

Designing a Hybrid System based on between Deep learning Network with Face Detection algorithm for Emotion Recognition through Facial Features Analysis

Mohamad Bashir Amir Nahas, Asst-Prof. Lara Fathi Kadid, lecturer \ Mohamad Ahmad Mounir Batikh*

Faculty of Informatic Engineering | Cordoba Private University | Syria

Received:

21/10/2024

Revised:

07/11/2024

Accepted:

28/11/2024

Published:

30/12/2024

* Corresponding author:

mohamedbt@gmail.com

Citation: Nahas, M. B., Kadid, L. F., & Batikh, M. A. (2024). Designing a Hybrid System based on a Deep Learning Network with a Face Detection algorithm for Emotion Recognition through Facial Features Analysis. *Journal of engineering sciences and information technology*, 8(4), 21 – 32. <https://doi.org/10.26389/AJSRP.B241024>

2024 © AISRP • Arab Institute of Sciences & Research Publishing (AISRP), Palestine, all rights reserved.

• Open Access



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license

Abstract: Objectives: In recent years, the demand for automatic emotion recognition systems has increased for use in various fields, including efforts to modify an individual's mood to improve mental health, as well as assisting in identifying the emotions of children with autism spectrum disorder who struggle to express their emotional states. Deep learning networks are linked with face Detection algorithms. If a deep learning network is used alone, any object detected in the image will be considered as a face and will be processed to determine whether it has emotions or not. This results in high computational complexity, very high response time, and low accuracy if there is more than one face in the same image. Methods: In this research, images were first introduced into the Fer-Net convolutional neural network after performing some preprocessing operations. The Fer-Net was selected after experimenting with several other CNN. Facial features were then extracted from the network, and these extracted features were classified into four basic emotions. Additionally, several standard databases were tested individually, such as FER-2013 and AffectNet, for training and evaluating the network. Subsequently, the previous databases were merged with other databases like RAF-DB and CK+ to increase the number of training samples and evaluation samples in order to avoid the issue of overfitting. Finally, we linked facial detection with the classification network obtained from the trained model using the MTCNN algorithm to identify the faces present in the image before analyzing the facial features and determining the emotions extracted from them. Results: First, Data Augmentation was implemented on the data in standard database (Fer-2013), and the overfitting problem was resulted.

Keywords: Convolution Neural network, FER-NET, Affect Net, MTCNN, Emotion Recognition, Hybrid System.

تصميم نظام هجين بالاعتماد على شبكات التعلم العميق وخوارزميات التعرف على الوجوه

للكشف عن المشاعر باستخدام التعابير الوجهية

محمد بشير أمير نحاس، الأستاذ المساعد / لارا فتحي قديد، المدرس / محمد أحمد منير بطيخ*

كلية الهندسة المعلوماتية | جامعة قرطبة الخاصة | سوريا

المستخلص: الأهداف: في السنوات الأخيرة تزايد الطلب على أنظمة التعرف التلقائي على العواطف للاستخدام في مجالات متعددة، بما في ذلك العمل على تعديل الحالة المزاجية للفرد لتحسين الصحة النفسية، وكذلك المساعدة في كشف مشاعر الأطفال الذين يعانون من طيف التوحد وذلك لعدم قدرتهم على التعبير عن حالتهم المزاجية. يقدم هذا البحث نظاماً (Mood Pulse) مصمماً لمحاولة الكشف عن المشاعر، حيث تم الدمج بين شبكات التعلم العميق وخوارزميات التعرف على الوجوه. حيث أنه في حال الاعتماد على شبكة تعلم عميق فإنه سيتم اعتبار أي كائن سيتم اكتشافه في الصورة كوجه ويتم التعامل معه وتحديد فيما إذا كان يمتلك مشاعر أو لا. وبالتالي سيتم الحصول على تعقيد حسابي كبير، وزمن استجابة كبير جداً، بالإضافة إلى الدقة القليلة في حال وجود أكثر من وجه في نفس الصورة. المنهجية: تم في هذا البحث أولاً إدخال الصور بعد إجراء بعض عمليات المعالجة الأولية عليها إلى شبكة الطي العصبونية Fer-Net والتي تم اختيارها بعد تجربة عدة شبكات طي عصبونية، والقيام ببعض التعديلات في بنية الشبكة. ومن ثم استخلاص السمات الوجهية من خلالها، ثم تصنيف هذه السمات المستخلصة إلى المشاعر الأربعة الأساسية. كذلك تم اختيار عدة قواعد بيانات قياسية كل منها على حدة مثل (Fer-2013, AffectNet) لتدريب الشبكة وتقييمها. لاحقاً تم دمج قواعد البيانات السابقة مع قواعد بيانات أخرى مثل (RAF-DB, CK+)، من أجل زيادة عدد عينات التدريب، وعينات التقييم لعدم الوقوع في مشكلة overfitting. أخيراً قمنا بالربط بين اكتشاف الوجه وشبكة التصنيف التي حصلنا عليها من النموذج المدرب من خلال خوارزمية (MTCNN) من أجل تحديد الوجوه الموجودة في الصورة فقط قبل أن يتم تحليل معالم الوجه وتحديد المشاعر المستخلصة منه. النتائج: في البداية قمنا بإجراء عملية Data augmentation للبيانات الموجودة في قاعدة البيانات القياسية (Fer-2013) فحصلنا على مشكلة Overfitting وذلك بسبب تكرار البيانات. من أجل ذلك تم البحث عن جميع قواعد البيانات المستخدمة في مجال الكشف عن المشاعر من أجل زيادة عدد البيانات للأصناف الأساسية المختارة. الكلمات المفتاحية: شبكات الطي العصبونية، Fer-Net، MTCNN، Fer-2013، التعرف على المشاعر، نظام هجين.

المقدمة:

يُعد تعرّف السلوك الانساني أحد أكثر المسائل تحديًا في مجال الرؤية الحاسوبية، ويعود ذلك إلى التشابه الكبير في المحتوى المرئي ضمن المشاهد المصورة، واختلاف وجهات النظر بالنسبة للفعل نفسه، ونتيجة التطور الكبير في تقنيات الذكاء الصناعي والتعلم العميق، أتاح لنا استكشاف مجالات جديدة في فهم السلوك البشري والتعامل معه. إن أحد أهم هذه المجالات هو نظام الكشف عن المزاج للشخص، والذي يهدف إلى المساعدة في فهم مشاعر الأشخاص وتقديم الدعم المناسب لهم وخاصةً للأطفال الذين يعانون من طيف التوحد. تم طرح العديد من الطرق التقليدية سواءً من خلال تحليل معالم الوجه أو الصوت والتي تكون محدودة في قدرتها على التقاط الطبيعة المعقدة لتعبيرات الوجه (Trevisan&Birmingham,2016). من أهم هذه الطرق المستخدمة في مجال التعرف على المشاعر البشرية:

- Active Appearance Model (AAM): تستخدم هذه الخوارزمية النماذج الإحصائية من أجل تمثيل تفاصيل الصورة، وخلال فترة التدريب يتم تعديل قيم البارامترات حتى نحصل على التوضيح المناسب لهذه التفاصيل، وقد أعطت نتائج جيدة في التعرف على الوجوه والتصوير الطبي (Cootes&Edward,1999).
- تحليل الصوت: يعد تحليل الصوت إحدى الطرق التقليدية لفهم المشاعر، حيث يتضمن ذلك تحليل نبرة الصوت وتردداته وضبطه، وقد تم استخدام تقنيات تحليل الخصائص الصوتية مثل (Acoustic Feature Extraction) لاستخراج ميزات الصوت المرتبطة بالعواطف (Kobayashi&Ye,2014).

مشكلة الدراسة:

يعد الكشف عن مزاج الشخص من خلال تحليل معالم الوجه وتعبيراته ذو أهمية كبيرة جداً في فهم مشاعر الشخص والعمل على تقديم الدعم المناسب له وخاصةً في مجال الرعاية الصحية للأطفال الذين يعانون من طيف التوحد، لذلك يمكن صياغة مشكلة الدراسة في التساؤلات التالية:

- 1- هل استخدام نوع معين من شبكات الطي العصبونية يعطي زيادة في دقة التصنيف، وتخفيض التكلفة الحسابية.
- 2- هل اعتماد قاعدة بيانات معينة خلال عملية التدريب والتقييم لشبكة الطي العصبونية يؤدي إلى زيادة الدقة.
- 3- هل استخدام خوارزميات الكشف عن الوجوه يساعد في تقليل زمن الاستجابة والسرعة في الاستجابة.

فرضيات الدراسة: تفترض الدراسة:

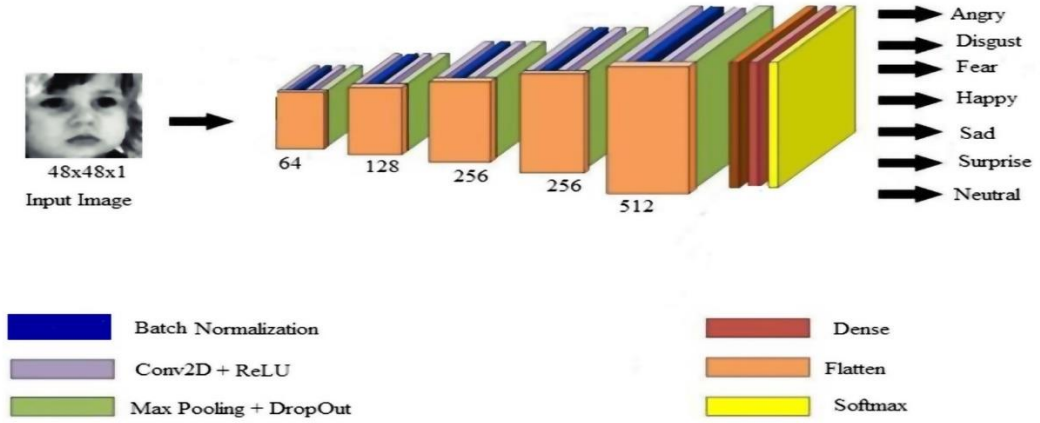
1. استخدام شبكة الطي العصبونية التقليدية Fer-Net وذلك لأنها تستخدم مرشحات ذات أحجام صغيرة تناسب استخلاص تفاصيل المعالم الوجهية.
2. استخدام أنواع مختلفة من قواعد البيانات الخاصة بالمعالم الوجهية + Fer-2013, AffectNet, RAF-DB, CK.
3. استخدام خوارزمية الكشف عن الوجه MTCNN لتحديد الوجوه الموجودة في الصورة فقط.

المبحث الأول- الإطار النظري والدراسات السابقة

أولاً- الإطار النظري

- مفهوم شبكات الطي العصبونية CNN: تعدّ شبكة الطي العصبونية CNN أحد أهم أنواع الشبكات العصبونية العميقة ذات التغذية الأمامية، ولها العديد من الاستخدامات في مجال تطبيقات الرؤية الحاسوبية (Computer Vision) وخاصةً في تحليل الملفات المرئية، ويعود ذلك إلى وجود طبقة طي (Convolution Layer) وهي الطبقة الأولى في شبكة الطي العصبونية. يتم تطبيق المرشح (Filter) في هذه الطبقة من أجل استخلاص السمات (Features) في الصور أو الملفات المرئية وبالتالي فهي تتضمن أكثر العمليات الحسابية، حيث يكون خرجها عبارة عن مصفوفة تسمى بخريطة السمات (Features Map)، ومن ثم تأتي طبقة التابع غير الخطي (Nonlinear layer) حيث يتم تطبيق (Rectified Linear Unit ReLU) على خرج طبقة الطي، تقوم هذه الطبقة باستخلاص السمات فقط والتخلص من جميع المكونات الأخرى. أما الطبقة الثالثة فهي طبقة الاختزال (Pooling Layer) التي تقوم بانتقاء أفضل السمات من خلال تطبيق توابع معينة مثل (Max, Avg, etc). وأخيراً طبقة الارتباط الكلي لإنتاج الخرج المطلوب، التي يمكن اعتبارها بمثابة مصنّف في شبكة CNN، كذلك يمكن أن يتم إضافة طبقات إلى شبكة CNN مثل طبقة Dropout التي تساعد في عدم الوقوع في مشكلة Overfitting، وطبقة BatchNormalize التي تساعد في تسريع عملية التدريب للشبكة وبالتالي التقليل من الدورات التدريبية. فمثلاً شبكة Fer-Net وهي إحدى شبكات الطي العصبونية والتي تم تقديمها من أجل الكشف عن المعالم الوجهية تتكون من خمس طبقات طي كلٍ منها تستخدم مرشحات (4x4)، تكون متبوعة بطبقة BatchNormalize، ومن ثم طبقة Maxpooling ذات

أبعاد (2x2)، كذلك يلها طبقة Dropout من أجل عدم الوقوع في مشكلة Overfitting. وأخيراً طبقة Dense Layer والتي تحوي تابع التنفيعل SoftMax من أجل عملية التصنيف النهائي (Bodapati&Veeranjaneyulu,2021). والشكل (1) يوضح بنية شبكة Fer-Net.

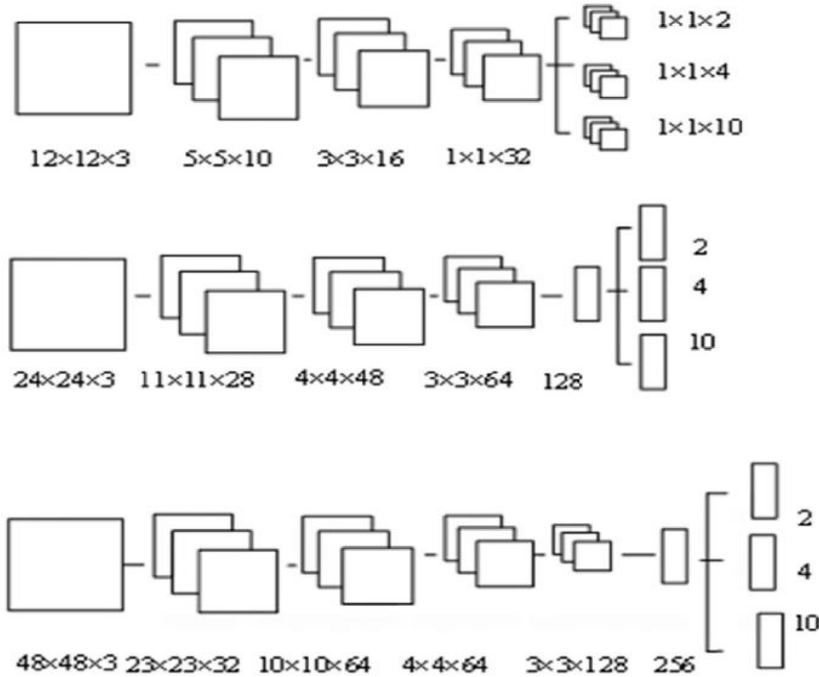


الشكل (1) بنية شبكة Fer-Net

• خوارزميات الكشف عن الوجه:

1. خوارزمية (MTCNN (Multi Task Cascade Convolutional Neural Network):

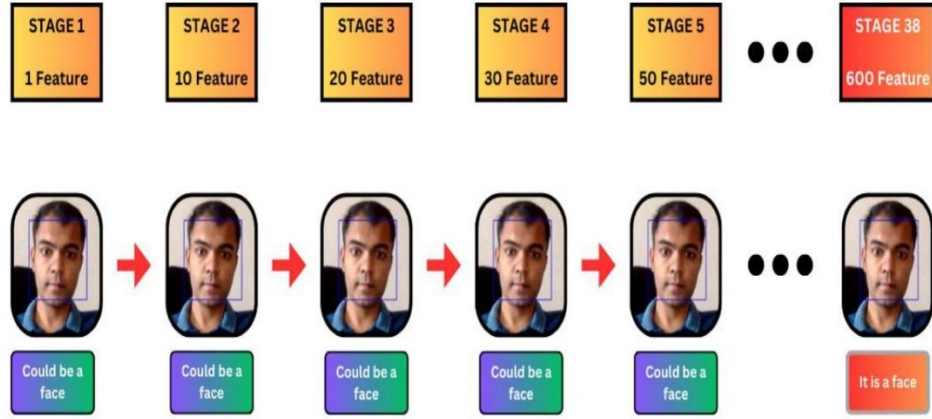
يعتمد مبدأ عمل خوارزمية MTCNN على ثلاثة مراحل هي كالتالي: اكتشاف الوجه والذي يتم من خلال استخدام شبكة PNET، ومن ثم إنشاء حدود للوجه المكتشف من خلال شبكة RNET، يليها استخدام خوارزمية NMS لانقاص حدود الوجه، وأخيراً وضع نقاط تحدد معالم الوجه وذلك باستخدام شبكة ONET والتي تعد الأكثر تعقيداً من الشبكات السابقة. والشكل (2) يوضح بنية شبكة MTCNN (Xie&Wang&Guo,2020).



الشكل (2) آلية عمل خوارزمية MTCNN

2. خوارزمية (Haar Cascade (Haar Feature-based Cascade Classifier):

تعد خوارزمية Haar Cascade من الخوارزميات الغير معقدة والتي يمكن تشغيلها في الزمن الحقيقي، يمكن تدريب هذه الخوارزمية من أجل الكشف عن كائنات مختلفة مثل السيارات، العجلات، المباني، الفواكه، الوجوه. كذلك لا تتطلب هذه الخوارزمية عمليات حسابية معقدة مثل خوارزمية MTCNN (Patel,2023). والشكل (3) يوضح آلية عمل خوارزمية Haar Cascade.



الشكل (3) آلية عمل خوارزمية Haar Cascade

ثانياً- الدراسات السابقة

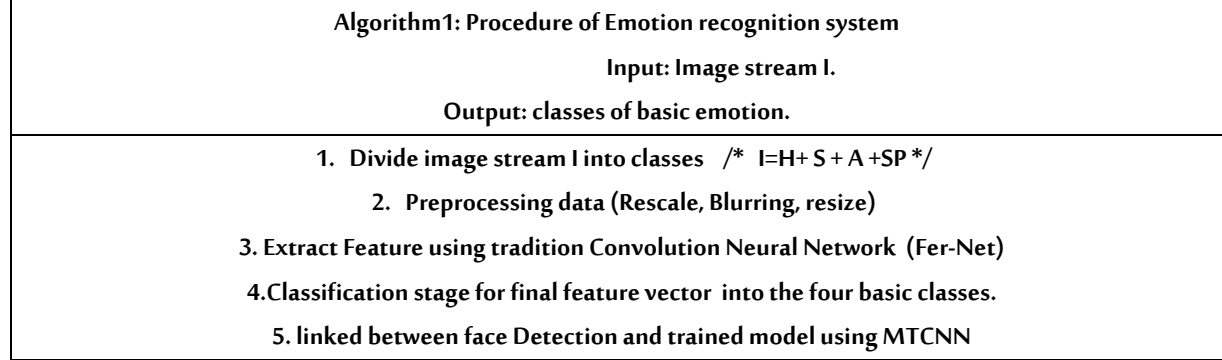
سيتم استعراض أهم الطرق التي اعتمدها الباحثون فيما على تحليل مشاعر الأشخاص سواءً بالاعتماد على التعابير الوجهية أو من خلال الصوت.

- يوجد العديد من الدراسات السابقة التي تمت مجال التعرف على المشاعر وقد تم اختيار بعض أهم الدراسات ومنها:
1. في هذه الدراسة التي تمت عام 2023، فقد تم استخدام شبكة (ResNet-18) والتي تعد إحدى شبكات CNN التقليدية الشهيرة، والتي تتميز بقدرتها على التغلب على مشكلة (Vanishing Gradients)، بلغ عدد البارامترات المستخدمة في هذه الشبكة ذات الثمانية عشر طبقة أكثر من (11,27) مليون بارامتر. كما تم في هذه الدراسة تجربة كلاً من قواعد البيانات القياسية (AffectNet, RAF-DB)، حيث تم تجميع الصور في قاعدة البيانات AffectNet وإعادة توليد صور جديدة حتى بلغ عدد الصور ما يقارب 156 ألف صورة. كذلك في قاعدة بيانات RAF-DB فقد بلغ عدد الصور ما يقارب 200 ألف صورة. ورغم هذا العدد من الصور في كل قاعدة بيانات فقد بلغت قيمة دقة التدريب (Training Accuracy) عند استخدام قاعدة البيانات AffectNet (79.08%)، وقيمة دقة الفحص (Validate Accuracy) ما يقارب (56.54%). أما عند استخدام قاعدة البيانات RAF-DB فقد بلغت قيمة دقة التدريب (76.51%)، وقيمة دقة التقييم (65.67%) (Hunag&Chiang&Chen,2023).
 2. في هذا البحث الذي تم إنجازه عام (2024)، تم تطوير نظام هجين يربط بين شبكة CNN والشبكة التكرارية RNN، من أجل الكشف عن المشاعر من خلال التعابير الوجهية. أُستخدمت في هذه الدراسة قاعدة البيانات Emotional Wearable Dataset 2020. تم اختبار ثلاثة أنظمة مختلفة يعتمد كل منها ربط شبكة CNN مع الشبكة التكرارية ذات المدى الطويل LSTM. حيث تم أولاً ربط شبكة MobileNetV2 ذات الأوزان الخفيفة مع شبكة LSTM والتي أعطت دقة في التصنيف بلغت 59%. كذلك تم ربط الشبكة التقليدية inception v3 مع شبكة LSTM وقد أعطت دقة بلغت 66%. أخيراً تم بناء شبكة طي عصبونية مخصصة وربطها مع شبكة LSTM، حيث حققت دقة بمقدار 63% (Manalu&Rifai,2024).
 3. اعتمدت هذه الدراسة التي أُجريت عام 2024 على استعراض العديد من الدراسات تلخيصها وذلك من خلال تحديد الطرق المستخدمة وقواعد البيانات في مجال التعرف على المشاعر. حيث تم استخدام العديد من شبكات الطي العصبونية CNN، وكذلك الشبكات التكرارية طويلة المدى LSTM، بالإضافة إلى الطرق التقليدية مثل SVM. كذلك تم التطرق إلى أهم المصنفات وأكثرها استخداماً مثل المصنف G-LMSM. أيضاً تم استعراض أهم قواعد البيانات التي تم استخدامها في هذا المجال مثل JAFFE, KDEF, CK+. كذلك تم في هذه الدراسة استعراض الطرق التي قامت بالتركيز على جزء معين من الوجه للكشف عن المشاعر مثل العين أو الفم. وقد بينت الدراسة أن شبكة CNN قد أعطت أفضل النتائج حيث حققت دقة بمقدار 84.4% مقارنةً بالطرق الأخرى (Pereira&Mendes&Ribeiro,2024).
 4. قام الباحثون في هذه الدراسة عام 2022، بإنشاء تطبيق EDR من أجل الكشف عن المشاعر من خلال اللجوء إلى استخدام شبكة CNN، وبالإضافة إلى استخدام مكتبة OpenCV. حيث تم الاعتماد في الكشف عن المشاعر من خلال تعابير الوجه الظاهرة أمام الكاميرا مباشرةً. ثم يتم مقارنة الوجه المكتشف مع الصور الموجودة في قاعدة بيانات Kaggle والقيام بعملية مطابقة بينهم لتحديد مزاجية الشخص الموجود أمام الكاميرا. وقد تم استخدام شجرة اتخاذ القرار Decision Tree لتحديد مدى أهمية المشاعر المكتشفة

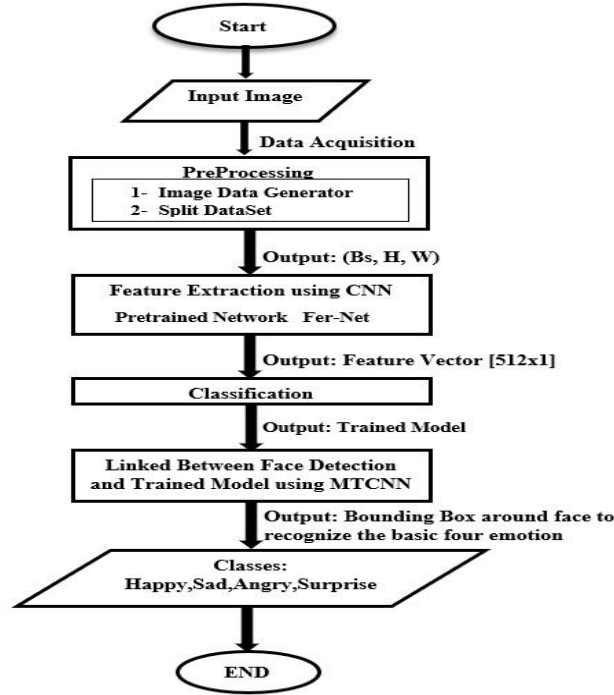
- أو في حال لم يتم تحديد أية مشاعر من الوجه. أما دقة التصنيف الناتجة من استخدام هذه الطريقة فقد بلغت 80% تقريباً (Giri&Singh&Kumar,2022).
5. تمت هذه الدراسة عام 2023، اعتمدت على مقارنة طريقتين للتعرف على المشاعر (تعبير الوجه، الصوت)، تم استخدام شبكة طي عصبونية (ResNet SSD) من أجل اكتشاف معالم الوجه، قبل أن يتم إدخالها إلى شبكة الطي العصبونية ذات البعدين (2D). أما بالنسبة للصوت فقد تم استخدام (Mel-Spectrogram) من أجل تحليل الصوت ومن ثم إدخالها إلى شبكة طي عصبونية ذات بعد واحد (1D). كذلك تم تنقية الفيديوهات المستخدمة في قاعدة البيانات (FER-2013) من الضجيج أو من أي أصوات في الخلفية لكي لا تؤثر على صوت الشخص في الفيديو، كذلك يجب ان يتم فصل صوت الذكور عن الإناث في حال استخدام الصوت للكشف عن المشاعر. وقد اكتفى الباحثين في هذه الدراسة بالحصول على المشاعر بالاعتماد إما على تحليل معالم الوجه أو من خلال الصوت. حققت هذه الدراسة دقة بمقدار 62.5% عند الكشف عن المشاعر باستخدام تعابير الوجه، ودقة بمقدار 82.3% في حال اللجوء إلى الصوت (Ng&Haw&Yoong,2023).
6. اعتمدت الدراسة عام 2022 على مراجعة وتحليل 15 طريقة للكشف عن المشاعر من خلال تعابير الوجه، و17 طريقة من خلال الصوت. تم مناقشة محاسن ومساوي كلاً من هذه الطرق، حيث أستخدم في هذه الطرق مثلاً (CNN, KNN, SVM)، بالإضافة إلى الطرق التي قامت بربط شبكة الطي العصبونية مع الشبكة التكرارية (CNN -LSTM). كذلك فقد قدمت بعض الطرق تحسناً في الدقة لكن أدى ذلك إلى التأخر في عملية المعالجة. أثبتت هذه الدراسة أنّ استخدام شبكة CNN كان الأكثر فعالية للكشف عن المشاعر سواءً باستخدام التعابير الوجهية أو الصوت، خاصةً عند دمجها مع الشبكة التكرارية طويلة المدى LSTM حيث حققت دقة وصلت إلى 88.01% (Lim&Ng,2022).
7. قام الباحثون في هذه الدراسة عام 2019، باختبار تطبيق ثلاثة مشاريع تم إنجازها في جامعات مختلفة من أجل استخدامها في الكشف عن المشاعر من خلال التعابير الوجهية والصوت. حيث تم تجربة كلاً من (FaceReader , AffdexSDK) من أجل التعابير الوجهية، كذلك من أجل الصوت تم تجربة (OpenVokaturi). وقد أثبتت النتائج التجريبية أنه عند اللجوء إلى استخدام التعابير الوجهية للكشف عن المشاعر فقد أعطت نتائج دقيقة في حالة التعابير السعيدة للوجه، بينما في حالة الصوت فقد حقق دقة أكبر في حالات الغضب (Ley&Egger&Hanke,2019).
8. اعتمد الباحثون في هذه الدراسة عام 2020 على تقسيم المشاعر إلى مشاعر أساسية مثل الفرح والحزن، ومشاعر ثانوية مثل الغضب والمفاجئة. تم اللجوء إلى استخدام شبكة CNN أولى من أجل الكشف عن المشاعر الأساسية، تضم ثلاث طبقات طي، وثلاثة طبقات تجميع وكذلك ثلاث طبقات ارتباط كلي. أيضاً تم استخدام شبكة CNN ثانية مع أجل المشاعر الثانوية والتي تضم خمس طبقات طي وخمس طبقات تجميع، بالإضافة إلى استخدام طبقتي Dropout من أجل تسريع عملية التدريب. أما بالنسبة لقاعدة البيانات المستخدمة في البحث، فقد تم تطبيق الدراسة على كلٍ من قواعد البيانات FER2013, JAFFE، حيث حققت دقة بمقدار 97.07% عند تطبيق قاعدة البيانات FER2013، ودقة بمقدار 94.12% عند استخدام قاعدة البيانات JAFFE (Verma,2020). وتعود قيمة الدقة العالية في هذه الدراسات نتيجة استخدام شبكة طي عصبونية خاصة فقط بمشاعر الفرح والحزن، وكذلك شبكة طي عصبونية أخرى من أجل الغضب والمفاجئة.
- نستنتج من الدراسات السابقة أن اختيار تقنية التعلم العميق يلعب دوراً كبيراً في الكشف عن المشاعر. وقد أثبتت شبكات الطي العصبونية تفوقاً على الأنواع الأخرى من الشبكات التكرارية والطرق التقليدية في استخلاص السمات الخاصة بالتعابير الوجهية وتصنيفها، وكذلك يساهم اختيار شبكة طي المناسبة في تحديد عدد البارامترات الناتج من عدد الطبقات الطي المستخدمة في الشبكة. فمثلاً في الدراسة رقم (1) بلغ عدد بارامترات التدريب أكثر من 11 مليون بارامتر تقريباً. كذلك فإن قاعدة البيانات المستخدمة في التدريب تؤثر بشكل كبير في دقة التصنيف سواءً في مرحلة التدريب أو مرحلة الفحص.
- أما بالنسبة للصوت فقد تبين من الدراسات أن دقة التصنيف الناتجة من خلال الصوت قد أعطت أفضل النتائج خلال مشاعر الغضب فقط، كذلك فإن استخدام قواعد بيانات الصوت يحتاج إلى فصل صوت الذكور عن الإناث.
- لذلك سنقوم في النظام المقترح التركيز على استخدام شبكة طي عصبونية مصممة للكشف عن المشاعر من خلال التعابير الوجهية، كذلك تمتلك عدداً قليلاً من البارامترات مقارنةً بالطرق الأخرى. أيضاً سنقوم باختبار قواعد البيانات القياسية خلال تدريب النظام المقترح من أجل الحصول على أفضل قيمة لدقة التصنيف. أيضاً سيتم ربط الشبكة المدربة بخوارزمية الكشف عن الوجوه MTCNN بدلاً من اللجوء إلى زيادة عدد الطبقات في الشبكة المستخدمة، أو اللجوء إلى استخدام الشبكات التكرارية للكشف عن المشاعر خلال الزمن، كما تم تنفيذه في الدراسات المرجعية.

المبحث الثاني- مواد البحث وطرائقه

قبل البدء بشرح النظام المقترح لا بد من التنويه إلى أن النظام قام بالتركيز على التعابير الوجهية فقط من أجل الكشف عن المشاعر الأساسية، وذلك لأن التعابير الوجهية تعد الأكثر دقة في الكشف عن المشاعر مقارنةً مع الصوت وحركات الجسم. كذلك فقد تم التركيز من أجل تجاوز السلبيات التي ظهرت في الدراسات المرجعية السابقة من حيث عدد البارامترات المستخدمة في الشبكة، كذلك اختيار قاعدة بيانات قياسية تضم عدد كبير من البيانات تكفي لعدم الوقوع في مشكلة overfitting. يعتمد النظام المقترح (Mood Pulse) على استخدام شبكة الطي العصبونية Fer-Net من أجل استخلاص السمات الوجهية بدقة ومن ثم تصنيفها إلى المشاعر الأربعة الأساسية (الفرح، الحزن، الغضب، المفاجئة). يمر النظام المقترح بمرحلتين، الأولى بناء الشبكة وتدريبها من أجل الكشف عن المشاعر وتصنيفها وذلك من خلال استخدام قواعد البيانات القياسية، ومن ثم تمرير الشبكة المدربة إلى خوارزمية الكشف عن الوجوه MTCNN لتحديد الوجوه الموجودة في الصور المراد اكتشاف المشاعر الخاصة بها. والخوارزمية التالية توضح خطوات العمل في النظام المقترح وهي على الشكل التالي.



والشكل (4) يوضح بنية النظام المقترح



الشكل (4) المخطط الصندوقي للنظام المقترح.

نلاحظ من المخطط السابق أن النظام المقترح يمر بعدة مراحل سيتم شرحها بالتفصيل من خلال الفقرات التالية.

1- المعالجة المسبقة لبيانات الصور

من أجل توليد العينات التي سيتم استخدامها لاحقاً في التدريب، تم العمل من خلال مكتبة (Keras) حيث قمنا باستخدام تابع (Image Data Generator) لإنتاج الصور من مسارات الأصناف التي تم تحديدها مسبقاً (الفرح، الحزن، الغضب، المفاجئة). حيث ينتج عن مولد الصور بيانات (H,W,N) حيث N حجم الدفعة (Batch Size BS). تمثل العرض والارتفاع. وفي بحثنا قمنا بتدريب سلسلة البيانات المنتقاة بأبعاد عرض وارتفاع 48x48 بيكسل وبحجم دفعة تصل إلى 64. بعد تجهيز مجموعة البيانات ننقل إلى الخطوة الثانية بتمريرها إلى شبكة (CNN) من أجل استخلاص السمات وتصنيفها إلى الأصناف الأربعة التي تم اختيارها.

2- استخلاص السمات وتصنيفها باستخدام شبكات الطي العصبونية التقليدية:

• شبكة Fer-Net :

تم أولاً اللجوء إلى استخدام شبكة (Face Expression Recognition Network) Fer-Net والتي تم طرحها عام 2020، تعد هذه الشبكة أهم شبكات الطي العصبونية التي أعطت دقة كبيرة في التعرف والتصنيف في مجال المشاعر مقارنة بالشبكات الأخرى السابقة المستخدمة في هذا المجال. تم تعديل بنية هذه الشبكة وأحجام المرشحات المستخدمة داخل طبقات الطي، حيث تم استخدام سبع طبقات موزعة على الشكل التالي: أربع طبقات طي (Convolution Layer)، تكون جميع الطبقات المطوية متبوعة بطبقة تجميع (Pooling Layer) يتم تطبيق تابع (Max) من خلالها. كذلك تم 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 تم استخدام طبقة (Flatten) من أجل تحويل شعاع السمات الناتج إلى شعاع ذو بعد واحد (512x1) ليتم تمريره إلى طبقة التصنيف (Classification Layer)، التي تتضمن ثلاث طبقات من الارتباط الكامل (Dense Layer)، تكون كلٍ منها متبوعة بطبقة (Dropout) لتجنب الوقوع في مشكلة Overfitting أثناء تدريب الشبكة. كذلك يتم استخدام تابع التفعيل (SoftMax) في الطبقة الأخيرة من طبقات الارتباط الكامل منها. والجدول (1) يوضح البنية التفصيلية لشبكة Fer-Net مع عدد المرشحات وأحجامها في كل طبقة، وكذلك الخرج الناتج عن كل طبقة وذلك بعد تحديد إطار الدخل للشبكة بأبعاد 46x46.

الجدول (1) الطبقات المستخدمة في شبكة Fer-Net

Layers	Filters	Filter size	Activate Function	Output	Params
Input	-	-	-	46x46x1	0
Conv2D	64	3x3	Relu	46x46x64	64
Batch Normalize	-	-	-	46x46x64	256
Maxpooling	-	2x2	-	23x23x64	0
Conv2D	128	3x3	Relu	21x21x128	73856
Batch Normalize	-	-	-	21x21x128	512
Maxpooling	-	2x2	-	10x10x128	0
Conv2D	256	3x3	Relu	8x8x256	295168
Batch Normalize	-	-	-	8x8x256	1024
Maxpooling	-	2x2	-	4x4x256	0
Conv2D	512	3x3	Relu	2x2x512	1180160
Batch Normalize	-	-	-	2x2x512	2048
Maxpooling	-	2x2	-	1x1x512	0
Flatten	-	-	-	512	0
Dense1	-	-	Relu	256	131328
Dropout	-	-	-	256	0
Dense2	-	-	Relu	128	32896
Dropout	-	-	-	128	0
Dense3	-	-	SoftMax	4	516

من الجدول (1) نلاحظ أن عدد البارامترات الكلي في شبكة Fer-Net وذلك بعد إضافة البارامترات المستخدمة في كل طبقة BN سيكون 1718404 بارامتر. أما عدد البارامترات التي تم تدريبها فقد بلغ عددها 1716484، والبارامترات الغير مدربة بلغت 1920 بارامتر. ويمكن تفسير هذا العدد من البارامترات الغير مدربة إلى وجود طبقات Dropout التي تقوم بإيقاف تدريب بعض البارامترات في كل مرحلة لعدم الوقوع في مشكلة overfitting.

• اختيار قاعدة البيانات لتدريب الشبكة واختبارها

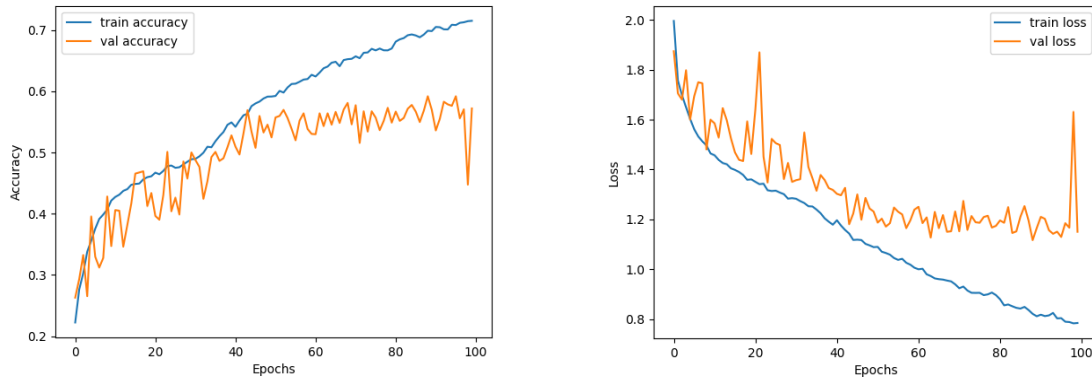
تم تدريب شبكة Fer-Net أولاً باستخدام قاعدة البيانات القياسية (Fer-2013) حيث تم تقسيم البيانات إلى 80% من أجل تدريب البيانات و 20% من أجل اختبار المنظومة. كذلك فقد تم تقسيم الصور في قاعدة البيانات والبالغ عددها 28227 صورة إلى أربع مجلدات يضم كل منها الصور الخاصة بكل صنف (الفرح، الحزن، الغضب، المفاجئة). والجدول (2) التالي يوضح كيفية توزيع الصور إلى الأصناف الأربعة الأساسية.

الجدول (2) توزيع بيانات Fer-2013 إلى الأصناف الأربعة

Category	Happy	Sad	Angry	Surprise
Num of image	10215	7830	5995	4171

كما تم تحديد عدد الدورات التدريبية 100 دورة تدريبية. كذلك فقد قمنا باستخدام خوارزمية التدريب (Adam Optimization Algorithm) حيث أثبتت هذه الخوارزمية كفاءتها بالنسبة لطيف واسع من بنى الشبكات العصبونية، أيضاً ضُبِطت قيمة معدل التعلم الذي تم استخدامه إلى (0.001).

كذلك فقد بلغ الزمن الوسطي للدورة التدريبية (160 s). أما بالنسبة لقيمة الدقة (Accuracy) والتي تمثل قدرة المنظومة على التصنيف الصحيح للصور المدخلة إلى النموذج خلال كل دورة تدريبية والتي يجب أن تتراوح قيمتها بين (0&1)، والخطأ (Loss) الذي يمثل الفرق بين القيمة المتوقعة والقيمة الفعلية الناتجة مقسماً على عدد العينات خلال الدورة التدريبية، فقد تم تسجيل قيمة الدقة خلال كل دورة تدريبية وكذلك قيمة الخطأ. والشكل (5) يوضح الفرق بين قيمة الدقة (Accuracy) وقيمة الدقة خلال كل مرحلة اختبار (val_Accuracy)، وكذلك الفرق بين قيمة الخطأ (Loss) والخطأ خلال كل مرحلة اختبار (Val_Loss)، حيث أن مرحلة الاختبار تتم بعد كل دورة تدريبية.



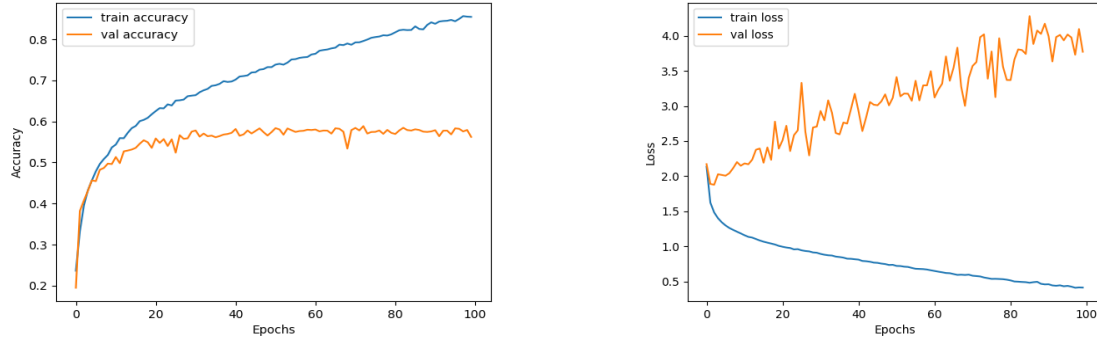
الشكل (5) قيمة الدقة والخطأ في شبكة Fer-Net

نلاحظ من الشكل (5) أن قيمة الدقة خلال الدورات التدريبية قد تجاوزت 70% بينما قيمة الدقة (Val_accuracy) لم تتجاوز أكثر من 55%، كذلك بالنسبة لقيمة الخطأ (Loss) بدأت بقيم كبيرة تجاوزت 1 ولم تنخفض أكثر من 0.7 وهي قيمة كبيرة بالنسبة للخطأ سواءً خلال التدريب أو الاختبار. ويمكن تفسير هذه النتائج إلى ظهور مشكلة Overfitting خلال تدريب المنظومة واختبارها، حيث يعود ذلك إلى العدد القليل من البيانات في قاعدة البيانات التي تم اختيارها، فقد تم أخذ نسبة 20% من قيمة البيانات الكلية. لذلك قمنا باستبدال قاعدة البيانات (Fer-2013) بقاعدة البيانات (AffectNet) والتي تضم بيانات للمشاعر الأساسية تجاوزت أكثر من 60 ألف صورة للأصناف الأربعة. والجدول (3) التالي يوضح توزيع البيانات في قاعدة البيانات AffectNet إلى الأصناف الأربعة الأساسية.

الجدول (3) توزيع بيانات AffectNet إلى الأصناف الأربعة

Category	Happy	Sad	Angry	Surprise
Num of image	21091	15044	13218	10039

وبعد تدريب المنظومة على البيانات الجديدة وبنفس القواعد المتبعة سابقاً سواءً من عدد دورات التدريب وخوارزمية التدريب ومعدل التعلم حصلنا على النتائج التالية الموضحة في الشكل (6) التالي.



الشكل (6) قيمة الدقة والخطأ عند استخدام قاعدة البيانات AffectNet

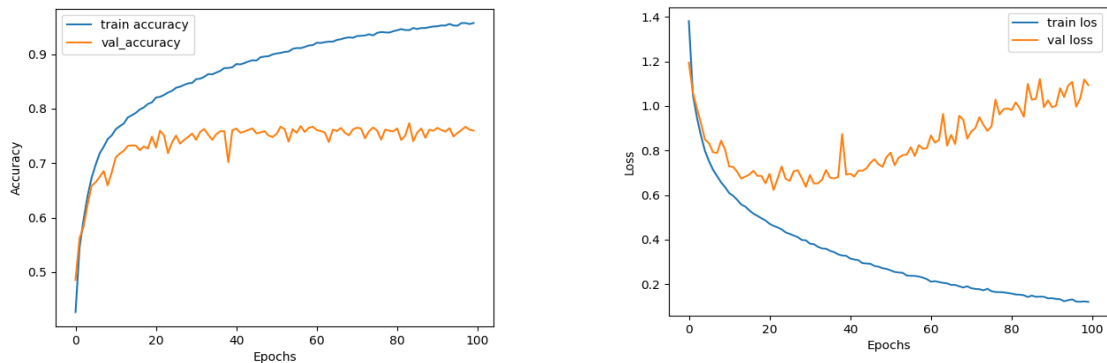
نلاحظ من الشكل (6) أن قيمة الدقة خلال التدريب قد بلغت أكثر من 80%، لكن قيمة الدقة خلال مرحلة الاختبار لم تتجاوز أكثر من 50%، أما بالنسبة للخطأ فقد بلغ قيم كبيرة جداً خلال مراحل التقييم، حيث بلغت قيمة الخطأ قيماً تجاوزت 4 وهي قيمة كبيرة جداً للخطأ الذي يجب أن لا تتجاوز قيمته الأعظمية 1. وهذا يدل على أن البيانات المستخدمة في الاختبار لم تكن كافية. تم تغيير نسبة توزيع البيانات بين التدريب والاختبار، حيث تم جعل بيانات التدريب 70% وبيانات التقييم 30%، لكن لم يطرأ أي تحسين سواءً في نتائج التدريب والتقييم.

أخيراً قمنا بدمج قواعد البيانات السابقة مع قواعد بيانات قياسية جديدة بهدف تجاوز السلبيات التي ظهرت خلال تدريب المنظومة، سواءً في الدقة أو الخطأ الناتج. كذلك فإنه عند تجربة تطبيق عملية Data augmentation للصور الموجودة في قواعد البيانات السابقة، ظهرت مشكلة overfitting بسبب تكرار البيانات. كما أنه في الدراسات المرجعية لم يتم تجربة دمج عدة قواعد بيانات قياسية تحوي صور تنتمي لنفس الأصناف. حيث تم استخدام كلاً من قاعدة البيانات (Fer-2013, AffectNet, RAF-DB, CK+) والتي تحوي كل منها صور خاصة بالتعبير الوجهية. وقمنا بتوزيع الصور حسب كل صنف مع بعضها. حيث بلغ عدد الصور الكلية في قاعدة البيانات المدمجة 125 ألف صورة. والجدول (4) التالي يوضح توزيع بيانات كل قاعدة بيانات إلى الأصناف الأربعة الأساسية.

الجدول (4) يوضح توزيع بيانات قواعد البيانات المستخدمة

Category	Num of image in Fer-2013	Num of image in AffectNet	Num of image of RAF-DB	Num of image of CK+
Happy	10215	21091	7772	6393
Sad	7830	15044	4982	3563
Angry	5995	13218	2305	1298
Surprise	4171	10039	1290	753

أخيراً تم إعادة تدريب المنظومة باستخدام قاعدة البيانات الناتجة عن الدمج وحصلنا على النتائج الموضحة في الشكل (7) التالي.



الشكل (7) قيمة الدقة والخطأ عند استخدام قاعدة البيانات الناتجة عن الدمج.

نلاحظ من النتائج الموضحة في الشكل (7) أن قيمة الدقة ازدادت خلال الدورات التدريبية بشكل كبير، حيث بلغت أكثر من 95%، كذلك فإن قيمة الخطأ انخفضت بشكل كبير حيث وصلت إلى (0.1). وبالتالي فإنه تم اعتماد هذه النتائج في تقييم المنظومة لاحقاً بعد ربطها مع خوارزمية اكتشاف الوجوه MTCNN.

مناقشة النتائج

في البداية تم تدريب الشبكة على جهاز يمتلك المواصفات التالية: معالج بسرعة 2.3GHz core I7، ومعالج رسومي Intel Iris plus بذاكرة رسومية 4GB، وذاكرة الجهاز 32GB. حيث تم أولاً تعديل بنية الشبكة بزيادة عدد طبقات الطي الى أربع طبقات بدلاً من ثلاث طبقات، وذلك لزيادة عدد السمات المستخلصة من الصور. كذلك فقد تم إضافة عدة طبقات Dropout بعد كل طبقة Dense من أجل عدم الوقوع في مشكلة Overfitting أثناء التدريب. وقد بلغ عدد البارامترات الكلي للشبكة بعد القيام بعملية التدريب (1,718,404) بارامتر، وهو عدد قليل مقارنةً باستخدام شبكات الطي الأخرى. تم تدريب الشبكة على قاعدة البيانات القياسية Fer-2013 والتي أعطت دقة منخفضة وخطأ ذو قيمة كبيرة. ثم تم القيام بعملية Data Augmentation والتي أدت إلى ظهور مشكلة Overfitting وذلك نتيجة تكرار البيانات. تم محاولة تغيير قاعدة البيانات المستخدمة إلى AffectNet واتبعتنا نفس الخطوات، لكن قيمة الخطأ بقيت كبيرة أيضاً. لذلك تم البحث عن قواعد بيانات قياسية تم استخدامها في دراسات سابقة يمكن دمجها للحصول على قاعدة بيانات تضم عدد كبير من البيانات تساعد في زيادة عدد عينات التدريب وعينات التقييم، وبالتالي تم دمج قواعد البيانات السابقة مع كلٍ من قاعدة البيانات القياسية RAF-DB,CK+ . تم مقارنة ما تم إنجازه في هذا البحث مع الدراسة المرجعية رقم (1) والتي قام بها الباحثون (Hunag&Chiang&Chen,2023) والتي اعتمدت استخدام شبكة ResNet-18، والتي بلغ عدد البارامترات فيها أكثر من 11 مليون بارامتر. وقد اعتمدت هذه الدراسة على استخدام عملية Data Augmentation لزيادة عدد البيانات الموجودة سواء في قاعدة البيانات AffectNet أو قاعدة البيانات RAF-DB. والجدول (5) يوضح ما تم إنجازه في بحثنا ومقارنتها مع النتائج في البحث المذكور أعلاه.

الجدول (5) مقارنة بين النظام المقترح مع الدراسة المرجعية رقم (1)

النظام المقترح بعد تدريبه باستخدام قاعدة البيانات النهائية	النظام في الدراسة المرجعية رقم (1)
Parameters	1,718,404
Dataset	11,27 Million
Number of Data	Collect Data from four Dataset
Training Accuracy	Data augmentation to increase data
Validate Accuracy	156,000 (AffectNet) after Data augmentation 200,000 (RAF-DB) after Data augmentation
	79.08% (AffectNet) 76.51%(RAF-DB)
	56.54% (AffectNet) 65.67%(RAF-DB)

من الجدول (5) نلاحظ أن استخدام عملية Data Augmentation لزيادة عدد البيانات، لم يساعد في زيادة دقة التصنيف للمنظومة في الدراسة المرجعية. بينما عند زيادة عدد البيانات وتجميعها في قاعدة بيانات واحدة واستخدامها في تدريب المنظومة أدى إلى نتائج جيدة وصلت إلى أكثر من 95%. كذلك فإن شبكة الطي المستخدمة في الدراسة المرجعية هي شبكة ResNet-18 والتي تحوي 18 طبقة وبالتالي سيؤدي ذلك إلى زيادة عدد البارامترات حتى وصل إلى 11,27 مليون بارامتر وبالتالي تكلفة حسابية كبيرة وزمن تدريب كبير، رغم أن ذلك لم يؤثر على زيادة دقة التصنيف. بينما الشبكة المستخدمة في النظام المقترح تحوي 4 طبقات طي فقط، حيث بلغ عدد البارامترات الكلي خلال التدريب 1,718,404 بارامتر، وبالتالي تخفيض في الكلفة الحسابية للشبكة وزمن تدريب غير كبير نسبياً. أخيراً نستنتج أن زيادة عدد البيانات من خلال توليد صور جديدة باستخدام Data Augmentation لم يؤدي إلى نتائج جيدة مثل زيادة عدد الصور من خلال دمج قواعد البيانات القياسية التي تتضمن صور تنتمي إلى نفس الأصناف الأربعة الأساسية.

كذلك سيتم مناقشة نتائج تمرير الشبكة المدربة في النظام المقترح إلى خوارزمية MTCNN من أجل تحديد الوجوه الموجودة في الصورة. عند استخدام الشبكة المدربة بدون اللجوء إلى خوارزميات الكشف عن الوجوه، فإن هذا النموذج سيكون مفيد في حال كانت الصورة تحتوي على وجه قريب فقط، وفي حال وجود أكثر من وجه أو وجود خلفيات متعددة سيكون من الصعب أن تقوم الشبكة المدربة باكتشاف باقي الوجوه لأنها ستعتبر الوجه القريب يمثل كامل الصورة. وقد تم تجربة تمرير مجموعة من العينات الجديدة بطريقتين، تمرير العينات الجديدة إلى الشبكة المدربة وبدون استخدام خوارزمية الكشف عن الوجوه والاكتفاء بالشبكة التي ستقوم باستخلاص السمات وتصنيفها. كذلك تم تمرير العينات الجديدة نفسها إلى الشبكة المدربة بعد ربطها مع خوارزمية MTCNN لتحديد الوجوه الموجودة في الصورة قبل أن يتم الكشف عن المشاعر. والشكل (8) يوضح نتيجة ربط خوارزمية MTCNN مع الشبكة المدربة، وعدم استخدامها مع الشبكة المدربة.

الشكل (8) مقارنة بين استخدام خوارزمية MTCNN وبدون استخدامها



من الشكل (8) نلاحظ أنه في حال ربط الشبكة المدربة مع خوارزمية MTCNN فقد أعطت دقة كبيرة في تحديد الوجوه في الصورة وتحديد حالة كل وجه على حدا. بينما في حال استخدام الشبكة المدربة لوحدها فإنها تأخذ الوجه الأقرب فقط وتعطي معالجة للوجه الأقرب فقط، بدون أن تأخذ بقية الوجوه الموجودة في الصورة. ومن هنا يتبين أهمية ربط الشبكة المدربة مع خوارزمية التعرف على الوجوه من أجل الحصول على الدقة في اكتشاف المشاعر لجميع الوجوه في الصورة.

أيضاً عند مقارنة ما تم إنجازه من ربط الشبكة المدربة مع خوارزمية MTCNN ، والدراسة المرجعية رقم (2) والتي تم فيها ربط شبكة CNN مع الشبكة التكرارية LSTM، فإن هذه المنظومة الهجينة أعطت دقة كبيرة في حال الصور المتسلسلة، أي أنها تعالج التغيرات الزمنية في العواطف أو تعبيرات الوجه الواحد. ولكنها ستأخذ زمن استجابة كبير في حال وجود أكثر من وجه في نفس الصورة. بالإضافة إلى العدد الكبير للبارامترات المرافقة للمنظومة نتيجة استخدام شبكة طي عصبونية وشبكة تكرارية. وبالتالي فإن استخدام الشبكة المدربة مع خوارزمية MTCNN أعطت دقة كبيرة في تحديد الوجوه الموجودة في الصورة، بالإضافة إلى السرعة في زمن الاستجابة والمعالجة. أخيراً قمنا بتقييم المنظومة من خلال تمرير صور جديدة للمشاعر الأساسية التي تم اختيارها، حيث تم تمرير (1000) صورة جديدة إلى المنظومة والقيام بتقييم عمل المنظومة من خلال عوامل التقييم (TP,TN,FP,FN) حيث يمكن عرض نتائج التقييم من خلال الجدول (2) التالي.

الجدول (6) عوامل تقييم المنظومة

	Happy	Sad	Angry	Surprise
Happy	Tp=200	Fn=10	Fn=20	Fn=5
Sad	Fp=15	180	10	10
Angry	Fp=25	5	190	10
Surprise	Fp=10	20	10	220

من الجدول (2) نلاحظ أن قيم عوامل التقييم الخاصة بالحالة (Happy) هي على الشكل التالي:

$$\text{True Positive(TP)}=200$$

$$\text{False Positive(FP)}=15+25+10=50$$

$$\text{False Negative(FN)}=10+20+5=35$$

$$\text{True Negative(TN)}= 1000 - (200 + 50+35)=715$$

وبنفس الطريقة نوجد قيم عوامل التقييم للحالات الأخرى (Sad , Angry, Surprise)

نقوم باستخدام قيم عوامل التقييم من أجل حساب عامل الدقة الناتجة خلال التنبؤ بالحالة (Happy) ، حيث يتم حسابه كما يلي.

$$\text{Accuracy}=(\text{Tp}+\text{TN}) / (\text{Tp}+\text{TN}+\text{Fp}+\text{FN})$$

بالتعويض بالقيم الناتجة نجد أن قيمة عامل الدقة هي (91.5%). وبنفس الطريقة نقوم بإيجاد قيمة عامل الدقة للمشاعر الأخرى.

$$\text{Accuracy (Sad)}=88.5\%$$

Accuracy (Angry)=86.5%

Accuracy (Surprise)=84.5%

وبالتالي نستنتج أن المنظومة أصبحت قادرة على التعرف على المشاعر الأساسية وبنسبة دقة عالية بالنسبة لجميع المشاعر التي تم اختيارها.

التوصيات:

بناءً على النتائج التي تم التوصل إليها توصي الدراسة بالتالي:

- 1- الزيادة الكبيرة في حجم قاعدة البيانات المستخدمة يساعد بشكل كبير في تدريب شبكة الطي العصبونية والحصول على نتائج عالية في دقة التصنيف وأقل قيمة ممكنة للخطأ الناتج، بدون الحاجة إلى استخدام أكثر من شبكة طي عصبونية من أجل الكشف عن المشاعر.
- 2- الاستفادة من خصائص شبكة الطي العصبونية Fer-Net والتي صُممت من أجل التعرف على التعابير الوجهية وذلك لأنها تستخدم في طبقات الطي مرشحات ذات أحجام صغيرة تساعد في استخلاص التفاصيل والميزات الدقيقة في الوجه وبالتالي تساعد في الحصول على دقة عالية في التصنيف.
- 3- يمكن أن يتم بناء نظام هجين يعتمد على اكتشاف المشاعر من خلال التعابير الوجهية والصوت وحركات الجسم، وبالتالي زيادة دقة التصنيف إلى حد كبير جداً.

References

- Bodapati, J.D., & Veernajaneyulu, N.(2021). FerNet: A deep CNN architecture for facial expression recognition in the wild. institution of engineers, India, 103(2):439–448.
- Cootes, T.F., Edward, G.j., & Taylor, C.J. (1999). Active Appearance Models. University of Manchester, Manchester M13, 9PT, U K.
- Huang, Z.Y., Chiang, C-C., & Chen, J-H.(2023). A study on computer vision for facial emotion recognition. Scientific reports, nature portfolio., ISSN: 13-8425.
- Giri, Sarwesh., Singh, Gurchetan., Kumar, Babul.(2022). Emotion detection with facial feature recognition using CNN&OpenCV. International conference on advance computing and innovative technologies in engineering(ICACITE). 2022IEEE.
- Kobayashi, T., Ye, J. (2014). Acoustic feature extraction by statistics based local binary pattern for environmental sound classification. IEEE international conference on Acoustics, speech and signal processing(ICASSP), May 2014, Florance, Italy.
- Lim, Yixen., Ng,K.W., &Naveen, Palanichamg. (2022) . Emotion recognition by facial expression and voice: review and analysis. Journal of informatics and web engineering, Vol.1 No.2 , ISSN: 2821-370x.
- Ley, Matthias.,Egger, Maria.,& Hanke, Sten.(2019). Evaluating methods for emotion recognition based on facial and vocal features. European conference on ambient intelligence, Rome, Italy.
- Manalu, H.V., Rifai, A.P.(2024). Detection of human emotions through facial expressions using hybrid convolutional neural network-recurrent neural network algorithm. Elsevier, intelligent system with applications 21.
- Ng, k.W., Haw, S-C., &Yoong, Y-J.(2023). Emotion recognition on facial expression and voice: analysis and discussion. International journal on advanced science engineering and information technology. Vol.13 No 5, ISSN: 2088-5334.
- Pereira, R., Mendes, Carla., Ribeiro, Jose.(2024). Systematic review of emotion detection with computer vision and deep learning. Sensors journal.
- Patal, Kruti.D., Patal, Nilen.A.(2023). Face recognition using Haar Cascade algorithm in python. International journal of current science, volume 13 issue 4, ISSN: 2250-1770.
- Trevisan, D.A., Birmingham, E. (2016). Are emotion recognition abilities related to everyday social functioning in ASD? A meta-analysis. Elsevier, research in Autism Spectrum Disorders, volume 32 , pages 24-42.
- Verma, Garima.,Verma, Hemraj. (2020). Hybrid-deep learning model for emotion recognition using facial expression. Springer Japan KK, part of springer nature.
- Xie, Y.G., Wang, Hui., & Guo, S.H.(2020). Research on MTCNN face recognition system in low computing power scenarios. Journal of internet technology volume 21 No.5.