

## تقييم خوارزميات التعلم الآلي لإنشاء خرائط الغطاء

### الأرضي من صور القمر الصناعي Sentinel-2

م. محمود قرنداش \* د.م. حنان كامل درويش \*\* د.م. عدنان أحمد محمد \*\*

#### الملخص:

يشير مفهوم خريطة الغطاء الأرضي إلى تمثيل الغطاء السطحي للأرض مثل الغطاء النباتي، البنية التحتية الحضرية، المياه، التربة العارية وما إلى ذلك. تقدم خرائط الغطاء الأرضي المعلومات الأساسية للعديد من التطبيقات مثل رسم الخرائط الموضوعية، تحليل اكتشاف التغيير. كما تمثل بيانات الغطاء الأرضي نقطة انطلاق يمكن من خلالها تنفيذ مجموعة متنوعة من الأنشطة، كالتخطيط الحضري وأنشطة المراقبة.

تستعرض هذه المقالة دراسة مقارنة بين ثلاث خوارزميات من التعلم الآلي هي خوارزمية شعاع الدعم الآلي (SVM)، خوارزمية الجار (K-NNC) وخوارزمية تعزيز التدرج للتعلم الآلي (GBM) لتصنيف 6 فئات من الغطاء الأرضي ضمن منهجية محددة للتصنيف تشمل الخوارزميات السابقة وتبدأ من كتابة الأكواد اللازمة باستخدام لغة البرمجة بايثون واعتماداً على مكتباتها مفتوحة المصدر، مروراً بإجراء المعالجة الحاسوبية، انتهاءً بإظهار النتائج والمقارنة.

تم اختيار صور حديثة للقمر الصناعي Sentinel-2 مأخوذة في 31 آذار 2022 لمنطقة القصير في محافظة حمص. تم تقييم التصنيف بناءً على مجموعة من العينات المرجعية، وأظهرت خوارزمية شعاع الدعم الآلي نتائج أفضل من الخوارزميات الأخرى، وخلص البحث إلى أنّ خوارزمية شعاع الدعم الآلي تعطي دقة أفضل حتى مع عدد صغير من بيانات التدريب وتتميز بالسرعة والبساطة في التنفيذ، كما أعطت جميع الخوارزميات نتائج مرضية.

**كلمات مفتاحية:** تصنيف صور الأقمار الصناعية، التعلم الآلي، الجار القريب، تعزيز التدرج، خرائط الغطاء الأرضي.

\*طالب دكتوراه في قسم الهندسة الطبوغرافية\_ كلية الهندسة المدنية\_ جامعة البعث\_ حمص\_ سورية [mabojodm@gmail.com](mailto:mabojodm@gmail.com).

\*\*أستاذ مساعد في قسم الهندسة الطبوغرافية\_ كلية الهندسة المدنية\_ جامعة البعث\_ حمص\_ سورية.

[hanan.darwish@hotmail.com](mailto:hanan.darwish@hotmail.com)

[ad.moh4@gmail.com](mailto:ad.moh4@gmail.com)

# Evaluation of machine learning algorithms for creating land cover maps from satellite images Sentinel-2

ENG. Mahmoud karndash\* Dr.Eng. Hanan Darwishe\*\* Dr. Eng. Adnan Mohammad\*\*

## Abstract:

The concept of a land cover map refers to the representation of the land surface cover such as vegetation cover, urban infrastructure, water, bare soil, etc.

Land cover maps provide essential information for many applications such as drawing thematic maps and change detection analysis, land cover data also represents a starting point in which a variety of activities can be carried out, such as urban planning and monitoring activities.

This article represents a comparative study of three machine learning algorithms, (SVM) algorithm, the (K-NNC), and the (GBM) for classifying 6 land cover classes within a specific methodology for classification using the previous algorithms, which starts from writing the codes using the Python programming language and based on its open-source libraries, passing through computer processing, ending with showing results and comparison.

Recent images of the Sentinel-2 satellite taken on March 31, 2022 have been selected for the Al-Qusayr area in Homs Governorate.

The classification was evaluated based on a set of reference samples, support vector machines algorithm showed better results than other algorithms, the research concluded that the SVM algorithm gives better accuracy even with a small number of training data and it is characterized by speed and simplicity in implementation, and all algorithms gave satisfactory results.

**Key Words:** Classification of satellite images, machine learning-Nearest Neighbor Classifier, Gradient Boosting Classifier, Land cover maps.

---

\*PhD student, Department of Topographic Engineering, Faculty of Civil Engineering, Al-Baath University, Homs, Syria □ [maboaljodm@gmail.com](mailto:maboaljodm@gmail.com).

\*\*Associate Professor, Department of Topographic Engineering, Faculty of Civil Engineering, Al-Baath University, Homs, Syria. □  
□ [hanan.darwishe@hotmail.com](mailto:hanan.darwishe@hotmail.com) □ [ad.moh4@gmail.com](mailto:ad.moh4@gmail.com)

---

## 1-مقدمة:

يعتبر تصنيف صور الأقمار الصناعية من ضمن المجالات المتنامية بشكل مستمر، حيث دخل الذكاء الاصطناعي وخوارزميات التعلم الآلي والرؤية الحاسوبية والتقيب عن البيانات لاستخراج خصائص وميزات الغطاء الأرضي [1]. ويمكن تعريف هذه الخوارزميات بأنها مجال لمعالجة وتحليل وفهم الصور للوصول إلى المعرفة لاستخدامها لاحقاً في اتخاذ القرار، انطلاقاً من مجموعة بيانات من الصور.

تعتبر صور الأقمار الصناعية على أنها أحد أهم مصادر البيانات لرسم خرائط الغطاء الأرضي بسبب تغطيتها الجغرافية الواسعة مع توفير معلومات كبيرة عن سطح الأرض وبكلفة منخفضة نسبياً مقارنة مع طرق المسح الطبوغرافي التقليدية الأخرى [2]. إن خرائط الغطاء الأرضي هي أكثر المنتجات المطلوبة التي أنتجت على أساس تصنيف الصور الفضائية. ومع ذلك، فإن الدقة ووقت المعالجة لا تزال تحدي لمجتمع الباحثين في هذا المجال.

من أجل تصنيف صور الأقمار الصناعية، تتمثل الخطوة الأولى هي جعل الحاسوب يتعرف على الفئات المطلوب تصنيفها ليستخرج لها ميزات وخصائص ثم يتدرب عليها ليقوم بإعطاء مشهد متكامل للغطاء الأرضي.

تم تطوير العديد من المصنفات من قبل مختلف الباحثين، بعض هذه الطرق هي شعاع الدعم الآلي SVM، وخوارزمية الجار القريب K-NNC وخوارزمية تعزيز التدرج GBS، تحتوي كل خوارزمية على بعض المتغيرات التي يتم اختيارها بما تناسب الحالة المدروسة.

وفقاً للعديد من الباحثين في [3]، فإنه ليس فقط حجم ودقة صور الأقمار الصناعية ولكن أيضاً الاختيار الصحيح لطريقة التصنيف سوف تؤثر على إنتاج خرائط الاستخدام/الغطاء الأرضي، حيث هناك مجموعة متنوعة من طرق التصنيف تم تطويرها واختبارها، وتتنوع هذه الأساليب من خوارزميات غير خاضعة للإشراف (مثل ISODATA) إلى الخوارزميات الخاضعة للإشراف البارامترية والتي كانت موضع دراستهم في هذا البحث.

في بحث علمي نشرته المجلة الأوروبية للاستشعار عن بعد تم مقارنة طريقة شعاع دعم الآلة مع المصنفات الإحصائية التقليدية حيث أعطت خوارزمية SVM نتائج جيدة متفوقة على الطرق الإحصائية التقليدية مثل (طريقة الاحتمالية العظمى) ولكن هناك تحذيرات من اختيار البارامترات المتعلقة بهذه الخوارزمية حيث بينت الدراسة أن مؤشر غاما ونوع النواة له تأثير على دقة التصنيف [4].

كما أشارت المجلة الدولية لنظم المعلومات الجغرافية في بحث علمي تضمن استخدام الذكاء الاصطناعي لتصنيف الغطاء الأرضي ورسم الخرائط، حيث رأى الباحث بأن الذكاء الاصطناعي يقترب باتجاه معالجة الصور وأصبح يعترف به كنمط بديل عن الطرق الإحصائية التقليدية [5].

ناقش [6] نظام لتصنيف صور الأقمار الصناعية باستخدام خوارزميتي شعاع الدعم الآلي وتحسين سرب الجسيمات SVM, PSO. وخلص البحث إلى أن طريقة دعم الآلة باستخدام التابع (RBF) أعطى أعلى دقة تصنيف وذلك باستخدام تابع (KERNEL) متعدد النواة.

في [7] تم تصنيف صور الأقمار الصناعية للمناطق الحضرية باستخدام النوى المركبة لخوارزمية شعاع الدعم الآلي وقد أظهرت الأعمال نتائج واعدة من خلال الاندماج المكاني والمعلومات الطيفية باستخدام شعاع الدعم الآلي وخاصة مع نواة التجميع المرجح Weighted Summation Kernel.

## 2-هدف البحث وأهميته:

الهدف الرئيسي من هذه الدراسة هو المقارنة بين ثلاث خوارزميات تعلم آلي مختلفة وهي خوارزمية شعاع الدعم الآلي (Support Vector Machines (SVM)، وخوارزمية الجار القريب (K-Nearest Neighbor Classifier) وخوارزمية تعزيز التدرج للتعلم الآلي (Gradient Boosting Classifier of machine learning (GBM). لاختيار أفضل طريقة مناسبة لتصنيف عدة فئات من الغطاء الأرضي، لاختيار الطريقة الأدق أولاً والمساهمة في وضع بعض التوجيهات المتعلقة باختيار بارامترات الإدخال والعينات التدريبية.

تكمن الأهمية الخاصة للبحث في استخدام خوارزميات الذكاء الاصطناعي لتوفير الوقت والجهد في رسم خريطة الغطاء الأرضي مع الوصول إلى دقة عالية في المنتج النهائي.

### 3- مواد وطرائق البحث:

#### 3-1 صور القمر الصناعي Sentinel:

هو أول كوكبة أقمار صناعية تابعة لبرنامج Copernicus تديرها وكالة الفضاء الأوروبية ومهمتها مراقبة الأرض وتعطي صوراً ضوئية بدقة مكانية عالية (10 م إلى 60 م) وتتكون حالياً من قمرين صناعيين Sentinel-2A و Sentinel-2B يخضع قمر صناعي ثالث Sentinel-2C حالياً للاختبار استعداداً لإطلاقه في عام 2024 [8]. تدعم الوكالة مجموعة واسعة من الخدمات والتطبيقات مثل المراقبة الزراعية، وإدارة حالات الطوارئ، وتصنيف الغطاء الأرضي وجودة المياه.

يتميز القمر Sentinel-2 بالخصائص الرئيسية التالية:

- ✓ بيانات متعددة الأطياف مع 13 نطاقاً.
- ✓ تغطية عالمية منهجية لأسطح الأرض من 56 درجة جنوباً إلى 84 درجة شمالاً، والمياه الساحلية، وكل البحر الأبيض المتوسط.
- ✓ إعادة زيارة المنطقة كل 10 أيام في نفس زوايا المشاهدة.
- ✓ توفر دقة مكانية 10 م و 20 م و 60 م.

تم الاعتماد على بيانات القمر Sentinel-2 الملتقطة بتاريخ 2022/3/31- تم تشكيل صورة ملونة لجزء من منطقة القصير في برنامج ArcGIS Ro وبدقة مكانية 10 م وباستخدام الأداة (Composite Bands) من 10 نطاقات طيفية: النطاق الأول (nm-60m/px443)، النطاق الأزرق (nm-10m/px490)، النطاق الأخضر (nm-10m/px560)، النطاق الأحمر (nm-10m/px665)، والنطاقات (5,6,7) على حافة الأحمر حيث النطاق الخامس (nm-20m/px705)، النطاق السادس (nm-20m/px740)، النطاق السابع (nm-20m/px783)، ونطاق الأشعة تحت

الحمراء (nm-10m/px842) ، والنطاق التاسع (nm-20m/px865) ، والنطاق العاشر (nm-60m/px1375).

### 3-2 خوارزميات التعلم الآلي لتصنيف الصور:

يعتمد التعلم الآلي في المبدأ العام على تصنيف البيانات (Classification) ضمن فئات مختلفة حسب خصائصها المشتركة بالاعتماد على خوارزميات متعددة وبالاستناد إلى نظريات التعلم الإحصائية [9].

تم في الآونة الأخيرة تطوير طرق وخوارزميات كثيرة، تحت عنوان عريض وهو الذكاء الاصطناعي لتتفرع منه علوم أخرى كان أبرزها التعلم الآلي Machine Learning حيث يعرف بأنه علم يعطي الحاسوب إمكانية التعلم دون وجود برمجة صريحة في اتخاذه القرار.

#### ✓ التعلم الآلي:

يعرف بأنه علم يعطي الحاسوب إمكانية التعلم دون وجود برمجة صريحة في اتخاذه القرار. حيث تتضمن عدة تقنيات منها

#### ✓ التصنيف Classification

يعتبر التصنيف في تعلم الآلة من قسم التعلم الخاضع للإشراف (supervisor) حيث من خلاله تصنف عدة عناصر مجهولة ضمن مجموعات منفصلة. يحاول التصنيف أن يتعلم العلاقة الموجودة بين مجموعة متغيرات (features) (والتي تمثل نطاقات صور الأقمار الصناعية) والهدف (target) (والذي يمثل فئات الغطاء الأرضي).

### 3-2-1 شعاع الدعم الآلي: Support Vector Machine

تعد Support Vector Machine (SVM) خوارزمية تعلم آلي خاضعة للإشراف تجمع بين النظرية الإحصائية والتعليم الموجه، طورت من قبل الباحث فابنيك Vabnik عام 1998 [10]، وتستخدم هذه الخوارزمية لتحليل البيانات من أجل تصنيفها تصنيفاً

إحصائياً أو تحليل الانحدار لها، يعمل مصنف (SVM) الخطي عن طريق رسم خط مستقيم بين فئتين. سيتم تصنيف جميع نقاط البيانات التي تقع على جانب واحد من الخط على أنها فئة واحدة وسيتم تصنيف جميع النقاط التي تقع على الجانب الآخر على أنها الفئة الثانية. يبدو الأمر بسيطاً، ولكن هناك عدد لا حصر له من الأسطر للاختيار من بينها. كيف نعرف الخط الذي سيعرف بالغرض في تصنيف البيانات؟ هذا هو المكان الذي تدخل فيه خوارزمية (SVM)، حيث ستحدد خوارزمية (SVM) خطأً لا يفصل بين الفئتين فحسب، بل يظل بعيداً عن أقرب العينات قدر الإمكان. في الواقع، سوف يتم تقسيم البيانات بخط يدعى المستوي الفائق (hyperplane) بغض النظر عن طبيعة البيانات سواء كانت قابلة للفصل الخطي أو لا وهنا تكمن مصدر قوتها [11].

#### آلية عمل الخوارزمية

تعتمد آلية العمل على التنقيب في المعطيات وهي عملية تهدف إلى تحليل المعطيات من أجل استخراج نماذج وقواعد تمثلها لا يمكن استخراجها بطرق التحليل المعروفة، يتم ذلك من خلال تصنيف المعطيات المتشابهة وبناء نموذج (أو نماذج) تعبر عنها من خلال تحليل الانحدار لهذه المعطيات [12].

تتضمن عملية التصنيف مرحلتين، مرحلة التدريب ومرحلة الاختبار:

يتم في مرحلة التدريب توفير مجموعة من البيانات تدعى بيانات التدريب (Training Dataset) إذ يعمل شعاع الدعم الآلي بالاعتماد على بيانات التدريب على إنتاج نموذج قادر على إعطاء الصنف حيث تقوم الخوارزمية بإيجاد حد فاصل بين خصائص كل من الصنفين أو أكثر.

**المرحلة الثانية** هي مرحلة اختبار أو توظيف شعاع الدعم Validation:

تلي عملية التدريب ويتم فيها (تبعاً للتعلم الذي تم اكتسابه خلال مرحلة التدريب) تصنيف مجموعة جديدة من البيانات، حيث يحدد موقع كل منها على أي جهة من الحد الفاصل. تكون مهمة شعاع الدعم الآلي إذاً في هذه المرحلة هي إيجاد المستوي الفائق. يتم في مرحلة التدريب حل مسألة الانحدار بالاعتماد على مجموعة من أشعة تدريب ضمن شعاع الدعم الآلي الذي يتم تطبيقه في قاعدة بيانات معينة وليكن الشعاع:

$$\{x_k, y_k\}: k = 1, 2, 3, \dots, s \quad x_k \in R^m, y_k \in R^n$$

حيث أن  $x_k$  هو شعاع المعطيات أو المدخلات  $y_k$  شعاع الخرج (Output). يستخدم التابع  $f(x)$  من أجل التعبير عن السطح الفاصل بين مجموعة النقاط (بافتراض شعاع أوزان  $w$  وثابت  $b$  كما يلي:

$$f(x) = \langle w, x \rangle + b \quad w, x \in R^m, b \in R^n \quad (1)$$

تتلخص المهمة الأساسية في إيجاد التابع  $f(x)$  الذي يملك انحراف مقداره  $\varepsilon$  عن الهدف المراد الوصول إليه  $y_i$  والمحدد ضمن معطيات التدريب، إذ يجب أن يملك المستوي الفائق أكبر هامش يفصل بين مجموعتي النقاط التي يصنفها (إلى مجموعتين في أبسط الحالات). من أجل ذلك يعرف تابع الفقد (Loss function) المرمز  $N_\varepsilon(y)$  والمرتبط بقيمة  $\varepsilon$  كما يلي:

$$N_\varepsilon(y) = |y - f(x)|_\varepsilon = \begin{cases} 0 & |y - f(x)| \leq \varepsilon \\ |y - f(x)| & \text{otherwise} \end{cases} \quad (2)$$

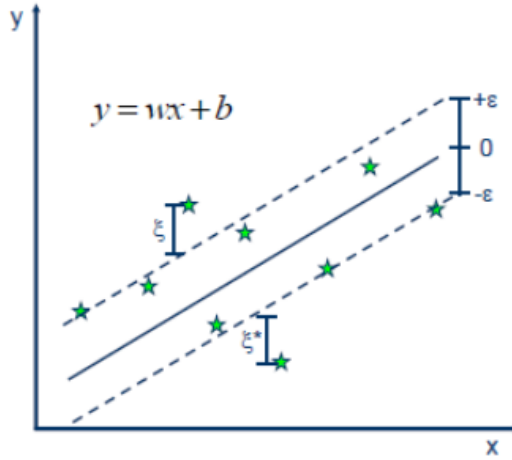
من أجل تحديد قيمة التابع الذي يعطي أكبر قيمة ل  $\varepsilon$  بعد معالجة كل بيانات التدريب الشكل (1) تتحول المسألة لمسألة (Optimization) تتطلب إيجاد الحل الأمثل للقضية التالية:

$$MinJ = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^s (\xi_i^* + \xi_i) \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i^* \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i \\ \xi_i^*, \xi_i \geq 0 \end{cases} \quad (3)$$

حيث أن الحد الأول  $\frac{1}{2} \|w\|^2$  يدعى حد الضبط (Regularized term)،  $c$  : هو ثابت أكبر من الصفر يحدد النسبة بين خطأ التدريب وأداء التنعيم كما أن  $\xi^*$  ,  $\xi$  هي متغيرات.

تتلخص الفكرة الأساسية إذاً في تخفيض قيمة الخطأ في تحديد موقع الحد الفاصل الذي يملك أكبر هامش بين النقاط المراد تصنيفها [13].





الشكل (1) البحث عن حل مسألة الانحدار ضمن شعاع الدعم الآلي.

يمكن لشعاع الدعم الآلي معالجة قضايا الانحدار غير الخطي بنفس الأسلوب من خلال مقارنة بيانات التدريب وتقسيم النماذج التي تحويها في مستوي ذو أبعاد أكبر (يكون فيه الانحدار الخطي ممكناً) ومن ثم تطبيق منهجية الانحدار الخطي في هذا المستوي.

### 3-2-2 خوارزمية الجار القريب KNN:

هي من خوارزميات التعلم الآلي الخاضع للإشراف والتي تعد من أبسط الخوارزميات نظراً لسهولة استخدامها واستهلاكها القليل من الوقت.

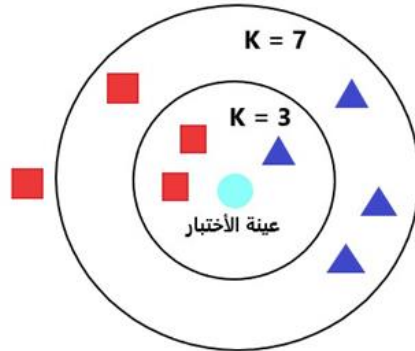
وتعرف بأنها خوارزمية غير بارمترية (non-parametric) أي لا تحوي بارامترات للتدريب بل تحتفظ ببيانات التدريب وتستخدمها في عملية التصنيف لنقطة الاختبار (test point) ويكون هذا النوع من الخوارزميات ذات أداء عالي بسبب عدم بناء فرضيات كثيرة عن دالة النموذج.

حيث تعتمد على ترتيب نقاط التدريب (training points) على أساس مقياس التشابه (similarity measures) بين مواقع نقاط البيانات (datapoints) أي أن عملية التدريب تتم أثناء تصنيف بيانات الاختبار [14].

### آلية عمل خوارزمية KNN:

تقوم الخوارزمية بعملية تصنيف نقطة الاختبار اعتمادا على نقاط التدريب المحيطة بها أي الجار الأقرب لنقطة الاختبار. نستطيع أن نقول إن الخوارزمية تعتمد على مقياس التشابه لنقطة الاختبار مع أقرب نقاط تدريب لها في الشكل نرى فضاء ثنائي البعد (Dimensional Space-2) يحتوي على فئتين من نقاط التدريب (training points) تعتمد خوارزمية KNN على معامل نقوم بتنظيمه وهو K حيث يمثل عدد الجيران الأقرب (nearest neighbors) لنقطة الاختبار الشكل (2).

الشكل (2) تغير المعامل k وتأثيره على عدد العينات المجاورة لعينة الاختبار



يتم اختيار القيمة المناسبة ل k والتي تسمى هذه العملية تنظيم المعامل ( parameter tuning) شيء أساسي في خوارزمية (KNN) للحصول على أفضل دقة ممكنة [15].

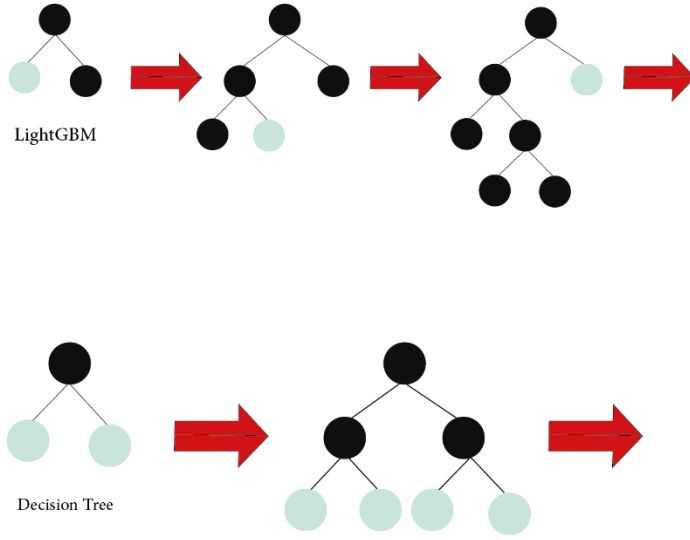
تصادفنا مشكلة في اختيار قيمة K فعند اختيار قيم متنوعة سنلاحظ اختلاف في نتيجة التصنيف كل مرة ففي الشكل (5) فلو اخترنا (K=3) فتنتمي نقطة الاختبار إلى فئة المربع الأحمر ولو اخترنا (K=7) فستنتهي إلى فئة المثلث الأزرق. ومن هنا يمكن القول إننا بحاجة إلى تجريب عدة قيم لـ K ونختبر نتيجة التصنيف ونختار الأفضل. لإيجاد الجيران الأقرب لنقطة الاختبار علينا استخدام المسافة الإقليدية (Euclidean Distance)

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \dots + (x_n - x'_n)^2} \quad (4)$$

### 3-2-3 خوارزمية تعزيز التدرج Gradient Boosting Classifier:

هو إطار عمل معزز متدرج سريع وموزع وعالي الأداء يعتمد على خوارزمية شجرة القرار (Decision Tree)، ويستخدم للترتيب والتصنيف والعديد من مهام التعلم الآلي الأخرى [16].

تنمو شجرة خوارزمية تعزيز التدرج للتعلم الآلي وترمز اختصاراً GBM عمودياً عن طريق إضافة عمق للشجرة باستخدام الخاصية (leaf-wise tree growth) بينما تنمو خوارزميات شجرة القرار الأخرى أفقياً (level-wise tree growth) مما يعني أن خوارزمية تعزيز التدرج تنمو على شكل أوراق الشجرة بينما تنمو الخوارزمية الأخرى على شكل مستويات الشكل (3).



### الشكل (3) توضيح الفرق بين خوارزمية تعزيز التدرج وشجرة القرار

تكتسب خوارزمية Light GBM شعبية كبيرة مؤخراً مع تزايد حجم البيانات يوماً بعد يوم حيث يصبح من الصعب على خوارزميات علوم البيانات التقليدية تقديم نتائج أسرع. يبدأ Light GBM بـ "Light" نظراً لسرعته العالية. يمكن لـ Light GBM التعامل مع الحجم الكبير للبيانات ويأخذ ذاكرة أقل للتشغيل. سبب آخر لشعبية Light GBM هو أنه يركز على تقليل دالة الخسارة مما يعطي نتائج أفضل [17].

إن استخدام خاصية leaf-wise tree growth في خوارزمية Light GBM تتضمن بعض المحاذير بالوصول إلى حالة فرط التدريب overfitting ويمكن حل هذه المشكلة عن طريق التحكم ببعض البارامترات.

الهدف من خوارزمية تعزيز التدرج في التعلم الآلي هو تحسين دالة الخسارة أو تقليلها. فكرة تعزيز التدرج هي إنشاء نماذج ضعيفة ( $M$ ) متعددة بشكل متكرر، ثم إضافة نتائج التنبؤ لكل نموذج ضعيف، حيث أن النموذج الأخير  $F_{m+1}(x)$  يكون بناءً على نموذج التعلم السابق  $F_m(x)$  وذلك من خلال بناء نموذج جديد يضيف مقدّر  $h(x)$  لتوفير نموذج أفضل [18].

وفق العلاقة التالية :

$$F_{m+1}(x) = F_m(x) + h(x) \quad (5)$$

حيث  $m$  تكون محصورة بين  $1 \leq m \leq M$

حيث  $h(x)$  تمثل التدرجات السلبية، حيث تتخفف عن طريقه قيمة الخطأ المتوسط التريبي تدريجياً

$$h(x) = y - F_m(x) \quad (6)$$

يتم تقليل دالة الخسارة عن طريق تقليل الخطأ المتوسط التريبي

$$\frac{1}{n} \sum_i (\hat{y}_i - y_i)^2 \quad (7)$$

حيث  $(i)$  رقم العينة من مجموعة التدريب ذي الحجم  $n$  ، القيمة المتوقعة  $\hat{y} = F(x)$  ، القيم الفعلية للمتغير  $y$  ،

### 3-3 تقييم صحة التصنيف: Accuracy Assessment of Classification

تستخدم عادةً مصفوفة الخطأ (confusion matrix) وهي مقياس شائع جداً يستخدم أثناء حل مشاكل التصنيف. يمكن تطبيقه على التصنيف الثنائي وكذلك على مشاكل التصنيف متعدد الفئات. يظهر مثال لمصفوفة الخطأ الخاصة بالتصنيف الثنائي في الشكل (4)، تمثل مصفوفات الأخطاء الفروقات بين القيم المتوقعة والفعلية [19].

#### الشكل (4) مصفوفة الخطأ

		التصنيف المتوقع	
		+	-
التصنيف الفعلي	+	TP True Positives	FN False Negatives <small>Type 2 error</small>
	-	FP False Positives <small>Type 1 error</small>	TN True Negatives

**True Positives:** يشير "TP" إلى (إيجابي صحيح) والذي يشير إلى عدد الأمثلة الإيجابية المصنفة بدقة.

**False Positives:** يظهر المصطلح "FP" قيمة (إيجابي خطأ) ، أي عدد الأمثلة السلبية الفعلية المصنفة على أنها إيجابية.

**False Negatives:** وتعني "FN" قيمة (سليبي خطأ) وهي عدد الأمثلة الإيجابية الفعلية المصنفة على أنها سلبية.

**True Negatives:** يشير "TN" يرمز إلى (سليبي صحيح) والذي يوضح عدد الأمثلة السلبية المصنفة بدقة.

تستخدم عدة مقاييس لاختبار كفاءة الخوارزمية نذكر منها [20]:

✓ دقة الصحة (precision): والتي تمثل دقة التوقعات الإيجابية (positive)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

✓ دقة الاسترجاع (Recall) : والتي تمثل تغطية عينات التوقعات الإيجابية الفعلية.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

✓ دقة القياس (F1 score): هو مقياس الذي يجمع بين دقة الاستدعاء ودقة الصحة ويمثل المتوسط التوافقي للإثنين معا:

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \quad (10)$$

تتطلب دائماً الخوارزمية حساب دقة الاسترجاع ودقة القياس معاً وذلك لأنه من السهل تحقيق دقة استدعاء 100% من خلال استخلاص فئة محددة، ولكن نحتاج إلى قياس عدد البكسلات التي تم استخلاصها مع هذه الفئة وهي من غير صنف [21].

من أجل التقييم النهائي للصورة المصنفة يتم مقارنة الصورة المصنفة بمعلومات بيانات الحقيقة الأرضية والتي هي النقاط المرجعية التي يتم إدخالها إلى الخوارزمية على شكل مصفوفة مصنفة حسب الفئات والتي يتم المقارنة بين قيمة البيكسل التي توقعتها الخوارزمية والقيمة الحقيقة الموجودة في المصفوفة المرجعية من أجل التحقق من الدقة. يتم حساب الدقة الإجمالية بتقسيم عدد البكسلات التي تم تصنيفها بشكل صحيح مقسومة على العدد الكلي للبكسلات وفق العلاقة (11) وفقاً لمصفوفة الخطأ التي تم الحصول عليها [22] [23].

$$\text{Overall accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (11)$$

من التقنيات الأخرى المستخدمة في حساب صحة التصنيف "معامل كبا ( Kappa ) (Analysis)، وتتراوح قيمة معامل كبا بين 0 و1، أو بين 0 و100 نسبة مئوية؛ حيث تشير القيمة صفر إلى عدم الاتفاق؛ بينما تشير القيمة 1 إلى الاتفاق التام بين نتائج التصنيف والبيانات المرجعية [24]. ويمكن حساب قيمة كبا ( K ) على النحو التالي: أولاً: نحتاج إلى معرفة احتمالية الاتفاق (PO) ويتم اشتقاق هذه الصيغة عن طريق جمع عدد التوقعات الإيجابية ثم تقسيمها على العدد الإجمالي.

$$P_0 = \frac{TP+TN}{TP+FN+FP+TN} \quad (12)$$

الخطوة التالية هي إيجاد احتمالية الاتفاق العشوائي  $P_e$ .

$$P_e = p_{correct} + p_{incorrect} \quad (13)$$

حيث:

$$p_{correct} = \left( \frac{TP+FN}{TP+FN+FP+TN} \right) * \left( \frac{TP+FP}{TP+FN+FP+TN} \right) \quad (14)$$

$$p_{incorrect} = \left( \frac{FP+TN}{TP+FN+FP+TN} \right) * \left( \frac{FN+TN}{TP+FN+FP+TN} \right) \quad (15)$$

أخيراً، صيغة معامل كبا Cohen's Kappa [25] تعطى بالعلاقة:

$$K = (P_0 - P_e)/(1 - P_e) \quad (16)$$

✓ مكتبات بايثون للتعلم الآلي:

تعتبر لغة البايثون لغة قوية (powerful) ومتعددة الأغراض (general-purpose)

فهي تحوي العديد من المكتبات والدوال ومن أبرزها:

**GDAL**: هي مكتبة برمجية لقراءة وكتابة تنسيقات البيانات الجيومكانية والمتجهية.

النقطية وتشكل الأساس لمعظم البرامج الخاصة بمعالجة البيانات الجغرافية المكانية

[26].

**Numpy**: للتعامل مع المصفوفات متعددة الأبعاد [27].

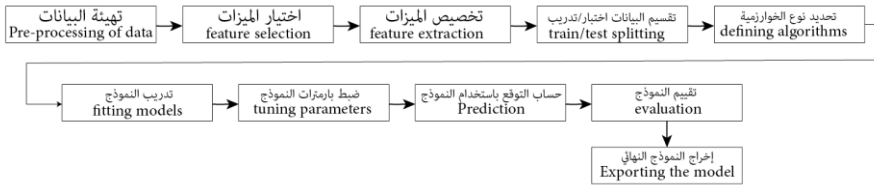
**SciPy**: تتضمن الإحصائيات والتحسين والتكامل ووحدات الجبر الخطي وتحويل فورييه ومعالجة الإشارات والصور وحل المعادلات التفاضلية العادية [28].

**Matplotlib**: للرسم ثنائي وثلاثي الأبعاد [29].

**Pandas**: تحليل هياكل البيانات والتحكم بطريقة عرضها بشكل رقمي [30].

**Scikit-learn**: أهم مكتبة تضم خوارزميات تعلم الآلة [31].

حيث يتم العمل باستخدام أغلب المكتبات وفق التسلسل التالي الشكل (5):



الشكل (5) منهجية عمل مكتبات الذكاء الصناعي

### 3-4 منطقة الدراسة:

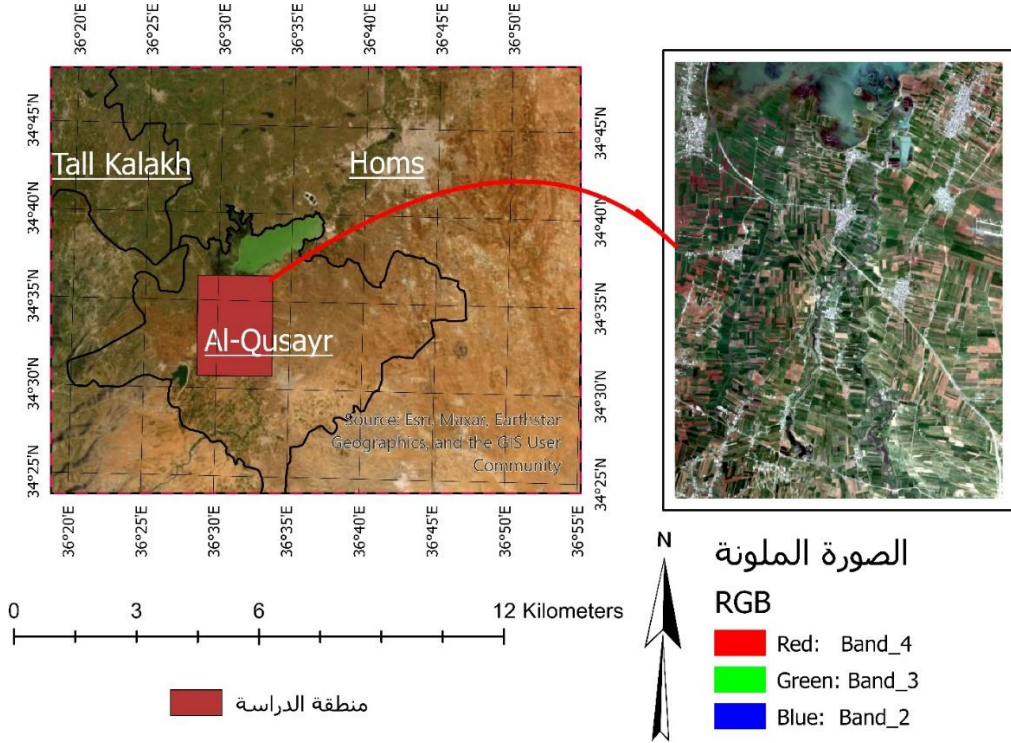
منطقة القصير إحدى المدن في محافظة حمص في سوريا تتميز بمساحتها الكبيرة وقربها من نهر العاصي، واحتوائها على مختلف أنواع النشاطات الزراعية والصناعية والتجارية، وكثرة القرى التابعة إدارياً لهذه المدينة.

تشمل منطقة الدراسة عدة قرى من ريف المدينة والتي تتميز بالأراضي الزراعية السهلية الخصبة الصالحة للزراعة والتي تشكل خزان زراعي هام للمحافظة حيث توفر عدة مزروعات موسمية وتمتاز بوفرة المياه الجوفية والمياه السطحية لقربها من بحيرة قطينة ومرور نهر العاصي بها.

تم تشكيل صورة ملونة لجزء من منطقة القصير الشكل (6).



الشكل (6) منطقة الدراسة (صورة ملونة مركبة بالألوان الطبيعية لجزء من منطقة القصير)



### 3-5 البرامج والأدوات المستخدمة:

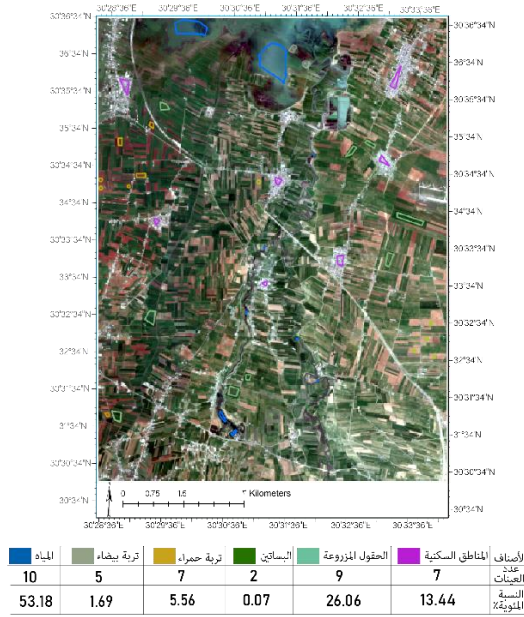
تم الاعتماد على برنامج ArcGIS Pro وهو برنامج نظم معلومات جغرافية مكتبي طوّره شركة ESRI، يتميز ArcGIS Pro ببنية 64 بت ودعم ثنائي الأبعاد وثلاثي الأبعاد ويتكامل مع ArcGIS Online ويدعم بايثون (3). وقد دمج الإصدار الجديد من البرنامج التطبيقات الأربعة القديمة التي كانت تتفرع من النسخة الأساسية ArcMap، Arc Scene، Arc Globe، ArcCatalog [32].

تم استخدامه من أجل استعراض النطاقات الطيفية لصور القمر الصناعي المستخدم وتشكيل الصورة المركبة الملونة وتشكيل العينات التدريبية الشكل (7). كما تم استخدام مواقع خاصة لتحويل عينات التدريبية من صيغة (TIFF) إلى صيغة مصفوفيه (MAT) [33].

تم كتابة الكود البرمجي بواسطة Google Collaboratory وتعرف بأنها خدمة سحابية من Research Google يسمح لأي مستخدم بكتابة كود في محرره وتشغيله من المتصفح. على وجه التحديد، يدعم لغة برمجة Python، وهو موجه لمهام التعلم الآلي، وتحليل البيانات، والمشاريع التعليمية.

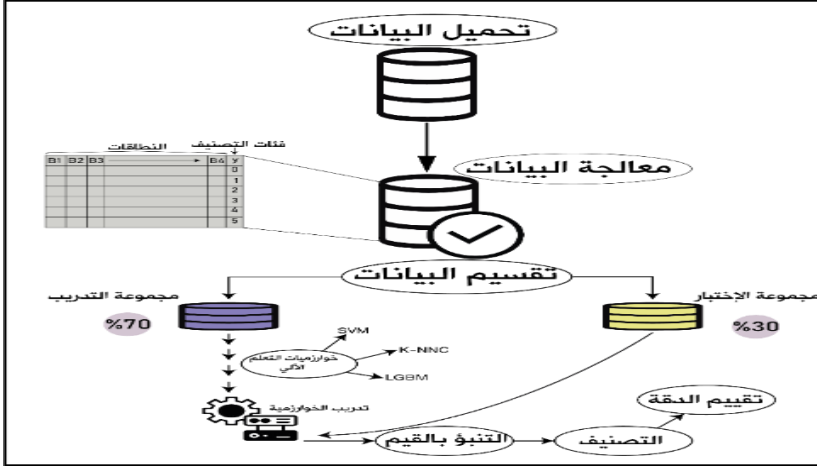
هذه الخدمة، تعتمد على أساس فكرة Jupiter الخاصة بلغة البرمجة Python، وتعتبر خدمة مجانية تماماً مع حساب (Gmail) الخاص بالمستخدم، ولا يتطلب أي تكوين، ولن تضطر إلى تنزيل Jupiter أو تثبيته. يوفر لنا موارد الحاسوب لتحرير واختبار التعليمات البرمجية، مثل GPGPUS لخوادمها.

### الشكل (7) صورة ملونة RGB=432 للعينات التدريبية باستخدام Arc GIS pro



#### 4- تطبيق خوارزميات التعلم الآلي:

تعتمد منهجية العمل العامة للتعلم الآلي الشكل (8) على مجموعة من الخطوات انطلاقاً من تحميل الصور الفضائية واختيار العينات التدريبية وتقسيم البيانات إلى بيانات تدريبية وبيانات اختبار التي سيجري التقييم على أساسها لاحقاً مروراً بكتابة الأكواد اللازمة للتصنيف واستخراج الدقة ومقارنة النتائج.



الشكل (8) توضيح خطوات منهجية العمل

### ✓ قراءة البيانات: Read Data

يتضمن الكود البرمجي لكافة الخوارزميات بخطوته الأولى قراءة النطاقات الطيفية باستخدام الأمر (rasterio) وهي إحدى تنسيقات مكتبة (GDAL) ويتم جمعها في مصفوفة ذات أبعاد n باستخدام طريقة ( ) Numpy. Stack. البيانات الناتجة بعد التكديس في مصفوفة هي من الشكل (10، 402، 534) حيث:

(534) عدد الأسطر، (402) عدد الأعمدة، (10) عدد المصفوفات (عدد النطاقات)، ويكون عدد السجلات الكلي (214,668).

تم إدخال البيانات الخاصة لعمليتي التدريب والاختبار على شكل مصفوفة باستخدام طريقة (loadmat) من حزمة (scipy.io) ذات أبعاد مساوية تماما للنطاقات الطيفية. والتي أخذنا منها العينات التدريبية (training data) بنسبة (70%) حيث استخدمنا 6 فئات للتصنيف تشمل الماء، الأراضي غير المزروعة (تربة بيضاء)، الأراضي غير المزروعة (تربة حمراء) والأراضي المزروعة، والبساتين، والبناء. حيث كل صنف مرمز له برقم من (0 إلى 5) بالترتيب.

### ✓ العرض البصري للبيانات Data Visualization :

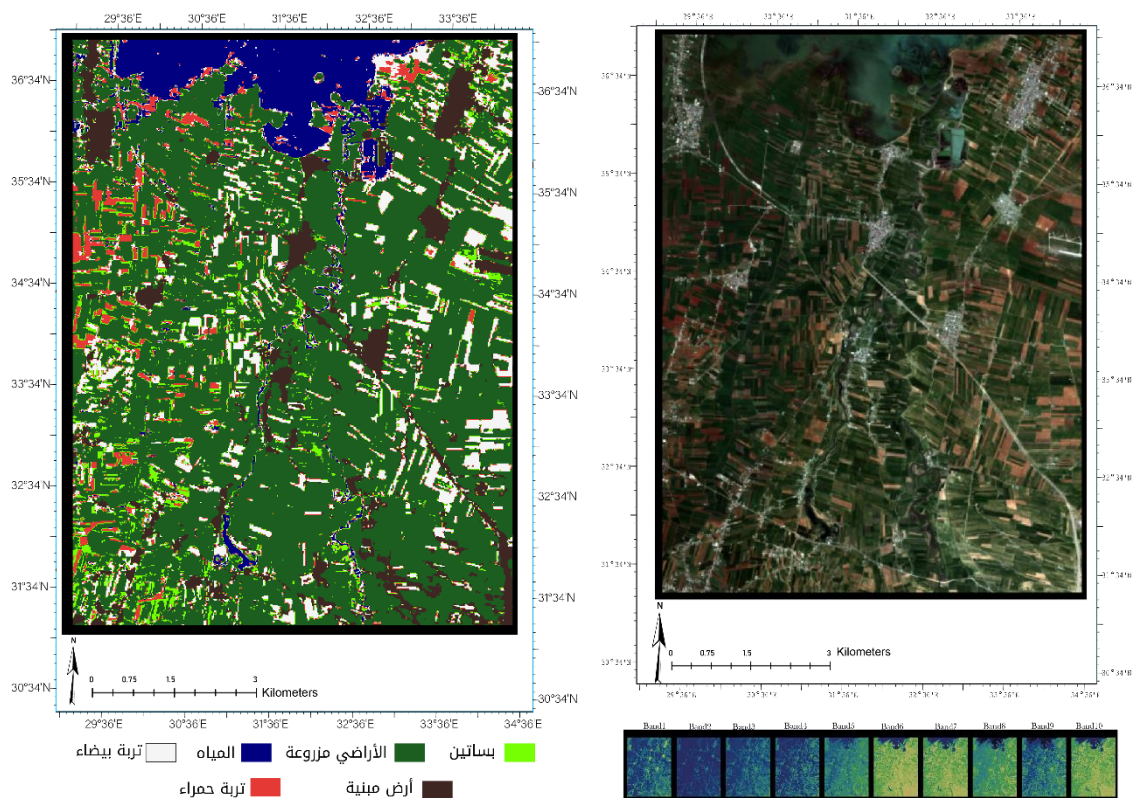
يتم عرض الصور من خلال إنشاء صورة RGB مركبة تسهل فهم البيانات بشكل فعال. لرسم صور RGB بالألوان الطبيعية سنستخدم النطاقات الأزرق والأخضر والأحمر، وهي النطاقات 2 و 3 و 4 على التوالي. نظراً لأن Python تستخدم نظام فهرس قائم على الصفر، فنحن بحاجة إلى طرح قيمة 1 من كل فهرس. لذلك، فإن مؤشر النطاق الأحمر هو 3، والأخضر هو 2، والأزرق هو 1.

يمكن أن تكون الصور المركبة التي أنشأناها مظلمة في بعض الأحيان إذا كانت قيم سطوع البيكسل مائلة نحو قيمة الصفر. يمكن حل هذا النوع من المشاكل عن طريق تمديد قيم سطوع البيكسل في صورة ما بتفعيل أمر تمديد التباين (= True) لتوسيع القيم إلى النطاق الكامل (0-255) لزيادة التباين المرئي للصورة. أيضاً، تسمح لنا (`str_clip`) بتحديد مقدار عدد البكسلات التي نريد تفتيحها أي زيادة قيمتها حيث حددت بقيمة 2% من إجمالي عدد البكسلات.

تحتوي البيانات على 10 نطاقات. لطباعة كل نطاق باستخدام حزمة (`EarhPy`) وهي عبارة عن حزمة تسهل رسم البيانات النقطية المكانية والبيانات المتجهية والعمل معها باستخدام أدوات مفتوحة المصدر.

يتم استخدام الأداة (`plot bands`) لطباعة النطاقات جنباً إلى جنب الشكل (9).

## تقييم خوارزميات التعلم الآلي لإنشاء خرائط الغطاء الأرضي من صور القمر الصناعي Sentinel-2



الشكل (9) الصورة المركبة وصورة الحقيقة الأرضية (العينات المرجعية) وصور النطاقات

## المعالجة Preprocessing:

يتم استخدام فكرة تقسيم البيانات عندما تكون مجموعة البيانات كبيرة بما يكفي للتقسيم إلى مجموعتين، فإذا كانت مجموعة البيانات المتاحة صغيرة فلن تكون هناك بيانات كافية في مجموعة بيانات التدريب للنموذج، ولن تكون هناك أيضاً بيانات كافية في مجموعة الاختبار لتقييم أداء النموذج بشكل فعال. قد يؤدي ذلك إلى نتائج مفرطة في التفاؤل (جيد) أو مفرطة في التشاؤم (سيء) [34].

تقسيم البيانات موضح بالجدول (1).

الجدول (1) تقسيم بيانات التدريب والاختبار

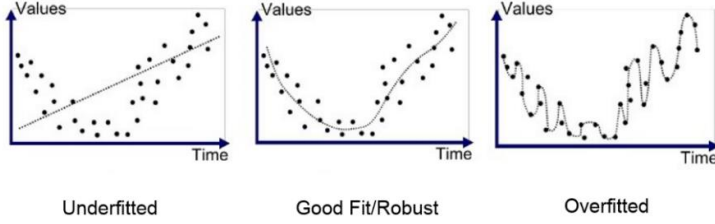
نوع البيانات	حجم البيانات (عدد السجلات، عدد النطاقات)	Y بيانات الهدف
بيانات X-train (70%) التدريب	(150267,10)	150267
بيانات X-test (30%) الاختبار	(64401,10)	64401

## نتائج تطبيق شعاع الدعم الآلي (SVM) Support Vector Machine:

من أجل التصنيف بطريقة شعاع الدعم الآلي، غالباً ما تُستخدم وظائف النواة غير الخطية (kernel) من أجل التصنيف غير الخطي. تستخدم خوارزمية SVM أنواعاً مختلفة من وظائف kernel، وتستخدم التوابع (RBF kernel، Sigmoid kernel، Polynomial kernel) يتم بعد ذلك إنشاء الحد الأقصى للهامش. يعتمد النموذج الذي تم إنتاجه على هذا النحو على مجموعة فرعية فقط من بيانات التدريب بالقرب من حدود الفصل والتي تدعى متجهات الدعم المبينة سابقاً.

يتم أيضاً اختيار قيمة المعامل C والمسمى معامل التنظيم أو معامل الحساسية والذي يؤثر على طبيعة التدريب حيث يؤدي اختيار قيم مناسبة للوصول إلى حالة ملائمة لتوزع البيانات (Good Fit) بينما اختيار قيم كبيرة لهذا المتغير يؤدي إلى حالة فرط تدريب

(Overfitting) بينما لو كانت قيمة  $C$  صغيرة هذا يؤدي إلى حساسية قليلة وحالة نقص تدريب (Underfitting) [12]، كما هو موضح بالشكل (10).



الشكل (10) توضيح حالات المختلفة لتأثير معامل الحساسية  $C$  في SVM

يعتمد اختيار معامل الحساسية وتابع النواة على الحالة المدروسة ويوضح الجدول (2) تأثير تابع النواة والجدول (3) تأثير معامل الحساسية على دقة التصنيف.

الجدول (2) نتائج تغيير وظيفة النواة على دقة التصنيف

تابع النواة مع ( $C=3$ )	الدقة الكلية (%)	معامل كايا (%)
RBF	98.39	97.11
Poly	91.61	83.93
Sigmoid	65.57	39.35

الجدول (3) نتائج تأثير معامل الحساسية ( $C$ ) على دقة التصنيف

معامل الحساسية ( $C$ ) مع التابع RBF	الدقة الكلية (%)	معامل كايا (%)
1	98.09	96.58
2	98.28	96.93
3	98.39	97.11
5	98.32	97.34

باعتقاد وظيفة النواة (RBF) ومعامل الحساسية ( $C=3$ ) تم الحصول على النتائج المبينة في الجدول (4) لدقة المصنف (SVM).



الجدول (4) معايير الدقة (%) لكل صنف وفق خوارزمية svm

الصنف	دقة الصحة	دقة الاسترجاع	درجة F1 score
المياه	98	98	98
تربة بيضاء	95	97	96
تربة حمراء	97	96	96
البساتين	95	88	91
الحقول المزروعة	99	100	100
المناطق السكنية	97	97	97

نتائج تطبيق خوارزمية الجار القريب (K-Nearest Neighbor Classifier (K-NNC):  
 يتم تطبيق خوارزمية الجار القريب باستخدام حزمة التعلم (scikit) مع 6 أصناف نلاحظ  
 تأثير المعامل (k) على دقة التصنيف في الجدول (5). حيث كانت النتائج متقاربة  
 وسنعمد المعامل (k=10) للمقارنة.

الجدول (5) تأثير المعامل k على دقة التصنيف ضمن خوارزمية K-NNC

المعامل k	الدقة الكلية (%)	معامل كبا (%)
6	96.71	94.08
4	96.55	93.81
10	96.96	94.53

يوضح الجدول (6) تفاصيل معايير الدقة في حالة (k=10):

الجدول (6) معايير الدقة (%) لكل صنف وفق خوارزمية K-NNC

الصنف	دقة الصحة	دقة الاسترجاع	درجة F1 score
المياه	96	98	97
تربة بيضاء	93	97	95
تربة حمراء	95	92	94
البساتين	87	74	80
الحقول المزروعة	98	99	99
المناطق السكنية	96	92	94

#### نتائج تطبيق خوارزمية تعزيز التدرج: Gradient Boosting Classifier

يوجد الكثير من البارامترات التي يمكن التحكم بها تدخل في عمل الخوارزمية وسنذكر أهمها:

دالة الخسارة loss function: والتي تستخدم للعثور على الميزات والعتبة المثلى لتقسيم البيانات وسنختار الدالة (gbdt) الخاصة بالخوارزمية المدروسة وهناك دالات أخرى تخص خوارزمية شجرة القرار Random Forest.

معامل التعلم Learning rate: تتحكم هذه المعلمة في حجم التغيير بكل مجموعة تدريبية، تُفضل القيم المنخفضة عموماً لأنها تجعل النموذج قوياً لميزات كل فئة وبالتالي تسمح له بالتعميم جيداً [18].

يبين الجدول (7) نتائج مصنف lightGBM مع دالة الخسارة (gbdt) وباستخدام قيم مختلفة لمعامل تعلم.

الجدول (7) معايير الدقة (%) لكل صنف وفق خوارزمية lightGBM

معامل كابا (%)	الدقة الكلية (%)	معامل التعلم
92.72	95.99	0.03
94.73	97.07	0.06
95.75	97.63	0.1
96.29	97.92	0.4
92.83	96.00	0.5

وباعتماد قيمة معامل التعلم (0.4) لأنه أعطى أفضل النتائج يبين الجدول (8) معايير الدقة لخوارزمية lightGBM.

الجدول (8) تأثير معامل التعلم على دقة التصنيف ضمن خوارزمية lightGBM

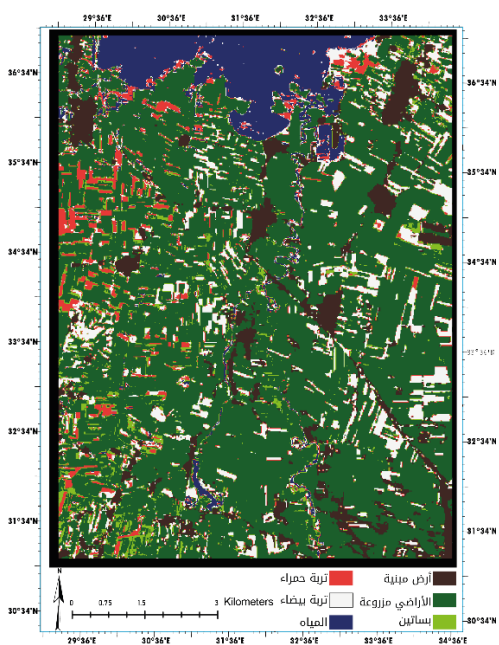
الصف	دقة الصحة	دقة الاسترجاع	درجة F1 score
المياه	98	98	98
تربة بيضاء	96	97	96
تربة حمراء	96	95	96
البساتين	89	85	87
الحقول المزروعة	99	99	99
المناطق السكنية	97	96	96

##### 5- المقارنة:

بمقارنة قيم الدقة المختلفة في الجدول (9) نلاحظ أن خوارزمية شعاع الدعم الآلي (SVM) كانت الأكثر مهارة في تصنيف الغطاء الأرضي ويوضح الشكل (11) خريطة التصنيف بواسطة خوارزمية SVM التي أعطت أعلى دقة.

الجدول (9) مقارنة بين خوارزميات التصنيف

الخوارزمية	معامل كابتا (%)	الدقة الكلية (%)
شعاع الدعم الآلي	98.39	97.11
خوارزمية الجار القريب	96.96	94.53
خوارزمية تعزيز التدرج	97.92	96.29



الشكل (11) خريطة التصنيف بواسطة خوارزمية SVM

## 6- الاستنتاجات والتوصيات:

أثبتت خوارزمية شعاع الدعم الآلي قدرتها على إجراء عملية تصنيف الغطاء الأرضي ضمن فئات محددة بدقة أعلى بالمقارنة مع خوارزميات التعلم الآلي الأخرى المقدمة في البحث.

تتميز خوارزميات التعلم الآلي بقدرتها على التعامل مع كميات كبيرة من البيانات ولكن هذه الخوارزميات تحتاج إلى تدخل بشري في تحديد بعض البارامترات اللازمة لهذه الخوارزميات حيث يكمن التحدي في اختيار الخوارزمية المناسبة مع متغيراتها بما يناسب الحالة المدروسة.

أعطت جميع الخوارزميات نتائج مرضية وتميزت خوارزمية تعزيز التدرج بقدرتها الحسابية وإعطاء نتائج أسرع. حيث تطلبت تشغيل ذاكرة منخفضة من ذواكر الجهاز المحمول المستخدم.

كما كان لحجم بيانات التدريب دوراً هاماً في زيادة دقة التصنيف فيجب إعطاء كمية كافية من عينات التدريب حيث نلاحظ أن الخوارزميات بشكل عام تدرجت بشكل أفضل في حالة الحقول المزروعة لوجود عينات تدريبية كبيرة نسبياً بينما أخطأت في التنبؤ في حالة البساتين المزروعة لعدم كفاية البيانات التدريبية.

لقد ساهمت الدقة المكانية التي يوفرها القمر sentinel-2 في زيادة دقة التصنيف بشكل ملحوظ مقارنة مع الدراسات السابقة، فالدقة المكانية تلعب دوراً هاماً في دقة التصنيف ويمكن عند استخدام صور عالية الدقة أن نحصل على نتائج في التصنيف قد تمكننا من استخدامها في مجالات أكبر وأكثر حساسية مثل حدود العقارات وحساب المساحات مما يساعد متخذي القرار في عدة مجالات (زراعية وتخطيطية وغيرها).

إن تطور تقنيات الذكاء الاصطناعي يجعل هذا البحث قابل للتطوير بشكل مستمر لذلك نوصي بالاستفادة من الإضافة المتتابة على خوارزميات التعلم الآلي التي تجعلها أكثر كفاءة وقدرة على القيام بمهامها للوصول إلى الآلية المثالية التي ستكون مرجع لكل إجراءات وعمليات تصنيف وتحليل صور الأقمار الاصطناعية.

7-المراجع:

- [1]: S. K. Dash and M. Panda 2016, Image Classification using Data Mining Techniques, **Advances in Computer Science and Information Technology (ACSIT)**, vol. 3, pp. 157–162.
- [2]: J. Kim, B. S. Kim, and S. Savarese 2005, Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines, **Applied Mathematics in Electrical and Computer Engineering**, pp.133-138.
- [3]: Lu, D.; Weng, Q.A., 2007- Survey of image classification methods and techniques for improving classification performance. **Int. J. Remote Sens.** pp823–870.
- [4]: Mustafa U, Fusun B and Barnali D, 2015-Application of Support Vector Machines for Land use Classification Using High-Resolution Rapid Eye Images: A Sensitivity Analysis. **European Journal of Remote Sensing** - 48: 403-422.
- [5]: Daniel L. Civco ,1993- Artificial intelligence for land-cover classification and mapping. **International Journal of Geographical Information Systems**, 7:2, 173-186.
- [6]: Omar S, Amira S, & Safaa M, 2012-Remote Sensing Satellite Images Classification using Support Vector Machine and Particle Swarm Optimization. **Third International Conference on Innovations in Bio-Inspired Computing and Applications.**
- [7]: Aissam Bekkari, Soufiane Idbraim, Driss Mammass and Mostafa El Yassa, 2014 SVM Classification of High Resolution Urban Satellites Images using Composite Kernels and Haralick Features. **Journal of emerging technologies in web intelligence**, VOL. 6, NO.1.pp79-80.
- [8]: European Space Agency, 2021-Gearing up for third Sentinel-2 satellite. **Gearing up for third Sentinel-2 satellite (phys.org).**
- [9]: Alex Smola and S.V.N. Vishwanathan, 2008- Introduction to Machine Learning. **Cambridge University**, First published. pp 40–100.
- [10] Vapnik V, Chervonenkis A. 1964. A note on one class of perceptrons\_ **Automation and Remote Control**, 25.

[11]: L.M. Griffel, D. Delparte, J. Edward, 2018- Using Support Vector Machines classification to differentiate spectral signatures of potato plants infected with Potato Virus Y. **Computers and Electronics in Agriculture** 153 318–324.

[12]: Candade, N and Dr. Dixon, B, 2004- Multispectral classification of landsat images: a comparison of support vector machine and neural network classifiers, **University of South Florida, FL 33701.**

[13]: Boubou Rim, 2019- Prediction of ground surface settlements induced by tunnels excavation using Support Vector Machine, **Al-Baath University Journal**, Vol.4. 6-8.

[14]: Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P, Toussaint G, Output 2005- sensitive algorithms for computing nearest neighbor decision boundaries, **Discrete and Computational Geometry**, pp. 593 – 604.

[15]: Vyoma Patel, G. J. Sahani, 2014 \_A Survey on Image Classification using Data Mining Techniques, **IJSRD - International Journal for Scientific Research & Development**, Vol. 2, Issue 10, pp. 746 – 750.

[16]: Machado, M.R.; Karray, S.; Sousa, August 2019 I.T.d. : An Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry. **In Proceedings of the 2019 14th International Conference on Computer Science & Education** (ICCSE), Toronto, ON, Canada, 19–21.

[17]: Bierman, L. ,1997- Arcing the edge **Statistics Department, University of California at Berkeley**.pp.1-5

[18]: Key, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y, 2017 -Light GBM: A highly efficient gradient boosting decision tree. **International Conference on Neural Information Processing Systems**, pp. 3149–3157.

[19]: Gao, Jay. ,2008. Digital analysis of remotely sensed imagery: **McGraw-Hill Professional.**

- [20]: Powers, 2010, David MW. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." **arXiv preprint arXiv2020**, 16061.
- [21]: Darwishe.Hanan, Chaaban. Fadi, Mohamad. Adnan,2021- Developing a Model of Deep Learning by ANNs for Urban Areas Extraction from Remote Sensing Images- Al-Baath University Journal V 43 Issue 7 PP 18-21.
- [22]: Story, Michael, & Congalton, Russell G, 1986. Accuracy assessment: a user's perspective. **Photogrammetric engineering and remote sensing**, 52(3), pp 397-399.
- [23]: Congalton, Russell G, & Green, Kass,2008. Assessing the accuracy of remotely sensed data: principles and practices: **CRC press**.
- [24]: Monserud, Robert A, & Leemans, Rik,1999- Comparing global vegetation maps with the Kappa statistic. **Ecological modelling**, 62(4), pp275-293.
- [25]: Sim, Julius; Wright, Chris C. (2005). "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements". **Physical Therapy**. 85 (3): 257–268.
- [26]: Frank Warmerdam, Even Rouault, and others, May 16, 2022- **GDAL Documentation**. CHAPTER (1) pp2-10.
- [27]: Numpy, 2022. [Online]. Available: <https://numpy.org> [Accessed: 17-May-2022].
- [28]: SciPy, 2022. [Online]. Available: <https://scipy.org/> [Accessed: 17- May -2022].
- [29]: Matplotlib, 2022. [Online]. Available: <https://matplotlib.org/> [Accessed: 17- May -2022].
- [30]: Pandas, 2022. [Online]. Available <https://pandas.pydata.org/> [Accessed: 17- May -2022].



- [31]: Scikit-learn, 2022. [Online]. Available: <https://scikit-learn.org/> [Accessed: 17- May -2022].
- [32]: ESRI, “Environmental Systems Research Institute,” 2022. [Online]. Available: <https://www.esri.com/>. [Accessed: 17- May - 2022].
- [33]: <https://www.vertopal.com/en/convert/tif-to-mat> [Accessed: 17- May -2022].
- [34]: Rácz, A.; Bajusz, D. Héberger, K. Effect of Dataset Size and, 2021- Train/Test Split Ratios in QSAR/QSPR Multiclass Classification. **Molecules journal** 26,1111.

