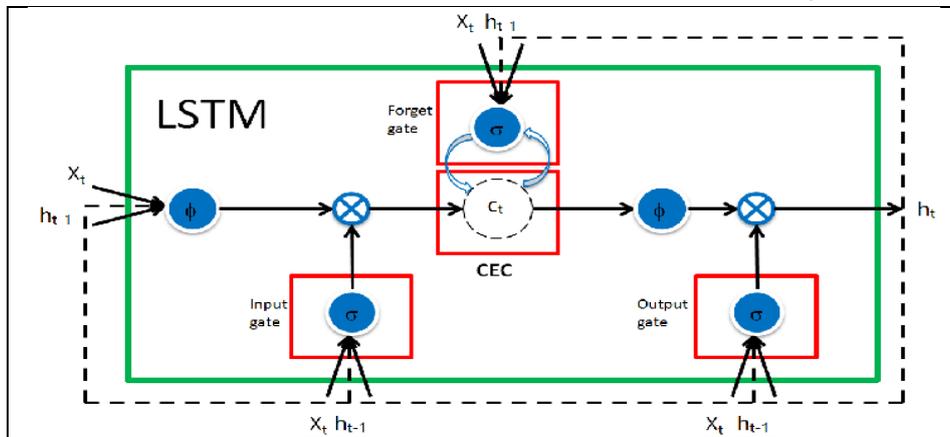
على عكس الشبكات العصبية المطوية حيث يكون الدخل مستقل تماماً عن الخرج، فإن الشبكات التكرارية تستخدم الخرج الناتج من المرحلة السابقة كدخل للمرحلة الحالية. كما تمتلك الشبكات التكرارية ميزة هامة وهي التسلسل المتتابع في الترابط سواء في إطارات الدخل أو الخرج. تمتلك الشبكات التكرارية ذاكرة قصيرة المدى بحيث لا تستطيع معالجة الاعتمادية الطويلة بين اطارات الدخل [11]. ولحل هذه المشكلة وخصوصاً في ملفات الفيديو تم إنشاء شبكات تكرارية ذات مجال زمني كبير لربط السمات المستخرجة من الإطارات للوصول الى حركة معينة خلال الزمن [7] ومن أهمها:

### 1- شبكة (Long Short Term Memory LSTM):

في الشبكة التكرارية RNN لا يتم حفظ المعلومات لمدة زمنية طويلة، حيث يمكن أن يتم ضياع معلومات مهمة للحركة وبالتالي سينتج خطأ في التصنيف. أما في شبكة LSTM فهي تحوي ما يسمى بالبوابات التي تقوم بتنظيم تدفق المعلومات، وأثناء التدريب تقوم البوابات بالتعلم وحفظ المعلومات المهمة والمعلومات التي يجب حذفها. بالإضافة إلى ذلك فإنه تم حل مشكلة انعدام قيم الأوزان في شبكة RNN خلال عملية التدريب [12]. تحتوي شبكة LSTM ثلاث بوابات رئيسية:

1. بوابة الإدخال Input Gate: مسؤولة عن تحديد المعلومات المضافة في المرحلة الحالية.
2. بوابة النسيان Forget Gate: مسؤولة عن تحديد المعلومات المحفوظة من المرحلة السابقة.
3. بوابة الخرج Output Gate: مسؤولة عن تحديد المعلومات من المطلوبة من المرحلة الحالية إلى المرحلة التالية.

والشكل (2) يوضح بنية شبكة LSTM وبواباتها الثلاثة [12].

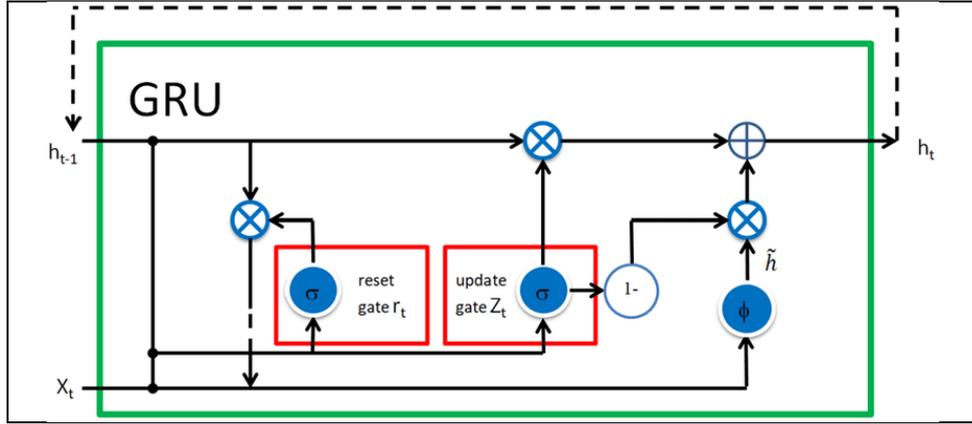


الشكل (2) بنية شبكة LSTM

## 2- شبكة (Gated Recurrent Unit GRU):

تم إصدار شبكة GRU في عام 2014 وتعتبر أحد أحدث إصدارات الشبكات التكرارية RNN وهي مشابهة في مبدأ عملها لشبكة LSTM لكنها تحوي بوابتين فقط.

1. بوابة التعديل Update Gate: تعمل بشكل مشابه لبوابة (Input+ Forget) في LSTM، حيث تحدد المعلومات المضافة في المرحلة الحالية والمعلومات المحفوظة من المرحلة السابقة.
  2. بوابة التصفير Reset Gate: مسؤولة عن تحديد المعلومات المحذوفة من المرحلة السابقة [4].
- والشكل (3) يوضح بنية وبوابات شبكة GRU [12].



الشكل (3) بنية شبكة GRU

## 6-2 الدراسات السابقة

بعد تدريب الشبكة المطوية العصبية نلاحظ أن هناك بعض الأوزان (البارامترات) غير مهمة في النموذج المدرب (Trained Model). لذلك يجب حذف هذه العناصر ومن ثم إعادة ربط طبقات الشبكة بشكل مناسب مع التأكيد على ضرورة استمرار تدفق المعلومات بشكل صحيح خلال النموذج.

يوجد العديد من الدراسات السابقة التي هدفت لإنقاص حجم الشبكة والتي يمكن تصنيفها كما يلي [13]:

- 1- الطرق المعتمدة على إنقاص حجم الشبكة: وذلك من خلال دراسة تأثير حذف العصبونات والأوزان على الطبقة التالية لها، حيث يتم حذف الأوزان ذات التأثير الأقل والتي لا تؤثر على جودة الدقة.
- 2- الطرق المعتمدة على التشابه والتجميع: حيث يتم حذف الأوزان المتشابهة والمكررة والتي تكون فائضة ولا تؤثر على الدقة في حال حذفها.
- 3- الطرق المعتمدة على مصفوفة البارامترات: والتي تعتمد على تقسيم مصفوفة البارامترات إلى ناتج مصفوفتين أصغر، يمكن استخدامهم لأداء وظيفة مكافئة بشكل أكثر كفاءة من مصفوفة الوزن الأكبر.
- 4- الطرق المعتمدة على التكميم (Quantization): حيث يتم تحويل الأوزان إلى قيم صحيحة (integer) بدلاً من القيم الأساسية التي تكون بصيغة الفاصلة العائمة (floating point) وهذا ما سيؤدي إلى استهلاك أقل للذاكرة والحسابات تتم بشكل أسرع.
- 5- الطرق المعتمدة على التصميم: وذلك من خلال استخدام الخوارزميات والطرق الذكية لبناء شبكة عصبونية تكون طبقاتها خفيفة وبنفس جودة الأداء للشبكات القياسية.
- 6- الطرق الهجينة (Hybrid): التي تعتمد على دمج أكثر من طريقة في بناء الشبكة العصبية لإنقاص عدد البارامترات فيها. وقد تم اختيار بعض الدراسات التي تمت باستخدام الطرق السابقة ومنها:

1. دراسة Debes Jha وآخرون [6] حيث قام الباحثون في هذه الدراسة ببناء شبكة طي عصبية تعتمد في تصميمها على طبقات خفيفة وبدون الاعتماد على نموذج مدرب مسبقاً (Transfer Learning). حيث تم استخدام طبقة طي (Convolution Layer)  $3 \times 3$  أي تستخدم مرشح أبعاده  $3 \times 3$ ، وطبقة تجميع (pooling Layer) من نوع  $2 \times 2$  Max Pooling وطبقة ارتباط كلي (Dense layer). وقد تم استخدام ثلاث بلوكات من هذه الطبقات في تصميم الشبكة المطوية. ما يميز هذا النوع هو استخدام طبقة طي مختلفة عن الطبقة القياسية حيث تم استخدام طبقة طي تضم كلا من (Depthwise Spatial Convolution DW)، التي يكون فيها كل مرشح Kernel مسؤول عن ميزة Feature واحدة من الدخل، ومن ثم سيتم استخدام (Pointwise Convolution  $1 \times 1$ ) على الخرج الناتج من DW. وبالتالي يكون قد تم اختزال في عدد البارامترات والأوزان. لكن تبقى مشكلة طبقة الارتباط الكلي التي يتم فيها ارتباط كل عصبون بجميع العصبونات من الطبقة السابقة.
2. دراسة Dawei Dai وآخرون [2] حيث قام الباحثون من أجل تخفيض عدد البارامترات في هذا البحث بتطبيق نفس الأوزان (البارامترات) مرة واحدة على الطبقات المطوية المتشابهة الموجودة في نفس الكتلة (Block)، بدلاً من تطبيق البارامترات على كل طبقة طي لوحدها. تم تطبيق هذه الدراسة على شبكة ResNet وهي أحد الشبكات المطوية العصبية التي تضم 152 طبقة، بالإضافة إلى 34 طبقة من البارامترات المدربة في كل كتلة من الشبكة. وقد تم إنقاص البارامترات بمقدار الثلث من العدد الكلي مع المحافظة على الدقة. تم التركيز في هذه الدراسة على البارامترات الموجودة في الطبقات المطوية فقط، أي العمل على تخفيض عدد البارامترات في الطبقات المطوية من الشبكة فقط، رغم أن العدد الأكبر من البارامترات يكون موجود في طبقات الارتباط الكلي في شبكات الطي العصبية.
3. دراسة Qiangui Huang وآخرون [5] حيث قام الباحثون بإنشاء خوارزمية ("Try and Learn") وذلك بغية إنقاص حجم الشبكة من خلال حذف المرشحات غير الضرورية في طبقات الشبكة مع المحافظة على جودة الأداء. فإذا كان لدينا مجموعات المرشحات في طبقة طي معينة من الشبكة ولتكن  $L$ ، حيث تكون على الشكل التالي:  $WL = \{WL1, WL2, WL3, \dots, WLN\}$ ، ويتم تنظيم مجموعة من القيم الثنائية لتحديد أهمية المرشح أو عدم أهميته كما يلي  $AL = \{aL1, aL2, aL3, \dots, aLN\}$  حيث تكون قيم  $a$  / 0 أو  $1/a$ ، فإذا كانت قيمة  $aLi = 0$  فهذا يعني أن المرشح غير مهم وبالتالي يتم حذفه، أما إذا كانت قيمة  $aLi = 1$  فهذا يعني أن المرشح مهم ويتم الاحتفاظ به. لكن المشكلة الأساسية في هذه الطريقة أنها تعتمد على التجريب ومن ثم التقييم لمجموعة المرشحات التي تم استخدامها في استخلاص السمات.

## 7- المبحث الثاني- مواد البحث وطرائقه

قبل البدء بالطريقة المقترحة يجب أن يتم التعريف بالشبكة العصبية المطوية المستخدمة في بحثنا.

### 7-1 شبكة Alex Net:

تم طرح هذه الشبكة عام 2012 من قبل الباحث Alex Krizhensky وزملاءه [8]، وقد كانت أول شبكة عصبية مطوية تعطي دقة كبيرة في التعرف والتصنيف مقارنة بالطرق التقليدية السابقة المستخدمة في هذا المجال. تتضمن هذه الشبكة ثماني طبقات موزعة على الشكل التالي: خمس طبقات طي (Convolution Layer)، تكون الطبقة المطوية الأولى والأخيرة متبوعة بطبقة تجميع (Pooling Layer) يتم تطبيق تابع (Max) من خلالها. كذلك يتم استخدام طبقة (Batch Normalization BN) بعد كل طبقة طي من أجل ضبط قيم الخرج ضمن مجال معين. أما

تابع التفعيل (Activation Function) المستخدم في هذه الشبكة فهو التابع (RELU) الذي يساعد في تسريع عملية التدريب بشكل كبير، أخيراً يتم استخدام طبقة (Flatten) من أجل تحويل شعاع البيانات إلى شعاع ذو بعد واحد ليتم تمريره إلى طبقة التصنيف (Classification Layer)، التي تتضمن ثلاث طبقات من الارتباط الكامل (Dense Layer)، حيث يتم استخدام تابع التفعيل (Softmax) في الطبقة الأخيرة منها. والجدول (1) يوضح البنية التفصيلية لشبكة Alex Net مع عدد المرشحات وأحجامها في كل طبقة، بالإضافة إلى حجم خطوة الانتقال للمرشح المستخدم (Stride)، وكذلك الخرج الناتج عن كل طبقة وذلك بعد تحديد إطار الدخل للشبكة يبدأ بأبعاد 60x60 وعدد قنوات الدخل هو 3.

الجدول(1): طبقات شبكة Alex Net

Layers	Filters	Filter size	Stride	Activate Function	Output	Params
Input	-	-	-	-	60x60x3	0
Conv2D	96	11x11	4	relu	15x15x96	34944
Maxpooling	-	3x3	2	-	7x7x96	0
Conv2D	256	5x5	1	relu	7x7x256	614656
Maxpooling	-	3x3	2	-	3x3x256	0
Conv2D	384	3x3	1	relu	3x3x384	885120
Conv2D	384	3x3	1	relu	3x3x384	1327488
Conv2D	256	3x3	1	relu	3x3x256	884992
Maxpooling	-	3x3	2	-	1x1x256	0
Flatten	-	-	-	-	256	0
Dense1	-	-	-	relu	4096	1052672
Dense2	-	-	-	relu	4096	16781312
Dense3	-	-	-	softmax	2	8194

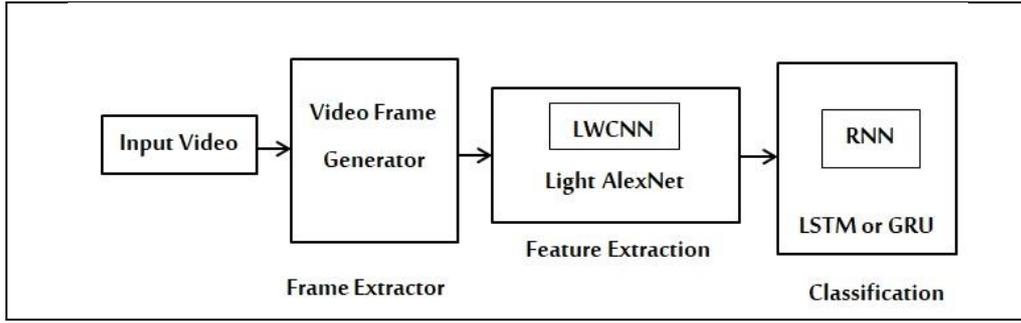
من الجدول (1) نلاحظ أن عدد البارامترات الكلي في شبكة Alex Net وذلك بعد إضافة البارامترات المستخدمة في كل طبقة BN سيكون 21, 591, 810 بارامتر. حيث يتم حساب عدد البارامترات في كل طبقة حسب القانون التالي:

$$p = ((m*n*d) + 1)*k \quad (1)$$

حيث: (m, n) أبعاد المرشح المطبق، d: عدد قنوات الطبقة السابقة، k: عدد قنوات الطبقة الحالية.

## 2-7 النظام المقترح:

يعتمد النظام المقترح على استخدام شبكة (Alex Net) ولكن بعد إجراء مجموعة من التعديلات عليها لتحويلها إلى شبكة عصبية مطوية خفيفة LWCNN، لأن استخدام هذه الشبكة مع عدد قليل من العينات سيؤدي إلى الوقوع في مشكلة (under fitting)، وهي استخدام شبكة كبيرة مع عينات قليلة، وعند التدريب سنحصل على خطأ كبير بحدود 40%. ثم سيتم ربط الشبكة الخفيفة مع إحدى الشبكات التكرارية (LSTM, GRU)، وذلك لأن شبكات الطي ذات الأوزان الخفيفة لا يمكن استخدامها في استخلاص السمات السلوكية وتصنيفها لأنها لا تمتلك طبقة للتزامن بين الإطارات، والشكل (4) يوضح بنية النظام المقترح



الشكل (4) بنية النظام المقترح

أما خطوات عمل النظام المقترح فهي على الشكل التالي:

- 1- إدخال الفيديو
  - 2- تقسيم الفيديو إلى إطارات وإجراء المعالجة المسبقة لإطارات الفيديو.
  - 3- استخراج السمات باستخدام شبكة Alex Net المعدلة الخفيفة LWCNN Alex Net.
  - 4- تصنيف السمات المستخلصة باستخدام الشبكات التكرارية RNN.
- علماً انه سيتم إعادة تنفيذ الخطوة (4)، وذلك بعد ربط خرج الشبكة الناتج في الخطوة (3) مع الشبكة التكرارية LSTM و مرةً أخرى مع شبكة GRU. في الفقرات التالية سيتم شرح خطوات العمل بالتفصيل.

1- المعالجة المسبقة لإطارات الفيديو:

بعد تمرير الفيديو يتم الاستعانة بمكتبة Keras الموجودة ضمن لغة Python لاستخدام التوابع الموجودة بداخلها، حيث تم استخدام تابع لتوليد الإطارات من الفيديو Video Frame Generator. وقد تم تحديد خواص هذه الإطارات على الشكل التالي:

1. أبعاد الإطارات أي العرض والارتفاع (60x60) بيكسل.
  2. عدد الإطارات التي سيتم تمريرها إلى الشبكة  $F=5$ .
  3. حجم الدفعة Batch Size BS، أي المقدار التسلسلي للانتقال بين الإطارات الخمسة التي تم اختيارها، وقد تم تحديده  $N=48$ ، أي إذا تم البدء بالإطار رقم 1 سيكون الإطار الثاني هو الإطار رقم 49.
  4. تم اختيار الإطارات RGB أي أن عدد القنوات Channels سيكون  $C=3$ .
- أي أن الشكل النهائي سيكون (5, 60, 60, 3, N)، علماً أن حجم الدفعة سيتم تبديله إلى القيمة المطبقة وذلك عند تمرير البيانات إلى الشبكة، ومن ثم ننتقل إلى الخطوة الثالثة وهي تمرير الإطارات التي تم اختيارها إلى شبكة Alex Net المعدلة من أجل استخراج السمات باستخدامها.

2- استخراج السمات باستخدام شبكة الطي الخفيفة:

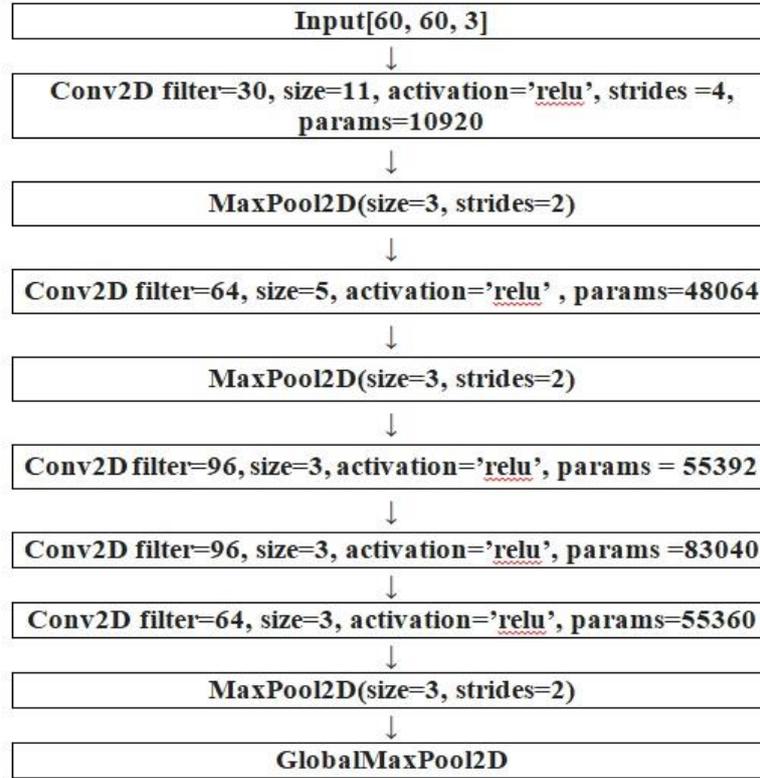
- شبكة Alex Net المعدلة (Alex Net Light Weight LW):

إن التعديل المطبق على شبكة (Alex Net) لنحصل على الشبكة الجديدة ذات الأوزان الخفيفة (Alex Net)

(LW) المستخدمة في النظام المقترح يكون على الشكل التالي:

1. إنقاص عدد المرشحات المستخدمة لاستخراج السمات في طبقات Conv2D، حيث تم تعديل مجال المرشحات ليكون من 30 إلى 96 مرشح، بدلاً من 96 إلى 384 مرشح وهو ما سيساعد بشكل كبير في تخفيض عدد البارامترات في الشبكة.

2. تم حذف كلاً من طبقات (Flatten, Dense)، واستبدالها بطبقة (Global Max Pool) التي لا تمتلك أيًا من البارامترات فيها. علماً أنه بعد إجراء هذه التعديلات انخفض عدد البارامترات من 21,591,810 إلى 254,176 في الشبكة الجديدة أي بمقدار 84%. والشكل (5) يوضح بنية الشبكة الجديدة.



الشكل (5) بنية شبكة Alex Net المعدلة

مع العلم أنه تم وضع طبقة BN بعد كل طبقة في الشبكة، وكذلك تم وضع طبقة (Dropout) لمساعدة الشبكة أثناء التدريب بعدم الوقوع في مشكلة (Overfitting)، وذلك من خلال إيقاف عمل بعض عصبونات الشبكة، حيث تم تحديد نسبة 10% من عصبونات الشبكة لتأخذ قيمة 0 (تتوقف عن العمل) خلال تدريب الشبكة.

3- التصنيف باستخدام الشبكات التكرارية (LSTM, GRU):

نلاحظ من الخطوة السابقة أن عدد البارامترات في الشبكة المعدلة قد انخفض بمقدار 84% وذلك قبل الانتقال إلى مرحلة التصنيف وهو ما سيساعد بشكل كبير في تسريع عملية التدريب وتخفيض كلفة الحسابات. في هذه المرحلة سنقوم بتجربة تمرير شعاع الخرج الواحد الناتج من طبقة (GlobalMaxPool2D) في شبكة Alex Net المعدلة والذي يبلغ طوله (64) قناة، والتي تتضمن أهم الميزات، إلى كلٍ من الشبكات التكرارية (LSTM, GRU) لاختبار واختيار الشبكة الأفضل من حيث عدد البارامترات الأقل، والشبكة ذات الدقة الأعلى والخطأ الأقل في التصنيف. حيث تقوم كلاً من هذه الشبكات بإضافة طبقات خاصة تساعد في عملية التصنيف على الشكل التالي:

[ Time distributed- LSTM- Dense- Dense- Dense- Dropout- BN- Dense- Dropout- BN- Dense]

حيث بلغ عدد البارامترات الكلي للشبكة المعدلة بعد ربطها مع شبكة LSTM (347, 522) حيث تم تدريب (346, 360) بارامتر، و(892) بارامتر لم يتم تدريبه، حيث يتم ضبط قيم بعض البارامترات في طبقة BN من أجل النتيجة النهائية للشبكة.

أما عند الربط مع شبكة GRU فقد تم إضافة الطبقات على الشكل التالي:

[ Time distributed- GRU- Dense- Dense- Dense- Dropout- BN- Dense- Dropout- BN- Dense]

وقد بلغ عدد البارامترات الكلي للشبكة المعدلة المربوطة مع GRU (339, 458) بارامتر وقد تم تدريب (338, 566) بارامتر منها. يمكن تفسير عدد البارامترات الأقل في شبكة GRU والذي بلغ حوالي العشرة آلاف بارامتر إلى بنية الشبكة التي تستخدم بوابتين فقط بينما تستخدم شبكة LSTM ثلاثة بوابات، مما يتطلب بارامترات أكثر.

## 8- مناقشة النتائج.

تم بناء الشبكة المعدلة وتدريبها واختبارها باستخدام Google Colaboratory (Colab)، الذي يسمح بكتابة وتنفيذ أي تطبيقات برمجية بلغة (Python)، بالإضافة إلى ما يمتلكه من إمكانيات عالية جداً (مخدمات - وحدة معالجة الرسومات GPU)، وبالتالي يساعد بشكل كبير في عدم الاعتماد الكامل على إمكانيات الجهاز الخاص بنا. كذلك فقد تم تقسيم قاعدة البيانات المؤلفة من 960/ فيديو وبشكل عشوائي إلى قسمين، القسم الأول للتدريب (Training) بنسبة بيانات 70% والقسم الثاني للاختبار (Testing) بنسبة 30%. وقد تم استخدام كاميرا لتصوير الأطفال بدقة (13Mp)، وقد كان بعد الكاميرا عن الطفل في معظم الفيديوهات (2m). كما أن متوسط حجم الفيديو (5MB)، ومتوسط طول الفيديو (45s)، ومتوسط عدد الإطارات في الفيديو الواحد (300) إطار. وقد بلغ عدد الدورات التدريبية لكل من الشبكتين (Alex Net LW- LSTM) و (Alex Net LW- GRU) 18/ دورة تدريبية، أي أنه عند هذه الدورة التدريبية تم الوصول إلى قيمة الدقة المثلى وثباتها وكذلك أقل قيمة للخطأ، أما الزمن المستغرق في كل دورة تدريبية في كلا الشبكتين فهو كما موضح في الجدول (2) التالي.

الجدول (2) الزمن المستغرق للدورات التدريبية في كلا الشبكتين

Epcho	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GRU Time(ms)	956	544	443	449	433	436	442	443	440	437	434	438	435	443	447	442	443	447
LSTM Time (ms)	729	497	497	410	400	395	393	397	394	397	399	399	405	405	397	400	395	391

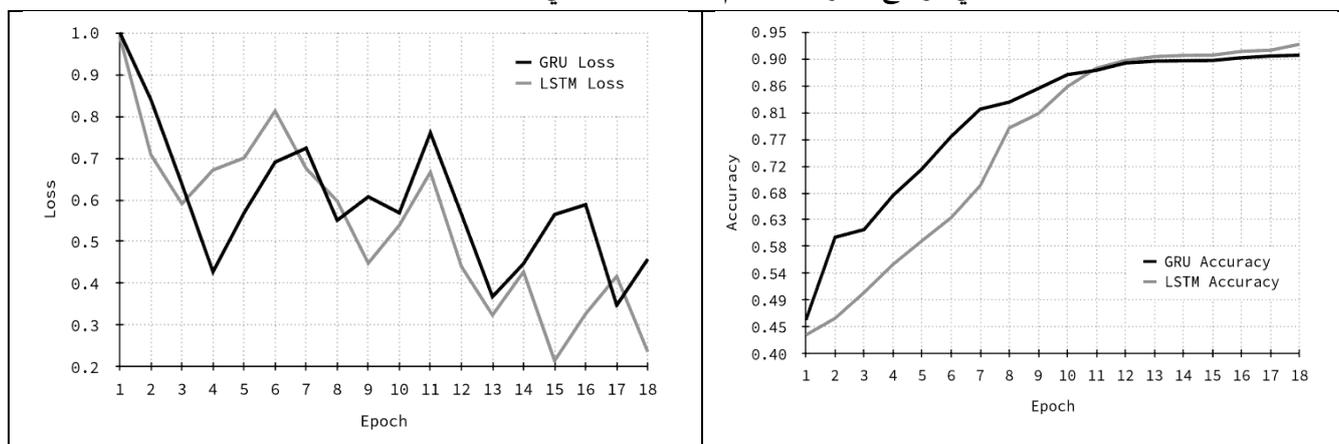
من الجدول (2) نلاحظ أن زمن الدورات التدريبية لشبكة LSTM أقل من الزمن المستغرق للتدريب في شبكة GRU. كما نلاحظ أن الزمن يبدأ بقيمة عالية ومن ثم يتدرج إلى أن ينخفض ويثبت عند مجال معين، أما بالنسبة لقيمة الدقة والخطأ التي تم تسجيلها خلال كل مرحلة من مراحل تدريب الشبكتين فهي كما موضح في الجدول (3) التالي.

جدول (3) قيم الدقة والخطأ خلال الدورات التدريبية للشبكتين

Epoch	Custom Alex Net With GRU		Custom Alex Net With LSTM	
	Loss	Accuracy	Loss	Accuracy
1	1.0003	0.4583	0.9906	0.4325
2	0.8399	0.5993	0.7079	0.4609

Epoch	Custom Alex Net With GRU		Custom Alex Net With LSTM	
	Loss	Accuracy	Loss	Accuracy
3	0.6374	0.6124	0.5907	0.5049
4	0.4278	0.6708	0.6715	0.5531
5	0.5687	0.7163	0.7001	0.5935
6	0.6904	0.7714	0.8126	0.6328
7	0.7236	0.8181	0.6750	0.6874
8	0.5511	0.8301	0.5970	0.7861
9	0.6074	0.8533	0.4485	0.8103
10	0.5689	0.8769	0.5392	0.8564
11	0.7603	0.8844	0.6663	0.8880
12	0.5657	0.8968	0.4404	0.9012
13	0.3682	0.9	0.3242	0.9078
14	0.4474	0.9007	0.4278	0.9098
15	0.5650	0.9012	0.2151	0.9102
16	0.5884	0.9058	0.3274	0.9167
17	0.3477	0.9088	0.4171	0.9187
18	0.4582	0.9102	0.2364	0.9289

والشكل (6) و(7) التالي يوضح الفرق بين قيم الدقة والخطأ في كلا الشبكتين



الشكل (7) قيم الخطأ الناتج في كلا الشبكتين

الشكل (6) قيم الدقة الناتجة في كلا الشبكتين

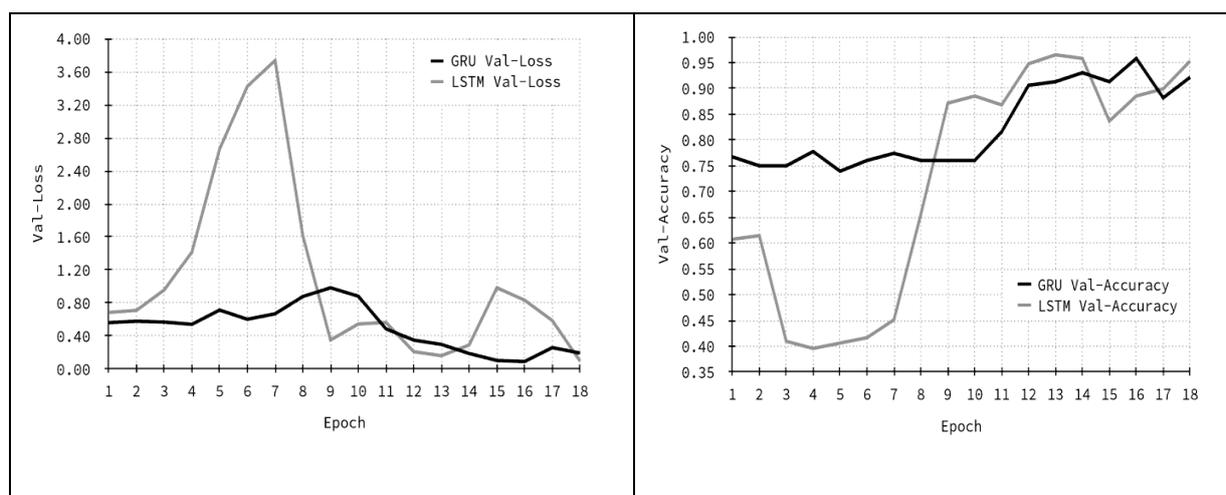
أما بالنسبة لقيم الخطأ والدقة الناتجة خلال مرحلة التقييم (Validate) التي تتم بعد كل دورة تدريبية للشبكتين، فقد تم تسجيل القيم التالية الموضحة من خلال الجدول (4).

جدول (4) قيم الدقة والخطأ خلال مراحل التقييم للشبكتين

Epoch	Custom Alex Net With GRU		Custom Alex Net With LSTM	
	Val- loss	Val- Accuracy	Val- loss	Val- Accuracy
1	0.5634	0.7674	0.6864	0.6076
2	0.5799	0.7500	0.7116	0.6146
3	0.5699	0.7500	0.9586	0.4097
4	0.5427	0.7778	1.4171	0.3958

Epoch	Custom Alex Net With GRU		Custom Alex Net With LSTM	
	Val- loss	Val- Accuracy	Val- loss	Val- Accuracy
5	0.7152	0.7396	2.6647	0.4062
6	0.6037	0.7604	3.4228	0.4167
7	0.6697	0.7743	3.7353	0.4514
8	0.8790	0.7604	1.6158	0.6562
9	0.9846	0.7604	0.3523	0.8715
10	0.8831	0.7604	0.5466	0.8854
11	0.4873	0.8160	0.5656	0.8681
12	0.3518	0.9062	0.2119	0.9479
13	0.2993	0.9132	0.1626	0.9653
14	0.1896	0.9306	0.2928	0.9583
15	0.1051	0.9132	0.9849	0.8368
16	0.0924	0.9583	0.8350	0.8854
17	0.2608	0.8819	0.5881	0.8993
18	0.1945	0.9218	0.0972	0.9532

نلاحظ من الشكلين (6) و(7) أن الخطأ (loss) كان في أعلى قيمة له في أول دورة تدريبية في كلا الشبكتين، ومن ثم بدأ بالانخفاض تدريجياً حتى تم الوصول إلى أقل قيمة له وثباته عند هذه القيمة. وكذلك بالنسبة للدقة فقد بدأت بأقل قيمة في الدورة التدريبية الأولى، ومن ثم ازدادت تدريجياً حتى أعطت أفضل قيمة لها في كلا الشبكتين عند الدورة التدريبية الأخيرة. كما نلاحظ أن قيمة الدقة النهائية في شبكة (Alex Net - LSTM) كانت أفضل من قيم الدقة في شبكة (Alex Net- GRU)، وكذلك بالنسبة لقيمة الخطأ في شبكة (Alex Net - LSTM) نلاحظ من القيم الناتجة في المرحلة الأخيرة أن الدقة في شبكة LSTM كانت أكثر من الدقة في شبكة GRU بمقدار 1% تقريباً، والخطأ الناتج في شبكة LSTM كان أقل من الخطأ في شبكة GRU بمقدار 2%. والشكل (8) و(9) التالي يوضح الفرق بين قيم الدقة والخطأ الناتجة خلال عملية التقييم في كلا الشبكتين.



الشكل (9) قيم الخطأ الناتج في كلا الشبكتين

الشكل (8) قيم الدقة الناتجة في كلا الشبكتين

نلاحظ من الشكل (8) و(9) قيم الدقة والخطأ خلال مراحل التقييم، كما نلاحظ من الشكل (9) أن شبكة (Alex Net- LSTM) قد أعطت خطأ كبير تجاوز قيمته الواحد في أكثر من دورة تدريبية، أي ظهرت مشكلة التعلم

المثالي (Overfitting) لكنه تم تجاوز هذه المشكلة لاحقاً. أما بالنسبة للقيم النهائية في كلا الشبكتين، فقد كانت قيم شبكة (Alex Net- LSTM) أفضل من قيم شبكة (Alex Net- GRU). حيث حققت شبكة LSTM تقدماً على شبكة GRU في قيمة الدقة عند التقييم النهائي بمقدار تقريبي 3% وبنسبة خطأ أقل بمقدار 9%.

## 9- خلاصة البحث:

تم في هذا البحث تخفيض عدد البارامترات وذلك من خلال استخدام شبكات الطي العصبية ذات الاوزان الخفيفة، حيث تم استخدام شبكة Alex Net وذلك بعد إجراء مجموعة من التعديلات عليها، من تخفيض عدد المرشحات المستخدمة في طبقات الطي، وكذلك استبدال الطبقات الاخيرة من الشبكة المستخدمة التي تكون مسؤولة عن التصنيف بأحد أهم أنواع الشبكات التكرارية، حيث تم استخدام كلاً من الشبكة التكرارية ذات الاعتمادية الطويلة (LSTM) والشبكة التكرارية المبوبة (GRU). وقد تم اختبار العمل خلال فترة البحث على مجموعة بيانات تضم 960 فيديو لأطفال طبيعيين وأطفال مصابين بطيف التوحد. وقد أثبتت النتائج التجريبية الانخفاض الكبير في عدد البارامترات في الشبكات العصبونية بمقدار 84%، كذلك فإن الشبكة التكرارية ذات الاعتمادية الطويلة LSTM أعطت نتائج أفضل من الشبكة التكرارية المبوبة GRU من حيث الدقة والخطأ، حيث زادت الدقة بمقدار 3% وانخفض الخطأ بمقدار 9%.

## 10- التوصيات والمقترحات.

بناءً على النتائج التي تم التوصل إليها توصي الدراسة بالتالي:

1. استخدام شبكات الطي العصبية الخفيفة LWCNN وذلك في حال كانت الشبكة تقوم بتصنيف محدود (صنفين أو ثلاثة) مما يوفر بشكل كبير في عدد البارامترات المستخدمة في الشبكة، حيث انخفض عدد البارامترات في النظام المقترح في بحثنا بمقدار 84%، وبالتالي زيادة السرعة في التدريب وتخفيض كلفة الحسابات واستهلاك الذاكرة.
2. الاستفادة من الشبكات التكرارية (LSTM, GRU) لما تمتلكه من ميزات تساعد في تصنيف السلوك البشري في الملفات المرئية، وذلك من خلال ربط الإطارات خلال الزمن، كما توفر في عدد البارامترات في الشبكة العصبونية، مع المحافظة على الدقة المطلوبة حيث حقق استخدام كلاً من الشبكتين في بحثنا دقة بمقدار 95% في شبكة LSTM و92% في شبكة GRU.

## 11- قائمة المراجع.

- 1- Alom, M.D.Z., Taha, T.M., (2018), "The History Began from Alex Net: A Comprehensive Survey on Deep Learning Approaches", University of Dayton, USA.
- 2- Dai, D., Yu, L. & Wei, H., (2019), "Parameters Sharing in Residual Neural Networks", Fudan University, Shanghai, China, Neural Processing Letters, Springer Science+ Business Media, LLC, part of Springer Nature 2019.
- 3- He, Y., Li, T., (2020), " A lightweight CNN model and its application in intelligent practical teaching evaluation", Dalian University of Technology, China, MATEC Web of Conferences 309.

- 4- Huang, Z., Yang, F., (2019), "Convolutional gated recurrent unit- recurrent neural network for state-of- charge estimation of lithium- ion batteries", university of hong kong, China, IEEE Access, volume 7, pages: 93139- 93149, ISSN: 2169- 3536.
- 5- Huang, Q., Zhou, K., (2018), "Learning to Prune Filters in Convolutional Neural Networks", University of California, Los Angeles, IEEE Winter Conference on Applications of Computer Vision (WACV), Volume: 1, Pages: 709- 718.
- 6- Jha, D., Yazidi, A., (2021) "Light Layers: Parameter Efficient Dense and Convolutional Layers for Image Classification", university of Norway, Norway, Lecture Notes in Computer Science, vol 12606. Springer, Cham.
- 7- Katte, T., (2018), " Recurrent Neural Network and its various architecture types", Institute of Technology, Pune, Maharashtra, India, International Journal of Research and Scientific Innovation (IJRSI), volume V, Issue III, ISSN 2321–2705.
- 8- Krizhevsky, A., Sutskever, I.& Hinton, G.E(2017), "ImageNet Classification with Deep Convolutional Neural Networks", university of Toronto, Communications of the ACM Journal, Volume 60, Issue 6, pp 84- 90.
- 9- Llamas, J., Leronés, P.M., (2017), " Classification of Architectural Heritage Images Using Deep Learning Techniques", University of Valladolid, Spain, Applied Science Journal, volume7, Issue10, 10.3390/app7100992.
- 10- Oktar, Y., Karakaya, D., (2019)- "Convolutional Neural Networks: A Binocular Vision Perspective" university of Economics, Izmir, Turkey.
- 11- Sherstinsky, A., (2020), " Fundamentals of Recurrent Neural Network (RNN) and Long Short- Term Memory(LSTM) Network, Elsevier journal, "Physica D: Nonlinear Phenomena", Volume 404, March 2020.
- 12- Sua, Y., Kuo, C.J., (2019), "On Extended Long Short- term Memory and Dependent Bidirectional Recurrent Neural Network", University of Southern California, USA, Neuro Computing Journal, volume 356.
- 13- Vadera, S., Ameen, S., (2020), " Methods for Pruning Deep Neural Networks", university of Salford, Salford M5 4WT, Greater Manchester.
- 14- Zhou, Y., Chen, S., (2020), "Review of research on lightweight convolution neural network", Naval university of engineering, Wuhan, China, 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference.