مقارنة أداء الشبكات العصبية الترشيحية والنماذج الهجينة في تصنيف فقاعات الإجابة

المهندس محمد أحمد علي[1]

إشراف: الدكتورة اليدا إسبر [2] الدكتور سهيل الحمود[3]

الملخص

تهدف هذه الدراسة إلى مقارنة أداء نموذج يعتمد على الشبكات العصبية الترشيحية (CNN) مع نموذج هجين يجمع بين بنية الشبكات العصبية الترشيحية CNN والمحولات Transformer في مهمة تصنيف فقاعات الإجابة (مملوءة، فارغة، أو مشطوبة). قمنا بتطوير شبكة CNN تتألف من خمس طبقات ترشيحية كنموذج أساسي، ونموذج هجين مكون من طبقات ترشيحية تليه وحدتا Transformer. تم تدريب النماذج على صور فقاعات إجابة حقيقية مقسمة إلى ثلاث فئات (مملوءة، فارغة، مشطوبة) باستخدام إعدادات تدريب موحدة. تم اختبار تأثير زيادة بيانات التدريب (Data Augmentation) على دقة التعميم في كلا النموذجين، أظهرت النتائج أن النموذج الهجين بدون زيادة البيانات حقق أفضل أداء بدقة تصنيف بلغت حوالي 99.96%. كما بينت النتائج أن معظم الأخطاء كانت في التفريق بين الفقاعات المملوءة والمشطوبة. تشير الدراسة إلى أن دمج خصائص محلية مستخرجة بواسطة الشبكة العصبية الترشيحية مع سياق عالمي يوفره المحول (Transformer) يحسّن دقة تصنيف الصور في المهام الدقيقة مثل تصنيف فقاعات الإجابة، كما تبيّن أن استخدام زيادة البيانات بشكل غير مدروس قد يؤدي في بعض الحالات إلى نتائج عكسية. تسلط هذه النتائج الضوء على منهجية فعّالة يمكن اعتمادها في أنظمة تصحيح الاختبارات الآلية لزيادة المرونة والدقة.

الكلمات المفتاحية: التعرف الآلي على العلامات(OMR)، المحولات(Transformers)، الشبكة العصبية الترشيحية(CNN)، زيادة البيانات(DataAugmentation)، فقاعات الإجابة(Bubbles).

[1] طالب دكتوراه - قسم هندسة البرمجيات ونظم المعلومات – كلية الهندسة المعلوماتية – جامعة حمص - سورية
[2] أستاذ مساعد – قسم هندسة البرمجيات ونظم المعلومات – كلية الهندسة المعلوماتية - جامعة حمص - سورية
[3] مدرس – قسم هندسة البرمجيات ونظم المعلومات – كلية الهندسة المعلوماتية - جامعة حمص - سورية

Comparative Performance Analysis of Convolutional Neural Networks and Hybrid Models in Answer Bubble Classification

Eng. Mohammed Ahmad Ali
Supervision: Dr. Alida Isber Dr. Suhel Hammoud

ABSTRACT

This study aims to compare the performance of a model based on Convolutional Neural Networks (CNN) with a hybrid model that integrates CNN architecture with Transformer layers in the task of answer bubble classification (filled, empty, or crossed-out). We developed a CNN composed of five convolutional layers as the baseline model, and a hybrid model consisting of convolutional layers followed by two Transformer units. The models were trained on real-world answer bubble images categorized into three classes—filled, empty, and crossed-out—using unified training settings.

The impact of data augmentation on generalization accuracy was also examined. Results showed that the hybrid model without data augmentation achieved the highest performance, with a classification accuracy of approximately 99.96%. The findings further indicated that most classification errors occurred in distinguishing between filled and crossed-out bubbles. The study demonstrates that combining local features extracted by CNNs with the global contextual awareness provided by Transformers enhances classification performance in fine-grained visual tasks such as answer bubble recognition. Additionally, it was found that unstructured use of data augmentation can, in some cases, lead to adverse effects. These results highlight an effective methodology that can be employed in automated exam grading systems to improve both accuracy and flexibility.

Keywords: Optical Mark Recognition (OMR), Transformers, Convolutional Neural Networks (CNN), Data Augmentation, Answer Bubble Classification.

1. مقدمة

تعتمد الكثير من القطاعات وخاصة المؤسسات التعليمية على تقنيات التعرف الآلي على العلامات (Optical Mark Recognition - OMR) في تصحيح الاختبارات متعددة الخيارات لما لها من دور في الحصول على نتائج سريعة وتقليل الجهد اليدوي [1]، لكن أغلب الأنظمة المستخدمة ما تزال تقليدية وتعتمد على خوارزميات معالجة صور ثابتة، كأن يتم تحديد ما إذا كانت الفقاعة مملوءة عن طريق تحليل مستويات الرمادي [2,3] أو حساب عدد البكسلات السوداء ضمن فقاعة الإجابة [4,5,6]، لذا عانت هذه التقنيات من محدودية في التعامل مع التنوع في شكل علامات التظليل أو وجود علامات شطب، وبالتالي قد يؤدي ذلك إلى انخفاض في دقة التصنيف، وبسبب هذا الأمر فرضت هذه الأنظمة شروط صارمة على الطالب بمنع إلغاء إجابة من خلال علامة X أو شطب على الفقاعة؛ فإذا قام بذلك قد تعجز أنظمة التصحيح عن التمييز بين الإجابة الملغاة والإجابة المؤكدة وتعتبر إجابته لاغية. ومع التقدم في مجال التعلم العميق برزت الشبكات العصبية الترشيحية (CNN) كأداة قوية في رؤية الحاسوب لقدرتها على استخلاص الخصائص المميّزة من الصور بشكل تلقائي، وقد بدأ اعتمادها في مشاكل تصنيف مشابهة لتجاوز التحديات التي تواجه الطرق التقليدية [7]. بالتوازي مع ذلك ظهرت نماذج المحولات (Transformer) في السنوات الأخيرة وبدأ تبنّيها في مجال الرؤية الحاسوبية حيث تمتاز بنيتها بقدرتها على نمذجة العلاقات طويلة المدى بين أجزاء الصورة [8]، وقد بينت أبحاث لاحقة أن دمج النموذجين معاً في نموذج هجين بإمكانه أن يحقق اكتشاف الميزات بأفضل ما يمكن حيث يستفيد من قدرة شبكات CNN على استخراج سمات محلية فعّالة ومن قدرة المحولات على إعادة توزيع الانتباه عبر الصورة بأكملها [9]. نقترح في هذه الورقة تصنيف فقاعات الإجابات باستخدام نموذج يعتمد على شبكات CNN ونموذج هجين يجمع بين بنية شبكات CNN والمحولات، سنقوم بمقارنة أداء النموذجين تحت ظروف تدريب متماثلة مع التركيز بشكل خاص على تأثير زيادة البيانات (Data Augmentation) على دقة النماذج. تكمن أهمية هذه الدراسة في فهم ما إذا كان إدخال وحدات

Transformer لتحسين استيعاب السياق الكلي للصورة يمكن أن يحسّن من تمييز النموذج بين الفقاعة المملوءة والفقاعة المشطوبة بعلامة، وكذلك بين الفقاعة المملوءة والفقاعة الفارغة، كما نهدف إلى فهم دور عمليات زيادة البيانات augmentation في مثل هذه المهام التي يكون فيها تنوع البيانات الحقيقية محدود (كون جميع الصور تأتي من عمليات مسح ضوئي قياسية). من المتوقع أن تساهم نتائج هذه الدراسة في تصميم أنظمة تصحيح آلي أكثر دقة ومرونة، خاصة في البيئات التعليمية التي تتطلب التعرف على أنماط تظليل غير تقليدية (مثل إلغاء الإجابة بعلامة).

2. أهداف وأهمية البحث

يهدف البحث إلى تحقيق ما يلي:

- تصميم نظام تصنيف دقيق لفقاعات الإجابة في أوراق OMR باستخدام بنى CNN و Transformer.
- تحليل تأثير استخدام تقنيات زيادة البيانات Augmentation على أداء النماذج في بيئات مستقرة.
- تقييم دقة النماذج في التمييز بين الفقاعات المملوءة والمشطوبة والفارغة.

تنبع أهمية هذا البحث من الحاجة المتزايدة إلى أنظمة تصحيح إلكترونية دقيقة وقابلة للتعميم، خاصة في المؤسسات التعليمية التي تعتمد على نماذج OMR لقياس أداء الطلبة.

ويساعد هذا البحث في فهم الفروق الدقيقة بين المزايا التي تتيحها بنى شبكات CNN التقليدية والبنى المحسنة منها باستخدام المحولات، وتأثير توليد زيادة في البيانات (augmentation) على دقة التصنيف، مما ينعكس على تحسين أدوات التصحيح المؤتمت وتقليل نسبة الخطأ البشري.

تُلخص مساهمات البحث بالنقاط التالية:

- تطوير مقارنة بين أربع نماذج للتصنيف تشمل نموذج CNN والنموذج الهجين (CNN+Transformer) (مع وبدون زيادة البيانات).

- استخدام إعدادات تدريب موحدة لجميع النماذج لتقديم مقارنة فعلية.
- تحليل مصفوفات الالتباس وحالات الخطأ في التصنيف.
- توضيح تأثير عمليات زيادة البيانات (Augmentation) على مهام التصنيف الدقيقة ضمن بيئة بيانات مستقرة.
- تقديم توصيات عملية لتحسين تصميم نماذج OMR المستقبلية بناءً على النتائج التجريبية.

3. دراسة مرجعية

ركزت العديد من الأعمال السابقة على تحسين عملية تصحيح الاختبارات المؤتمتة، يمكن تصنيف هذه الأعمال ضمن فئتين رئيستين: طرق تقليدية قائمة على معالجة الصور، وطرق حديثة قائمة على التعلم الآلي العميق. فيما يلي نستعرض أبرز ما قدمته الدراسات في كلا الاتجاهين:

3.1. نماذج تقليدية Traditional Models:

اعتمدت الطرق الأولى لتحديد الإجابات على خطوات معالجة صورة بسيطة، مثل تحويل الصورة إلى الأبيض والأسود ثم تحديد مناطق العلامات ومقارنتها بعتبة معينة. قدم Deng وآخرون [10] حلاً منخفض التكلفة لتحديد الإجابات المملوءة باستخدام تقنيات معالجة متوازية، كما اقترح Sanguansat [11] في دراسته نظام آلي لقراءة العلامات يعتمد على أساليب معالجة الصور التقليدية، تضمنت أغلب هذه الأساليب خطوات معايرة (لتصحيح ميلان ورقة الإجابة) ثم تقسيم الورقة إلى خانات كل سؤال، وأخيراً تصنيف كل فقاعة عبر قياس كثافة اللون أو عدد البكسلات الغامقة فيها [6-2]. نجحت هذه الطرق إلى حد كبير في بيئات منتظمة، لكنها وجدت صعوبة في التعامل مع الحالات غير الاعتيادية (مثل وجود علامات شطب داخل الفقاعة أو اختلاف أشكال التظليل من طالب لآخر) مما أثر على موثوقية نتائجها.

3.2. نماذج شبكاتٍ عصبيةٍ عميقةٍ Deep Neural Networks Models:

مع ظهور الشبكات العصبية وازدياد القدرة الحاسوبية، اتجه الباحثون لتوظيف التعلم العميق في مشكلة تصنيف فقاعات الإجابة. من أوائل الأبحاث في هذا السياق دراسة عفيفي وحسين [12] حيث قدموا خلالها منهجية لتصنيف مربعات الإجابات في أوراق الاختبار متعددة الخيارات، عملوا على استخراج صورة كل فقاعة إجابة بشكل مستقل وإدخالها إلى نموذج تصنيف يقوم بتحديد حالة

الفقاعة (الإجابة المؤكدة (مملوءة)، والإجابة الملغاة بعلامة (مشطوبة)، والإجابة الفارغة)، جرّب الباحثان في هذه الدراسة طريقتين: الأولى باستخدام السمات التقليدية (مثل طريقة Bag of Visual Words) والثانية باستخدام شبكة CNN مدرّبة لهذا الغرض، بيّنت نتائجهم أن نموذج الـCNN استطاع تحقيق دقة تصنيف أعلى ومرونة أفضل في التعامل مع الاختلافات بين أنماط التظليل المختلفة وفي التعرف على الحالات التي يضع فيها الطالب علامة (×) لإلغاء إجابة معينة، وتُعد مجموعة البيانات التي نشروها (MCQ Dataset) إحدى أوائل المجموعات المتاحة التي تحتوي على أمثلة حقيقية لحالات إلغاء الإجابات بعلامات شطب، مما ساهم في تحفيز أعمال لاحقة.

وفي السنوات الأخيرة اقترح Mondal وآخرون [13] نموذج OMRNet يستخدم بنية MobileNetV2 خفيفة الوزن لتصنيف الفقاعة إلى ثلاث فئات، وقد حقق دقة عالية جداً بعد تدريبه على مجموعة بيانات فقاعات الإجابة التي قدمها عفيفي في بحثه سابقاً. أثبتت هذه النماذج المقترحة أن شبكات CNN بإمكانها تحقيق دقة عالية في تصنيف فقاعات الإجابات عند توفر بيانات تدريب كافية وعند تمثيل جميع حالات الإجابات الممكنة (مملوءة بالكامل، مملوءة جزئياً، مشطوبة، ...إلخ) في البيانات.

3.3. نماذجٌ مبنيةٌ على المحوّلات Transformer-based Models والنماذج الهجينة:

في الوقت الذي حققت فيه نماذج CNN نجاحات ملموسة، بدأت الأبحاث المهتمة بمجال الرؤية الحاسوبية باستكشاف نماذج مبنيةٌ على المحوّلات Transformer. يُعتبر نموذج Vision Transformer (ViT) المقترح من قبل Dosovitskiy وزملائه [14] أحد أبرز التطورات، حيث أظهر إمكانية منافسة نماذج CNN عبر تقسيم الصورة إلى بقع وتمثيلها بتسلسل مدخلات للمحول .Transformer، يتمتع Transformer بقدرته على النظر إلى الصورة كوحدة متكاملة والتعلم من علاقات بعيدة المدى بين أجزائها.

وقد أشارت بعض الدراسات إلى تكاملية دور كل من CNN و Transformer؛ حيث تتفوق الـCNN في استخلاص ملامح محلية (كحافة أو زاوية ضمن فقاعة) بينما يُبرز الـTransformer العلاقات السياقية الشاملة (مثل نمط العلامة داخل الفقاعة بالنسبة لإطارها الكلي). بناءً على ذلك اتجهت بعض الأبحاث الحديثة إلى بناء نماذج هجينة تمزج المعماريتين معاً، حيث قدم Nie

وآخرون [15] مثال ناجح لذلك في المجال الطبي، إذ دمجوا CNN مع Transformer لتصنيف صور الأورام الجلدية وحققوا تفوق على نماذج CNN التقليدية في الدقة. في ذلك النموذج تتولى نماذج CNN استخراج خصائص أولية من الصورة، ثم يقوم المحول Transformer بمعالجة هذه الخصائص على مستوى أعلى لاكتساب فهم شامل للصورة قبل اتخاذ قرار التصنيف. هذه النتائج الواعدة تدفعنا لاستكشاف ما إذا كان نهج مشابه يمكن أن يفيد في مشكلة تصنيف فقاعات الإجابة، خاصة في التمييز الدقيق بين الفقاعة المملوءة وتلك المشطوبة اللتان قد تبدوان متشابهتين للوهلة الأولى.

4. مواد وطرائق البحث

يستعرض هذا القسم توصيف مجموعة البيانات والنموذج والخوارزميات المستخدمة ومقاييس التقييم المُعتمدة وتفاصيل بيئة التجربة.

1.4. توصيف مجموعة بيانات البحث Dataset

تم استخدام مجموعة بيانات تحتوي على صور لفقاعات الإجابات المقصوصة من أوراق إجابة حقيقية عددها 18000 صورة (6000 صورة لكل فئة). كل صورة بالأبيض والأسود أو التدرج الرمادي تمثل فقاعة إجابة واحدة مقتطعة (من ورقة إجابة مسحوبة ضوئيًا) ومصنفة يدويًا إلى إحدى ثلاث فئات:

- فئة "مملوءة " (Filled): وتشمل الفقاعات التي تم تظليلها بشكل واضح كدلالة على اختيار الإجابة.
- فئة "فارغة" (Empty): وتشمل الفقاعات التي تُركت بدون أي علامة أو تظليل.
- فئة "مشطوبة" (Crossed-out): وتشمل الفقاعات التي وضع عليها علامة إلغاء (مثل علامة X) أو تم شطبها بخطوط مائلة للدلالة على أن الطالب اختار إلغاء تلك الإجابة.

2.4. النماذج والخوارزميات المستخدمة

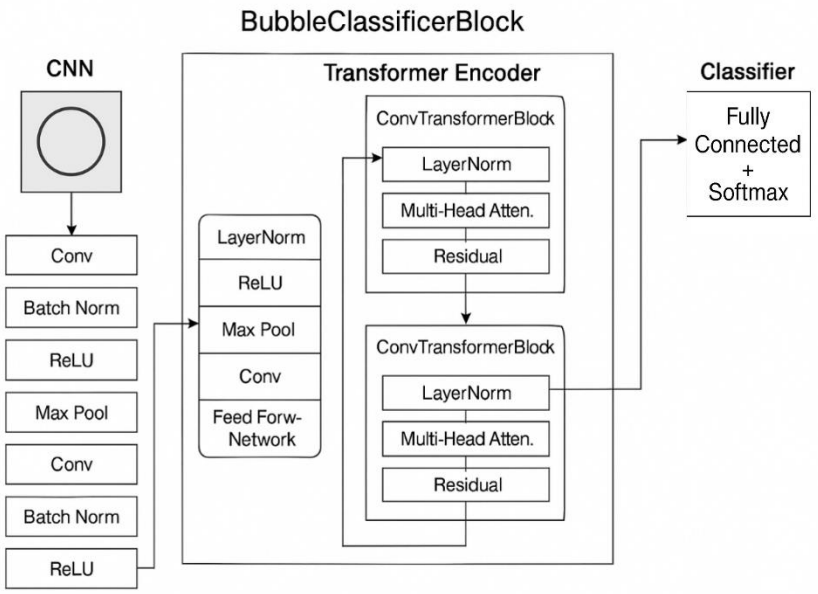
طورنا نموذجين رئيسيين للمقارنة:

1.2.4 نموذج CNN:

عبارة عن شبكة عصبية ترشيحية مكونة من خمس طبقات ترشيحية متتابعة، تتخللها طبقات تفعيل غير خطية (ReLU) وطبقات تجميع (Pooling) لخفض الأبعاد تدريجيًا. ينتهي النموذج بطبقة كاملة الاتصال (Fully Connected) مع دالة softmax بثلاث وحدات تمثل احتمالات الفئات الثلاث. تم اختيار بنية CNN بسيطة نسبياً (5 طبقات) لتكون مماثلة من حيث التعقيد للنموذج الهجين اللاحق، وضمان أن الفرق في الأداء يرجع لوجود وحدة Transformer من عدمه وليس لحجم النموذج. هذا النموذج لا يحتوي على أي مكوّن Transformer أو آلية انتباه، ويُفترض أنه يلتقط الملامح المحلية مثل مساحات الظل داخل الفقاعة أو أجزاء من حدودها.

2.2.4 النموذج الهجين(CNN+Transformer) :

يتكون من قسمين أساسيين: قسم أمامي Extractor CNN وقسم خلفي Transformer Encoder. القسم الأول هو طبقات من شبكة CNN (بضع طبقات ترشيحية) بهدف استخراج خرائط الخصائص المميزة من صورة الفقاعة. بعد المرور عبر الطبقات، نحصل على تمثيل ذو أبعاد منخفضة للصورة (مثلاً خارطة خصائص بحجم 4×4×عمق معيّن). يتم تسطيح (Flatten) هذه الخارطة وتحويلها إلى سلسلة من المتجهات (patch embeddings) تمهيداً لإدخالها إلى القسم الثاني. القسم الثاني هو المحوّل (Transformer) مكوّن من طبقتين من طبقات Encoder الخاصة بالمحول Transformer. كل طبقة Transformer تتألف من آلية انتباه متعدد الرؤوس (Multi-Head Self-Attention) يليها طبقة تغذية أمامية (Feed-Forward Network) مع إضافة اتصالات الالتفاف (skip connections) والتطبيع (Layer Norm) كما هو متبع في بنية Transformer الأصلية. يقوم Transformer بمعالجة متجهات الخصائص المستخرجة من شبكة CNN بطريقة تأخذ بالاعتبار العلاقات المتبادلة بين جميع أجزاء الصورة (أي أنه يمكنه مثلاً ملاحظة علامة الشطب كاملة وعلاقتها بحواف دائرة الفقاعة). مخرج آخر طبقة Transformer يمر عبر طبقة تصنيف نهائية (fully connected + softmax) لإنتاج احتمالات الفئات الثلاث. الشكل العام لهذا النموذج الهجين موضح ضمن المخطط في الشكل 1 أدناه.



الشكل *1* – بنية النموذج الهجين المقترح

3.4. استراتيجيات زيادة البيانات(Augmentation) :

في البداية تم الاعتماد على مجموعة متنوعة من تقنيات زيادة البيانات أثناء تدريب النماذج لمحاولة تحسين قدرتها على التعميم. شملت هذه التقنيات تدوير الصور بزاوية عشوائية صغيرة (مثلاً ±5 درجات) وقص بسيط أو إزاحة طفيفة للصورة، بالإضافة إلى تغيير مستويات الإضاءة أو التباين بشكل طفيف. الغرض من استخدام augmentation هو إعطاء النموذج حالات متنوعة من نفس الفقاعة حتى لا يحفظ تفاصيل محددة جداً للصورة خلال التدريب، ولكن بخلاف مهام الرؤية التقليدية، تبيّن أن بيانات فقاعات الإجابة تتمتع بقدر عالٍ من الثبات، فكل الصور ناتجة عن ماسح ضوئي بنفس الدقة، والفقاعات دائماً في مواضع متشابهة وبتخطيط موحّد، ولا توجد اختلافات كبيرة في الإضاءة أو الميلان (بعد عملية المعايرة)، لذا بدت بعض عمليات زيادة البيانات وكأنها تولّد حالات غير واقعية بالنسبة لهذه المهمة (مثل تدوير الفقاعة بزاوية كبيرة أو تغيير السطوع بشكل مبالغ فيه). وقد أشار بعض الباحثين إلى أن الإفراط في زيادة البيانات augmentation قد يؤدي إلى إدخال أنماط لا تعكس التوزيع الحقيقي للبيانات، مما يربك النموذج

بدلاً من تحسينه [16]. بناءً على ذلك قمنا بإعداد تجربة خاصة ضمن منهجيتنا: درّبنا نفس النموذجين (CNN والهجين) دون أي عمليات Augmentation على الإطلاق، وقارنا أداءهما مع النماذج نفسها لكن المدرّبة بعد زيادة البيانات augmentation. الهدف من ذلك هو رصد ما إذا كانت عمليات زيادة البيانات تساعد النموذج في التعميم في هذه المهمة أم أنها تتسبب في تشويش قد يضر بالأداء نظراً لطبيعة البيانات المستقرة. باختصار لدينا أربع نماذج نقارن بينها في التجارب:

1) نموذج CNN بدون استخدام augmentation.
2) نموذج CNN مع استخدام augmentation.
3) نموذج هجين CNN+Transformer بدون استخدام augmentation.
4) نموذج هجين CNN+Transformer مع استخدام augmentation.

4.4. بيئة التجربة

استُخدم عدد من مكتبات لغة بايثون Python في تجارب البحث، وأجريت جميع التجارب ضمن بيئة Google Colab pro باستخدام Python 3 على وحدة معالجةٍ رسوميةٍ (GPU) من نوع NVIDIA A100-SXM4-40GB ووحدة معالجةٍ مركزيةٍ (CPU) من نوع Intel(R) Xeon(R) CPU @ 2.20GHz , 6 cores, 12 threads وذاكرة وصولٍ عشوائيٍ (RAM) بحجم 83.5GB.

5. التجارب والنتائج ومناقشتها

مرت عملية الحصول على نموذج بخطواتٍ عدة يمكن تلخيصها وفق الآتي:

1.5. تهيئة وتقسيم مجموعة البيانات

تم تقسيم البيانات إلى مجموعات تدريب وتقييم باستخدام أسلوب (5− fold cross-validation) لضمان الاستفادة القصوى من البيانات المحدودة وضمان أن يتم اختبار النموذج على بيانات لم يرها أثناء التدريب، تمت موازنة التوزيع بين الفئات الثلاث في مجموعات التدريب والاختبار (بعد التجزئة) بقدر الإمكان بحيث لا يطغى عدد عينات فئة على أخرى.

تم تقسيم مجموعة البيانات إلى ثلاث مجموعاتٍ وهي مجموعة التدريب Training (70% من مجموعة البيانات الكلية) ومجموعة التحقق Validation (15%) ومجموعة الاختبار Testing (15%).

2.5. الضبط الدقيق للنموذج

تم تدريب كل النماذج باستخدام نفس إعدادات التدريب لضمان مقارنة عادلة، بداية تم تحجيم الصور كلها لتكون موحّدة الحجم في تدريب كل النماذج (بحجم 67×44) لضمان تناسق المدخلات للنموذج، تم استخدام دالة خسارة (CrossEntropyLoss) مدمجة مع softmax على المخرجات للتصنيف متعدد الفئات، تم اعتماد المُحسّن (Optimizer: Adam) بقيمة تعلم ابتدائية مناسبة (learning rate= 1e-3). كما تم تطبيق ضبط لمعدل التعلم لتخفيضه تدريجياً أثناء التدريب عند ثبات الأداء وعدم تحسن دقة التحقق لفترة محددة (ReduceLROnPlateau)، تم تدريب النماذج لعدد من الأجيال (epochs=50)، وتم تحديد حجم الدفعة (Batch Size=128). وخلال التدريب تمت مراقبة الدقة على مجموعة التحقق (validation) لتتبع أداء النموذج وتجنب حدوث (overfitting) باستخدام الإيقاف المبكر (Early Stopping).

3.5. مقاييس التقييم:

تم تحديد أخطاء التصنيف من قبل كل من النماذج باستخدام مصفوفات الالتباس (Confusion Matrix) [17,18]:

		Assigned Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

الشكل *2* – مصفوفة الالتباس

TP تعني تصنيفاً إيجابياً صحيحاً وFP تعني تصنيفاً إيجابياً خاطئً وTN تعني تصنيفاً سلبياً صحيحاً وFN تعني تصنيفاً سلبياً خاطئً.

يستخدم البحث أيضاً عدداً من المقاييس وهي: الدقة Accuracy والإحكام Precision والاستدعاء Recall ومقياس إف1 F1-Score. تُحسب هذه المقاييس وفق المعادلات التالية [18,19]:

$$\text{الدقة} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{الإحكام} = \frac{TP}{TP + FP}$$

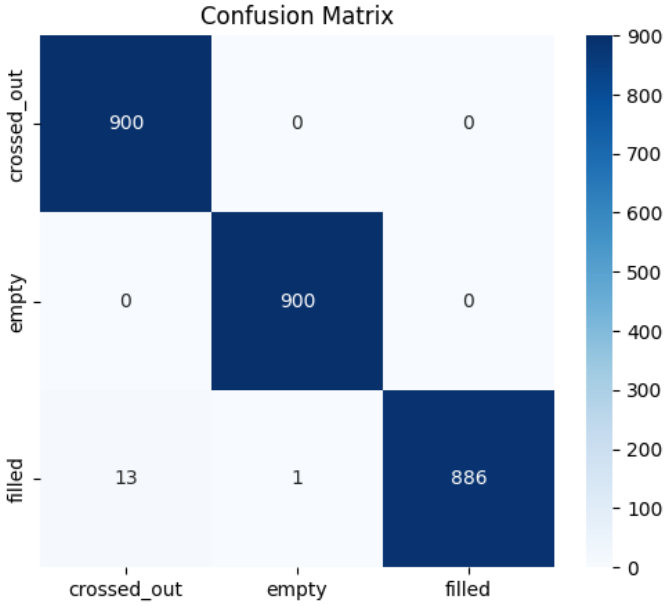
$$\text{الاستدعاء} = \frac{TP}{TP + FN}$$

$$\text{مقيا إف1} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4.5. النتائج والتقييم

بعد تدريب النماذج الأربعة تم تقييمها على مجموعة الاختبار، وبشكل عام أظهرت جميع النماذج دقة تصنيف عالية جداً لكن مع وجود فروق دقيقة فيما بينها. نستعرض فيما يلي نتائج كل نموذج على حدا:

1.4.5 نموذج CNN بدون Augmentation



الشكل *2* – مصفوفة الالتباس لنموذج CNN بدون زيادة البيانات

يعرض الشكل 2 مصفوفة الالتباس لنموذج CNN المدرب بدون أي زيادة للبيانات. صنّف النموذج جميع الفقاعات بشكل صحيح (900 صحيحة)، أيضاً بالنسبة للفقاعات الفارغة تم تصنيفها جميعها بشكل صحيح على أنها فارغة، بينما وقع 14 خطأ عند تصنيف الفقاعات المملوءة حيث اعتبر فقاعة مملوءة على أنها فارغة و 13 منها تم تصنيفها على أنها مشطوبة.

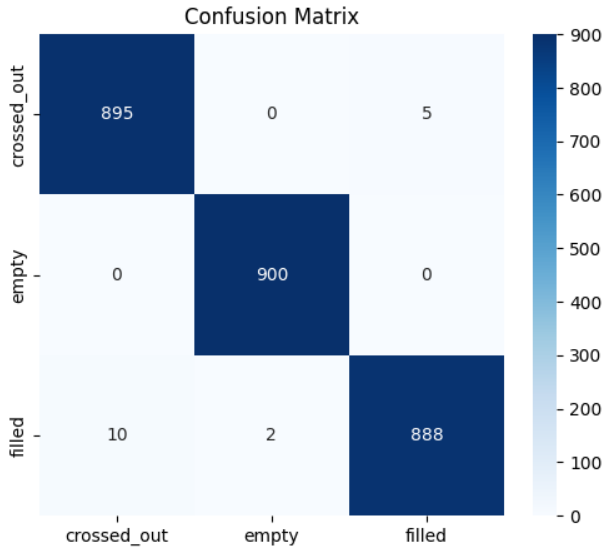
يبين الجدول (1) نتائج اختبار أداء نموذج CNN المدرب (بدون زيادة البيانات) وقد بلغت الدقة الكلية تقريباً 99.48%:

الجدول *1* –نتائج اختبار أداء نموذج CNN بدون زيادة البيانات

	عدد العينات Support	الإحكام Precision	الاستدعاء Recall	مقياس إف1 F1-score
فقاعات مشطوبة	900	0.9858	1.0000	0.9928
فقاعات فارغة	900	0.9989	1.0000	0.9994

فقاعات مملوءة	900	1.0000	0.9844	0.9922
إجمالي الفقاعات	2700	**0.9949**	**0.9948**	**0.9948**

2.4.5 نموذج CNN مع Augmentation



الشكل *3* – مصفوفة الالتباس لنموذج CNN مع زيادة البيانات

يوضح الشكل 3 مصفوفة الالتباس لنموذج CNN. المدرب مع تطبيق زيادة البيانات augmentation. يمكن ملاحظة أن النموذج حقق أداء جيد جداً بشكل عام، حدث خطأ في تصنيف (5 فقاعات من 900) من فئة المشطوبة تم تصنيفها على أنها مملوءة، وبالنسبة لفئة الفقاعات الفارغة صُنفت جميعها بشكل صحيح على أنها فارغة ولم يتواجد أي خطأ. أما فئة الفقاعات المملوءة، فقد تم تصنيف 888 منها بشكل صحيح وحدث 12 خطأ، خطأين صنفا على أنهما فارغتين و10 حالات صنفت بأنها مشطوبة.

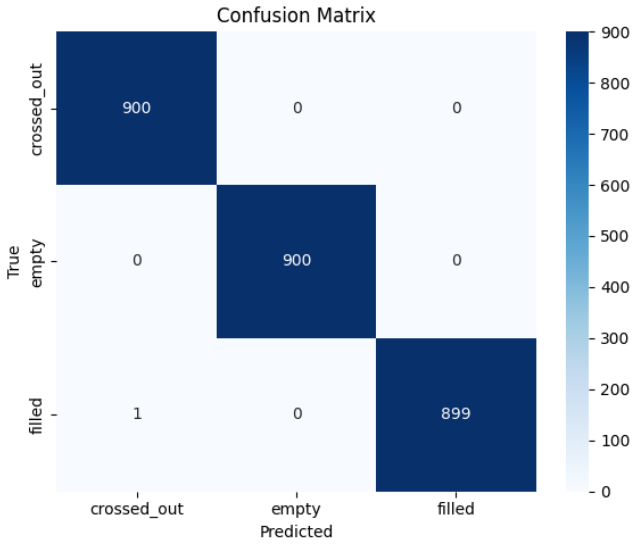
يبين الجدول (2) نتائج اختبار أداء نموذج CNN المدرب (مع زيادة البيانات):

الجدول *2* –نتائج اختبار أداء نموذج CNN مع زيادة البيانات

	عدد العينات Support	الإحكام Precision	الاستدعاء Recall	مقياس إف1 F1-score
فقاعات مشطوبة	900	0.9890	0.9944	0.9917
فقاعات فارغة	900	0.9978	1.0000	0.9989
فقاعات مملوءة	900	0.9944	0.9867	0.9905
إجمالي الفقاعات	2700	**0.9937**	**0.9937**	**0.9937**

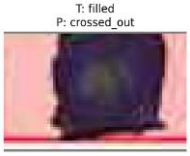
تبين النتائج أن معدل الدقة الكلي لهذا النموذج قد بلغ 99.37%، لكن النموذج (بعد زيادة البيانات) أظهر خطأ أكبر في التصنيف بين الفقاعات المملوءة والمشطوبة.

3.4.5 النموذج الهجين (CNN+Transformer) بدون زيادة البيانات



الشكل *4* – مصفوفة الالتباس للنموذج الهجين بدون زيادة البيانات

يبين الشكل 4 مصفوفة الالتباس للنموذج الهجين عند تدريبه بدون زيادة البيانات نلاحظ تحسن في تصنيف الفقاعات مقارنة بنماذج CNN، حيث لا توجد أي أخطاء في تصنيف الفقاعات ضمن فئة الفقاعات الفارغة، وبالنسبة للفقاعات الفارغة تمكن النموذج من تصنيفها بشكل صحيح، ظهر خطأ وحيد فقط في تصنيف فقاعة مملوءة حيث صنفت على أنها مشطوبة. العدد الإجمالي للأخطاء هو خطأ واحد فقط من أصل 2700 عينة مبين في الشكل 5:



الشكل *5* –حالة الخطأ في تصنيف الفقاعات باستخدام النموذج الهجين بدون زيادة البيانات

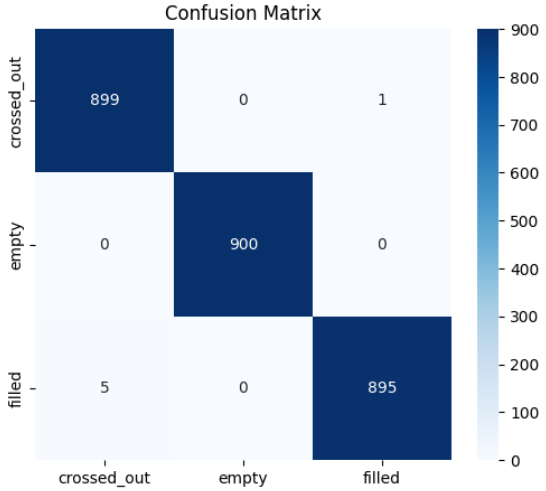
حقق النموذج دقة إجمالية حوالي 99.96% كما هو مبين في الجدول 3:

الجدول *3* –نتائج اختبار أداء النموذج الهجين بدون زيادة البيانات

	عدد العينات Support	الإحكام Precision	الاستدعاء Recall	مقياس إف1 F1-score
فقاعات مشطوبة	900	0.9989	1.0000	0.9994
فقاعات فارغة	900	1.0000	1.0000	1.0000
فقاعات مملوءة	900	1.0000	0.9989	0.9994
إجمالي الفقاعات	2700	**0.9996**	**0.9996**	**0.9996**

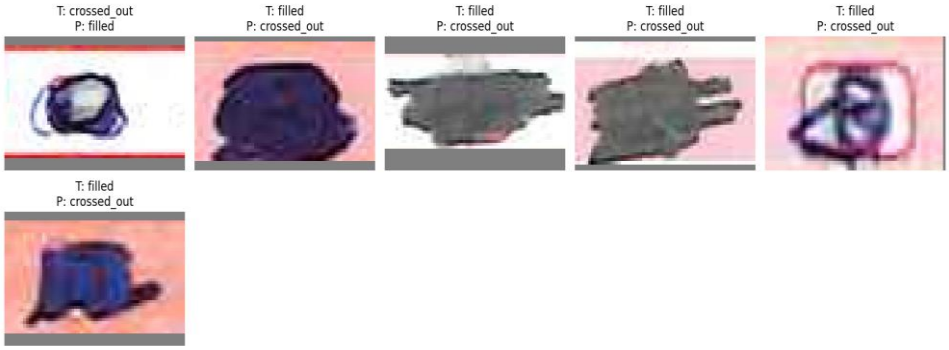
نلاحظ أن النموذج الهجين بدون زيادة البيانات (augmentation) كان الأفضل من حيث تقليل أخطاء التصنيف ودقته، يعود ذلك إلى دمج المحول Transformer الذي مكّنه من التقاط أنماط العلامات بشكل أشمل، فمثلاً استطاع النموذج الانتباه إلى وجود خطوط مائلة مميزة في الفقاعة المشطوبة تميّزها عن الفقاعة المملوءة كلياً، مما قلل من خلطه بينهما مقارنة بالنماذج السابقة.

4.4.5 النموذج الهجين(CNN+Transformer) مع زيادة البيانات



الشكل *6* – مصفوفة الالتباس للنموذج الهجين مع زيادة البيانات

يعرض الشكل 6 مصفوفة الالتباس للنموذج الهجين عندما تم تدريبه مع تطبيق زيادة البيانات. نلاحظ من الشكل أن جميع عينات الفقاعات المشطوبة صنفت صحيحة عدا واحدة صنفت على أنها مملوءة (899 من 900)، وكذلك جميع عينات الفقاعات الفارغة صنفت صحيحة (900 من 900) دون أي خطأ. أما بالنسبة لعينات الفقاعات المملوءة صُنف 895 منها بشكل صحيح وال 5 عينات المتبقية صنفت بشكل خاطئ على أنها مشطوبة. عدد الأخطاء الكلية في التصنيف 6 أخطاء مبينة في الشكل 7:



الشكل *7* –حالات الخطأ في تصنيف الفقاعات باستخدام النموذج الهجين مع زيادة البيانات

حقق النموذج دقة تقارب 99.78% كما هو موضح في الجدول 4:

الجدول *4* –نتائج اختبار أداء النموذج الهجين مع زيادة البيانات

	عدد العينات Support	الإحكام Precision	الاستدعاء Recall	مقياس إف1 F1-score
فقاعات مشطوبة	900	0.9945	0.9989	0.9967
فقاعات فارغة	900	1.0000	1.0000	1.0000
فقاعات مملوءة	900	0.9989	0.9944	0.9967
إجمالي الفقاعات	2700	**0.9978**	**0.9978**	**0.9978**

يبين الجدول 5 مقارنة أداء النماذج الأربعة:

الجدول 5 –مقارنة أداء النماذج المقترحة في تصنيف الفقاعات

النموذج	Augmentation	الدقة الكلية (Accuracy)	عدد العينات الصحيحة من 2700	عدد الأخطاء الكلي
CNN	✗	0.9948	2686	14
CNN	✓	0.9937	2683	17
CNN + Transformer	✗	0.9996	2699	1
CNN + Transformer	✓	0.9978	2694	6

6. الاستنتاجات والتوصيات

تشير نتائج الدراسة إلى أن النموذج الهجين المقترح الناتج عن دمج بنية الشبكة العصبية الترشيحية CNN مع المحولات Transformer حقق أعلى أداء في تصنيف فقاعات الإجابات دون تطبيق تقنيات زيادة البيانات Augmentation، حيث سجّل دقة كلية بلغت 99.96% مع خطأ واحد فقط من أصل 2700 عينة. يُعزى هذا النجاح إلى قدرة المحولات Transformer على تمثيل السياق الكامل للصورة ومساعدة CNN في التقاط العلاقات الدقيقة داخل الفقاعة، مثل نمط التظليل أو وجود خطوط الإلغاء.

وبخصوص زيادة البيانات augmentation بهدف تعميم النماذج على حالات مختلفة مولدة من مجموعة البيانات الحقيقية، توضح هذه الدراسة أن استخدام هذه التقنيات في بيئة بيانات حساسة قد يؤدي إلى إدخال أنماط مشوشة وغير واقعية تؤثر سلباً على فعالية النموذج، خاصة عند التعامل مع تصنيفات دقيقة تتطلب حساسية عالية لاختلافات طفيفة في نمط الحبر أو التظليل. بينت نتائج اختبار النماذج أن عملية زيادة البيانات أضعفت دقة بعض الفئات، كحالة فقاعات OMR التي تم تدويرها أو قلبها دون مراعاة "السلامة التصنيفية" (label safety)، إذ إن أي تدوير للصورة يمكن أن يغيّر من نمط الحبر داخل الفقاعة بطرق غير واقعية مما يُدخل تشويش لا داعي له ويتسبب في تعلم النموذج لأنماط زائفة.

أحد أبرز أخطاء التصنيف في جميع النماذج تمثلت في التمييز بين فقاعات مملوءة وأخرى مشطوبة، وهذا ما أظهرته مصفوفات الالتباس حيث أن غالبية الأخطاء نتجت عن خلط بين هاتين الفئتين، وبالعودة إلى حالات الفقاعات التي صنفت بشكل خاطئ وجدنا تشابه بين تظليل الفقاعة بخطوط أشبه بالشطب ومنها مملوءة بخطوط مائلة حتى خارج زوايا الفقاعة، يزداد التصنيف الخاطئ عند تطبيق تحويلات هندسية مثل التدوير أو قلب الصورة حيث يجعل التظليل المشوَّه أكثر تشابهاً مع الشطب، يشير ذلك إلى ضرورة تجنب زيادة البيانات بشكل غير مدروس عندما يتعلق الأمر بفئات تتداخل بصرياً بشكل كبير.

بناءً على ما سبق، نوصي بما يلي:

- تبني نموذج هجين (CNN + Transformer) عند تطوير أنظمة تصحيح OMR لضمان دقة أعلى في التمييز بين الفئات.
- الامتناع عن زيادة البيانات Augmentation بشكل عشوائي في بيئات بيانات حساسة، واعتماد تحويلات مدروسة فقط تعكس الواقع الحقيقي.
- التركيز على تحليل الأخطاء الناجمة عن التمييز بين filled و crossed-out لتحسين المعالجة الدقيقة للفروقات.

تؤكد هذه النتائج أن التصميم الدقيق للبنية وتخصيص استراتيجيات زيادة البيانات بما يتوافق مع طبيعة المهمة يلعب دور حاسم في تحسين الأداء، وتفتح الطريق نحو تطوير أدوات تصحيح مؤتمتة أكثر ذكاءً وموثوقية.

7. المراجع:

[1] DE ELIAS,E., TASINAFFO,P. , HIRATA,R. 2021 Optical mark recognition: Advances, difficulties, and limitations, SN Computer Science, Vol. 2(5), 367.

[2] PATEL, R., SANGHAVI, S., GUPTA, D., RAVAL, M.S. 2015 CheckIt–A low cost mobile OMR system, TENCON IEEE Region 10 Conference, pp. 1–5.

[3] ZAMPIROLLI, F., GONZALEZ, J.A.Q., DE OLIVEIRA NEVES, R. 2010 Automatic correction of multiple-choice tests using digital cameras and image processing, Universidade Federal do ABC, Brazil, 34p.

[4] HUSSMANN, S., DENG, P.W. 2005 A high-speed optical mark reader hardware implementation at low cost using programmable logic, Real-Time Imaging, Vol. 11(1), 19–30.

[5] GOROKHOVATSKYI, O. 2016 Neocognitron as a tool for optical marks recognition, IEEE International Conference on Data Stream Mining & Processing (DSMP), 169–172.

[6] PARUL, H.M., KAUR, M. 2012 A novel optical mark recognition technique based on biogeography based optimization, International Journal of Information Technology and Knowledge Management, Vol. 5(2), 331–333.

[7] KHAN, S., NASEER, M., HAYAT, M., ZAMIR, S.W., KHAN, F.S., SHAH, M. 2022 Transformers in vision: A survey, ACM Computing Surveys (CSUR), Vol. 54(10s), 1–41.

[8] BHATT, D., PATEL, C., TALSANIA, H., PATEL, J., VAGHELA, R., PANDYA, S., GHAYVAT, H. 2021 CNN variants for computer vision: History, architecture, application, challenges and future scope, Electronics, Vol. 10(20), 2470.

[9] SHEN, Z. 2024 A comparative study of hybrid CNN and Vision Transformer models for facial emotion recognition, IEEE

International Conference on Dependable Systems and Their Applications (DSA), 401–408.

[10] DENG, H., WANG, F., LIANG, B. 2008 A low-cost OMR solution for educational applications, IEEE International Symposium on Parallel and Distributed Processing with Applications, 967–970.

[11] SANGUANSAT, P. 2015 Robust and low-cost Optical Mark Recognition for automated data entry, IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 1–5.

[12] AFIFI, M., HUSSAIN, K.F. 2019 The achievement of higher flexibility in multiple-choice-based tests using image classification techniques, International Journal on Document Analysis and Recognition (IJDAR), Vol. 22, 127–142.

[13] MONDAL, S., DE, P., MALAKAR, S., SARKAR, R. 2024 OMRNet: A lightweight deep learning model for optical mark recognition, Multimedia Tools and Applications, Vol. 83(5), 14011–14045.

[14] DOSOVITSKIY, A., BEYER, L., KOLESNIKOV, A., WEISSENBORN, D., ZHAI, X., UNTERTHINER, T., HOULSBY, N. 2020 An image is worth 16×16 words: Transformers for image recognition at scale, arXiv preprint, arXiv:2010.11929.

[15] NIE, Y., SOMMELLA, P., CARRATÙ, M., O'NILS, M., LUNDGREN, J. 2022 A deep CNN transformer hybrid model for skin lesion classification of dermoscopic images using focal loss, Diagnostics, Vol. 13(1), 72.

[16] MAI, J., GAO, C., BAO, J. 2025 Domain generalization through data augmentation: A survey of methods, applications, and challenges, Mathematics, Vol. 13(5), 824.

[17] TING, K.M. 2011 – Confusion Matrix. In: **Sammut**, C., Webb, G.I. (eds), Encyclopedia of Machine Learning. Springer, Boston, USA.

[18] OBI, J.C. 2023 A comparative study of several classification metrics and their performances on data, World Journal of Advanced Engineering Technology and Sciences, Vol. 8(1), 308–314.

[19] AL-KURDI, R.M. 2024 Interpreting a model trained for detecting hate speech in Arabic tweets, Homs University Journal – Series of Engineering Science, Vol. 46(1), 133–168.