

استخدام خوارزميات التعلم العميق وعمليات التشكيل لتصنيف صور الاستشعار عن بعد

جعفر سلمان *

ميرنا سليمان **

(تاريخ الإيداع 2023/1/3 . قَبْلَ للنشر في 2023/4/6)

□ ملخص □

مع تطور تقنيات مراقبة الأرض تتزايد صور الاستشعار عن بعد المكتسبة بشكل كبير، ويقترّب عصر جديد من البيانات الضخمة في مجال الاستشعار عن بعد، وتشكل كيفية التعامل بفعالية مع هذه الكميات الهائلة من البيانات تحديات جديدة. يوفرّ التعلّم العميق طريقة حديثة لتحليل صور الاستشعار عن بعد وكأحد نماذج التعلم العميق يمكن للشبكات العصبية الالتقافية ((Convolutional neural Network (CNN) استخراج الميزات مباشرة من عدد هائل من صور الاستشعار عن بعد؛ إذ أنّها أبدت أداءً جيداً في استغلال الميزات الدلالية لبيانات الصور تلك. في السنوات الأخيرة درس عدد غير قليل من الباحثين تصنيف صور الاستشعار عن بعد باستخدام خوارزميات (CNN) ولكن معظم هذه الدراسات واجهت مشكلة الخلط بين الفئات الدلالية لنماذج الصور في قواعد البيانات، وعليه تهدف هذه الورقة إلى تقديم نموذج على ثلاث مراحل لتطبيق نماذج من خوارزميات (CNN) باستخدام تقنيات نقل التعلم مع إضافة معالجة مسبقة للصور وصولاً لتصنيف وتمييز أوضح للمشاهد مع تجنّب الخلط بين الفئات الدلالية للصور. نأمل أن يسهّل البحث المقدم أبحاث تصنيف صور الاستشعار عن بعد ويساعد علماء الاستشعار على معالجة مهام التصنيف باستخدام خوارزميات وتقنيات التعلم العميق الحديثة.

الكلمات المفتاحية: تعلم الآلة، التعلم العميق، الشبكات العصبونية الالتقافية، معالجة الصورة، نقل التعلم، صور الاستشعار عن بعد.

*مدرس في قسم هندسة تكنولوجيا المعلومات - كلية هندسة تكنولوجيا المعلومات والاتصالات - جامعة طرطوس.
** طالبة دراسات عليا (ماجستير) في قسم هندسة تكنولوجيا المعلومات والاتصالات - كلية هندسة تكنولوجيا المعلومات والاتصالات - جامعة طرطوس .

Remote sensing images classification using deep learning algorithms and forming processes

Jaafar Salman*
Mirna Souliman**

(Received 3/1/2023 . Accepted 6/4/2023)

□ ABSTRACT

With the development of earth observation technologies, the acquired remote sensing images are increasing dramatically, and a new era of big data in remote sensing is coming. How to effectively deal these massive volumes of remote sensing data are appears a new challenges. Deep learning provides a new approach for analyzing these remote sensing data. As one of the deep learning models, convolutional neural networks (CNN) can directly extract features from massive amounts of imagery data and is good at exploiting semantic features of that imagery data. In recent years, quite a few researchers have studied remote sensing image classification using CNNs, However, most of these studies faced the problem of confusing the semantic categories of image models in databases. Accordingly, this paper aims to present Three-phase model for applying models from (CNN) algorithms using transfer learning techniques with add pre-processing of images in order to get clearer classification and distinction for the viewer while avoiding confusion between the semantic categories of images. We hope that the study can facilitate the advancement of remote sensing image classification research and help remote-sensing scientists to tackle classification tasks with the state-of-art deep learning algorithms and techniques.

Keywords: Machine learning, deep learning, convolutional neural networks, image processing, transfer learning, remote sensing images.

* Teacher, Information Technology Engineering Department, Information and Communication Technology Engineering, Tartous University.

** Student Master, Information Technology Engineering Department, Information and Communication Technology Engineering, Tartous University.

1. مقدمة:

التعلم العميق Deep Learning هو أحد فروع تعلم الآلة Machine Learning الذي يعتمد على الشبكات العصبية الاصطناعية إلى جانب التعلم التمثيلي. لذلك، فإنه عبارة عن تقنية حاسوبية تحاكي العقل البشري من خلال تصميم خوارزميات مستوحاة من بنية القشرة الدماغية ووظيفتها حتى يمكنها تقليد جميع قدرات الدماغ مثل فهم اللغة الطبيعية والأصوات ومحتوى الصور والقدرة على تنفيذ العديد من الأوامر والتعليمات كما يفعل الإنسان. وهذه الشبكة الاصطناعية تحتوي على خلايا عصبية تضم مجموعة من الطبقات المتصلة والمتراكمة فوق بعضها البعض، والتي تبدأ بطبقة الإدخال المستوحاة من حواس الإنسان وتستقبل إشارات البيانات والمعلومات وتنتهي بطبقة الإخراج التي تظهر النتيجة النهائية للمعلومات، وبين هاتين الطبقتين توجد طبقات مخفية Hidden layer تكون مسؤولة عن تحليل البيانات للوصول إلى الاستنتاجات قبل استخراجها على هيئة معلومات مفيدة [1].

كل طبقة في الشبكات العصبية العميقة تكون مسؤولة عن التقاط أنماط معينة من البيانات أو المعلومات المحددة ومعالجتها باستخدام خوارزميات التعلم العميق، ويطلق على الخلية العصبية في هذه الشبكة اسم (عصبون) وكل خلية لها وزن يحدد أهمية كل عنصر على حدة، إضافة إلى إمكانية تعديل الأوزان أكثر من مرة لضمان الحصول على نتائج دقيقة في نهاية المطاف، وبذلك يمكن لخاصية التعلم العميق معالجة كمية هائلة من البيانات.

تعد الشبكات العصبية الالتفافية Convolutional Neural Network التي تعرف اختصاراً CNN، من أبرز الشبكات العصبية العميقة المستخدمة في التعلم العميق، فإنها تشمل على طبقات متعددة ذات بنية استثنائية بهدف معالجة بيانات غير منظمة غالباً مثل الصور للتعرف بها فيها وقراءتها واستخراج الميزات الفريدة فيها.

يعتبر تصنيف صور الاستشعار عن بعد مجالاً بحثياً نشطاً، إذ تعد صور الاستشعار عن بعد مصدراً قيماً للبيانات ويزيد حجم هذه الصور بشكل كبير بفضل التقدم في تكنولوجيا مراقبة الأرض، الأمر الذي أعطى إلهاماً خاصاً للبحث عن كيفية الاستفادة الكاملة من هذه الصور وصولاً لرصد الأرض الذكي.

تصنيف مشهد صورة الاستشعار عن بعد هو تسمية هذه الصور بشكل صحيح بفئات دلالية محددة مسبقاً، وعلى مدى العقود القليلة الماضية، تم إجراء أبحاث مكثفة حول تصنيف مشهد صورة الاستشعار عن بعد مدفوعاً بتطبيقاته في العالم الحقيقي، مثل التخطيط الحضري، الكشف عن المخاطر الطبيعية، مراقبة البيئة، رسم خرائط الغطاء النباتي، وكشف الأجسام الجغرافية المكانية، ونظراً لتقدم التصوير في مجال الاستشعار عن بعد، زادت الدقة المكانية للصور الملتقطة بحيث فقدت وحدات البكسل الفردية معانيها الدلالية. في مثل هذه الحالة، أصبح تحليل صور الاستشعار عن بعد على مستوى الكائن أكثر كفاءة من التحليل على مستوى البكسل. ومع ذلك، قد تحتوي صور الاستشعار عن بعد على كائنات من فئات مختلفة ومتميزة وفي هذه الحالة يتطلب تحليل صور الاستشعار عن بعد العمل على مستوى المشهد بهدف تصنيف المشهد ككل إلى فئة دلالية واحدة [2].

2. مشكلة البحث:

في السنوات الأخيرة، وبسبب التزايد المستمر لبيانات مشهد صور الاستشعار عن بعد، حظي تصنيف المشهد لمثل هذه الصور باهتمام واسع النطاق، إذ يعتبر الاستشعار عن بعد من التقنيات الفعالة لتصوير سطح الأرض ورصدها بتكلفة منخفضة وفي وقت قصير. ومع ذلك، لا يزال تصنيف مشهد صورة الاستشعار عن بعد يواجه العديد من التحديات المتعلقة باختلاف زوايا التصوير وحجم الكائنات الصغيرة والدقة المنخفضة التي تؤدي لتشويه وطمس

معالم الصورة وغيرها من التحديات التي نجم عنها خلط بين الفئات الدلالية للصور وعدم القدرة على تمييز الأشكال في أغلب الدراسات المرجعية مما جعل من تطوير أساليب التحليل والتفسير أمراً حتمياً، ومن هنا انطلق البحث من تحسين الصورة وهي أكثر خطوات المعالجة الأولية أهمية في تحليل الصور وعالج البحث أغلب مشاكل صور الاستشعار عن بعد مبيناً معالم هذه الصور وصولاً لفصل دقيق بين الفئات الدلالية.

3. الدراسات المرجعية:

➤ في العام 2018 [3] قام الباحث Zeng بإجراء دراسة بعنوان Improving Remote Sensing Scene Classification by Integrating Global-Context and Local-Object Features، حيث استخدم نموذج VGG-16 لبناء شبكة CNN بفرعين الأول يمثل فرع الكائن المحلي (Local Object) لاستخراج الميزات على مستوى الكائن والفرع الثاني هو الفرع الدلالي العالمي (Global semantic) لاستخراج الميزات على مستوى الصورة ككل وصولاً لتمييز أكثر دقة للمشاهد، وقام بتطبيقه على عدة قواعد بيانات أحدها قاعدة تم الحصول عليها من موقع Kaggle (نفس القاعدة المستخدمة في الدراسة)، وتم الحصول على دقة %96.

➤ في العام 2019 [4] قام الباحث Qayyum بإجراء بحث بعنوان Designing deep CNN models based on sparse coding for aerial imagery: a deep-features reduction approach، استخدمت الدراسة قاعدة بيانات مجهزة يدوياً. اقترح الباحث نهج CNN الذي يشمل معمارياً تقنية sparse coding لتصنيف المشاهد وحل مشكلة التجهيز الزائد (Overfitting)، وقرن نتائج تصنيف التقنية المستخدمة مع تقنية تقليل التجهيز (Principle component analysis (PCA)، تقوّت التقنية المقترحة على تقنية PCA المبنية أيضاً على هياكل CNN المدربة مسبقاً. حققت الدراسة دقة تصنيف %93.64.

➤ في العام 2020 [5] قام الباحث Byju بإجراء دراسة بعنوان Remote-Sensing Image Scene Classification With Deep Neural Networks in JPEG 2000 Compressed Domain، حيث قام ببناء نموذج شبكة عصبية عميقة DDNS مع استخدام ضغط JPEG_2000 على الصور وتطبيق تقنية لتقريب النطاقات الفرعية الموجية للصورة (wavelet sub bands) من خلال سلسلة طبقات التفاضلية تستخدم مجموعة من المرشحات بهدف تقليل الحاجة إلى عمليات فك الضغط المطلوبة لتصنيف صور الاستشعار عن بعد، تم تطبيق النموذج المقترح على قاعدة البيانات AID المستخدمة في دراستنا نفسها، حيث كانت أعلى دقة تم الحصول عليها %92.24 عند استخدام شبكة Resnet_50 في تقريب النطاقات.

➤ في العام 2021 [6] قام الباحث Xiaowei بإجراء دراسة بعنوان A novel approach for scene classification from remote sensing images using deep learning methods، استخدمت الدراسة الشبكة العصبية التكرارية RNN إلى جانب خوارزمية الغابات العشوائية Random Forest وتم تطبيق النموذج على مجموعة البيانات (UC Merced)، أنتجت الطريقة المقترحة دقة بنسبة %87.

➤ في العام 2022 [7] قام الباحث Nisia بإجراء بحث بعنوان Ensemble of features for efficient classification of high-resolution remote sensing image، حيث قام ببناء نموذج شبكة CNN معتمداً طريقة جديدة تعمل على استخراج ثلاث أنواع مختلفة من الميزات وهي الميزات العميقة

وميزات Gabor وميزات النمط الثنائي المحلي RLBP، واستخدمت الدراسة إصدار (IRS P6 Satellite) للحصول على بيانات التدريب، وتم الحصول على دقة 94.6%. وفيما يلي ملخص للدراسات المرجعية:

الدراسة	العام	الإيجابيات	السلبيات
الأولى	2018	استخراج الميزات على مستوى الكائن والمشهد معاً مما حقق تمييز أكثر دقة للصورة، حققت دقة 96%	ركزت الدراسة على تصنيف المشهد دون العمل على تحديد الكائنات باستخدام ميزات الكائن التي تم استخراجها
الثانية	2019	تفوق النهج المقترح في الدراسة على تقنيات تقليل الإفراط في التجهيز التقليدية وحققت دقة تصنيف وصلت إلى 93.64%	ضبطت الدراسة عدد قليل من المعلمات الفائقة Hyper Parameter للتدريب مما نتج عنه دقة تصنيف منخفضة
الثالثة	2020	استخدام تقنية ضغط الصور JPEG-2000 مع تقريب النطاقات الموجية للصور مما قلل زمن المعالجة وحققت دقة 92%	يمكن تطبيق هذه التقنية مع نهج الضغط القائم على الموجات (wavelet) فقط
الرابعة	2021	بناء نموذج يجمع بين الشبكات العصبية التكرارية وخوارزمية الغابات العشوائية مما قلل من تكلفة التدريب والاختبار، وحققت دقة 87%	دقة النموذج منخفضة
الخامسة	2022	استخراج ثلاث أنواع من الميزات وهي الميزات العميقة وميزات Gabor وميزات النمط الثنائي المحلي، وحققت دقة 94%	الحصول على العديد من الميزات نتج عنه وقت طويل لتدريب نموذج الشبكة العصبية

جدول (1) ملخص بعض الدراسات المرجعية

4. أهمية البحث وأهدافه:

تهدف هذه الدراسة للاستفادة من التعلم العميق في مجال تصنيف صور الأقمار الصناعية متعددة الأطياف ومتوسطة الدقة المكانية وصولاً لمعالجة أفضل للصورة وبناء نموذج تعلم عميق للعمل على تحسين دقة التصنيف مقارنة بالطرق التقليدية، وتكمن الأهمية الخاصة للبحث في تطبيقاته الواسعة في مجال الكشف عن الكائنات وتجزئة الصورة وصولاً لتتبع وتفسير صور الاستشعار عن بعد مع توفير الوقت والجهد اللازم لإجراء التصنيف.

5. طرق البحث ومواده

تم تنفيذ هذا البحث باستخدام لغة البرمجة Python ضمن منصة Google Colab والتي تقدم نسخة سحابية مجانية من Jupyter، تعمل على خوادم غوغل السحابية، وتتيح الاستفادة من العديد من الميزات مثل وحدة معالجة الرسومات GPU والذاكرة الإضافية والأجهزة الخلفية، بالإضافة للقيام بكل الوظائف التي يتم تنفيذها عادةً على برنامج Jupyter Notebook المثبت على جهاز الحاسوب دون الحاجة إلى تهيئة أي إعدادات. يتضمن البحث عدة مراحل بدءاً من الحصول على الصورة وصولاً لمرحلة تحديد الفئة الدلالية للمشهد من خلال بناء نموذج تعلم عميق يعتمد على الشبكات العصبية الالتفافية ونماذج نقل التعلم مع الاستعانة بمكتبات Keras و Numpy و Tensorflow من أجل بناء النموذج.

6. مراحل البحث:

6.1 الحصول على الصور:

تم تحميل قاعدة البيانات من موقع Kaggle وهو عبارة عن مجتمع عبر الإنترنت لعلماء البيانات ومهندسي التعلم الآلي، يسمح بالعثور على مجموعات البيانات ونشرها واستكشاف نماذج الذكاء الاصطناعي [AID: A scene classification dataset | Kaggle](#). تتضمن هذه القاعدة 10000 صورة استشعار من google earth بأبعاد 600*600 مقسمة إلى 30 فئة دلالية وضمن كل فئة 200-400 صورة.

6.1.1 المعالجة المبدئية:

يظهر في قاعدة بيانات المختارة وجود فئات غير متوازنة في أعداد الصور إذ نلاحظ وجود فئة تحوي 198 صورة مقابل فئة تحوي 378 صورة والتي ستؤدي حتماً لنتائج تصنيف غير متوقعة ومنحازة لفئة دون الأخرى، وعليه كانت الخطوة الأولى للعمل على هذه القاعدة هي موازنة الفئات لتكون الصور ضمن كل فئة [190-200] صورة، وتعديل حجم الصور إلى 224*224 لتناسب النموذج المستخدم.

6.1.2 تحديات معالجة صور الاستشعار عن بعد مقارنة بالصور العادية:

تعاني معالجة صور الاستشعار عن بعد مقارنة بالصور العادية من صعوبة في الفصل بين الفئات الدلالية نتيجة الاختلافات الكبيرة في مظاهر الكائنات داخل نفس الفئة الدلالية؛ إذ تختلف الكائنات الأرضية بشكل عام في الشكل والمقياس والتوزيع، مما يجعل من الصعب تصنيف صور المشهد بشكل صحيح. بالإضافة إلى ذلك، عندما تلتقط المنصات المحمولة جواً أو الفضاء صور الاستشعار عن بعد، فقد تكون هناك اختلافات كبيرة في اللون وكثافة الإشعاع تظهر ضمن نفس الفئة الدلالية بسبب ظروف التصوير، والتي يمكن أن تتأثر بعوامل مثل الطقس والسحابة والضباب، إلخ. قد تؤدي الاختلافات في إضاءة المشهد أيضاً إلى تنوع داخل الفئة، على سبيل المثال، يظهر مشهد فئة "الشاطئ" اختلافات كبيرة في ظل ظروف التصوير المختلفة [2].

6.1.3 مرحلة معالجة الصورة:

استخدم البحث عمليات التحويلات من الناحية الشكلية للصورة (morphological transformations) والتي تسند إلى نظرية المجموعات الرياضية الحديثة [8]، وهي عبارة عن سلسلة من عمليات معالجة الصورة بناءً على الأشكال، والتي تستخدم التشكيل الرياضي المعتمد على البكسل لاستخراج مكونات الصورة المفيدة في تمثيل شكل المنطقة، والحدود، وما إلى ذلك. وتعد معالجة الصورة بالعمليات المورفولوجية أهم عمليات المعالجة الأولية للصور الرقمية، وقد وضعت أساساً لتحليل الصور الثنائية (أبيض وأسود) وتوسع نطاقها في وقت لاحق إلى الصور ذات السويات الرمادية ومن ثم الصور الملونة [9].

تعتمد العمليات المورفولوجية على ربط صورة الإدخال بعنصر هيكلية (structuring element) أو ما يسمى بالنواة (Kernel) وهو عبارة عن مصفوفة أو قالب صغير الحجم يستخدم لمسح الصورة إذ يتم وضع النواة في جميع المواقع الممكنة في الصورة وتحوي هذه النواة نقطة ربط قابلة للتحديد وعادة ما تكون هي نقطة مركز النواة نفسها.

تشبه معالجة الصور المورفولوجية عملية التصفية المكانية إذ يتم تمرير عنصر الهيكله عبر كل بكسل في الصورة الأصلية لإعطاء كل بكسل في صورة معالجة جديدة. تعتمد قيمة هذا البكسل الجديد على العملية المورفولوجية التي يتم إجراؤها، إذ أنّ العمليتان الأكثر استخداماً هما التآكل والتمدد [10].

التمدد (Dilation): الهدف منه توسيع المناطق الساطعة (ذات القيم الرمادية الكبيرة) في الصورة لملء الثقوب في الكائن وفق التالي: يتم فيه تمرير النواة على الصورة واستخراج أقصى قيمة بكسل للمنطقة التي تغطيها النواة واستبدال قيمة نقطة الربط بالقيمة القصوى التي تم الحصول عليها بعد مسح الصورة بالنواة.

التآكل (Erosion): تمثل عملية تقلص الكائنات في الصورة وهي عكس عملية التمدد، تهدف لتحسين وتوسيع المنطقة بقيم رمادية صغيرة (أعمق بصرياً) والتي تستخدم بشكل أساسي لإزالة الضوضاء إذ تؤدي هذه العملية إلى تنعيم حدود السطح الأمامي.

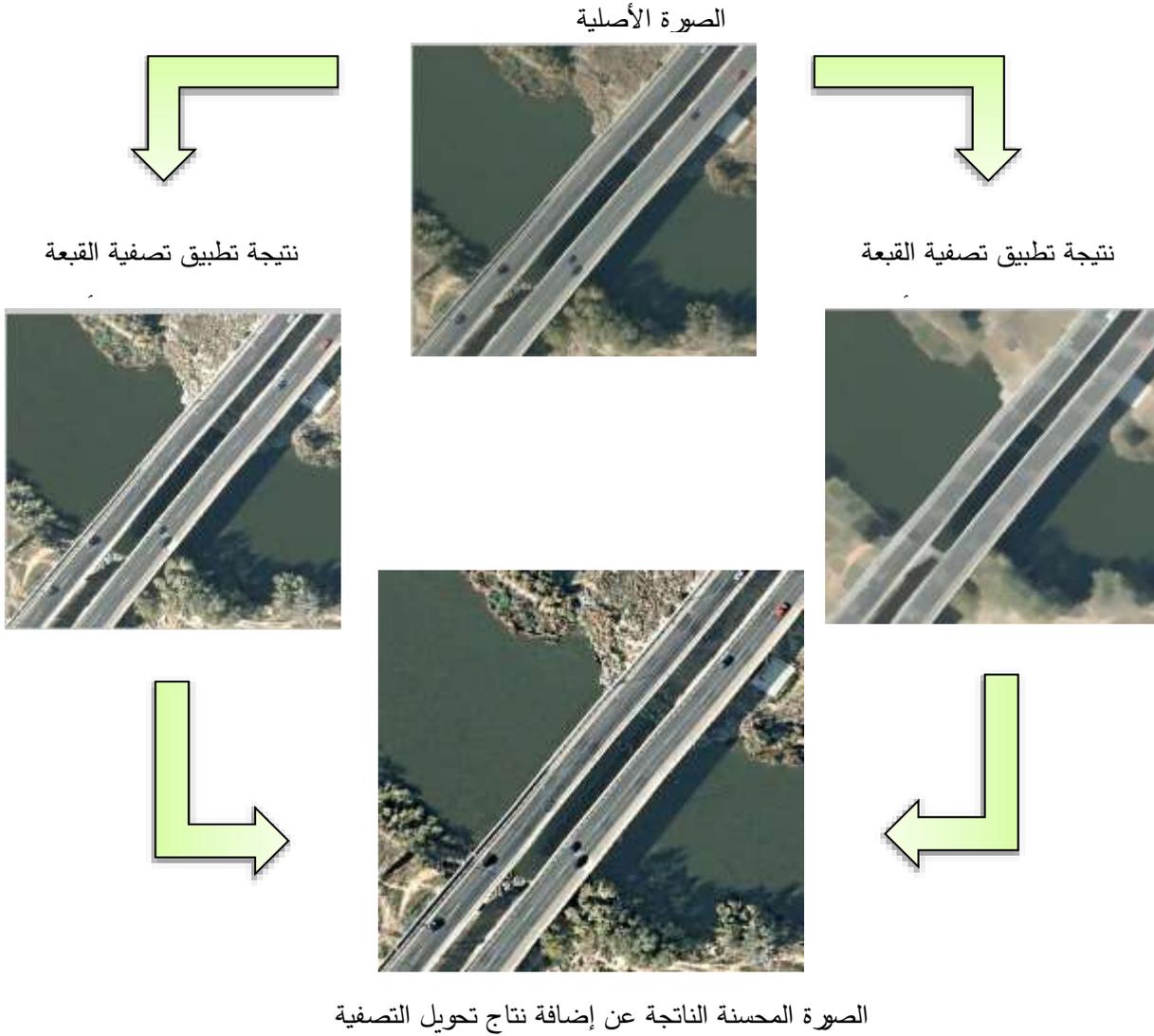
لا يتم تنفيذ معظم العمليات المورفولوجية باستخدام التمدد أو التآكل فقط؛ بل يتم إجراؤها باستخدام كليهما بدلاً من ذلك. وأكثر العمليات التركيبية استخداماً هي الإغلاق المورفولوجي (عن طريق إجراء التمدد أولاً ثم التآكل)، والفتح المورفولوجي (عن طريق إجراء التآكل أولاً ثم التمدد).

يتم تعريف تصفية القبة العلوية على أنها الفرق بين صورة الإدخال ونتاج الفتح المورفولوجي للصورة في حين أن تحويل القبة السفلية هي الفرق بين صورة الإغلاق والصورة الأصل؛ وعليه يمكننا الاستفادة من هذين المشغلين عن طريق إضافة نتيجة تحويل القبة العلوية للصورة وطرح نتيجة تحويل القبة السفلية منها للحصول على صورة محسنة من خلال إبراز المناطق الساطعة في الصورة وإزالة المناطق المظلمة.

وتعطي العلاقات الرياضية المورفولوجية بالعلاقات التالية [11]:

$$\begin{array}{l}
 (A \circ B) = (A \ominus B) \oplus B \\
 (A \bullet B) = (A \oplus B) \ominus B
 \end{array}
 \quad (1) \quad \longrightarrow \quad
 \begin{array}{l}
 A_{top} = A - (A \circ B) \\
 A_{bot} = (A \bullet B) - A \\
 A_{enhance} = (A + A_{top}) - A_{bot}
 \end{array}
 \quad \begin{array}{l}
 (3) \\
 (4) \\
 (5)
 \end{array}$$

حيث: A الصورة الأصلية، B: النواة (Kernel)، \oplus : عملية التمدد، \ominus : عملية التآكل، $(A \circ B)$: عملية الفتح المورفولوجي $(A \bullet B)$: عملية الإغلاق المورفولوجي، A_{top} : تصفية القبة العلوية، A_{bot} : تصفية القبة السفلية، $A_{enhance}$: عملية تحسين الصورة بإضافة التصفية العلوية وطرح التصفية السفلية. يوضح الشكل (1) تحسين صورة من فئة الجسر (من قاعدة البيانات المستخدمة في الدراسة) عن طريق تطبيق معالجة للصورة باستخدام عملية تحويل شكلي مركب من التصفية بنوعيه (تصفية القبة العلوية والسفلية) وفق تسلسل العمليات الوارد ذكره سابقاً في المعادلة رقم (5).



الشكل (1)، تطبيق تقنية معالجة الصور باستخدام عمليات التشكيل المركبة (التصفية العلوية والسفلية معاً)

6.2 مرحلة بناء النموذج وتدريبه Model building :

6.2.1 تحميل البيانات: تم تقسيم قاعدة البيانات إلى ثلاث مجموعات:

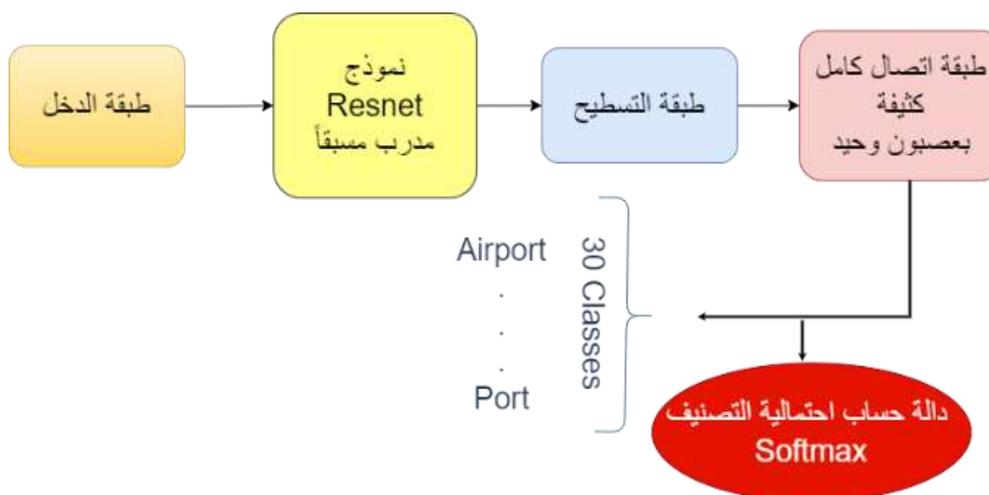
- ✓ مجموعة التدريب Training set: وهي مجموعة من الصور المستخدمة لضبط الأوزان أثناء تدريب الشبكة.
 - ✓ مجموعة التحقق Validation set: هي مجموعة من الصور تستخدم لضبط معاملات الشبكة أثناء التدريب.
 - ✓ مجموعة الاختبار Test set: مجموعة صغيرة من الصور لتقييم أداء النموذج بعد انتهاء التدريب.
- تم تقسيم البيانات إلى 90% للتدريب، 5% للتحقق، 5% للاختبار أي 9000 صورة للتدريب و 500 صورة للتحقق و 500 صورة للاختبار .

6.2.2 تحديد بنية النموذج:

في هذا البحث تم اقتراح منهجية العمل على عدة مراحل إلى أن توصلنا إلى الحل الأمثل لتحقيق هدف الدراسة، إذا تم بناء نموذج الشبكة تبعاً وفق المراحل التالية:

6.2.3 النموذج الأول:

يقوم النموذج الأول الموضح في الشكل (2) على استخدام تقنية نقل التعلم؛ والتي تشير كما يوحي الاسم إلى نقل معلمات النموذج المدربة إلى النموذج الجديد وتطبيقها، ويعرف أيضاً على أنه نقل المعرفة التي تعلمها النموذج إلى النموذج الجديد لزيادة الكفاءة وتقليل وقت التدريب وتكلفته. في هذا البحث تم استخدام نقل التعلم من شبكة ResNet-50 وهي عبارة عن نموذج شبكة مدرب مسبقاً على قاعدة البيانات ImageNet (وهي قاعدة بحثية ضخمة تضم أكثر من 14 مليون صورة مقسمة إلى 20000 فئة دلالية)، يتألف نموذج ResNet-50 من 50 طبقة وحجمه 102 ميغابايت، استخدم النموذج تقنية residual block وهي كتلة من طبقتين التفاضليتين لهما نفس عدد المرشحات، حيث يضاف خرج الطبقة الثانية مع دخل الطبقة الأولى وبعدها يجمع دخل هذه الوحدة مع خرجها ويسمى وصلة الاختصار shortcut connection.



الشكل (2)، النموذج الأول لنقل التعلم من شبكة ResNet

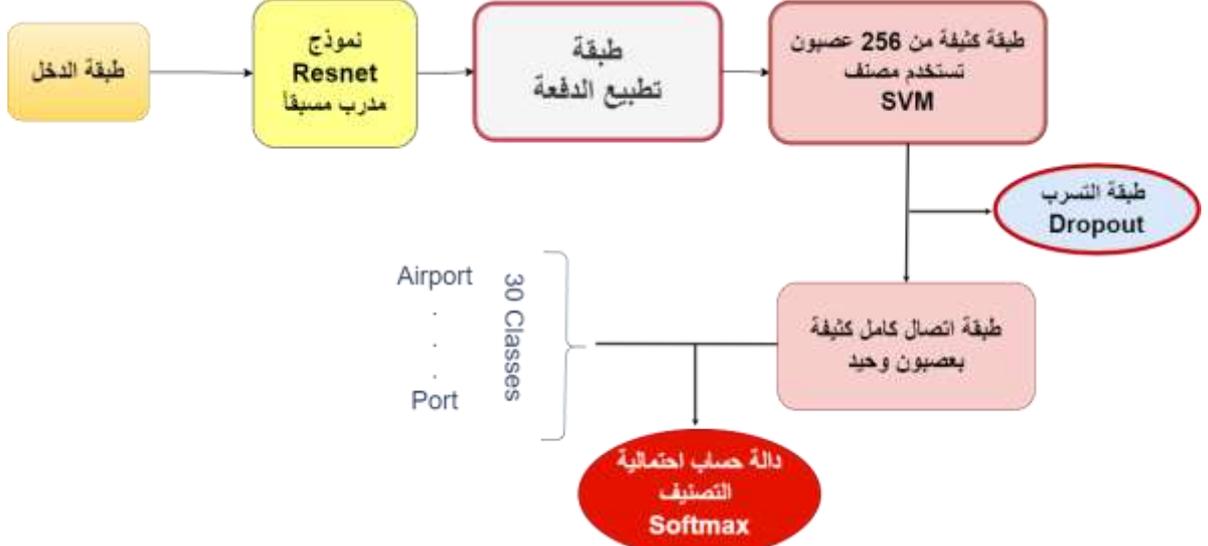
تعمل الطبقات الالتفافية في النموذج كمستخرج للخصائص يتلوهما طبقة التسطيح (Flatten)؛ إذ تحول هذه الطبقة خريطة الميزات المستخرجة من طبقات الالتواء ذات N بعد إلى طبقة ببعد واحد (1D) ليتم الحصول على خريطة ميزات قوية، والتي يتم سحبها كناقل أحادي البعد كمدخلات للطبقة كاملة الاتصال المتمثلة بالـ (Dense) لإجراء التصنيف عليها؛ إذ تعمل هذه الطبقة كمصنف (تم استخدام المصنف التقليدي في هذا النموذج)، مع دالة SoftMax التي تعمل على تحويل ناتج قيمة الجزء الخطي من العصبون إلى قيمة احتمالية للحصول على خرج التصنيف (يتبع التصنيف لفئة العصبون الذي يحمل الاحتمالية الأكبر) [12].

استخدم هذا البحث منهجية نقل التعلم من نموذج Resnet_50 المبني على CNN، مع إعادة تدريب طبقات الطي بالأكمل لتعديل الأوزان بما يلائم قاعدة بيانات البحث (AID)؛ تم التوصل إلى دقة 80% من تطبيق هذا النموذج على قاعدة بيانات البحث لكن الدقة الكلية للنموذج (F1_Score) كانت منخفضة 67% مما يشير إلى

وجود مشكلة Overfitting؛ والتي تتمثل بانخفاض قدرة الشبكة في تحديد الميزات العامة ليس لمجموعة التدريب فحسب بل يتعداها لمجموعات الاختبار والتنبؤ أيضاً [13].

6.2.3.1 النموذج الثاني:

بناءً على ما نتج من النموذج الأول اعتمد البحث المقدم على بناء نموذج ثاني للتغلب على مشكلة الـ Overfitting من خلال إضافة طبقات جديدة وصولاً للنموذج الموضح في الشكل (3):



الشكل (3)، النموذج الثاني، إضافة طبقات لنموذج الشبكة ResNet

إذ يحدد الهدف من هذه الإضافات وفق التالي:

✓ **طبقة Batch Normalization [14]:** طبقة تطبيع الدفعة؛ الهدف الأساسي منها تقليل تأثير covariate shift على الشبكة والذي يعني أنه أي تغيير في قيم الدخل يلزمه إعادة تدريب الشبكة من جديد لتجنب زيادة قيمة خطأ النموذج، إذ أنه في الحالة العادية التغيير في البدايات، سيصاحبه تغيير حتمي في النهايات وعليه تسوية البيانات تقلل من التغيير في النهايات مما يسهل عمل الشبكة، ويجعلها أقرب لتحقيق نتائج أفضل دون لزوم إعادة التدريب من أجل كل دخل مختلف.

$$Z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (6) \quad Z_{tilde}^{(i)} = \gamma Z_{norm}^{(i)} + \beta \quad (7)$$

حيث: z تمثل الخرج الفعلي للشبكة، μ المتوسط الحسابي، σ^2 الانحراف المعياري، ϵ قيمة مضافة لتجنب القسمة على صفر، Z_{norm} الخرج بعد تسوية البيانات، β , γ قيم مضافة غير مرتبطة بالدخل (x) لزيادة التحكم بالخرج الناتج، Z_{tilde} : الخرج بعد إضافة قيم غير مرتبطة بالدخل.

✓ **تم إضافة المصنف Support Vector Machine (SVM) إلى نموذج Resnet**، وهو مشابه لنظام التصنيف التقليدي لكن مع عدد من الاختلافات، الهدف الأساسي منه هو العثور على مستوى له أقصى هامش أي المسافة القصوى بين نقاط البيانات من الفئات المتعددة [15]، إذ تقوم فكرته على إجراء

تصنيف للبيانات بخط وهمي، بشرط أن يكون هناك حاجز مسبق يفصل النقاط عن بعضها البعض، يمنع هذا الحاجز التداخلات بين فئات التصنيف عن طريق مقارنة الخرج الناتج عن كل صورة وتجميع الصور التي تحمل نتائج خرج متقاربة معاً لتحمل نفس الفئة الدلالية مما ينتج عنه نتائج تصنيف أدق وأسرع.

✓ **إضافة تقنية تنعيم البيانات [16] (Regularization)** التي تعمل على تقليل حساسية دالة الخطأ في النموذج للأوزان، وذلك عن طريق إضافة معامل تنعيم λ بهدف تقليل حساسية النموذج ككل للمدخلات الجديدة والذي يعطى بالعلاقة:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m (L(\hat{y}^{(i)}, y^{(i)})) + \frac{\lambda}{2m} \|w\|_2^2 \quad (8)$$

$$\|w\|_2^2 = \sum_{j=1}^n w_j^2 = w^T w \quad (9)$$

حيث: $J(w, b)$ تمثل تابع الكلفة، m : عدد عناصر الدخل لحساب المتوسط، $L(\hat{y}^{(i)}, y^{(i)})$: تابع الخسارة والذي يمثل الفرق بين القيم الحقيقية والقيم المتوقعة، λ : معامل التنعيم، $\|w\|_2^2$: قيمة مرتبطة بالوزن وفق العلاقة (9).

✓ **طبقة Dropout [16]**: والتي تعمل على حذف عدد من الخلايا في الطبقات الخفية بشكل عشوائي، بعد كل مسار أمامي وخلفي ليصبح عدد الخلايا المستخدم في كل مرة اقل، وهو ما يعني شبكة أبسط، مما يقلل فرص حدوث الـ OF، تم تطبيق نسبة إسقاط 0.4 في شبكتنا.

✓ **وأخيراً تم استخدام تابع التنشيط (SoftMax)**: تقوم هذه الدالة على تشكيل شعاع بقيم تنحصر بين الصفر والواحد بحيث يكون ناتج جمع هذه القيم يساوي الواحد لتتسجم النتائج مع النظرية الاحتمالية وتعطى الدالة بالصيغة التالية:

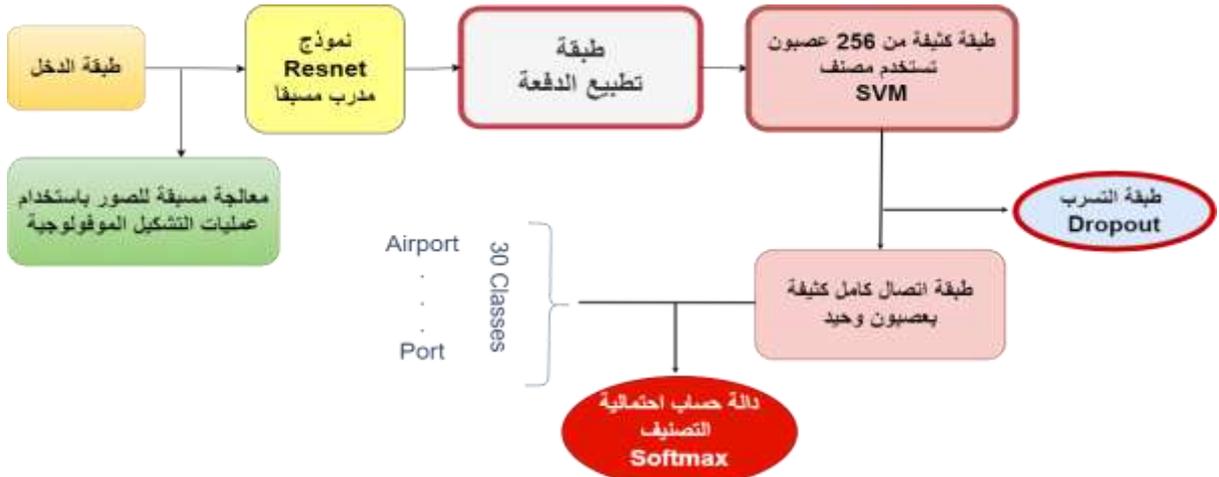
$$\text{softmax}(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (10)$$

حيث: i شعاع الدخل من z_0 إلى z_k ، z_i يمثل أي عنصر من شعاع الدخل، e^{z_i} يمثل تطبيق التابع الأسّي على كل عنصر من عناصر الدخل، k عدد العناصر في شعاع الدخل.

حقق النموذج الثاني دقة وصلت إلى 82% من أجل F1_Score وتم التغلب على مشكلة الـ overfitting إذ حسن استخدام المصنف (SVM) إلى جانب نموذج (ResNet) بدلاً من المصنف التقليدي تحسناً ملحوظاً لكونه يحقق فصل واضح بين الفئات الدلالية دون الخلط بينها، إلا أن دقة النموذج ما زالت منخفضة وعليه اعتمد البحث المقدم استخدام تقنيات معالجة الصور بالتحويلات الشكلية.

6.2.3.2 النموذج الثالث:

يتمثل النموذج الثالث الذي تم تطبيقه في الدراسة على استخدام المعالجة المسبقة للصور باستخدام عمليات التشكيل (تحويل القبعات العلوية والسفلية) وإعادة تدريب النموذج وعليه تم التوصل إلى النموذج التالي:



الشكل (4)، النموذج الثالث (إضافة المعالجة المسبقة للصور إلى نموذج الدراسة المقترح)

حقق النموذج الأخير دقة وصلت إلى 96%، كما أن الدقة الكلية للنموذج (F_1 _Score) وصلت إلى 88%؛ إذ ساعد تطبيق التحولات الشكلية لمعالجة الصور على إظهار معالم الصور والحدود والارتفاعات وخفف من مشكلة طمس المعالم التي تعاني منها صور الاستشعار عن بعد الأمر الذي جعل استخراج الميزات أسرع وأدق.

6.2.4 تدريب النموذج:

بعد تصميم نموذج نقل تعلم من شبكة CNN نوع Resnet_50 تم تدريبه باستخدام الخوارزمية الأكثر شيوعاً وهي الانتشار الخلفي (Back propagation) التي تعمل على تعديل الأوزان بصفة تكرارية مع كل مسار أمامي خلفي بما يتناسب مع ناتج دالة الخطأ والتي تعطى على أنها الفرق بين الخرج المتوقع والخرج الحقيقي الذي نحصل عليه، كما تم اختيار خوارزمية التحسين AdamaX لتسريع تقارب النموذج وهي خوارزمية فعالة تستخدم على نطاق واسع في تطبيقات التعلم العميق، كما تم اختيار دالة الخسارة (Categorical_cross_entropy) التي تقوم بحساب الخطأ وقياس الدقة (accuracy) لاختبار أداء النموذج.

6.2.5 تقييم النموذج وحساب الدقة

يقوم النموذج بتصنيف الصور إلى 30 فئة إذ أنه بعد بناء النموذج واختباره على بيانات الاختبار يتم حساب مصفوفة الارتباك (Confusion Matrix)، تظهر المصفوفة عدد الحالات السلبية والايجابية الصحيحة والخاطئة لاستخدامها في تقييم النموذج كما في الجدول (2):

		Predicted Values	
		Positive	Negative
Actual Values	Positive	True Positive [TP]	False Negative [FN]
	Negative	False Positive [FP]	True Negative [TN]

الجدول (2) مصفوفة الارتباك Confusion Matrix

تستخدم مصفوفة الارتباك في حساب مقاييس الأداء للنموذج إذ يتم استخلاص المقاييس التالية منها [17]:
 1. الدقة: تمثل عدد الحالات التي تم تصنيفها بشكل صحيح على جميع الحالات وتعطى

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

2. الاسترجاع
 3. المقياس دقة
 4. المقياس دقة

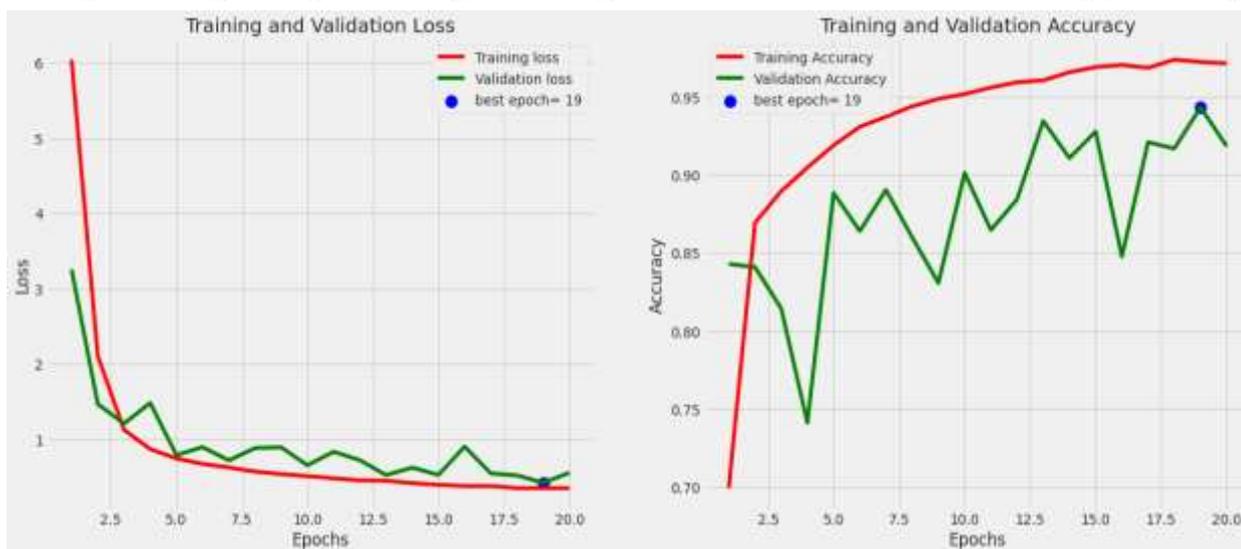
$$F1_score = \frac{2*(precision*recall)}{precision+recall} = \frac{2TP}{2TP+FP+FN} \quad (14)$$

6.2.6 النتائج ومناقشتها:

أثبتت النتائج المقدمة أفضلية المنهجية المقدمة في تصنيف هذا النوع من الصور إذ اعتمد النموذج المقترح على معالجة الصور باستخدام عمليات التشكيل ومن ثم تطبيق نقل التعلم مع مراعاة تدريب طبقات النموذج كاملة وتطبيق تقنيات لتسوية البيانات والتنعيم مع استخدام مصنف SVM مما حقق دقة في الأداء وهو أمر بالغ الأهمية للمساعدة في تحليل صور الاستشعار عن بعد. يظهر الجدول (3) مقارنة بين مجموعة من مقاييس الأداء للنهج المقترح المبني على النماذج الثلاث والتي تم تطبيقها تبعاً في الدراسة.

الجدول (3)، مقارنة مقاييس الأداء المستخدمة في مراحل بناء النموذج المقترح الثلاث

مقاييس الأداء				منهجية النموذج	النموذج
F1_Score	recall	precision	Accuracy		
67%	68%	73%	80%	نقل التعلم من نموذج ResNet_50	النموذج /1/
82%	83%	87%	96%	إضافة مصنف SVM إلى نموذج /1/ لنقل التعلم	النموذج /2/
88%	88%	90%	96%	إضافة معالجة الصور بطريقة التحويل الشكلي (Top Bottom Hat &) إلى النموذج /2/ لنقل التعلم	النموذج /3/



الشكل (5)، منحنيات الدقة والخسارة الناتجة عن تطبيق النموذج /3/ المحسن

يوضح الشكل (5) مخططات الدقة والخسارة التي تم الحصول عليها بعد التدريب إذ نلاحظ من الشكل تناقص منحني الخسارة في كل مرحلة تباعاً بالتزامن مع زيادة دقة التصنيف الناتجة عن التدريب.

7. مقارنة النموذج المقترح مع الدراسة المرجعية:

يظهر الجدول (4) مقارنة مع الدراسة المرجعية الخامسة عام 2022 والتي تستخدم قاعدة بيانات مصممة يدوياً لفصل الفئات الدلالية لصور الاستشعار عن بعد:

الدراسة	منهجية الدراسة المرجعية	منهجية الدراسة الحالية
Ensemble of features for efficient classification of high-resolution remote sensing image	استخراج ثلاث أنواع من الميزات وهي الميزات العميقة وميزات Gabor وميزات النمط الثنائي المحلي، وحققت دقة 94%	اعتماد تقنية نقل التعلم مع معالجة الصور باستخدام التحويلات الشكلية؛ إذ حقق النموذج دقة وصلت لـ 96% دون الحاجة لتعقيد مستويات استخراج الميزات

الجدول (4)، مقارنة النموذج المحسن مع الدراسة المرجعية

8. الاستنتاجات:

- ✓ استخدام تقنيات نقل التعلم من هيكل CNN المدربة مسبقاً (نموذج Resnet_50) والتي نتج عنها دقة نموذج محسنة بعد 20 مرحلة (Epoch) تدريب فقط مما قلل الوقت والجهد اللازم لبناء نموذج CNN من الصفر وتدريبه.
- ✓ حقق مصنف SVM تحسین واضح على النموذج المطبق إذا زادت الدقة من 80% في النموذج الأول (الذي يستخدم المصنف التقليدي للشبكات الالتفافية) إلى 82% في النموذج الثاني (الذي يستخدم مصنف SVM) وعليه كان اختيار هذا المصنف قائماً على التجريب.
- ✓ طبقت معالجة الصورة باستخدام عمليات التشكيل المورفولوجية على قاعدة بيانات البحث باستخدام التحويل الشكلي المركب من تحويل القبعات العلوية والسفلية مما حسن دقة الصورة كما هو موضح سابقاً في الشكل (1)، فضلاً عن كونه حسن من دقة التصنيف إذ ساعد على تحسين التباين وإخماد تراكيب الخلفية في صور الاستشعار عن بعد للتركيز على عنصر الكائن وإبراز المعالم والحدود وتحسين التفاصيل الدقيقة.
- ✓ تمت مقارنة أداء هذا النموذج مع نموذج الدراسة المطبق عام 2022 إذ تفوق النموذج المقدم على نموذج الدراسة المرجعية بمقدار 2% دون الحاجة لبناء نموذج معقد مع مستويات استخراج ميزات متعددة.

9. الآفاق المستقبلية:

العمل على دمج النموذج الذي تم التوصل إليه مع نماذج تحديد الكائنات في خطوات لاحقة للاستفادة من دقة الصور التي توصلت إليها الدراسة بعد المعالجة.

10. المراجع

- [1] Janiesch, c., Zschech, p. & Heinrich, k. 2021. *Machine learning and deep learning*. Electron Markets.
- [2] Cheng, G.; Xie, X.; Han, J. & Guo, L. 2020, *Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities*. IEEE. vol. 13, pp. 3735-3756.
- [3] Zeng, D.; Chen, S.; Chen, B. & Li, S. 2018. *Improving Remote Sensing Scene Classification by Integrating Global-Context and Local-Object Features*. Remote Sensing. 10, 734.
- [4] Qayyum, A.; Malik, A. & Mazher, M.2019, *Designing deep CNN models based on sparse coding for aerial imagery: a deep-features reduction approach*. European Journal of Remote Sensing.
- [5] Byju, A. P.; Sumbul, G.; Demir, B. & Bruzzone, L. 2021, *Remote-Sensing Image Scene Classification with Deep Neural Networks in JPEG 2000 Compressed Domain*. IEEE. Vol. 59, pp. 3458-3472.
- [6] Xiaowei, Xu.; Chen, Y.; Zhang, J.; Chen, Y.; Anandhan, P. & Manickam, A. 2021. *A novel approach for scene classification from remote sensing images using deep learning methods*. European Journal of Remote Sensing, Vol. 54, pp. 383-395.
- [7] Nisia, G. & Rajesh, S. 2022. *Ensemble of features for efficient classification of high-resolution remote sensing image*. European Journal of Remote Sensing, vol. 55:1, pp. 326-337.
- [8] Gonzalez, R.C.; Woods, R.; Masters, B.R. 2008. *Digital Image processing, Third Edition*. Semantic scholar.
- [9] serra, j. 1983. *Image Analysis and Mathematical Morphology*. Semantic scholar.
- [10] chhikara, p. 2022. *Understanding Morphological image Processing and Its Operations*. Towards Data Science.
- [11] Srisha, R.; khan, A. 2013. *Morphological operations for image processing: understanding and its application*. ResearchGate.
- [12] Albawi, S.; Mohammed, T. A. & Al-Zawi, S. 2017. *Understanding of a convolutional neural network*. IEEE.
- [13] O'Shea, k.; Nash, R. 2015. *An Introduction to Convolutional Neural Networks*. Semantic scholar.
- [14] Loffe, s. & Szegedy, c. 2015. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. Semantic scholar.
- [15] Rohith, G. 2018. *support vector machine-introduction to machine learning algorithms*, Towards Data Science.
- [16] Karagiannakoa, S. 2021. *Regularization techniques for training deep neural network*. AI Summer.
- [17] Padilla, R.; Netto, S. L. & Eduardo, A. B. 2020. *A Survey on Performance Metrics for Object-Detection Algorithms*. ResearchGate, PP. 237-242.