مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٨)العدد (٨) ٢٠٢٤

Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (8) No. (8) 2024

تقييم أداء عمل الشبكات العصبونية الالتفافية CNN والمتكررةRNN في تحديد مشاعر الأطفال

د. حسن البستاني *

بلال حسن **

(تاريخ الإيداع ٢٠٢٤/٥/٩ . قُبِل للنشر في ٢٠٢٤/٧/٢٨)

□ ملخّص □

مرض التوحد أو الذاتوية هو أحد الاضطرابات التابعة لمجموعة اضطرابات التطور المسماة باللغة الطبية اضطرابات في الطيف الذاتوي Autism Spectrum Disorders ASD

بالرغم من اختلاف خطورة وأعراض مرض التوحد من حالة إلى أخرى، إلا أن جميع اضطرابات الذاتوية تؤثر على قدرة الطفل على التواصل مع المحيطين به وتطوير علاقات متبادلة معهم.

من الممكن للعلاج المكثف والتشخيص المبكر أن يحدثا تغييراً ملحوظاً وجدياً في حياة الأطفال المصابين بهذا الاضطراب.

في هذه الدراسة، تم تطبيق خوارزميات التعلم العميق في التعرف على مشاعر الوجه عند الأطفال في سن مبكرة، وعند اكتمال تجميع جميع المشاعر يتم مقارنتها مع الحالات المرجعية الأخرى للأطفال التي تنمو بشكل طبيعي والأطفال التي تبدي مشاعر حيادية أخرى، حيث أن من أهم صفات وخصائص مريض التوحد أنه لا يعبر عن مشاعره وعواطفه ويبدو أنه غير مدرك أيضاً لمشاعر الأخرين.

استخدمنا في هذا البحث مفهوم الشبكات العصبونية الالتفافية CNN والشبكات العصبونية المتكررةRNN لتصنيف وتحديد مشاعر الأطفال في سن مبكرة واعتمدنا في ذلك على قواعد بيانات تم تحضيرها يدوياً ولشعورين أساسيين فقط من المشاعر وهما السعادة والحزن، وقارنا بين أداء الشبكتين العصبيتين المذكورتين من خلال الدقة لإيجاد أفضل نموذج لتقييم المشاعر واختبار الشبكتين على مجموعة من الصور ومقاطع الفيديو.

الكلمات المفتاحية:مرض التوحد، الذاتوية، التعلم العميق، شبكة عصبونية التفافية، شبكة عصبونية متكررة.

*مدرس في قسم هندسة النظم الحاسوبية والالكترونية – كلية هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس – سوريا.

**طالب دراسات عليا (ماجستير) في قسم هندسة النظم الحاسوبية والالكترونية – كلية هندسة تكنولوجيا المعلومات والاتصالات – جامعة طرطوس – سوريا.

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٨)العدد (٨) ٢٠٢٤

Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (8) No. (8) 2024

Evaluating the performance of convolutional neural networks (CNN) and recurrent neural networks (RNN) in identifying children's emotions

Dr. Hassan Al-bustani*

Belal hassan**

(Received 9/5/2024 . Accepted 28/7/2024)

□ ABSTRACT □

Autism is one of the disorders belonging to the group of developmental disorders called in medical language, Autism Spectrum Disorrders (ASD).

Although the severity and symptoms of autism vary from one case to another, all autism disorders affect the child's ability to communicate with those around him and develop mutual relationships with them.

Intensive treatment and early diagnosis can make a noticeable and serious change in the lives of children with this disorder.

In this study, deep learning algorithms were applied to recognize facial emotions in children at an early age, and when the collection of all emotions is complete, they are compared with other reference cases of children who grow normally and children who show other neutral emotions, as this is one of the most important characteristics and characteristics of a patient. Autism is that he does not express his feelings and emotions and also seems to be unaware of the feelings of others.

In this research, we used the concept of convolutional neural networks (CNN) and recurrent neural networks (RNN) to classify and identify children's emotions at an early age. We relied on manually prepared databases for only two basic emotions, which are happiness and sadness, and we compared the performance of the two aforementioned neural networks through accuracy to find the best modelTo evaluate emotions and test the two networks on a set of images and video clips.

Keywords: autism, deep learning, convolutional neural network, recurrent neural network.

*Teacher, Computer and Electronic System Engineering Department, Information and Communication Technology Engineering, Tartous University, Syria.

**Student Master, Computer and Electronic System Engineering Department, Information and Communication Technology Engineering, Tartous University, Syria.

١- المقدمة:

تعد الشبكات العصبونية من الخوارزميات المهمة في مجال تعلم الألة، حيث استقت أفكارها من محاولة محاكاة الدماغ البشري وتشابك الأعصاب، وترجع بدايات فكرة الشبكات العصبونية لعام ١٩٤٣ ببحث نشره كل من وارن مكولش (warren mcculloc) ووالتر بيتز (Walter pitts)، تم في هذا البحث وضع نموذج رياضي لكيفية عمل العقل وشرح نموذج للعصبونات التي تستخدم كمكون أساسي للشبكات العصبونية. [1]

وفي عام ١٩٤٩ قام دولاند هيب (Donald hebb) بنشر بحث يوضح كيف يتم التعلم من الدماغ، وشرح كيف أن الروابط بين العصبونات البيولوجية التي تستخدم مع بعضها البعض تقوى مع كل استخدام.[2]

وبالتالي فإن أربعينيات القرن الماضي شهدت نشأة مفهوم التعلم العميق، حيث توصل العلماء إلى فكرة أن الخلايا العصبية هي وحدات ذات قيم محددة وحالتي تشغيل وإطفاء، ويمكن بناء دارة منطقية من خلال ربط تلك الخلايا مع بعضها البعض وإجراء استدلال منطقي بناءاً على ذلك.

عاود التعلم العميق للانطلاق بنشاط أكبر عام ١٩٨٥ مع ظهور فكرة الانتشار العكسي للأخطاء [3]، وصولاً إلى العام ٢٠١٠ حيث تم البدء باستخدام شبكات الخلايا العصبية للتعرف على الكلام أولاً وحصل تحسن كبير في الأداء حينها.

في التعلم العميق تعتبر الشبكة العصبونية الالتفافية CNN فئة من الشبكات العصبونية العميقة وهي الأكثر شيوعاً لتحليل الصور المرئية [4]، وغالباً ما تتم مقارنة CNN بالطريقة التي يحقق بها الدماغ معالجة الرؤية في الكائنات الحية [5].

ونعتبر أيضاً الشبكات العصبونية المتكررة RNN شبكات متخصصة مع ردود فعل مستمرة لمعالجة البيانات المتسلسلة مع الزمن، حيث يتم استخدام المخرجات التي يتم الحصول عليها مرة أخرى كمدخلات مع المدخلات الجديدة، وقد تم اقتراح مفهوم RNN لأول مرة بواسطة مجموعة من الباحثين في رسالة تم نشرها عام ١٩٨٦ لوصف إجراء تعليمي جديد مع شبكة عصبية ذاتية التنظيم.[6]

٢- الدراسات المرجعية:

- في الدراسة [7]أكد الباحثون أن تعبيرات الوجه واحدة من الإشارات الرئيسية المستخدمة للكشف عن قدرات المعوقين في البيئة الاجتماعية بين الأطفال الذين يعانون من ارتفاع طيف التوحد، و قامت هذه الدراسة بتحليل تعابير الوجه المعقدة لدى الأطفال الذين يعانون من ارتفاع طيف التوحد باستخدام بعض التقنيات الحسابية ولم تستخدم خطوات وخوارزميات متطورة في التعلم العميق، وقد لاحظت الدراسة أن تعبيرات الوجه لدى الأطفال الذين يعانون من طيف التوحد أقل تعقيداً من تعبيرات الأطفال التي تنمو بشكل طبيعي، حيث أنه يوجد تتبع متسلسل بين مناطق الوجه المختلفة عبر ظروف التعبير المتنوعة، وبالتالي يوجد درجة منخفضة من ديناميكية الحركة في الألية المستخدمة، لاحظت الدراسة أيضا أن الأطفال الذين يعانون من التوحد يفتقرون إلى التنوع في تعبيرات الوجه حيث تقل درجة حرية التنقل بين المناطق المختلفة. وهذه التبعية واضحة بشكل خاص للتعبير عن حالات الغضب والاشمئزاز والفرح والحزن وتكون أقوى بين المناطق المتجاورة كمنطقة الخد والفم وأضعف في المناطق غير المتجاورة كالمنطقة بين العين والفم.
- اقترح الباحثون في الدراسة [8]طريقة لاستخراج ميزة الوجه واستخدامها لاحقاً في تحديد الجنس والعمر، حيث أكدت الدراسة أهمية السمات الهندسية لصور الوجه مثل العيون والأنف والفم وما إلى ذلك في تحديد السمة الأساسية، إلا أنه توجد عوامل أخرى مهمة لتحديد السمة غير الوجه قد تتضمن الصوت وحركة الشفاه واليد وقزحية العين

وبصمات الاصبع وغيرها من الخصائص النفسية والسلوكية الهامة جدا والتي تسمى بشكل أساسي المقاسات الحيوية للشخص، وفي النتيجة تم التأكيد على أن التعرف على الوجه واحدة من هذه القياسات حيث تم قياس مساحة العين والأنف والشفة وقد تم اقتراح خوارزمية تساهم في تقليل الضجيج والتشويش ضمن الصورة واكتشاف الحواف بشكل أفضل لزيادة جودة الصورة.

- وفي الدراسة [9]تم التأكيد على أن ضعف التواصل الاجتماعي هو أحد السمات التشخيصية لاضطرابات طيف التوحد حيث يوجد نسبة كبيرة من ضعف الادراك الاجتماعي لدى الأشخاص المصابين بالتوحد، وإن إحدى أكثر قدرات التواصل الاجتماعي هي القدرة على الحكم على عاطفة شخص ما من خلال تعابير وجهه، وأن الأشخاص المصابين بالتوحد يصدرون قرارات عاطفية غير محددة، وبالتالي يجب تحديد فيما إذا كان المصابون يعانون من هذه التشوهات لأن لديهم مفاهيم عاطفية غير طبيعية، وما هي الركائز العصبية أو الشعورية لهذه العيوب.
- تم في الدراسة [10] استخدام شبكة عصبونية التفافية CNN لتحليل التعبير المتنوع للأطفال المصابين بالتوحد وقد بينت أن الأطفال المصابين يظهرون المشاعر الحيادية والاشمئزاز كشعور سائد على بقية العواطف، بينما بقية الأطفال غير المصابين يظهرون السعادة بشكل أساسي ثم المشاعر الحيادية الأخرى بمرتبة ثانية،وأوصت هذه الدراسة بضرورة تطبيق شبكة عصبونية متكررة في نماذج متتابعة للوصول إلى دقة أكبر، وهذا ما تم العمل عليه خلال هذا البحث.
- تم الاعتماد في الدراسة [11] على خوارزمية فيولا – جونز لإجراء اكتشاف الوجه ومن ثم استخراج الميزات حيث يتم الفحص في وقت مبكر ومن ثم استكشاف الميزة الأساسية للشعور، حيث اقترحت هذه الدراسة طريقة لوصف ميزات الوجه المختلفة باستخدام أسلوب التعلم الألي، مما ساهم في منح المستخدم ميزة البحث في الصورة بمساعدة السمات المرئية لها، و يلعب اكتشاف واستخراج ملامح الوجه دوراً مهماً في التعرف على الوجه وتحديد النمط. إذ يعتبر البحث في الجزء الهام أو المحدد في صورة واحدة تحدياً كبيراً في مجال الرؤية الحاسوبية والتعرف على الأنماط حيث يتم تعيين السمة لميزة الوجه وفق المعنى الدلالي للميزة وهذا يمنحنا مزايا للبحث في الصورة بمساعدة السمات المرئية فيها.
- في الدراسة [12] تم التعرف على الحركة المستندة إلى فيديو من خلال استخدام شبكة هجينة CNN – RNNحيث تتخذ RNN ميزات مظهر الحركة المستخرجة عن طريق الشبكة العصبونية CNNمن إطارات الفيديو كمدخلات وترميز الحركة في وقت لاحق، وبالتالي فإن هذه الدراسة اعتمدت على نمذجة مظهر وحركة الفيديو عن طريق إطاراته المتتابعة حيث تم تصوير كل جزء من أجزاء الحركة وفق تقنية ثلاثية الأبعاد وبزوايا وأبعاد مختلفة، بالإضافة إلى دراسة عوامل أخرى كسطوع الجسم المتحرك لتمييز الحركة الكلية، إلّا أن هذه الدراسة لا تتعرف في النهاية على عمل الأشياء ضمن الفيديو، وهذا ما قمنا به في هذا البحث إذ تم التعرف على طبيعة المشاعر الموجودة على وجه الطفل من خلال التركيز على تفاصيل الوجه الأساسية والمعبرة أساساً عن الشعور كحركة الفم والشفاه وتعبيرات العين مثلاً.

٣- أهمية البحث وأهدافه:

تبرز أهمية هذا البحث في استخدام خوارزميات التعلم العميق في كشف وتحديد مشاعر الأطفال في بيئة غير خاضعة للمؤثرات الخارجية والتي من الممكن أن تؤثر على سلوك الطفل وتعابير وجهه، حيث تم تصميم نموذجين باستخدام خوارزميتين مختلفتين وتقييم أيها أكثر كفاءةً وأعلى دقة في الوصول إلى النتائج، حيث تكتفي الشبكات

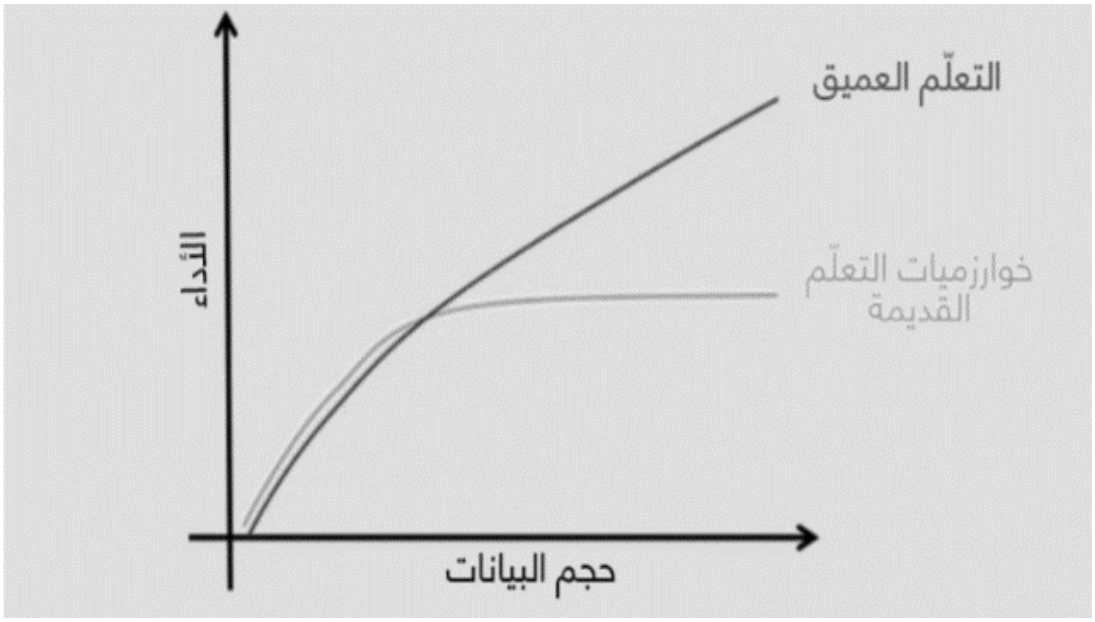
العصبونية الالتفافية بالخصائص والميزات المكانية للصورة فقط، وهذا ما تم العمل على تحسينه في الشبكات العصبونية المتكررة حيث تبني نتائجها اعتماداً على الخصائص المكانية والزمنية معاً، وتم تطبيق هذه النماذج على بيئة الحوسبة السحابية Google Colab باستخدام لغة البرمجة Pythonوالتي تعد رائدة في مجال الذكاء الاصطناعي والشبكات العصبونية.

٤- طرق البحث ومواده:

٤,١ التعلم العميق:

يعد التعلم العميق مفهوماً حديثاً في مجال التكنولوجيا، حيث يشير إلى تفوق النظم الحاسوبية في فهم وتحليل البيانات باستخدام شبكات عصبونية متقدمة كما أنه يتيح استخدام تقنيات تعلم الالة لفهم وتحليل البيانات بشكل أعمق وأكثر تعقيداً وشمولاً وهو ما يميزه عن تقنيات الذكاء الاصطناعي الأخرى، وتفتحت أعين الباحثين أكثر على التعلم العميق في بعض المجالات كرؤية الحاسب ومعالجة اللغة الطبيعية عام ٢٠١٢ حيث فازت خوارزمية التعلم العميق Alexnet بتحدي Large Scale Visual Recognition Challenge (ILSVRC) باستخدام مجموعة البيانات Image net وحصلت على المركز الأول في السباق بفارق كبير عن المركز الثاني [13].

يوضح الشكل (١)الدور الكبير للتعلم العميق في زيادة وتحسين قيم الأداء (دقة النموذج بشكل أساسي) مع الازدياد الكبير في حجم البيانات في الوقت الحالي مقارنةً مع خوارزميات التعلم التقليدية الأخرى، ويحتل التعلم العميق صدارة تكنولوجيا الذكاء الاصطناعي، إذ أنه يساهم في تشكيل الأدوات المناسبة لتحقيق مستويات كبيرة من الدقة وفيه تتعلم نماذج التعلم العميق كيفية القدرة على تصنيف المهام مباشرةً من خلال نص أو صور أو صوت أو أي وسيلة بيانات أخرى.[14]



الشكل – ١ – تحسنأداء (دقة) التعلم العميق عند ازدياد حجم البيانات

٤,١,١ مفاهيم وأساسيات التعلم العميق:

توجد مجموعة من المفاهيم الأساسية الخاصة بالتعلم العميق من أهمها:

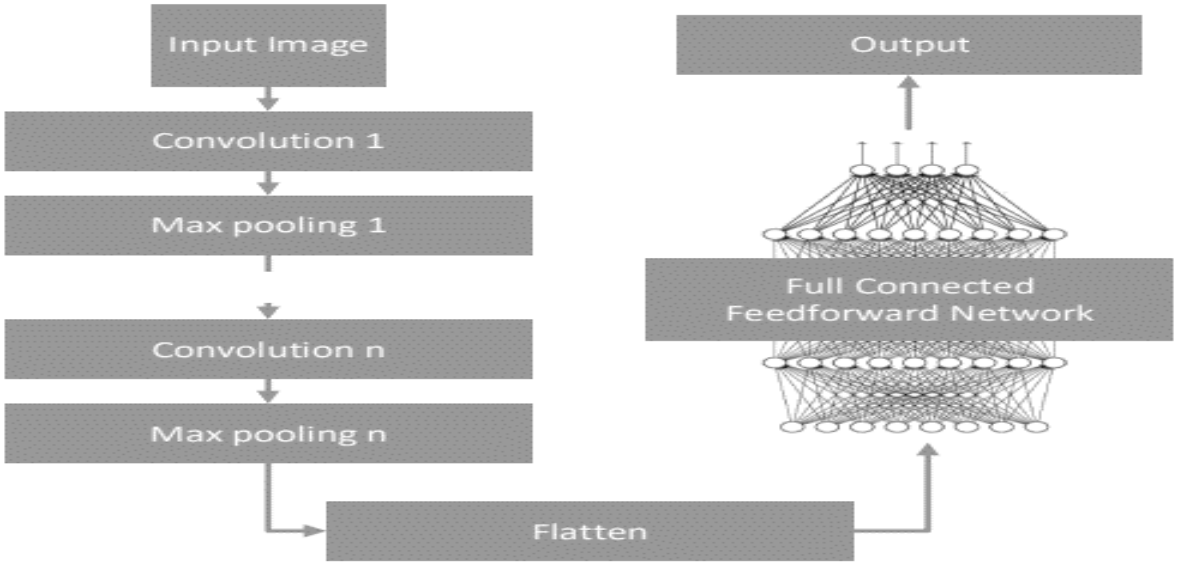
- **البيانات:** هي المدخلات التي تستخدمها الشبكات العميقة للتعلم، ويمكن أن تكون البيانات صور أو نص أو أرقام أو فيديو أو أي نوع أخر من أنواع البيانات ...

- **الوزن:** هي قيمة تحدد مقدار التأثير الذي يمارسه إدخال العقدة على الناتج حيث يتم ضبط الأوزان أثناء عملية التعلم.
- **الطبقة:** هي مجموعة من العقد المتصلة مع بعضها البعض، حيث تتكون الشبكات العصبونية عادةً من عدة طبقات.
- **الانحدار التكراري:** وهي خوارزمية تستخدم لضبط أوزان الشبكات العصبونية.

٤,٢ الشبكات العصبونية الالتفافية CNN:

تستخدم الشبكات العصبونية الالتفافية عملية رياضية تسمى الالتفاف، وهي نوع متخصص من الشبكات العصبونية العميقة التي تستخدم الالتفاف بدلاً من مضاعفة المصفوفة العامة في طبقة واحدة على الأقل من طبقاتها [15].

يوضح الشكل (٢) بنية الشبكات العصبونية الالتفافية حيث تتكون من طبقة إدخال وطبقة مخفية (أو أكثر) وطبقة إخراج، تسمى أي طبقة وسطى بأنها مخفية لأن مدخلاتها ومخرجاتها محجوبة بتابع التنشيط او الالتفاف النهائي.[16]



الشكل – ٢ – المخطط العام لشبكات CNN

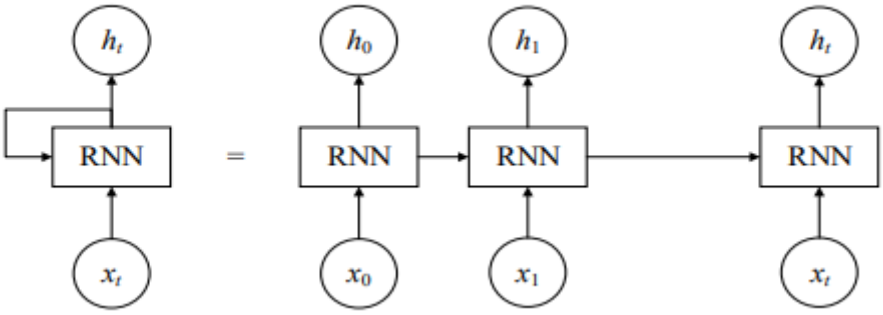
تقوم الطبقات الالتفافية الخفية بتطبيق عملية الالتفاف الرياضي على المدخلات المقدمة للشبكة وتمرير نتيجتها إلى الطبقة التالية، وتبني هذه الشبكة نتيجتها النهائية اعتماداً على الصورة المرئية (أي كما يبدو للإنسان بشكل تقريبي)، وقد وجد الباحثون أن فكرة تصميم وتخطيط هذه الشبكة مشابهه إلى حد كبير لتنظيم القشرة البصرية عند الكائنات الحية عند الاستجابة لمنبه معين [17].

قد تتضمن الشبكات الالتفافية طبقة تجميع حيث تعمل على تقليل أبعاد البيانات من خلال دمج مخرجات مجموعات الخلايا العصبونية في طبقة واحدة من خلية عصبية أخرى من الطبقة التالية.

٤,٣ الشبكات العصبونية المتكررة RNN:

هي نوع من الشبكات العصبية التي يمكنها معالجة البيانات المتسلسلة مع الحفاظ على الحالة الداخلية لذلك تعتبر هذه الشبكات مثالية لمهام التعرف على الكلام وترجمة اللغة وتحليل الفيديو وغيرها .....

يبين الشكل (٣) نموذج عام لعمل الشبكات العصبونية المتكررة:[16]



الشكل – ٣ – بنية شبكات RNN

تعتمد الفكرة الأساسية على تناولالبيانات المتسلسلة (إطارات الفيديو المتتابعة في دراستنا)، حيث يتم إدخال أول إطار في الفيديو X(0)إلى الشبكة العصبونية، ومن ثم استخدام دالة معينة لتحديد الميزة الأساسية للإطار حيث تقوم الدالة بإخراج الميزة h (0) من الإطار الأول، وحين تقوم الشبكة بتناول الإطار الثاني X(1)لا تعتمد في استخراج الميزة على هذا الإطار فقط، وإنما تقوم باستخدام دالة التنشيط الناتجة من معالجة الإطار الأول وإدخال الميزة السابقة المستخلصة h (0) لاستخلاص ميزة الإطار الثاني h(1)وهكذا ...، وبالتاليفإن الدالة التي تقوم بتقييم الإطار الثاني معتمدة على نتيجة الإطار الأول وهذا ما يسمى بنماذج التتابع .

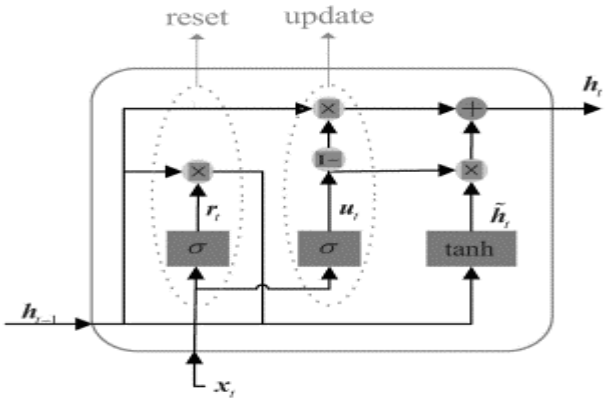
الميزة الرئيسية والأكثر أهمية للطبقة المخفية في RNN هو تذكر بعض المعلومات حول التسلسل، حيث تتمتع هذه الوحدة بقدرة فريدة على الحفاظ على حالة مخفية مما يسمح للشبكة بالتقاط التتابعات المتسلسلة من الأحداث وتذكر المدخلات السابقة أثناء عملية المعالجة، تعمل إصدارات مثل الذاكرة طويلة – قصيرة الأمد LSTM [18] ووحدات إرجاع البوابة GRU[19] على التعامل مع الأحداث طويلة المدى مع مرور الزمن وتذكرها .

٤-٤ شبكات GRU:

تعتبر شبكات Gate Recurrent Gate نوع خاص من أنواع الشبكات العصبونية المتكررة والتي تستخدم في مجالات البرمجة اللغوية الطبيعية ومهام تحليل الفيديو وغيرها ...

تستخدم هذه الشبكات مجموعة من الوظائف الحيوية والتي تتيح لها تحويل المدخلات إلى سلسلة من الأرقام التي يمكن للحواسيب فهمها بسهولة حيث تحتوي هذه الوظائف على بوابات تمكن الشبكة من تقليل عدد البيانات التي يجب معالجتها وتحليلها، وباستخدام هذه البوابات يمكن للشبكة تحديد متى يجب حفظ المعلومات في الذاكرة ومتى يجب حذفها.

تتوضح بنية الشبكة العصبونية المتكررة من النوع GRU[19]في الشكل(٤):



الشكل – ٤ – بنية شبكات GRU

ونلاحظ تكون هذه الشبكة من الطبقات التالية:

- **طبقة الإدخال:** تأخذ هذه الطبقة البيانات المتسلسلة كتتالي إطارات الفيديو بشكل متسلسل مع الزمن وتغذيها في الشبكة.

- **الطبقة المخفية:** تحدث في هذه الطبقة العمليات الحسابية المتكررة حيث يتم تحديث الحالة المخفية بناءاً على الإطار الحالي المدخل والحالة المخفية السابقة.

- **بوابة إعادة الضبط:** تحدد هذه الطبقة مقدار الحالة المخفية السابقة التي يجب نسيانها.

- **بوابة التحديث:** تحدد مقدار متجه التنشيط والذي سيتم دمجه في الحالة المخفية الجديدة.

- **ناقل تنشيط المرشح:** يعتبر بمثابة نسخة معدلة من الحالة المخفية السابقة والتي يتم إعادة ضبطها ودمجها في الإدخال الحالي.

- **طبقة الإخراج:** تأخذ هذه الطبقة الحالة المخفية النهائية كدخل وتنتج مخرجات الشبكة على شكل توزيع احتمالي يعبر عن النسبة المئوية لكل شعور من المشاعر المطبقة.

٥- مراحل العمل:

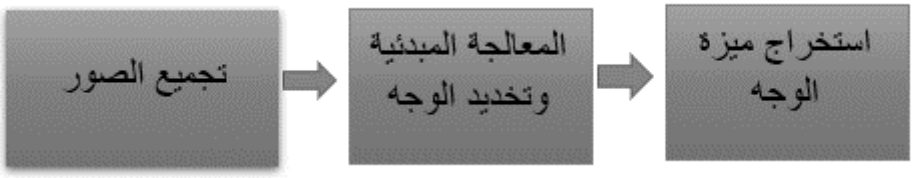
يقسم العمل في هذا البحث إلى قسمين:

١- ٥العمل مع الشبكات العصبونية الالتفافية:

١-١-٥ تحديد هدف النموذج: إن الهدف الأساسي لهذه الشبكة هو تحديد مشاعر الأطفال بناءاً على مجموعة من صورهم المرئية في مراحل مبكرة ووفق قاعدة البيانات المتاحة لدينا.

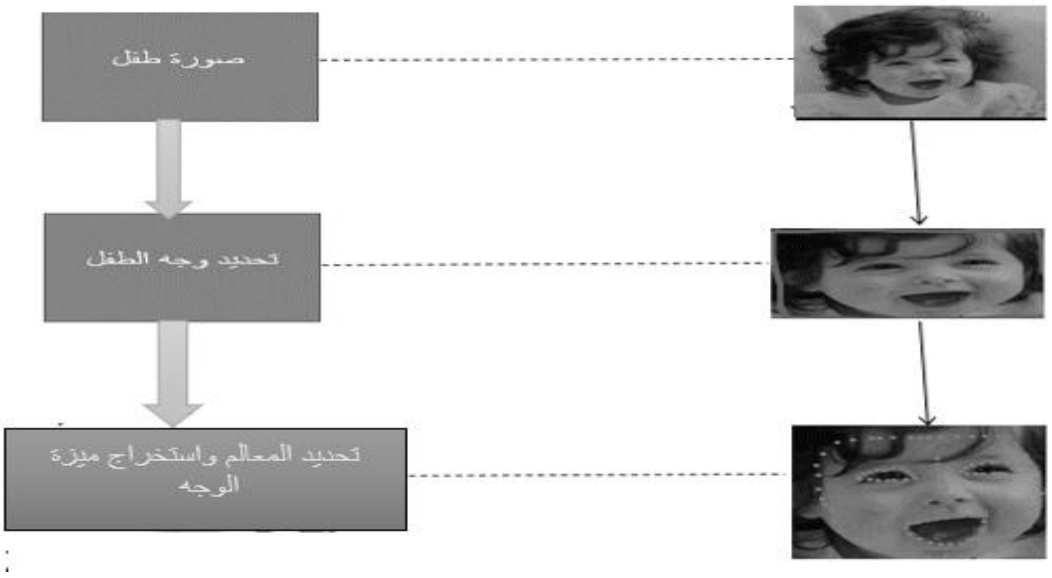
٢-١-٥تجميع بيانات النموذج:تم تحضير قاعدة البيانات يدوياً لهذا النموذج، عن طريق تجميع ١١٧٢ صورة ملونةjpgبقياسات مختلفة لأطفال في سن مبكرة تحمل شعورينأساسيينمن المشاعر وهما السعادة والحزن

٣-١-٥تحضير بيانات النموذج:يكمن العمل الأساسي في هذه الخطوة بمرحلة معالجة الصورة عن طريق تطبيق مجموعة من العمليات عليها بهدف تحسينها وتقليل كمية البيانات غير الضرورية عن طريق سلسلة من العمليات، يوضح الشكل (٥) الخطوات الأساسية للمراحل في هذه العملية:



الشكل –5– الخطوات الأساسية لتحضير البيانات في CNN

- **المعالجة المبدئية وتحديد الوجه:** يتم في هذه المرحلة توحيد حجم الصور والقيم اللونية لكل صورة، فبعد الحصول على الصورة يتم تحديد وجه الطفل فقط عن طريق عملية Anchor box [20] حيث يعتبر الجزء الأساسي من الصورة الذي يحدد العاطفة
- **استخراج ميزة الوجه:**يتم في هذه الخطوة تحديد معالم الوجه Facil landmark[21]حيث يتم رسم ٦٨ نقطة في الوجه الأمامي للطفل حول العينين والحاجبين والفم وبقية المعالم في الوجه كما هو موضح في الشكل (٦):



الشكل –6– كيفية تحديد معالم الوجه

٤–١–٥ تقسيم بيانات النموذج: تم تقسيم البيانات في هذا النموذج إلى بيانات تدريب وبيانات اختبار

حيث تحتوي مجموعة بيانات التدريب على ٩٧٦ صورة منها ٥١٦ صورة لأطفال بشعور السعادة، و ٤٦٠صورة لأطفال بشعورالحزن.

كما يوجد ضمن مجموعة بيانات الاختبار ١٩٦ صورة منها ١١٢ صورة لطفل سعيد و ٨٤ صورة لطفل حزين، كما تحتوي قاعدة البيانات على ٢٩ صورة أغلبها صور لأطفال مصابين أساساً بداء التوحد،

ونريد تطبيق نموذج الشبكة العصبونية بعد تدريبها واختبارها لتحديد ماهية الشعور السائد على وجوههم.

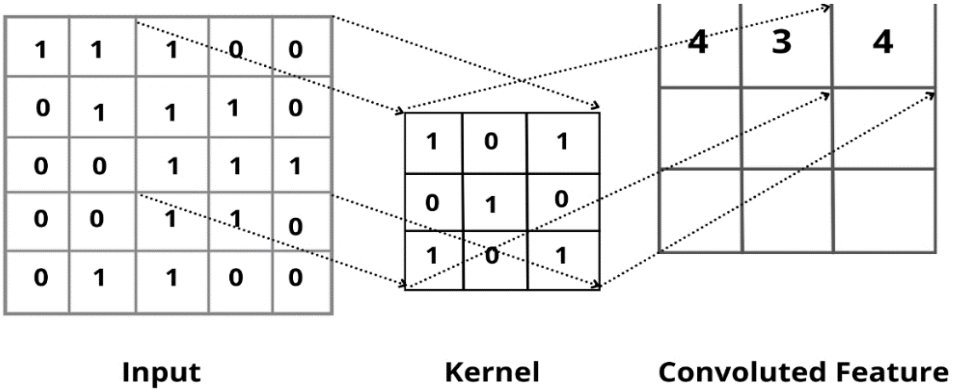
٥–١–٥تصميم نموذج الشبكة:تم تصميم شبكة عصبونية التفافية تتمثل مدخلاتها بمصفوفة الصور التي تم تجهيزها مسبقاً لإدخالها للشبكة، وطبقتين مخفيتين وطبقة متصلة بالكامل وطبقة إخراج، يوضح

الشكل (٧) خطوات بناء هذه الشبكة :

الشكل – ٧ – خطوات بناء نموذج CNN

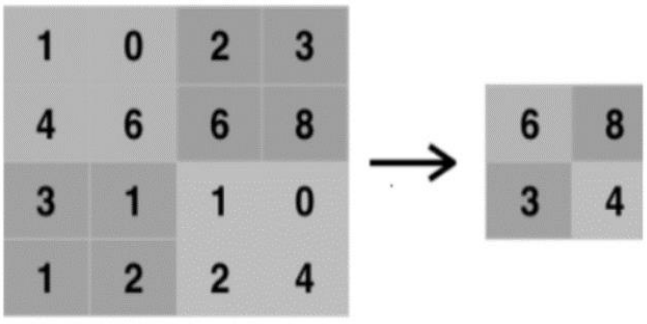
○ **إدخال الصور إلى الشبكة:** حيث يتم تحويل مجموعة الصور إلى مصفوفة بقياس معين لإدخالها إلى الشبكة، يتم في هذه الخطوة تحويل كل من بيانات النموذج إلى مصفوفة رباعية الأبعاد يتضمن البعد الأول عدد الصور ضمن الفئة، والبعدين الثاني والثالث هما قياس الصورة حيث يجب الانتباه لأن تكون جميع الصور بنفس القياس، والبعد الرابع يحدد ما إذا كانت الصورة ملونة أم رمادية.

○ **طبقة conv2D:** تحتوي هذه الطبقة أولا على ٢٠٠ مرشح، وحجم نواة المرشح ٣ *٣ ودالة التنشيط RELU، حيث يقوم المرشح بمسح وحدات البكسل الخاصة بالصورة في المرة الواحدة وإنشاء خريطة ميزات تتنبأ بالفئة التي تنتمي إلى كل ميزة، يوضح الشكل (٨) الية العمل ضمن هذه الطبقة :



الشكل – ٨ – ألية العمل ضمن طبقة conv2D

○ **طبقة max pooling:** يتم استخدام هذه الطبقة بقياس ٤ *٤ في المرحلتين، حيث تعمل على تقليل كمية المعلومات في كل ميزة تم الحصول عليها من الطبقة السابقة، مع الحفاظ على المعلومات الأكثر أهمية، يوضح الشكل (٩) كيفية القيام بهذه العملية عن طريق اختيار القيمة الأعلى للبكسل ضمن القياس المحدد:



الشكل – ٩ – الية العمل ضمن طبقةmax pooling

- **طبقة متصلة بالكامل:** تأخذ هذه الطبقة المدخلات من تحليل الميزات والمعالم وتطبق الأوزان للتنبؤ بالعلامة الصحيحة.
- **طبقة الإخراج:** تعطي الاحتمالات النهائية لكل ميزة.

٦-١-٥تدريب النموذج: يتم في هذه الخطوة استخدام مجموعة التدريب المحددة وفق قاعدة البيانات لتدريب النموذج المقترح، حيث يتم ضبط الاوزان والارتباطات بين الوحدات لتعلم المعرفة من البيانات.

٢-٥العمل ضمن الشبكات العصبونية المتكررة RNN:

١-٢-٥ تحديد هدف النموذج: إن الهدف الأساسي لهذه الشبكة هو تحديد نسبة كل شعور من المشاعر المتاحة في قاعدة البيانات المصممة لطفل بناءاً على مجموعة من مقاطع الفيديو في مراحل عمرية مبكرة.

٢-٢-٥تجميع بيانات النموذج: تم تحضير قاعدة البيانات يدوياً لهذا النموذج، عن طريق تجميع ٢٨٠ مقطع فيديو لأطفال في سن مبكرة تحمل أيضاً شعورين أساسيين وهما السعادة والحزن.

٣-٢-٥تحضير بيانات النموذج: يوضح الشكل (١٠) عمليات التحضير المستخدمة للبيانات قبل إدخالها إلى الشبكة:

الشكل – ١٠ – تحضير البيانات ضمن RNN

- يتم أولا قراءة الفيديو، ومن ثم تقسيمه إلى مجموعة من الإطارات المتتابعة كما في الشكل (١١) ، حيث يتم تعريف مصفوفة فارغة تخزن ضمنها الإطارات الفردية بعد إعادتها إلى حجم معين (٢٢٤ ، ٢٢٤) لكي تكون جميع الإطارات بنفس القياس .

الشكل – ١١ – تقسيم الفيديو إلى إطارات

○ **استخلاص الملامح:**يتم استخدام النموذج Inception v3 من أجل استخلاص الملامح، حيث يتألف من ٤٨ طبقة ومن وحدات بناء أساسية تسمى Inception module وهي كتلة من طبقات التفافية متوازية ذات أحجام مرشحات مختلفة وطبقة تجميع وفق القيمة للكبيرة مما يؤدي لتقليل كمية العمليات الحسابية المطلوبة [22]، ويتم في هذه الخطوة استخلاص الملامح الخاصة بكل إطار من الإطارات المقسمة.

٤-٢-٥تقسيم بيانات النموذج: تم تقسيم البيانات في هذا النموذج إلى بيانات تدريب وبيانات اختبار

حيث تحتوي مجموعة بياناتالتدريب على ٢٢٦ فيديو منها ١١٢ فيديو لأطفال بشعور السعادة، و١١٤ فيديو لأطفال بشعور الحزن.كما يوجد ضمن مجموعة بيانات الاختبار ٥٤ فيديو منها ٢٢ فيديو لطفل سعيد و٣٢ فيديو لطفل حزين.

كما تحتوي قاعدة البيانات على مجموعة من مقاطع الفيديو المنوعة لأطفال مصابين وأطفال سليمين، ونريد تطبيق نموذج الشبكة العصبونية المتكررة بعد تدريبها واختبارها لتحديد ماهية الشعور السائد على

وجوه هؤلاء الأطفال.

٥-٢-٥تصميم نموذج الشبكة:تم تصميم شبكة عصبونية متكررة من نوع GRU تتمثل مدخلاتها بمصفوفة إطارات الفيديوهات والتي تم تحضيرها في الخطوات السابقة، يوضح الشكل (١٢) خطوات بناء النموذج:

الشكل – ١٢ – خطوات تصميم نموذج RNN

○ **تضمين الإطارات:**يتم في هذه الخطوة تحديد عدد الإطارات الداخلة والتي تم تجهيزها مسبقاً، وتحديد عدد المعاني التي ستخرج وطول البيانات المستخدمة للإدخال كل مرة.

○ **تحديد الشبكة المستخدمة:**يتم في هذه الخطوة تحديد نوع الشبكة التي تستخدم لمعالجة البيانات، وقد تم استخدام شبكة عصبونية متكررة من النوع GRU ويتم تحديد عدد الخلايا ضمن الشبكة، في الطبقة الأولى نستخدم ١٦ خلية، وفي الثانية ٨ خلايا، بالإضافة إلى تحديد دالة التفعيل المناسبة.

○ **طبقة الخرج:**يتم من خلالها الحصول على الاحتمالية النهائية لكل من الميزات ويتم فيها تحديد عدد خلايا الخرج (٢ في دراستنا) ودالة التنشيط المناسبة.

٦-٢-٥ تدريب النموذج:يتم في هذه الخطوة استخدام مجموعة التدريب المحددة وفق قاعدة البيانات لتدريب النموذج المقترح، حيث يتم ضبط الاوزان والارتباطات بين الوحدات لتعلم المعرفة من البيانات.

٦- تقييم النماذج:

تقوم النماذج المقترحة بتحديد طبيعة الشعور السائد على وجه الطفل، طبقاً لقاعدة البيانات المصممة، وتحديد النسبة المئوية لكل شعور سائد، تشمل التنبؤات الصحيحة التحديد الصحيح لشعور السعادة ونرمز له بالرمز True happy (TH)، والتنبؤات الصحيحة لشعور الحزن ونرمز له بالرمز True sad (TS)، وتشمل التنبؤات الكلية بالإضافة إلى التنبؤات الصحيحة السابقة التنبؤ الخاطئ لشعور السعادة ونرمز له بالرمز False happy (FH) ، والتنبؤ الخاطئ لشعور الحزن ونرمز له بالرمز False sad (FS)وعليه يتم تقييم النماذج وفق المعيار التالي :

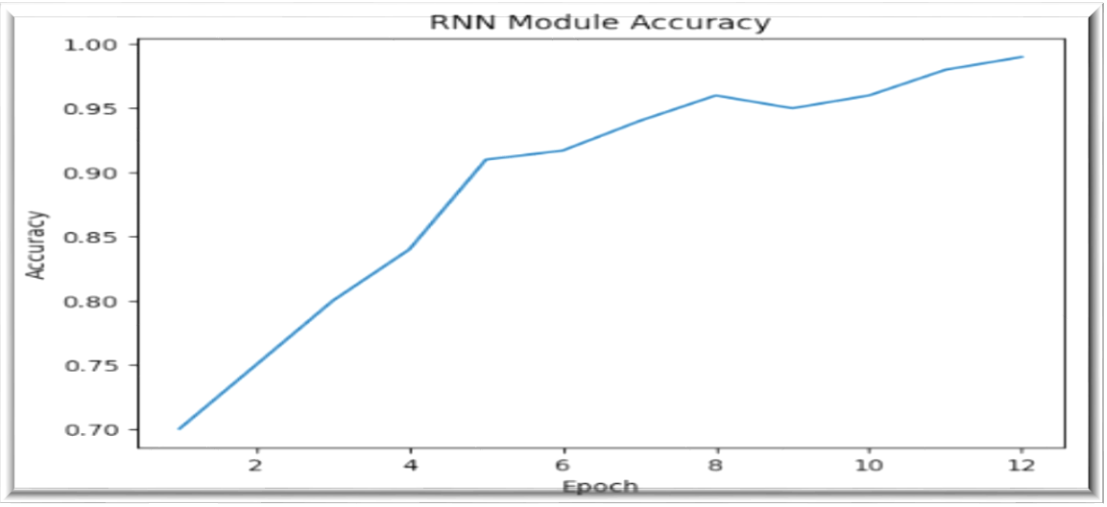
معيار الدقة Accuracy: الدقةهي مقياس لمدى قدرة الشبكة العصبونية على التنبؤ بالنتائج، ويتم حسابها كنسبة مئوية للتنبؤات الصحيحة من إجمالي عدد التنبؤات [23] :

$$\textbf{Accuracy} = \frac{\textbf{Number of correct assessments}}{\textbf{Number of all assessments)}} = \frac{TH+TS}{TH+TS+FH+FS} \quad (1)$$

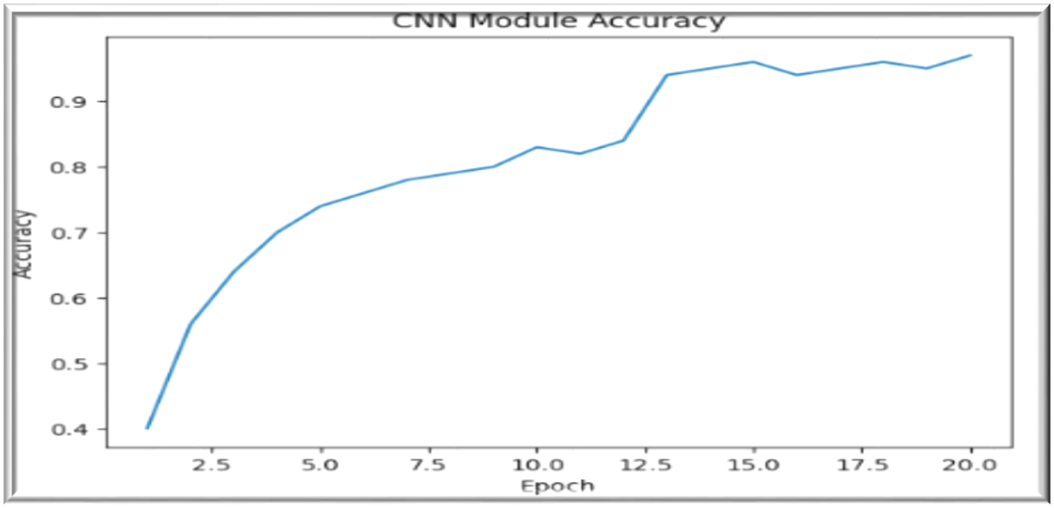
بالإضافة إلى دراسة متوسط دقة الاختبار عند كل من الشبكتين، والتحقق من أداء وعمل كل منها عن طريق إدخال مجموعة من الصور والفيديو لتقييم عمل النموذج.

٧- النتائج والمناقشة:

يبين الشكلان (١٣) و (١٤) نتائج الدقة التي تم الحصول عليها من تدريب الشبكتين، حيث تبدأ بقيم منخفضة ثم تزداد تدريجياً نتيجة زيادة عدد البيانات التي يتم أخذها مع كل تكرار ، حيث تم اعتماد عدد التكرارات epoch = 12



الشكل – ١٣ – الدقة في RNN



الشكل – ١٤ – الدقة في CNN

تلعب عمليات المعالجة المسبقة للبيانات دوراً كبيراً في تحسين الدقة حيث تصل في الشبكة المتكررة إلى ٩٩ % وفي الشبكة الالتفافية إلى ٩٧% بالإضافة إلى الخطوات المطبقة ضمن الشبكة والتي تساهم في تقليل حجم البيانات غير الضرورية.

يوضح الجدول التالي متوسط دقة الاختبار عند كل من الشبكتين، يعود الارتفاع في متوسط الدقة عند الشبكة العصبونية المتكررة كونها تعتمد على بيانات متسلسلة مع الزمن، وتبني نتيجة التنبؤ اللاحقة بناءاً على النتيجة التي تم الحصول عليها في اللحظة السابقة وفق تتابعات زمنية متتالية لإعطاء الخرج النهائي، قد يلعب عدد البيانات المستخدمة في نموذج الشبكة الالتفافية دوراً في انخفاض دقتها نوعاً ما حيث تم الاعتماد على عدد محدود نسبياً من البيانات لعملية التدريب وهذا ما يساهم في انخفاض عام لأداء الشبكة

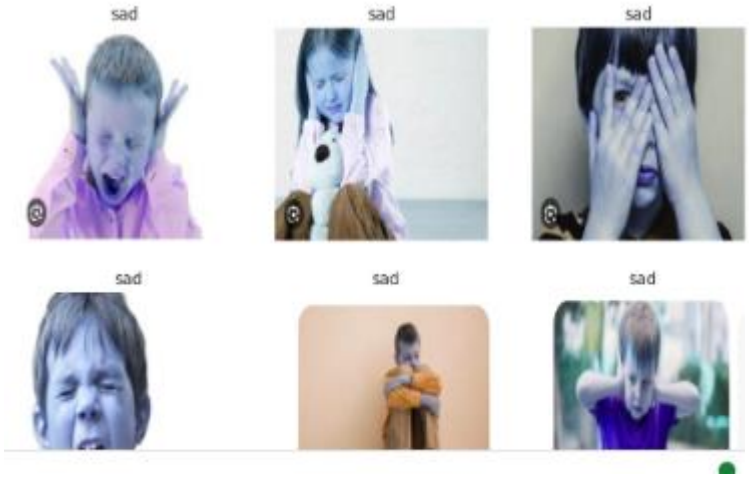
يوضح الجدول التالي (الجدول١) متوسط دقة الاختبار عند استخدام الشبكتين:

الجدول – ١ – متوسط دقة الاختبار في CNN , RNN

Module	CNN	RNN
Test accuracy	82.13	89.12

تعتبر هذه النسب جيدة نوعاً ما في الشبكتين، إلا أنه من الممكن تحسين هذه القيم إما عن طريق زيادة عدد البيانات المستخدمة، والاعتماد على بيانات أكثر دقة بالإضافة إلى استخدام عدة طبقات مخفية ضمن النماذج المقترحة وتغيير قيم بعض البارامترات والدوال التي تؤدي إلى تحسين أداء الشبكة.

تم استخدام نموذج الشبكة العصبونية الالتفافية لتحديد مشاعر بعض الأطفال كما هو موضح في الشكل (١٥)، تم الاعتماد على مجموعة من صور الأطفال الذين لديهم إصابة بداء التوحد، وقد أظهر هؤلاء الأطفال شعور الحزن كشعور أساسي



الشكل – ١٥ – استخدام نموذج CNN لتحديد مشاعر الأطفال

بالتأكيد لا يمكن اعتبار أي طفل يظهر شعور الحزن بأنه طفل مصاب بالتوحد، هنالك العديد من المشاعر الأخرى التي يجب إخطارها وإضافتها إلى قاعدة البيانات لتحديد الشعور الحقيقي على وجه الطفل ومن ثم ربطه بالحالات والمعدلات الطبيعية.

كما تم اختبار نموذج الشبكة العصبونية المتكررة على مجموعة من مقاطع الفيديو لأطفال بحالات معينة وفي بيئة غير خاضعة للإشراف لمعرفة نسبة الشعور السائد وكانت النتائج وفق الجدول التالي (الجدول ٢):

الجدول – ٢ – نسب الشعور السائد عند بعض الأطفال في حالات معينة

	Happy	Sad
Case (1)	85.72	14.28
Case (2)	89.51	10.49
Case (3)	٥٠,٨٨	٤٩,١٢
Case (4)	٥٣,٧٤	٤٦,٢٦

في الحالتين الأولى والثانية مقاطع فيديو لأطفال سليمين، حيث يتضح ارتفاع نسبة شعور السعادة عند هؤلاء مقارنةً مع شعور الحزن كما تبين ملامح الوجه.

أما في الحالتين الثالثة والرابعة حالات لأطفال مصابة بداء التوحد، حيث انخفضت نسبة شعور السعادة مقارنةً مع الأطفال السليمين، وزادت نسبة شعور الحزن لديهم كتعبير عن المشاعر الحيادية.

٨- الاستنتاجات والتوصيات:

خلصت هذه الدراسة إلى أنه من الممكن استخدام نموذج الشبكة العصبونية المتكررة والالتفافية لتحديد ملامح مشاعر الطفل ومعرفة نسب هذا الشعور وتوزعها، وتبين أنه يمكن استخدام الشبكات العصبونية الالتفافية بشكل أساسي للتعامل مع الصور المرئية الثابتة، أما الشبكات العصبونية المتكررة فإنها تصبح أكثر كفاءةً عند التعامل مع التسلسلات الزمنية للأحداث، مع الأخذ بعين الاعتبار ضرورة التقاط الصور ومقاطع الفيديو في بيئة غير خاضعة للتأثير الخارجي وضمن نطاق حركة وحياة الطفل العادية لتكون النتائج أكثر دقة.

يمكن اعتبار اعتمادنا على شعورين فقط فجوة بحثية ونقطة انطلاق لاحقة للتطوير حيث واجهنا صعوبة في تأمين قاعدة بيانات كافية (خصوصاً مقاطع فيديو) لأطفال في سن مبكرة تحمل بقية المشاعر الأساسيةالأخرى

(الاشمئزاز، القرف، الغضب، الحياد .....) ، وعند اكتمال هذه البيانات يمكن دراسة توزع المشاعر الكاملة ونسبتها على وجه الطفل في سن مبكرة حيث يعتبر تحديد ومعرفة هذه المشاعر أمراً ضرورياً وملحاً للغاية، حيث يمكن ربط نسب هذه المشاعر الكاملة مع النسب الطبيعية وغير الطبيعية للأطفال المصابين والسليمين ومحاولة اكتشاف احتمالية الإصابة بهذا المرض في سن مبكرة لمحاولة اتخاذ خطوات علاجية أولية قدر الإمكان.

٩- المراجع العلمية:

[1]: Mcculloch, W.,&Walter,P.(1943). A Logical calculus of ideas Immanent in Nervous Activity . Bulletin of Mathematical Biophysics, 115 – 133 .

[2]: Hebb , D.(1949). The organization of Behavior . New York: Wiley .

[3]: Rummelhart, DE ,Hinton , GE. , &Williams, R. J .(1986). Learning internal Representation by error propagation . In Parallel Distributed processing : Explorations in the Microstructure of cognition , 318-362 .

[4]: valueva,M.V.;Nagornov,N.N,;Lyakhov,^P.A.;Valuev,G.V.;Chervyakov,N.I. (2020) .”Application of the residue number system to reduce hardware costs of the convolutional neural network implementation “. Mathematics and computer in simulation . Elsevier BV.177 : 232-243 .

[5]:Mahapattanakul,puttatida(November^11,2019).”from human vision to computer vision - convolutional neural networks (part ¾)”.Medium

[6]: Rumelhart DE , Hinton GE , &Williams, R. J .(1986). Learning Representation by back – propagation errors. Neture 323(6088) : 533 – 536 .

[7] : Tanaya Guha, Zhaojuns Yang, Ruth B. Grossman, Shrikanth S. Narayanan, A computational study of expressive facial dynamics in children with autism, IEEE Transactions on Affective Computing (March 2016).

[8]: K. Ramesha, K.B. Raja, K.R. Venugopal, L.M. Patnaik, Feature extraction based face recognition gender and age classification, International Journal on Computer Science and Engineering 02 (01S) (2010) 14–23.

[9]: D.P. Kennedy, R. Adolphs, Perception of emotions from facial expressions in high-functioning adults with autism, Neuropsychologia 50 (14) (2012) 3313–3319.

[10] : S.P. Abirami, ME, G. Kousalya, ME, PhD, Balakrishnan, ME, PhD, R. Karthick, BOT , Varied Expression Analysis of Children With ASD Using Multimodal Deep Learning Technique , Deep Learning and Parallel Computing Environment for Bioengineering Systems , (2019) - page 225 – 235 .

[11] : N.K. Bansode, P.K. Sinha, Facial feature extraction and textual description classification using SVM, in: International Conference on Computer Communication and Informatics, (2014), pp. 1–5

[12]: Yin Fan, Xiangju Lu, Dian Li, Yuanliu Liu iQIYI Co. Ltd, Beijing, 10080, China , Video-Based Emotion Recognition using CNN-RNN and C3D Hybrid Networks , (2016).

[13]:Krizhevsky, A. , Sutskever, I. , &Hinton , G.E.(2012). ImageNet Classification with Deep convolutional neural networks . In proceedings of the 25th international Conference on Neural Information processing system , (pp. 1097 – 1105).

[14]: https://www.noonpost.com/27163/Access: 8 / 5 / 2024

[15]:Tsantekidis,Avraam;Passalis,Nikolaos;Tefas,Anastasios;Kanniainen,Juho; Gabbouj,Moncef;Iosifidis,Alexandros(July 2017). "Forecasting stock prices from the limit order book using convolutional neural networks". 2017 IEEE 19th Conference on Business informatics (CBI). Thessaloniki,Greece: IEEE: 7 – 12 . doi:10.1109/CBI.2017.23 ISBN 978-1-5386-3035-8.S2CID 4950757.

[16]:Keji Zheng , A project report submitted to Auckland University of Technology in partial fulfillment of the requirements for the degree of Master of Computer and Information Sciences (MCIS) ' Video Event Capturing Using RNN and CNN in Deep Learning ' (2017) .

[17]: convolutional neural networks " (LeNet) – Deep Learning 0.1 documentation". Deep Learning 0.1. LISA Lab. Retrieved 31 August (2013).

[18]: Hochreiter S, Schmidhuber J (1997) Long short-term memory . Neural Comput 9 (8) : 1735 – 1780 .

[19]: Cho K , Van Merrienboer B, Gulcehre C ,Bahdanau D , Bougares F , Schwenk H , Bengio Y (2014) Learning phrase representation using RNN encoder-decoder for statistical machine translation .Preprint . arXiv: 14061078 .

[20]: ASTON ZHANG, ZACHARY C. LIPTON, MU LI, AND ALEXANDER J. SMOLA , Dive into Deep Learning , section 14.4 Anchor Boxes 610 – 622 (2013) .

[21]: Kostiantyn Khabarlak , Larysa Koriashkina , Fast Facial Landmark Detection and Applications: A Survey . Dnipro University of Technology (2022).

[22] : Jaafar Salman, Ghada saad , Marie Suliman : Using deep learning algorithms and computer visionin detecting human brain tumor , Tartous University Journal for Research and Scientific Studies(2021).

[23]: Stralen, Karlijn J. van; Stel, Vianda S; Reitsma, Johannes; Dekker, Friedo; Zoccali, Carmine; Jager, Kitty (2009), *Diagnostic methods I: sensitivity, specificity, and other measures of accuracy*, Kidney International,V75,pp 1257–1263.