



تقييم صحة تصنيف خوارزميات تعلم الآلة لمرئيات سنتينل ٢-

لمحافظة جدة باستخدام محرك جوجل الأرضي

Assessment of the Classification Validity of Machine Learning Algorithms for Jeddah Governorate Sentinel-2 Imagery Using the Google Earth Engine

إعداد

أسماء بنت علي الزهراني

Asmaa Ali Al Zahrani

ماجستير جغرافيا - كلية العلوم الإنسانية والاجتماعية - جامعة الملك سعود

د. فهد بن عبد العزيز المطلق

Dr. Fahad Abdulaziz Almutlaq

أستاذ مشارك بقسم الجغرافيا - كلية العلوم الإنسانية والاجتماعية - جامعة الملك سعود

Doi: 10.21608/jasg.2025.458757

استلام البحث: ١٢ / ٣ / ٢٠٢٥

قبول النشر: ٦ / ٥ / ٢٠٢٥

الزهراني، أسماء بنت علي و المطلق، فهد بن عبد العزيز (٢٠٢٥). تقييم صحة تصنيف خوارزميات تعلم الآلة لمرئيات سنتينل ٢- لمحافظة جدة باستخدام محرك جوجل الأرضي. *المجلة العربية للدراسات الجغرافية*، المؤسسة العربية للتربية والعلوم والآداب، مصر، ٨(٢٥)، ٤٣ - ٧٢.

<https://jasg.journals.ekb.eg>

تقييم صحة تصنيف خوارزميات تعلم الآلة لمربئيات سننل - ٢ لمحافظة جة باستخدام محرك جوجل الأرضي

المستخلص:

توفر منصة محرك جوجل الأرضي (Google Earth Engine) الوصول إلى كمية كبيرة من بيانات الأقمار الصناعية المتاحة مجاناً ومعالجتها بسلاسة، إضافةً إلى توفير مجموعة من أحدث المصنفات للتصنيف القائم على الخلية التي يمكن استخدامها لرسم خرائط الغطاء الأرضي. يتم استخدام طرق مختلفة للتصنيف الرقمي للصور، ومع ذلك قد تقدم نتائج مختلفة ومن المهم اختبار صحتها للتحقق من فعاليتها في رسم الخرائط الموضوعية للغطاء الأرضي من أجل اختيار المصنف الذي يقدم أفضل النتائج، تقوم هذه الدراسة برسم الخرائط الموضوعية للغطاء الأرضي لمحافظة جة، وباستخدام بيانات القمر الصناعي سننل- 2 للفترة الزمنية (0504/2024 - 0104/2024)، في ضوء ذلك تهدف الدراسة إلى تقييم أداء ثلاث خوارزميات تصنيف خاضعة للإشراف لصحة تصنيف خرائط الغطاء الأرضي لمنطقة الدراسة، وكانت الخوارزميات التي تم اختبارها هي: آلة المتجهات الداعمة (SVM)، والغابة العشوائية (RF)، وشجرة التصنيف والانحدار (CART). تم استخدام مصفوفة الخطأ والدقة الإجمالية ومعامل كبا لمقارنة نتائج تصنيف تعلم الآلة، تبينت النتائج فحصنا على أفضلها من خلال تصنيف (RF) بصحة إجمالية (94.5%)، والأقل صحة كان تصنيف (SVM) بصحة إجمالية بلغت (86.7%). وأوصت الدراسة بالاعتماد على الأقمار الصناعية ذات وضوح مكاني مناسب، مع الأخذ بالحسبان عند إجراء عملية تصنيف الغطاء الأرضي: حجم منطقة الدراسة، كذلك توصي باستخدام تصنيف آلة الغابات العشوائية في المناطق ذات الكتلة العمرانية نظراً لما توفره من دقة تصنيف عالية، خصوصاً في تمييز المناطق الحضرية مقارنة بالتصنيفات الأخرى.

الكلمات المفتاحية: محرك جوجل الأرضي، الخرائط الموضوعية، تصنيف الغطاء الأرضي، تقييم الصحة، خوارزميات تعلم الآلة.

Abstract:

The Google Earth Engine platform provides access to a vast amount of freely available satellite data and facilitates seamless processing, in addition to offering a set of the latest classifiers for cell-based classification that can be used for land cover mapping. Different methods for digital classification of images are employed, but they may yield different results, making it crucial to test their validity to verify their effectiveness in objective land cover mapping and to choose the classifier that provides the best results. This study is based



on deriving objective land cover maps for Jeddah Governorate, using Sentinel-2 satellite data from the time period of (2024/04/01) to (2024/04/5). In this context, the study aims to evaluate the performance of three supervised classification algorithms for the validity of classifying land cover maps for the study in question. The algorithms tested were: Support Vector Machine (SVM), Random Forest (RF) and Classification and Regression Tree (CART). An error matrix, overall accuracy and Kappa coefficient were used to compare the results of machine learning classification. The results varied, with the best outcome obtained through RF classification with an overall validity of 94.5%, whilst the least validity was the SVM classification with an overall validity of 86.7%. The study recommended the reliance on satellites with appropriate spatial resolution, taking into consideration the size of the study area during land cover classification, the use of the Random Forest algorithm is recommended in built-up areas due to its high classification accuracy, particularly in distinguishing urban regions compared to other classification methods.

Keywords: Google Earth Engine, thematic maps, classification land cover, assessment validity, machine learning algorithms

المقدمة:

تُعَدُّ الخرائط تمثيل رمزي للواقع الجغرافي، فهي تسلط الضوء على ظواهر، أو جوانب معينة يتم تصميمها عندما تكون العلاقات المكانية هي الأكثر أهمية، فإن الخرائط ما هي إلا نتاج الجهد الإبداعي للمؤلف (Robinson et al., 1984)، وينبغي أن تخضع الخرائط الموضوعية المبنية من بيانات الاستشعار عن بعد لتقييم دقيق إحصائيًا، كما يحدد تقييم الصحة جودة البيانات حتى يتمكن مستخدمي الخريطة من تقييم فائدة الخريطة الموضوعية للتطبيقات المعنية (Stehman & Czaplewski, 1998).

وتتضمن صحة الخرائط الموضوعية تقييم مصداقية ظواهر المعالم المحددة بمقارنتها بمرجع يُفترض أن يكون صحيحًا (Zhang, 2020) وبناءً على ذلك يُعَدُّ تقييم الصحة، أو التحقق من مدى صحة خرائط الغطاء الأرضي، أو الغطاء النباتي المستخلصة من بيانات

الاستشعار عن بعد من الأمور الهامة، فيجب أن يؤخذ بالاعتبار قبل أي عملية صنع قرار معرفة صحة الخرائط المستخدمة. (Congalton, 2005)
فإن التصنيفات التي تنتجها خرائط استخدامات الغطاء الأرضي (Land Use and Land Cover) ذات الصحة العالية مطلوبة دائماً عندما تكون هناك حاجة لمعلومات جديرة بالثقة من صور الاستشعار عن بعد، ومما لا شك فيه فإن عملية التصنيف بواسطة خوارزميات التعلم الآلي على وجه الخصوص تنتج تصنيفات جيدة حتى على البيانات المعقدة متعددة الأبعاد، كما تتأثر صحة الخرائط المصنفة بعدة عوامل مختلفة مثل: حجم عينات التدريب، وجودتها، والصحة الموضوعية، وحجم منطقة الدراسة، وغيرها. (Shelty, 2019).

وتوفر منصة جوجل الأرضي (GEE) لمستخدميها العديد من الخوارزميات المختلفة للقيام بعملية التصنيف الموجه أو غير الموجه، وسيتم التركيز بهذه الدراسة على بعض أنواع الخوارزميات الخاضعة للإشراف مثل: أشجار التصنيف والانحدار (CART)، وتصنيف الغابة العشوائية (RF)، وتصنيف آلة المتجهات الداعمة (SVM)، وتقييم أداء هذه الخوارزميات على صحة تصنيف الخرائط الموضوعية المستخلصة بواسطة مصفوفة الخطأ Error Matrix التي تشتمل على: دقة المنتج Producer Accuracy ، دقة المستخدم User Accuracy ، والدقة الإجمالية الشاملة overall Accuracy. ومعامل كابا Coefficient Kappa.

أهمية الدراسة:

في الوقت الراهن ومع التقدم في التصوير بالأقمار الصناعية المختلفة وزيادة مصادر البيانات والبرمجيات المستخدمة في تحليل هذه البيانات، تتجلى أهمية الدراسة في تقييم أداء خوارزميات تعلم الآلة على صحة تصنيف خرائط الغطاء الأرضي باستخدام بيانات القمر الصناعي سنيتينل-2 عن طريق محرك جوجل الأرضي (GEE)، لتوفير الوقت، والجهد، وتوفير البيانات المجانية، ستسهم الدراسة في تقديم قاعدة بيانات تحمل معلومات صحيحة ودقيقة وموثوقة بها للغطاء الأرضي، قائمة على تحليل علمي تقود أصحاب القرار في استخدام الغطاء الأرضي.

مشكلة الدراسة:

تتمحور الدراسة حول تقييم أداء خوارزميات تعلم الآلة على صحة تصنيف خرائط الغطاء الأرضي المعتمدة على بيانات القمر الصناعي قمر سنيتينل-2 والمقارنة نتائج بين

الخوارزميات المستخدمة. فدرجة الوضوح المكاني للمرئيات الفضائية وطبيعة منطقة الدراسة له تأثير كبير على نوعية وجودة البيانات المستخلصة الممثلة على الخريطة. فقد يؤثر حجم الخلية في تمييز الظاهرة المستخلصة حيث قد تكون الظاهرة أصغر من حجم الخلية، وبهذا سيكون هناك عدم وضوح لتمثيل الظاهرة وإخراجها على خريطة. وتعد عملية جمع المعلومات الجغرافية القائمة على الرصد عرضة لعدم اليقين بعدة أشكال، وبشكل عام هناك نوعان من تقييمات صحة الخرائط: الموضوعية والموضوعية (Congalton & Green, 2009). وتركز الدراسة على الصحة الموضوعية فهي تتعلق بصحة الفئة، أو التصنيف.

أهداف الدراسة:

تسعى هذه الدراسة إلى تحقيق الأهداف التالية:

١. تقييم أداء الخوارزميات على صحة تصنيف خرائط الغطاء الأرضي بواسطة منصة محرك قوئل الأرضي (GEE).
٢. مقارنة نتائج تصنيف خوارزميات تعلم الآلة.

تساؤلات الدراسة:

- ١- ما مدى صحة أداء الخوارزميات المستخدمة في الدراسة في تصنيف خرائط الغطاء الأرضي؟

- ٢- ما الفرق بين نتائج تصنيف خوارزميات تعلم الآلة المستخدمة في الدراسة؟

منطقة الدراسة:

تقع محافظة جدة على الساحل الغربي من المملكة العربية السعودية بين دائرتي عرض $21^{\circ}25'$ و $21^{\circ}45'$ شمالاً وخطي طول $39^{\circ}5'$ و $39^{\circ}20'$ شرقاً، كما في الشكل (1)، عند منتصف الشاطئ الشرقي للبحر الأحمر جنوب مدار السرطان، ويحيطها من الشرق سهول تهامة، وتمثل منخفضاً لمرتفعات الحجاز، ومن الغرب يوجد على مسافة الشاطئ سلاسل متوازية من الشعب المرجانية (أمانة محافظة جدة، 2022م).

شكل (1) موقع منطقة الدراسة.



المصدر: امانة محافظة جدة، ٢٠١٧.

مراجعة الادبيات:

بعد الاطلاع على مجموعة من الأبحاث السابقة التي لها صلة بموضوع الدراسة تُعدُّ دراسة روانجا وندامبوكي (Rwanga & Ndambuki, 2017) مصدرًا أساسيًا للمعلومات حيثُ تمكن للمخططين وصناع القرار استخدامها للتخطيط المستدام للبيئة، حيثُ ذكر بدراسته من المهم؛ لكي تنجح عملية تصنيف المرئية أن ننظر في عدة عوامل، ومنها: توفر صور Landsat عالية الجودة، وعملية تصنيف دقيقة. وجاء الهدف من هذه الدراسة، هو تصنيف ورسم خريطة لاستخدام الأراضي بمحافظة ليمبوبو بجنوب أفريقيا. واحتوت الدراسة على قسمين 1- تصنيف الغطاء الأرضي. 2-تقييم الدقة. وتمَّ إجراء التصنيف الموجه، وكانت التصنيفات الرئيسة هي: (الزراعة 65%)، (المسطحات المائية 4.5%)، (المناطق المبنية 18.3%)، (الغابات المختلطة 52%)، (الأشجار، 70%)، (الأراضي القاحلة 0.5%). وكانت دقة التصنيف الإجماليَّة للدراسة (81.7% ومعامل كبا (0.722)، تمَّ تصنيف معامل كبا على أنه حقق نسبةً عاليةً، وبالتالي الصور المصنفة مناسبة للقيام بدراساتٍ أخرى.

- هدفت دراسة وهاب وشفري (Wahap & Shafri, 2020) إلى تقييم جدوى (Google Earth Engine) كم منصة مفتوحة المصدر قائمة على السحابة من خلال إجراء تصنيف

لمنطقة وادي كلانج بماليزيا، من مرئيات Landsat (لمدة ثلاث سنواتٍ مختلفةٍ -1988) (2018-2003 باستخدام خوارزميات متعددة للتعلم الآلي (MLA). تمَّ بعد ذلك استيراد أفضل نتائج التصنيف، ومعالجتها بشكلٍ أكبر لتحديد التغيرات على مرَّ السنين باستخدام البرامج التجارية. على الرغم من نتائج التصنيف عالية الدقة، ولكن أظهر تصنيف CART أفضل دقةً بنسبة (94.71% و (97.72% و (96.57% في الأعوام (1988 و2003 و(2018 مقارنةً بتصنيف RF و(SVM). وتوصّلت النتائج إلى أن أداء GEE (كان جيّدًا وسريعًا من حيث الوقت، وتعقيد معالجة مجموعة البيانات المتعددة بأقل قدرٍ من التفاعل والتداخل البشري، وبشكلٍ عامٍ أثبتت منصة GEE (موثوقيتها في تحقيق أهداف هذه الدراسة لتقييم جدوى GEE (من خلال إجراء التصنيف وتحديد استخدام الأراضي، وتوفير قاعدة جيّدة لمزيد من التحليل باستخدام منصة مختلفة.

- اقترح تشانغ وآخرون (Zhang al el., 2020) العديد من خوارزميات التصنيف؛ لاستخراج فئات استخدام الأراضي، استنادًا إلى صور الاستشعار عن بعد. تمَّ تطبيق الدراسة في المنطقة الهامشية الشرقية في جيان بالصين باستخدام بيانات (Lansat8) لعام (2019م) لمقارنة أداء خوارزميات التصنيف الخاضعة للإشراف وهي بشكلٍ خاص: (RF, ANN, SVM, MLC). وأظهرت النتائج أن الدقة الكلية لـ (RF) هي ((86% ومعامل كبا (0.8) والدقة الكلية لـ (SVM) هي: ((83.2% وكبا (0.75) والدقة الإجمالية لـ (ANN) هي: ((81% وكبا (0.72) والدقة الإجمالية لـ (MLC) هي (73.6% ومعامل كبا (0.63) ويدلُّ ذلك على أن خوارزمية (RF) حققت أفضل نتيجة تصنيف تليها (SVM, ANN) وتوصّلت خوارزمية (MLC) على أقل دقة.

- هدفت دراسة غيور وآخرون (Ghayour et al., 2021) في كردستان إلى تقييم أداء خوارزميات تعلم الآلة: آلة المتجهات الداعمة (SVM) والشبكات العصبية الاصطناعية (ANN) و تصنيف الاحتمال الأقصى (MLC) والحد الأدنى للمسافة (MD) وخوارزمية المسافة (MH) ومقارنتها؛ لإنشاء خريطة للغطاء الأرضي واستخدام الأرض (LULC) باستخدام بيانات الأقمار الصناعية (Sentinal2) و (Landsat8). تمَّ إنشاء بيانات التدريب، والتحقق من الصحة من صور (Google Earth) وبيانات (GPS) وبعد ذلك تمَّ تصنيف الصور باستخدام الخوارزميات، و بيانات التدريب، وأخيرًا لتقييم النتائج تمَّ استخدام بيانات التحقق من الصحة لتوليد مصفوفة الارتباك للصور المصنفة، وتوصّلت النتائج إلى أن التصنيف باستخدام خوارزمية (SVM) حققت أعلى دقة إجمالية بلغت (94%) وأداءً أفضل لكلٍ من بيانات الأقمار الصناعية مقارنةً بالطرق الأخرى، كما كانت صور (Sentinal2) أكثر دقة قليلًا مقارنةً بـ (Landsat8).

- تناولت دراسة لوكيكا، وآخرون (Loukika al el., 2021) تصنيف الغطاء الأرضي، واستخدام الأراضي (LULC) على منصة (Google Earth Engine) باستخدام

التصنيف الخاضع للأشراف لحوض مونبرو بالهند، وتمّ استخدام ثلاث خوارزميات مختلفة للتعلّم الآليّ وهي آلة المتجهات الداعمة (SVM) والغابات العشوائية (RF) وأشجار الانحدار الطبقيّة (CART) ومقارنة أدائها باستخدام تقيّمات الدقّة. ولتحسين دقة التصنيف تمّ اشتقاق وأدراج مؤشر الفرق المعياريّ للغطاء النباتيّ (NDVI) ومؤشر اختلاف المياه المعياريّ (NDWI) باستخدام بيانات (Sentinel-2 و Landsat-8) متعددة الأزمان بدقة مكانية تبلغ 10 أمتار و 30 مترًا، وكانت الفئات الرئيسة لاستخدام الأراضي: المسطحات المائية والغابات، والأراضي القاحلة، والنباتات، والمباني، توصلت إلى أنّ متوسط الدقّة الإجمالية لتصنيف صور Landsat-8 بلغت (90.88%) و(94.85%) و(82.88%) على التوالي، وبلغت صور Sentinel-2 (93.8%) و(95.8%) و(86.4%)، وتشير النتائج إلى أنّ تصنيف RF يتفوق على تصنيف SVM و CART من حيث الدقّة.

يتبين من خلال تناول الدراسات السابقة بأنّ الدراسات التي استخدمت خوارزميات تعلّم الآلة؛ لتقييم أدائها على صحة تصنيف الخرائط الموضوعيّة للغطاء الأرضيّ، قد طبقت بدراسات إقليمية وعالمية، أمّا منطقة الدراسة فمن خلال اطلاع الباحثان على الدراسات المنشورة لم يسبق القيام بتقييم أداء خوارزميات تعلّم الآلة (Machine Learning) على صحة تصنيف الخرائط الموضوعيّة للغطاء الأرضيّ على منطقة محدودة المساحة من بيانات القمر الصناعيّ سنتينل-2 لمحافظة جدة باستخدام منصة الحوسبة السحابية (GEE)).

المنهجية:

اتبعت هذه الدراسة المنهج الاستقرائيّ التحليليّ، باستخدام الأسلوب التحليليّ للمرئية الفضائية المستخدمة بالدراسة من خلال محرك منصة جوجل الأرضيّ (GEE)، بالاعتماد على صور الأقمار الصناعية سنتينل-2 مصحح جويًا بتاريخ 05-07-2024, 01-07-2024) باستخدام الباند 2 و 3 و 4 لتقييم أداء أساليب التصنيف المختلفة بواسطة خوارزميات تعلّم الآلة على فئات الغطاء الأرضيّ.

البيانات:

تتميز البيانات الجيومكانية الضخمة بالوقت الراهن باهتمام كبير محليًا وعالميًا، وتعدّ المرئيات الفضائية من أهم البيانات المستخدمة في الدراسة، بواسطة الأقمار الصناعية تمّ القيام بتصنيف الغطاء الأرضيّ لمنطقة الدراسة على خمس فئات: المياه، الأراضي الجرداء، الغطاء النباتيّ، الطرق، المناطق العمرانيّة. تمّ ذكر تصنيف ووصف لهذه الفئات في جدول (1). وتمّ تقييم أداء خوارزميات تعلّم الآلة على صحة التصنيف بواسطة مصفوفة الخطأ بحساب الصحة الشاملة وصحة المنتج وصحة المستخدم، ومعامل كابتا؛ لتقييم أداء كلّ تصنيف. بالاعتماد على بيانات القمر الصناعيّ سنتينل-2.



قمر سنتينيل ٢: (Sentniel-2)

يتكوّن برنامج كوبرنيكوس سنتينيل-2 من قمرين صناعيين يدوران في مدار قطبيّ يوضعان في نفس المدار المتزامن مع الشمس، على مراحل عند 180 درجةً لبعضهما بعضاً، وهو برنامج تصوير أوروبّي واسع النطاق، وعالي الدقّة ومتعدد الأطياف، بهدف رصد التباين في ظروف سطح الأرض، وعرض مساحته الواسعة (290 كم)، ويبلغ الوضوح الزمني 10 أيام عند خط الاستواء مع قمر صناعيّ واحد، و5 أيام مع القمرين الصناعيين، وفي ظل ظروف خالية من السحابة يؤدي عمله من 2-3 أيام عند خطوط العرض الوسطى) لدعم رصد التغيّرات السطحيّة للأرض. يحصل برنامج (Sentinel-2) بشكلٍ منهجيّ على بيانات الأراضي والمناطق الساحليّة في نطاق خط عرض يمتد من 56 درجة جنوباً (جزيرة هورنوس، كيب هورن، أمريكا الجنوبيّة) إلى 82.8 درجة شمالاً (فوق جرينلاند)، ويحمل (Sentinel-2) جهاز بصريّ لأخذ عينات من 13 نطاقاً طيفيّاً: حيث يمثل الوضوح المكاني أربع نطاقاتٍ منها على ارتفاع 10 أمتار، وستة نطاقات على ارتفاع 20 متراً، وثلاث نطاقاتٍ بدقّة مكانيّة تبلغ 60 متراً، وثلاث نطاقاتٍ بدقّة مكانيّة تبلغ 60 متراً، ويوضّح جدول (2) النطاقات الطيفيّة للقمر الصناعي (Sentinel-2)

جدول: (1) تصنيف فئات الغطاء الأرضي

نوع الغطاء الأرضي	المعرّف	الوصف
المياه	1	البحيرات والجداول والمستنقعات.
الأراضي جرداء	2	الأراضي ذات التربة، أو الرمال، أو الصخور المكشوفة، والتي لا يزيد فيها الغطاء النباتي عن ١٠% بمختلف فصول السنة.
الغطاء النباتي	3	الأشجار، والمناطق الزراعيّة، والحدائق، وغيرها.
المناطق الحضرية	4	المناطق التي تتواجد بها المباني وغيرها التي من صنع الإنسان، مثل: الأراضي السكنيّة، أو التجاريّة، أو الصناعيّة، والمسارات الخاصة بالسيارات والشوارع التي تكون ضمن المناطق السكنيّة والطرق.

جدول: (2) بيانات القمر الصناعي MSI Sentinel-2

النطاقات	الطول الموجي	درجة الوضوح المكاني
Bands	"Nanometers"	"Meters"
B1 السواحل	442.7	60
B2 الأزرق	492.4	10
B3 الأخضر	559.8	10
B4 الأحمر	664.6	10
B5 الحافة الحمراء "الغطاء النباتي"	704.1	20
B6 الحافة الحمراء "الغطاء النباتي"	740.5	20
B7 الحافة الحمراء "الغطاء النباتي"	782.8	20

10	832.8	B8 تحت الحمراء القريبة
20	864.7	B8A تحت الحمراء الضيقة
60	945.1	B9 بخار الماء
20	1373.5	B10 تحت الحمراء القصيرة

المصدر: موقع وكالة الفضاء الأوروبية، ٢٠٢٤.

تصنيفات تعلم الآلة:

يُعد التعلم الآلي الأكثر موثوقية من بين أساليب تصنيف الأنظمة غير الخطية، فهو يساعد على فهم سلوك النظام استنادًا إلى ملاحظات المدخلات ولديه القدرة على تقريب القيم دون معرفة مسبقة بالعلاقة بين البيانات (Lary et al., 2016) فإن تقنية التعلم الآلي تُعد خيارًا مناسبًا لتصنيف صور الاستشعار عن بعد لصعوبة المعرفة التامة بخصائص منطقة الدراسة بأكملها.

توفر منصة محرك الأرض (GEE) الدعم المتكامل لإجراء عمليات تصنيف الصور المختلفة، وتستخدم الدراسة للتصنيف المراقب تصنيفات التعلم الآلي التالية:

Classification And Regression Tree (CART), Random Forest (RF),
Support vector machine (SVM).

شجرة التصنيف والانحدار (CART):

تم نشر خوارزمية (CART) لأول مرة في عام 1984 بواسطة (Leo Breiman)، فهي عبارة عن شجرة تصنيف للقرار تسهل في عملية اتخاذ القرارات المباشرة وتحليلات الانحدار (Shetty, 2019). وتعمل (CART) عن طريق فصل العقد بناءً على قيم محددة مسبقًا حتى تصل إلى العقد الطرفية، وتقسّم هذه الخوارزمية بيانات الإدخال إلى عدة مجموعات، ولإنشاء الأشجار يتم ذلك بواسطة جمع جميع المجموعات باستثناء مجموعة واحدة (Basheer et al., 2022)، ولتطبيق هذه الخوارزمية تم تطبيق أداة (classifier.smileCart) الذي يُعد جزءًا من مكتبة منصة جوجل الأرضي (GEE).

تصنيف الغابة العشوائية (RF):

تعتمد كل شجرة في الغابة العشوائية على قيم ناقلات عشوائية، يتم أخذ العينات منها بشكل منفصل وبفصل التوزيع لجميع الأشجار في الغابة، كما أنه مع زيادة عدد الأشجار في الغابة، يتقارب خطأ التعميم إلى حد أقصى. وتؤثر قوة كل شجرة فردية في الغابة والعلاقة بينهما على خطأ التعميم لغابة مصنّفات الأشجار. (Breiman, 2001)

تم إجراء تصنيف الغابة العشوائية (RF) على منصة محرك الأرض (GEE) باستخدام أداة (Classifier.smileRandomForest) المتاحة في مكتبة منصة جوجل الأرضي (GEE).

تصنيف آلة المتجهات الداعمة (SVM)

تم وصف هذه الخوارزمية لأول مرة في (Cortes and Vapnik) عام 1995 استناداً على عمل (Vapink 1982)، وهي تُعد واحدة من تقنيات التعلم الآلي الخاضعة للإشراف، فهي تُستخدم على نطاق واسع في تطبيقات الاستشعار عن بعد لحل المشكلات المختلفة التي تتعلق بالتصنيف والانحدار (Abdi. 2020). كما تقوم آلة المتجهات الداعمة (SVM) بنمذجة بيانات الإدخال باستخدام وظائف النواة الخطية، أو غير الخطية، لتعيينها إلى أبعاد أعلى، ثم يتم تركيب مستوى فائق باستخدام عينات التدريب. تبحث خوارزمية (SVM) عن عينات من كل فئة في مساحة فائقة الأبعاد، وتناسب المستوى الخطي بطريقة تجعل المسافات المحددة بين المستويين هي الحد الأقصى. (Hosseiny et al., 2022). لتنفيذ تصنيف آلة المتجهات الداعمة (SVM) تم استخدام أداة (Classifier.libsvm) من مكتبة منصة جوجل الأرضي (GEE).

تقييم صحة خرائط الغطاء الأرضي:

بمجرد الانتهاء من عملية التصنيف الموجه، تم إجراء تقييم للصحة لتقييم فاعلية الطريقة التي يتم بها تعيين وحدات الخلية بشكل صحيح لفئات الغطاء الأرضي المناسبة (Rwang & Ndambuki, 2017)، كما يُعد تقييم الصحة هو أهم خطوات التصنيف، حيث يتم التحقق من صحة منتج تصنيف المخرجات وجودة البيانات نفسها، ويحدث ذلك عن طريق مقارنة وحدات الخلية الخاصة بالصورة المصنفة بالبيانات الأرضية الحقيقية (Geib et al., 2017). يتم استخدام تقييم الصحة في كل من خوارزميات التصنيف الخاضعة للإشراف وغير الخاضعة للإشراف، وفي الخوارزميات الخاضعة للإشراف المستخدمة في هذه الدراسة، وعن طريق الكود الذي تم كتابته في جافا سكريبت تم تقسيم بيانات التدريب التي تم جمعها إلى عینتين فرعيتين، الأولى بيانات تدريب الخوارزمية 70%، والأخرى بيانات الاختبار 30%، بحيث لا يتم استخدام نفس العينة أبداً لكل من التدريب والاختبار. طرق تقييم دقة التصنيف:

يُعد تقييم صحة الخرائط أمراً هاماً جداً فيما يتعلق بخرائط استخدامات الأرض، لذلك فإن الطريقة الأولى لتقييم الصحة هي إعداد البيانات المرجعية، تم تحديد (786) نقطة بيانات للتحقق من الصحة من إجمالي (2698) نقطة تم إنشاؤها ممثلة في الشكل (2)، تم اختيار هذه النقاط على مواقع مختلفة تمثل فئات مختلفة من الغطاء الأرضي. للتحقق من صحة النتائج التي تم التوصل إليها، تم إجراء مقارنة بين الصور المصنفة وبيانات العينات المرجعية باستخدام مصفوفة الخطأ ((Error matrix نظراً؛ لأنها الطريقة الأكثر شيوعاً لتقييم صحة تصنيف خرائط صور الأقمار الصناعية لجميع فئات الغطاء الأرضي. وتم ذلك من خلال مصفوفة الخطأ بحساب دقة المستخدم (UA) ودقة المنتج (PA) وحساب مؤشرات أخطاء الحذف، وأخطاء الإضافة والدقة الإجمالية (OA) ومعامل كبا (Kappa).

على النحو التالي:

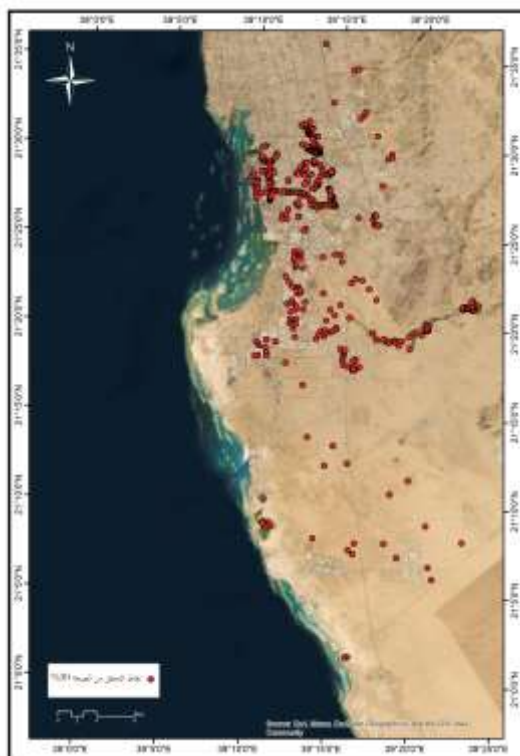
$$UA = \frac{x_{ij}}{x_j} \times 100\%$$

$$PA = \frac{x_{ij}}{x_i} \times 100\%$$

$$OA = \frac{s_d}{n} \times 100\%$$

حيث يمثل S العدد الإجمالي للخلايا المصنفة بشكل صحيح، ويمثل n العدد الإجمالي لخلايا التحقق من الصحة، ويمثل Xij مجموع النقاط الصحيحة للفئة على مجموع الصف xj في مصفوفة الارتباك، ويمثل xi المجموع الحقيقي للعمود في مصفوفة الارتباك. (Petroopoulos et al. 2015).

شكل (٢): نقاط التحقق من الصحة



المصدر: من إعداد الباحثان اعتمادًا على مرئية Sentinel-2

١- مصفوفة الخطأ (Error matrix):

من الطرق المستخدمة لتقدير صحة رسم الخرائط هي بناء مصفوفة الخطأ كنقطة انطلاق لها، وهي تُعدُّ من أكثر الطرق المستخدمة شيوعاً في تقييم تصنيف المرئيات في استخدامات الغطاء الأرضي. ويميل بعضهم إلى تسمية هذه الطريقة في تقييم التصنيف بمصفوفة الشك (Confusion matrix)؛ لأنها لا تقوم فقط بتحديد الأخطاء الكلية في تصنيف كل فئة، بل أنها تقوم أيضاً بتحديد أخطاء التصنيف بين المجموعات، وذلك نتيجة تشابه الخصائص الانعكاسية لبعض مكونات الأرض المختلفة في طبيعتها. (الغامدي، 2006). وتقوم هذه المصفوفة بعمل مقارنة لعدد من الخلايا لكل فئة من فئات الغطاء الأرضي في المرئية المصنفة، لعرض عدد الخلايا التي صُنفت بشكلٍ صحيحٍ وعدد الخلايا التي تمَّ تصنيفها بشكلٍ خاطئٍ. (Congalton, 1991) يعتمد بناء مصفوفة الخطأ على الخرائط المصنفة، والبيانات المرجعية، وبالتالي فإنَّ العينات الموجودة على البيانات المرجعية تمثل الحقيقة الأرضية (Stehman & Czaplewski, 1998).

٢- دقة المنتج (PA):

إنَّ دقة المنتج بالأصل، هي: دقة الخريطة من وجهة نظر الصانع "منتج الخريطة"، ويتمُّ حسابها على أنَّها وحدات الخلايا المحددة بشكلٍ صحيحٍ لفئةٍ معينةٍ مقسومةً على العدد الإجمالي للبكسلات الفعلية في تلك الفئة. ومن الشائع جداً التحدُّث عن نوعين من الأخطاء عند معالجة دقة التصنيف وهي: أخطاء الحذف وأخطاء الإضافة، وتُعدُّ أخطاء الحذف مكملَّةً لدقة المنتج، فهي تشير إلى وحدات الخلايا المرجعية التي تمَّ حذفها من الفئة الصحيحة في الخريطة المصنفة. خطأ الحذف = 100% - دقة المنتج. (Cardille et al., 2024).

٣- دقة المستخدم (UA):

هي: دقة الخريطة من وجهة نظر مستخدم الخريطة، ويتمُّ حسابها على أنَّها وحدات الخلايا المحددة بشكلٍ صحيحٍ لفئةٍ معينةٍ مقسومةً على العدد الإجمالي للبكسلات الموجودة بالفعل في تلك الفئة. وتُعدُّ أخطاء الإضافة مكملَّةً لدقة المستخدم، فهي تشير إلى وحدات بكسل الفئة المرجعية التي تمَّ تصنيفها عن طريق الخطأ في الخريطة. خطأ الإضافة = 100% - دقة المستخدم. (Cardille et al., 2024).

٤- الدقة الإجمالية (OA):

للتحقُّق من صحة النتائج المصنفة التي تمَّ التوصل إليها عن طريق التحليلات الإحصائية بمصفوفة الخطأ فإنَّ النتيجة بدقة 100% تدلُّ على أنَّ جميع وحدات الخلايا في الصورة قد تمَّ تصنيفها بشكلٍ صحيحٍ أثناء مرحلة التقييم. (Arantest et al., 2016).

٥- معامل كبا (Kappa):



هي طريقةٌ مستخدمةٌ على نطاق واسع للجمع بين الاتفاق بين الحقيقة الأرضية والخريطة "الموضوعية، وما يميّز استخدام معامل كابا عن الدقة الإجمالية، في أنّ حساب معامل كابا يشمل جميع عناصر مصفوفة الخطأ، وليس فقط عناصر القطر الرئيسي التي تمثل العينات صحيحة التصنيف، كما يحدث ذلك في الدقة الإجمالية. (Moreira, 2001) تتراوح قيمة معامل كابا من 0 إلى 1 كلما اقترب من 1 يمثل الاتفاق أكبر بين النقاط المرجعية والفئات المصنفة (Zafar et al., 2024) بالجدول التالي (3) تظهر التصنيفات النوعية المختلفة لقيم كابا القياسية لكوهين التي تحدد درجة الاتفاق، تمّ حساب معامل كابا باستخدام المعادلة 1.

جدول (3): عاير تصنيف إحصائيات كابا (Kappa)

قوة الاتفاق	إحصائيات كابا	
سيئ	<0.00	1
ضئيل	0.00 – 0.20	2
مقبول	0.21-0.40	3
متوسط	0.41-0.60	4
قوي	0.61-0.80	5
شبه مثالي	0.81-1.00	6

المصدر: (Santos & Lima.2018)

$$K = (1) \frac{N \sum_i^r x_{ii} - \sum_i^r x_{i+} x_{+i}}{N^2 - \sum_i^r x_{i+} x_{+i}} = 1 (x_{i+} x_{+i})$$

حيث: r = عدد الصفوف والأعمدة في مصفوفة الخطأ

X_{ii} = عدد العينات في الصف والعمود.

i, x_{i+} = الإجمالي الهامشي للصف.

i, x_{+i} = الإجمالي الهامشي للعمود.

$i, and N$ = العدد الإجمالي للعينات.

الشكل التالي يوضح هيكلية العمل رقم (٣)



شكل (3) هيكل العمل نتائج تصنيف الغطاء الأرضي:



لغرض دراسة رسم الخرائط الموضوعية للغطاء الأرضي بمحافظة جدة، وتقييم أداء خوارزميات تعلم الآلة في صحة تصنيف هذه الخرائط وبعد مقارنة النتائج ببعضها بعضاً تمّ حساب المساحة لفئات الغطاء الأرضي بمختلف تصنيفاتها كما هو واضح في جدول (4).
يتبين لنا أنّ مساحة الأراضي الجرداء كان لها النصيب الأكبر من جملة المساحة الكلية لمنطقة الدراسة البالغة (4721.09 كم²) حيث بلغت مساحة الأراضي الجرداء بالنسبة لتصنيف آلة المتجهات الداعمة (2720.89 (SVM كم²) بنسبة بلغت ((57.63%، وتعدّ هذه المساحة الأقل مقارنةً بالتصنيفات التالية، حيث بلغت مساحتها في تصنيف الغابة العشوائية (2793.93 (RF كم²) بنسبة بلغت (59.18%) أمّا خوارزمية شجرة التصنيف

والانحدار ((CART) فإنها المساحة الأكبر، حيث بلغت (3153.10 كم²) بنسبة بلغت (66.79%).

نستنتج كذلك أن مساحة المناطق الحضرية تبدو متقاربة نسبياً في جميع التصنيفات، حيث بلغت المساحة في تصنيف الغابة العشوائية (1370.20 كم²) بنسبة بلغت (29.02%)، مقارنةً بتصنيف آلة المتجهات الداعمة فإن المساحة أعلى من التصنيف السابق قليلاً بمساحة بلغت (1668.35 كم²) وبنسبة بلغت (35.34%) في حين جاءت مساحة المناطق الحضرية في شجرة التصنيف والانحدار الأقل فيهم حيث بلغت (1094.12 كم²) بنسبة بلغت (23.17%).

إن مساحة الغطاء النباتي متقاربة بشكل كبير بجميع التصنيفات، حيث بلغت المساحة في تصنيف الغابة العشوائية (472.23 كم²) بنسبة بلغت (10.00%) في حين بخوارزمية شجرة التصنيف والانحدار بلغت أقل من ذلك بمساحة (469.64 كم²) بنسبة بلغت (9.95%) أما تصنيف آلة المتجهات الداعمة، فبلغت المساحة الأقل فيهم مقارنةً بالتصنيفات السابقة، حيث بلغت المساحة (296.50 كم²) بنسبة بلغت (6.28%).

تُعد مساحة المياه الأقل في جملة المساحة الكلية لمنطقة الدراسة، حيث بلغت المساحة في تصنيف آلة المتجهات الداعمة (35.35 كم²) بنسبة بلغت (0.75%) في حين مساحة تصنيف الغابة العشوائية تُعد المساحة الأكبر مقارنةً بتصنيفات الدراسة الأخرى بمساحة بلغت (84.73 كم²) وبنسبة (1.80%) ثم تأتي المساحة الأقل في مختلف فئات الغطاء الأرضي بخوارزمية شجرة التصنيف والانحدار حيث مثلت فقط (4.23 كم²) بنسبة بلغت (0.64%).

جدول (4) مساحة الغطاء الأرضي ونسبها المئوية للخوارزميات المستخدمة في منطقة الدراسة

الفئات	نوع الصنف	خوارزمية ((RF		خوارزمية ((SVM		خوارزمية ((CART	
		النسبة %	المساحة كم ²	النسبة %	المساحة كم ²	النسبة %	المساحة كم ²
1	الأراضي الجرداء	59.18	2793.93	57.63	2720.89	66.79	3153.10
2	المناطق الحضرية	29.02	1370.20	35.34	1668.35	23.17	1094.12
3	الغطاء النباتي	10.00	472.23	6.28	296.50	9.95	469.64
4	المياه	1.80	84.73	0.75	35.35	0.09	4.23
	المجموع	100	4721.09	100	4721.09	100	4721.09

المصدر: من إعداد الباحثان اعتماداً على مرئية 2-Sentniel

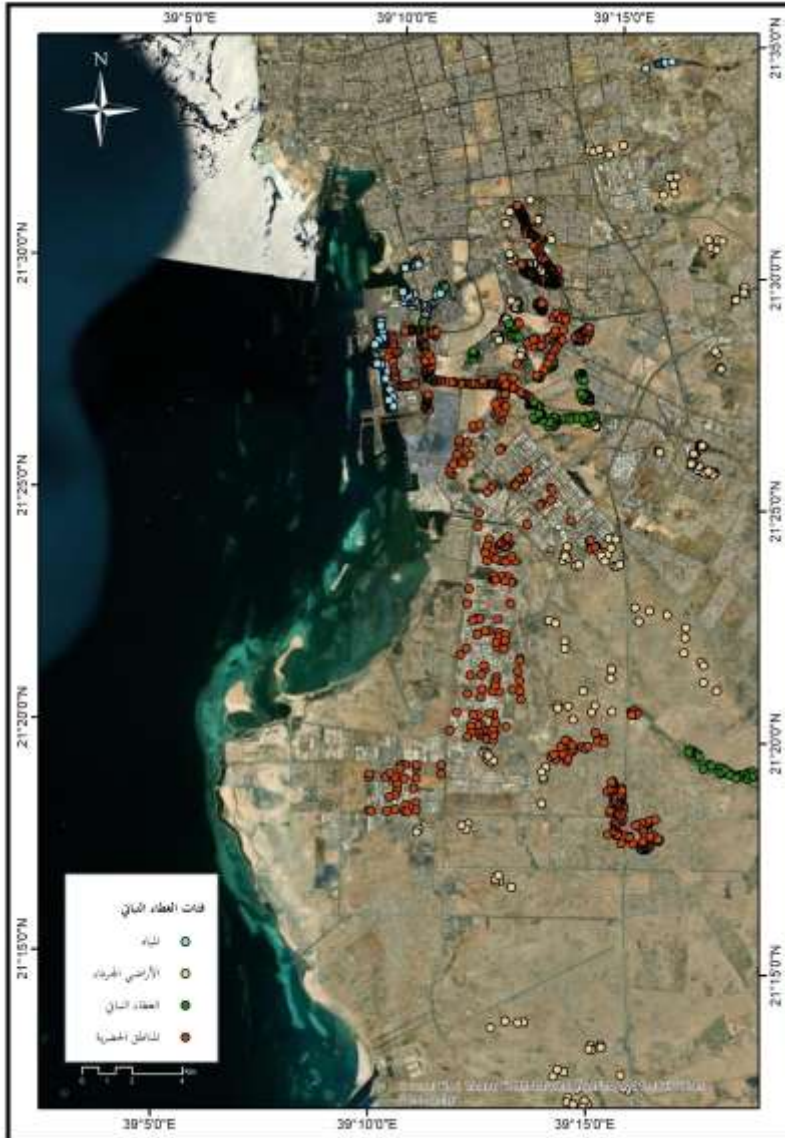


تقييم نتائج تصنيف خوارزميات التعلم الآلي:

يوضح الشكل (4) جزء من نقاط العينات التي تمّ انشاؤها بواسطة ((GEE وبلغ عددها (2698) تضمنت (219) عينةً من المياه، و(763) عينةً من الأرض الجرداء، و(716) عينةً من الغطاء النباتي، و(1000) عينةً من المناطق الحضرية، كما يوضح شكل (5) نقاط بيانات التدريب التي بلغ عددها (1757)، و(786) نقطة بياناتٍ للتحقق من الصحة التي تمّ اختيارها آلياً بواسطة منصة محرك جوجل الأرضي (GEE)، نلاحظ الاختلاف في عدد نقاط العينات (2698) ومجموع نقاط بيانات التدريب وبيانات التحقق من الصحة التي بلغ مجموعها معاً (2543)، وبالتالي لدينا (155) نقاط عينات مفقودة؛ يعود السبب في ذلك عند استخدام دالة (sampleRegions)) لم تتمكن من استخراج بيانات طيفية لها، وبالتالي تمت إزالتها من مجموعة البيانات، ويعدّ ذلك أمر طبيعي في معالجة بيانات الاستشعار عن بعد بسبب التباين في جودة البيانات ومدى التغطية الطيفية، وتمّ فقد جميع هذه البيانات من فئة المياه حيث تم إزالة (155) عينة منها لتصبح فقط (64) عينة لذلك نلاحظ بفصل النتائج لاحقاً تباين في نتيجة المياه.

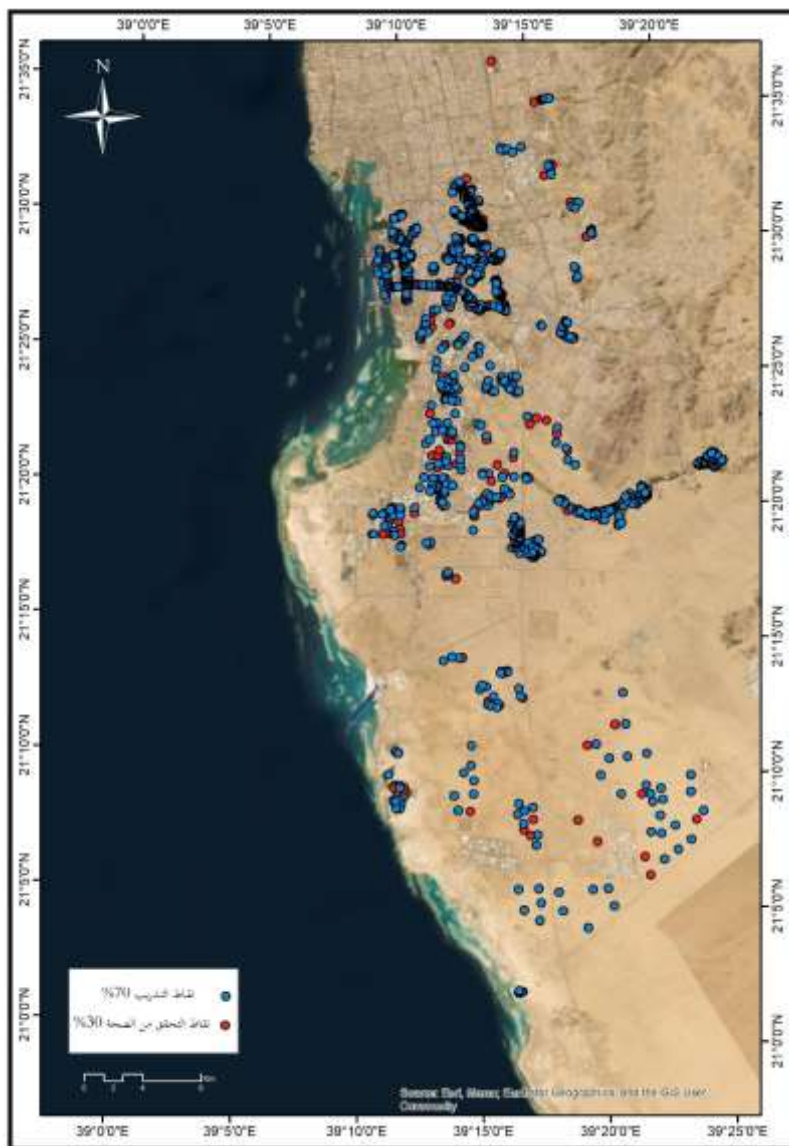
تظهر صور الغطاء الأرضي لمحافظة جدة المستمدة من تصنيفات آلة المتجهات الداعمة وشجرة التصنيف والانحدار والغابة العشوائية في الأشكال (8-6)، بينما تمثل جداول (7-5) مصفوفة الخطأ التي تتضمن أربع فئات: المياه، الأراضي الجرداء، الغطاء النباتي، المناطق الحضرية. يوضح جدول (8) الدقة الإجمالية (OA) وقيم معامل كبا (Kappa) لكل خوارزمية مستخدمة في هذه الدراسة.

شكل (٤): عينات منطقة الدراسة



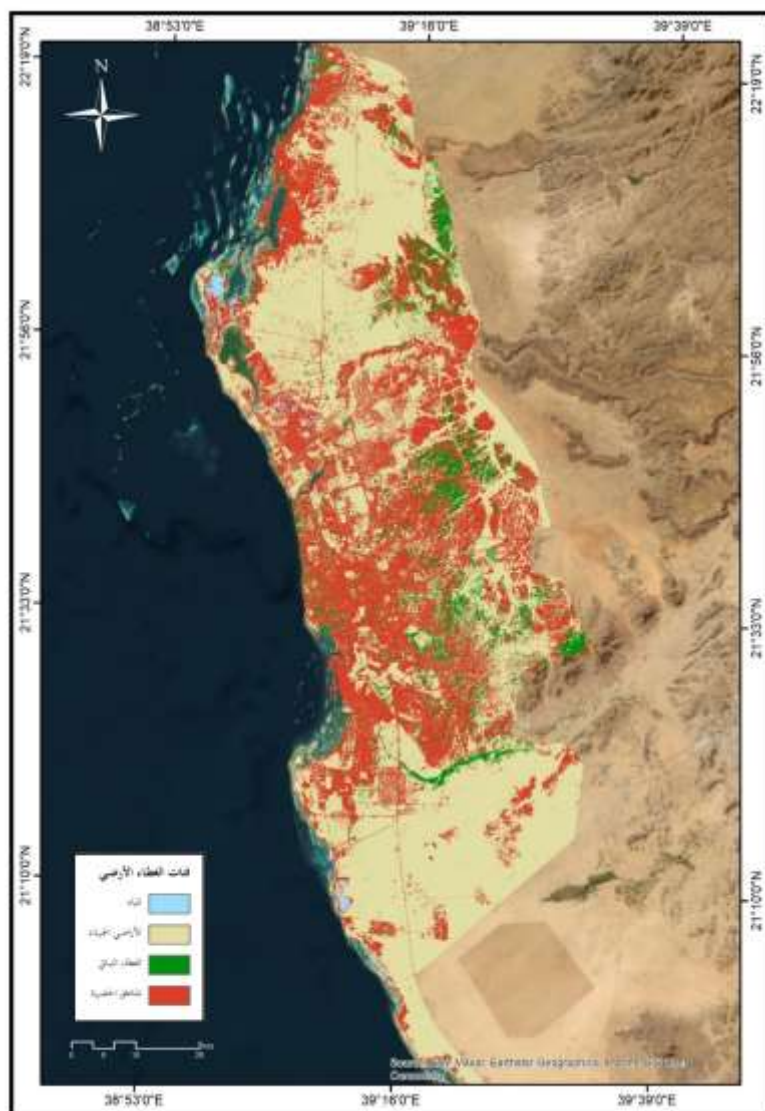
المصدر: من إعداد الباحثان اعتماداً على مرئية Sentinel-2

شكل (٥): نقاط بيانات التدريب والتحقق من الصحة



المصدر: من إعداد الباحثان اعتماداً على مرئية Sentinel-2

خوارزمية آلة المتجهات الدّاعمة (SVM)
شكل (6): تصنيف آلة المتجهات الدّاعمة (SVM) للغطاء الأرضي لمحافظة جدة



المصدر: من إعداد الباحثان اعتماداً على مرئية Sentinel-2

يوضح جدول (5) مصفوفة الخطأ لخوارزمية آلة المتجهات الداعمة (SVM) حيث العينات المصنفة بشكل صحيح والمصنفة بشكل خاطئ حسب خوارزمية آلة المتجهات الداعمة (SVM) في مصفوفة الخطأ.

كانت الدقة الإجمالية لخريطة الغطاء الأرضي (86.7%)، وتتمتع فئة المناطق الحضرية بدقة أعلى في التصنيف حيث بلغت (95.6%) في دقة المنتج و (78.8%) في دقة المستخدم، تليها فئة الغطاء النباتي حيث بلغت دقة المنتج (88.4%) في حين مثلت دقة المستخدم لنفس الفئة أعلى دقة بلغت (98.5%)، فيما ظهرت نسبة دقة المنتج لفئة الأراضي الجرداء (82.2%) ودقة المستخدم (88.8%)، وجاءت فئة المياه أقل فئات التصنيف على مستوى دقة المنتج والمستخدم كليهما (0%).

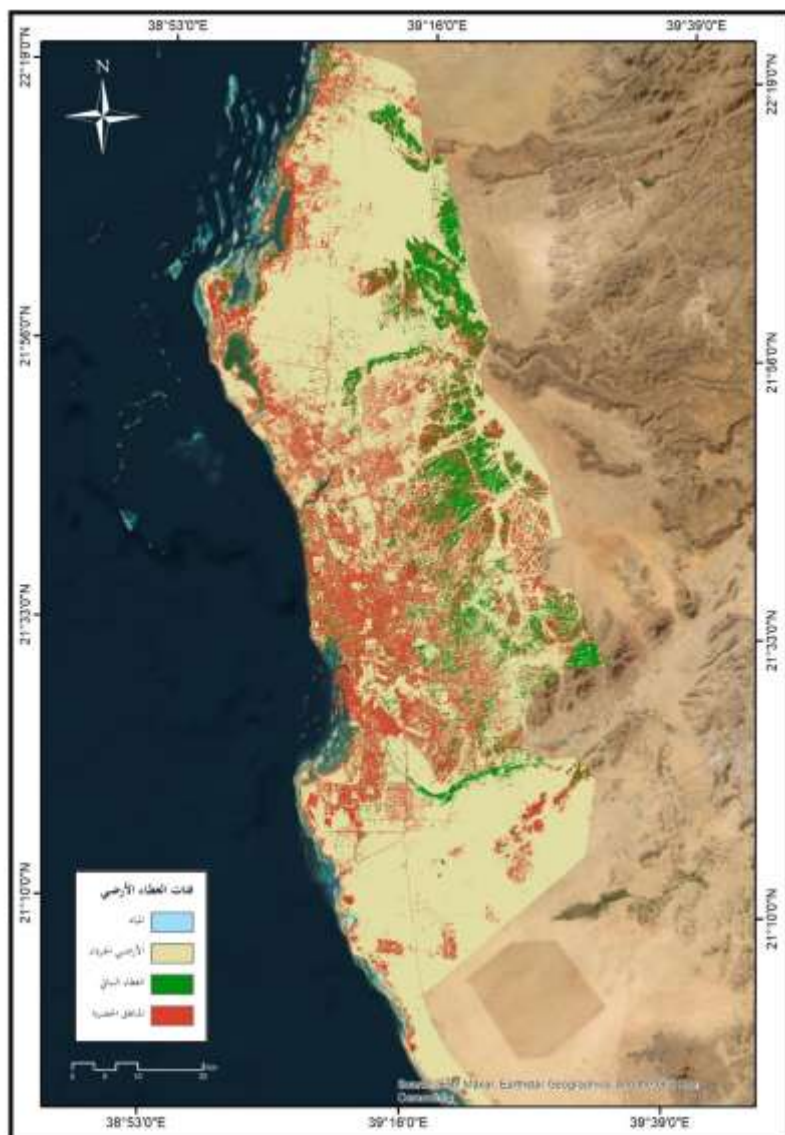
نلاحظ بهذا التصنيف أنه مناسب جداً لفئة الغطاء النباتي من ناحية دقة المنتج حيث كان أعلى دقة مقارنة ببقية الفئات، وكذلك المناطق الحضرية تعد مناسبة حيث ظهرت فيه دقة المنتج أعلى من بقية الفئات.

جدول (5) مصفوفة الخطأ لخوارزمية آلة المتجهات الداعمة (SVM)

فئات الغطاء الأرضي	المياه	أرض جرداء	غطاء نباتي	مناطق حضرية	المجموع	دقة المستخدم %	خطأ الإضافة %
المياه	0	0	1	21	22	0	100
الأراضي الجرداء	0	199	2	41	242	88.8	11.2
الغطاء النباتي	0	12	199	14	225	98.5	1.5
المناطق الحضرية	0	13	0	284	297	78.8	21.2
المجموع	0	224	202	360	786		
دقة المنتج %	0	82.2	88.4	95.6			
خطأ الحذف %	100	17.8	11.6	4.4			

المصدر: من إعداد الباحثان اعتماداً على مرئية 2-Sentinel

خوارزمية شجرة التصنيف والانحدار: (CART)
شكل: (7) شجرة التصنيف والانحدار (CART) للغطاء الأرضي لمحافظة جدة



المصدر: من إعداد الباحثان اعتمادًا على مرئية 2-Sentinel

بناءً على جدول (6) مصفوفة الخطأ لخوارزمية شجرة التصنيف والانحدار (CART) ظهرت نسبة الدقة الإجمالية لخوارزمية شجرة التصنيف والانحدار (88.8%)، ونلاحظ بنتائج التصنيف أن فئتي الأراضي الجرداء والغطاء النباتي أعلى الفئات في نتيجة دقة المستخدم، حيث جاءت نتيجة الأراضي الجرداء (90%) والغطاء النباتي (89.4%) فيما كانت النسبة الأعلى في دقة المنتج للغطاء النباتي (90.6%)، وبلغت دقة المنتج لفئة المناطق الحضرية (89.8%)، ومن ثم الأراضي الجرداء (87.1%)، ومن حيث دقة المستخدم فهي تأتي بعد فئتي الغطاء النباتي والمناطق الحضرية بدقة تصل (86.3%)، تأتي بعد ذلك الأقل فيهم وهي فئة المياه، حيث تباينت بدرجة متفاوتة نتائج دقة المنتج مقارنةً بدقة المستخدم، فإن نسبة دقة المنتج بلغت (81.8%) فهي ذات دقة إنتاج منخفضة بلغت الأقل فيهم مقارنةً ببقية الفئات، فيما كانت دقة المستخدم (75%)، وهي كذلك تُعد ذات دقة منخفضة مقارنةً ببقية الفئات.

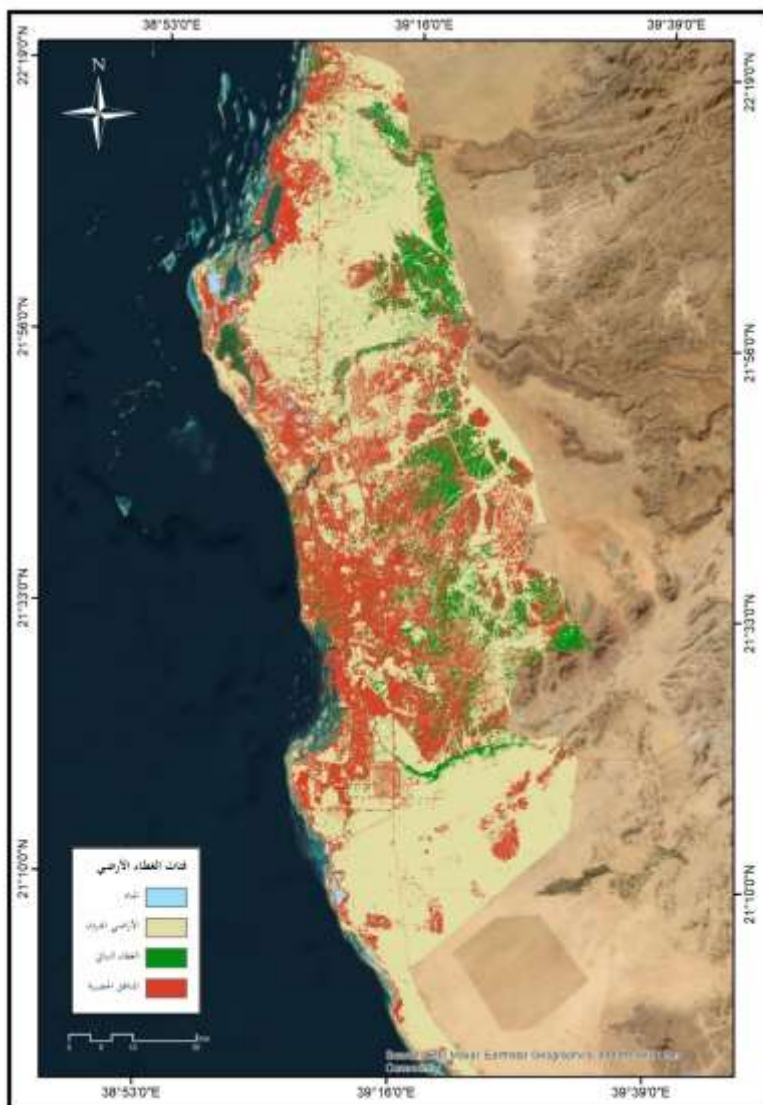
نلاحظ أن شجرة التصنيف والانحدار (CART) غير مناسبة كثيراً للتعامل مع تصنيف المياه، حيث تباينت فيها دقة النتائج مقارنةً ببقية فئات الغطاء الأرضي.

جدول (6): مصفوفة الخطأ لخوارزمية شجرة التصنيف والانحدار (CART).

فئات الغطاء الأرضي	المياه	أرض جرداء	غطاء نباتي	مناطق حضرية	المجموع	دقة المستخدم %	خطأ الإضافة %
المياه	18	0	3	1	22	75	25
الأراضي الجرداء	2	209	10	21	242	90	10
الغطاء النباتي	1	7	204	13	225	89.4	10.6
المناطق الحضرية	3	16	11	267	297	88.4	11.6
المجموع	24	232	228	302	786		
دقة المنتج %	81.8	86.3	90.6	89.8			
خطأ الحذف %	18.2	13.7	9.4	10.2			

المصدر: من إعداد الباحثان اعتماداً على مرئية 2-Sentniel

خوارزمية الغابة العشوائية (RF)
شكل (8): تصنيف الغابة العشوائية (RF) للغطاء الأرضي لمحافظة جدة



المصدر: من إعداد الباحثان اعتماداً على مرئية 2-Sentinel

جدول (7): مصفوفة الخطأ لخوارزمية الغابة العشوائية (RF).

فئات الغطاء الأرضي	المياه	أرض جرداء	غطاء نباتي	مناطق حضرية	المجموع	دقة المستخدم %	خطأ الإضافة %
المياه	19	1	2	0	22	95	5
الأراضي الجرداء	1	223	3	15	242	93.3	6.7
الغطاء النباتي	0	10	213	2	225	95.5	4.5
المناطق الحضرية	0	5	4	288	297	94.4	5.6
المجموع	20	239	222	305	786		
دقة المنتج %	86.3	92.1	94.6	96.9			
خطأ الحذف %	13.7	7.9	5.4	3.1			

المصدر: من إعداد الباحثان اعتماداً على مرئية 2-Sentniel

تميّزت خوارزمية الغابة العشوائية بالنتائج ذات الدقة العالية حيث تفوّقت على خوارزمية آلة المتجهات الدّاعمة وخوارزمية أشجار التصنيف والانحدار بدقّة إجمالية بلغت (94.5%) مقابل ذلك، فإنّ أداء دقة المستخدم ودقة المنتج ظهرت بنسبٍ عاليةٍ لا تقل عن (93%) باستثناء دقة المنتج لفئة المياه حيث بلغت (86.3%).

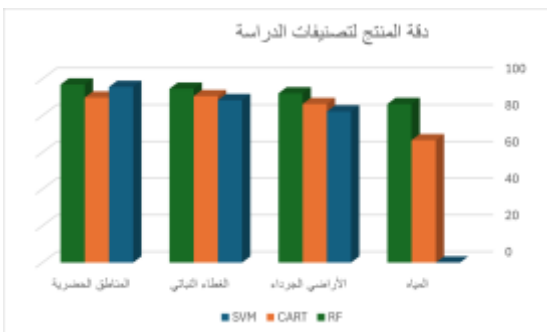
جدول (8): الدقة الإجمالية وقيم معامل كبا. (Kappa)

تصنيف الخوارزمية	الدقة الإجمالية %	معامل كبا
آلة المتجهات الدّاعمة (SVM)	86.7	80.1
شجرة التصنيف والانحدار (CART)	88.8	83.5
الغابة العشوائية (RF)	94.5	91.9

مقارنة نتائج تصنيف خوارزميات التعلم الآلي:

يتضح لنا من خلال تقييم نتائج تصنيف الخوارزميات السابقة، ومُتمثلةً كذلك في الجدول التالي (8) أنّ خوارزمية الغابة العشوائية (RF) امتازت بشكلٍ كبيرٍ عن تصنيف خوارزمية آلة المتجهات الدّاعمة وشجرة التصنيف والانحدار. حيث كانت أكثرهم دقةً. بلغت الدقة الإجمالية (94.5%) (قيمة كبا = 91.9)، بينما تتبعها خوارزمية شجرة التصنيف والانحدار (CART) بنسبة (88.8%) (قيمة كبا = 83.5) نلاحظ أنّ هذه الخوارزمية غير مناسبةٍ كثيرًا للتعامل مع تصنيف المياه، حيث تباينت فيها دقة النتائج مقارنةً ببقية فئات الغطاء الأرضي، بلغت دقة المستخدم لفئة المياه (75%) ودقة المنتج ((81.8%)، وتحتل خوارزمية آلة المتجهات الدّاعمة (SVM) المركز الأخير بدقّة إجمالية تبلغ (86.7%) (قيمة كبا = 80.1) كانت فيها فنتي الغطاء النباتي والمناطق الحضرية أعلى دقة من بقية الفئات، فيما جاءت فئة المياه الأقل دقةً حيث ظهرت الدقة (0%)، وتمّ تمثيل نتائج دقة المنتج والمستخدم في الشكل البياني. (9-10)

شكل (9): دقة المنتج.



شكل (10): دقة المستخدم.



المصدر: من إعداد الباحثان اعتمادًا على جدول الدقة الإجمالية وقيم معامل كابا. (Kappa)
المصدر: من إعداد الباحثان اعتمادًا على مرئية Sentinel-2
الخاتمة:

في ختام هذه الدراسة لا يدعني إلا أن أقول إنه قد تحقق فيها العديد من الأهداف، ممثلة أولاً في استخدام منصة محرك جوجل الأرض ((GEE ومن خلال لغة البرمجة جافا سكريبت تم إنتاج الخرائط الموضوعية للغطاء الأرضي لمنطقة الدراسة من بيانات القمر الصناعي سنيتل-2 والحصول عليها بدقة عالية، ومن ثم إجراء التصنيف الموجه باستخدام الخوارزميات الثلاثة: الغابة العشوائية ((RF، آلة المتجهات الداعمة (SVM)، أشجار التصنيف والانحدار (CART). وتقييم صحة النتائج عن طريق اختبار معامل كابا (Kappa Coefficient) والدقة الإجمالية (overall Accuracy). بواسطة دقة المنتج (Producer Accuracy)، ودقة المستخدم (User Accuracy)، ومقارنة نتائج التصنيف لهذه الخوارزميات، حيث تميزت بشكل واضح خوارزمية الغابة العشوائية (RF) بدقتها الإجمالية العالية ودقة جميع فئات الغطاء الأرضي بما لا يقل عن (3.86%).
أبرز النتائج:

سهل محرك جوجل الأرض ((GEE بشكل كبير في عملية التصنيف، حيث تم في هذه الدراسة استخدام ثلاثة طرق تصنيف من خوارزميات تعلم الآلة لتصنيف الغطاء الأرضي لمحافظة جدة بهدف تقييم صحة أداء الخوارزميات المستخدمة، ومقارنة النتائج ببعضها بعضاً، باستخدام صور سنيتل-2 بدقة مكانية تبلغ (10) أمتار.

تم تنفيذ عملية التصنيف باستخدام خوارزميات ((CART, SVM, RF لتصنيف أربع فئات من الغطاء الأرضي: المياه، الأراضي الجرداء، الغطاء النباتي، المناطق الحضرية، ثم بعد ذلك تم استخدام الدقة الإجمالية ومعامل كابا لتقييم أداء الخوارزميات، ومقارنة نتائج

طرق التصنيف الثلاثة. ووفقاً لمصفوفة الخطأ كانت الدقة الإجمالية لتصنيف (RF) أفضل من تصنيف (SVM) و (CART). في حين كان تصنيف (SVM) هو الأقل دقة. كذلك بالنسبة لإحصاءات كبا كانت قيم الدقة لخوارزمية أيضاً كالتالي: 91.9% (94.5) 80.1% (86.7) 83.5% (88.8) على التوالي.

التوصيات:

١- نظراً للتعامل مع المحافظة بشكل كامل كان هناك وضوح كبير للخلية الواحدة ذات وضوح مكاني عشرة متر، ولكن عند الدخول لتفاصيل الأحياء كان هناك ضعف في الوضوح؛ لهذا توصي الدراسة بالاعتماد على الأقمار الصناعية ذات الوضوح المكاني المناسب، مع الأخذ بالحسبان عند إجراء عملية تصنيف الغطاء الأرضي: حجم منطقة الدراسة، فعند التعامل مع المدينة ربما تكون الأقمار الصناعية ذات الوضوح المكاني (10 أمتار) مناسبة لحجمها، لكن عند التعامل مع الأحياء، أو الشوارع، أو الطرق يجب أن يكون الوضوح المكاني للقمر الصناعي من متر إلى مترين.

٢- إن تصنيف آلة الغابات العشوائية (RF) قدمت نتائج أكثر دقة بالنسبة لفئة المناطق الحضرية بالتالي توصي الدراسة باستخدام هذا التصنيف إذا كانت منطقة الدراسة ذات كثلة عمرانية.

٣- حث الجهات المعنية بإنتاج خرائط للمدن محدثة بشكل سنوي، وذلك للتغيرات التي تطرأ على غطاء الأرض بناءً على التطور الكبير الذي تشهده المملكة العربية السعودية في كثير من مجالات التنمية الحضرية، وهذا يؤثر بشكل واضح على التغير في الغطاء الأرضي، لذلك يتوجب على الجهات الحكومية مواكبة هذه التغيرات وتحديث خرائط المدن بشكل مستمر.

المراجع:

المراجع العربية:

الغامدي، سعد أبو راس. (٢٠٠٦). تصنيف استخدامات الأراضي في مدينة مكة المكرمة .
المجلة الجغرافية العربية، س٣٨، ع٤٧، ٣٣-٥٥. مسترجع من

<http://search.mandumah.com/Record/1348818>

المراجع الأجنبية:

Arantes, L. T.; Arantes, B. H. T.; Tondato, C. (2016). *Analise multitemporal para o monitoramento de desmatamento através da image landsat-5*. Revista do Agronegocio, v.5, p. 14-23.

Basheer. S., Wang. X., Farooque. A.A., Nawaz. R.A., Liu. K., Adekanmbi. T., Liu. S. (2022) *Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques*. Remote Sens. 2022, 14, 4978.

<https://doi.org/10.3390/rs14194978>

Breiman, L. (2001). *Random Forests*. Machine Learning, 45, 5-32, 2001. Manufactured in The Netherlands.

Cardille, J., Crowley, M., Saah, D., & Clinton, N. (2024). "*Cloud-Based Remote Sensing with Google Earth Engine*". One Edition. Springer Cham.

Congalton, R. G., Green, K. (1991). *A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data*. Remote Sensing of Environment, 37, 35-46.

[http://dx.doi.org/10.1016/0034-4257\(91\)90048-B](http://dx.doi.org/10.1016/0034-4257(91)90048-B)

Congalton, R. G., Green, K. (2009). "*Assessing the Accuracy of Remotely Sensed Data*". Second Edition. CRC press, Taylor & Francis Group. New York.

Congalton, R. G. (2005). *Thematic and Positional Accuracy Assessment of Digital Remotely Sensed Data*. Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium.

Geib, C., P. Aravena Pelizari, H. Schrade, A. Brenning, and H. Taubenbock. (2017). *On the Effect of Spatially Non-Disjoint Training and Test Samples on Estimated Model Generalization*

- Capabilities in Supervised Classification With Spatial Features.** *IEEE Geoscience and Remote Sensing Letters* 14 (11): 2008-12.
- Ghayour, L., Neshat, A., Paryani, S., Shahabi, H., Shirzadi, A., Chen, W., Al-Ansari, N., Geertsema, M., Amiri, M., Gholamnia, M., et al. (2021). **Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms.** *Remote Sens.* 2021, 13, 1349. <https://doi.org/10.3390/rs13071349>
- Hosseiny, B., Abdi, A., Jamali, S. (2022). **Urban land use and land cover classification with interpretable machine learning – A case study using Sentinel-2 and auxiliary data.** *Remote Sensing Applications: Society and Environment* 28 (2022) 100843. <https://doi.org/10.1016/j.rsase.2022.100843>
- Lary, J. D., Alavi, H. A., Gandomi, H. A., Walker, L. A. (2016). **Machine learning in geosciences and remote sensing.** *Geoscience Frontiers*, 7(1), 3-10.
- Moreira, M. A. (2001). **"Fundamentos de Sensoriamento Remoto e Metodologias de Aplicação"**. 1 ed. INPE, São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE.)
- Petropoulos, G., Kalivas, D., Georgopoulou, I., Srivastava, P. (2015) **Urban vegetation cover extraction from hyperspectral imagery and geographic information system spatial analysis techniques: case of Athens, Greece.** *J Appl Remote Sens* 9(1). <https://doi.org/10.1117/1.JRS.9.096088>
- Rwanga, S. S., Ndambuki, J. M. (2017) **Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS.** *International Journal of Geosciences*, 8, 611-622. <https://doi.org/10.4236/ijg.2017.84033>
- Shetty, S. (2019) **Analysis of Machine Learning Classifiers for LULC Classification on Google Earth Engine.** Master's Thesis, University of Twente, Enschede, The Netherlands.

- Stehman, S.V., Czaplewski, R. L. (1998). ***Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles.*** Remote Sens. Environ, Elsevier Science Inc, New York.
- Wahap. A. N., Shafri. H. Z. (2020). ***Utilization of Google Earth Engine (GEE) for land cover monitoring over Klang Valley, Malaysia.*** Earth and Environmental Scienc. <https://doi:10.1088/1755-1315/540/1/012003>
- Zafar, Z., Zubair, M., Zha, Y., Fahd, Sh., Nadeem, A. (2024). ***Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data.*** The Egyptian Journal of Remote Sensing and Space Sciences 27(2), 216-226. <https://doi.org/10.1016/j.ejrs.2024.03.003>
- Zhang, Y., Zhao, H., Ni, J. (2020). ***A Comparison of Four Algorithms for Land-Use Classification Based on Landsat 8 OLI Image.*** Journal of Physics: Conference Series 1631 (2020) 012075 doi:10.1088/1742-6596/1631/1/012075

المصادر الإلكترونية:

امانة محافظة جدة: <https://www.jeddah.gov.sa/Jeddah/About/index.php>

NASA, Landsat Science

<https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-mission-details/>

Esa, Sentinel Online

<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>

