Journal of Libyan Academy Bani Walid 2026
مجلة الأكاديمية الليبية بني وليد
e-ISSN: 3104-3860
Volume 2, Issue 1, 2026, Pages: 248-269
Website: https://journals.labjournal.ly/index.php/Jlabw/index

A Comparative Study for Enhancing Classification Accuracy Using (SVM) Models: An Applied Study

Hussein Mohammed Al-Mahdi Al-Sharif *
Department of Statistics, Faculty of Science, Sebha University, Libya
*Email: hussien@gmail.com

دراسة مقارنة لتحسين دقة التصنيف باستخدام نماذج *SVM* (دراسة نطبيقية)

حسين محمد المهدي الشريف *
قسم الإحصاء، كلية العلوم، جامعة سبها، ليبيا

Received: 08-12-2025	**Accepted:** 15-01-2026	**Published**: 29-01-2026
Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract

This study aimed to evaluate the performance of Support Vector Machine (SVM) models in data classification, examining the effect of the type of kernel function and the nature of the response on classification accuracy, using the statistical software ORANGE. Three different performance evaluation methods were applied: Test On Train Data, Cross Validation, and Random Sampling. The results indicated that classification accuracy in SVM models varies depending on the kernel function used and the response type. For binary-response models, the Radial Basis Function (RBF) achieved the highest classification accuracy compared to Polynomial and Sigmoid functions, with the evaluation methods ranked as follows: Test On Train Data (highest), followed by Cross Validation, and Random Sampling (lowest). In multi-response models, the highest accuracy was obtained using the Polynomial function compared to RBF and Sigmoid, maintaining the same ranking of evaluation methods.

Based on these findings, the study recommends adopting multi-response SVM models in future statistical studies involving multi-response data due to their high classification accuracy. It also emphasizes the use of a confusion matrix as an effective tool to evaluate model performance by comparing correctly and incorrectly classified positive and negative cases. Additionally, the study highlights the importance of enhancing healthcare services for heart disease patients by providing specialized hospitals and medical centers and ensuring access to necessary treatment, alongside supporting and encouraging research and studies related to cardiovascular diseases to improve therapeutic outcomes and deepen understanding of the condition.

Keywords: Binary-response Support Vector Machine – Multi-response Support Vector Machine – Radial Basis Function (RBF) – Polynomial Function – Sigmoid Function .

الملخص

هدفت هذه الدراسة إلى تقييم أداء نماذج آلة المتجه الداعم (SVM) في تصنيف البيانات، مع مقارنة تأثير نوع الدالة المستخدمة وطبيعة الاستجابة على دقة التصنيف، باستخدام البرنامج الإحصائي ORANGE. تم تطبيق ثلاث طرق مختلفة لتقييم الأداء: Test On Train Data، وCross Validation، و Random Sampling. أظهرت نتائج الدراسة، أن دقة التصنيف في نماذج آلة المتجه الداعم (SVM) تختلف باختلاف نوع الدالة المستخدمة وطبيعة الاستجابة. فقد بينت النماذج ثنائية الاستجابة أن استخدام دالة الإشعاع (RBF) حقق أعلى دقة تصنيف مقارنة بالدوال متعددة الحدود (Polynomial) وSigmoid، مع ترتيب طرق التقييم من حيث الدقة كالتالي :

Test On Train Data الأعلى، يليه Cross Validation، ثم Random Sampling الأقل دقة. أما في النماذج متعددة الاستجابة، فقد كانت الدقة الأعلى باستخدام الدالة متعددة الحدود (Polynomial) مقارنة بدالة الإشعاع (RBF) وSigmoid، مع الحفاظ على نفس ترتيب طرق التقييم. وبناءً على هذه النتائج، توصي الدراسة بالاعتماد على نماذج SVM متعددة الاستجابة في الدراسات الإحصائية المستقبلية الخاصة بالبيانات متعددة الاستجابة نظرًا لدقتها العالية، مع استخدام مصفوفة التشويش كأداة فعّالة لتقييم أداء النموذج من خلال مقارنة الحالات الإيجابية والسلبية المصنفة بشكل صحيح أو غير صحيح. كما أكدت الدراسة على أهمية تعزيز الرعاية الصحية لمرضى القلب عبر توفير المستشفيات والمراكز الطبية المتخصصة، وضمان توفير العلاج اللازم لهم، بالإضافة إلى دعم وتشجيع الدراسات والبحوث المتعلقة بأمراض القلب لتحسين النتائج العلاجية وفهم أفضل للمرض.

الكلمات المفتاحية: آلة المتجه الداعم ثنائي الاستجابة – آلة المتجه الداعم متعدد الاستجابة – دالة الإشعاع RBF -الدالة متعددة الحدود Polynomial - دالة Sigmoid .

المقدمة

تُعتبر نماذج آلة المتجه الداعم (SVM) إحدى أبرز تقنيات التعلم الآلي (Machine Learning) المشرف، نظراً لقدرتها العالية على مواجهة تحديات تصنيف البيانات الكبيرة والمعقدة وتحقيق دقة مرتفعة مقارنة بالطرق التقليدية مثل الأشجار القرار أو الشبكات العصبية البسيطة ، تعتمد SVM على مفهوم الحدود الفاصلة (Hyperplanes) التي تُمكِّن من فصل الفئات في فضاء متعدد الأبعاد، ويمكن تحسين أدائها عبر استخدام الدوال النواة (Kernel Functions) التي تحول البيانات إلى فضاءات أعلى قابلية للفصل بشكل أكثر فاعلية (Okfalisa et al., 2021).

منذ اقتراحها الأصلي بواسطة Vapnik وزملائه، تستند SVM إلى نظرية التعلم الإحصائي (Statistical Learning Theory) لتحقيق أفضل توازن بين تعميم النموذج والدقة، سواء في مهام الانحدار أو التصنيف.

يمكن تصنيف نماذج SVM وفقًا لاستخدامها إلى نوعين رئيسيين: (Sholihah et al., 2023).

1. ***آلة المتجه الداعم للانحدار (Support Vector Machine for Regression-SVR)،***
2. ***آلة المتجه الداعم للتصنيف (Support Vector Machine for Classification-SVC).***

كما أن اختيار نوع الدالة النواة (مثل RBF, Polynomial, Sigmoid) وطبيعة الاستجابة (ثنائية أو متعددة) يُعدّ من العوامل الحاسمة التي تؤثر في أداء النموذج وتحديد دقته. وقد أظهرت دراسات حديثة أن الأداء يختلف بشكل ملحوظ بين هذه الدوال: فبعض التطبيقات مثل التشخيص الطبي تفضّل دالة RBF، بينما في حالات أخرى مثل تصنيف بيانات معقدة متعددة الاستجابة قد تعطي دالة Polynomial أداءً أفضل. **(سالم،2025)**

لهذا، تهدف هذه الدراسة إلى تحليل أداء نماذج SVM في تصنيف البيانات مع مقارنة تأثير نوع النواة (RBF، متعددة الحدود Polynomial، Sigmoid وطبيعة الاستجابة (ثنائية مقابل متعددة) على دقة التصنيف. كما تسعى إلى تقديم توصيات عملية لتطبيق هذه النماذج في الأبحاث الإحصائية المستقبلية، واستخدام مصفوفة التشويش (Confusion Matrix) كأداة موثوقة لتقييم الأداء عن طريق مقارنة عدد الحالات الإيجابية والسلبية المصنّفة بشكل صحيح أو خاطئ. هذا البحث يُمكن أن يسهم في تحسين تطبيقات SVM في المجالات التطبيقية، وخصوصاً في الرعاية الصحية وأمراض القلب، من خلال تقديم نموذج تصنيف دقيق وموثوق يُساعد في تعزيز جودة التشخيص واتخاذ القرار العلاجي.

دراسات سابقة

1- El Maouaki, W., Said, T., & Bennai, M. (2024). *Quantum Support Vector Machine for Prostate Cancer Detection: A Performance Analysis.*

هدفت الدراسة إلى تقييم فعالية نموذج Quantum Support Vector Machine (QSVM) في تحسين دقة تشخيص سرطان البروستاتا، مقارنة بالنموذج التقليدي SVM. اعتمد البحث على معالجة البيانات الطبية المتعلقة بسرطان البروستاتا باستخدام خريطة ميزات كمومية (Quantum Feature Map)، والتي تقوم بتحويل البيانات الأصلية إلى فضاء عالي الأبعاد يسمح للنموذج بالتقاط الأنماط غير الخطية والمعقدة في البيانات بشكل أكثر كفاءة من النماذج الكلاسيكية.

تم تقييم أداء النموذج باستخدام مؤشرات إحصائية قياسية في التصنيف الطبي، شملت: الحساسية (Sensitivity)، الدقة (Accuracy)، ومقياس F1 (F1-Score). أظهرت النتائج أن نموذج QSVM تفوق على SVM التقليدي في جميع هذه المؤشرات، مما يعكس قدرته الأعلى على التمييز بين الحالات المصابة وغير المصابة بسرطان البروستاتا. وعزت الدراسة هذا التفوق إلى قدرة الخريطة الكمومية على التعامل مع العلاقات غير الخطية بين المتغيرات، وتقليل احتمال حدوث أخطاء التصنيف المرتبطة بالنماذج الخطية أو شبه الخطية. تشير نتائج هذه الدراسة إلى أن دمج الأساليب الكمومية مع تقنيات SVM التقليدية يوفر أداة واعدة لتحسين الأداء التشخيصي في التطبيقات الطبية، خاصة عند التعامل مع مجموعات بيانات كبيرة ومعقدة، حيث يمكن للنموذج الكمومي تحسين فعالية الاكتشاف المبكر للأمراض الحيوية وتقليل معدلات الخطأ في التشخيص. كما توصي الدراسة بمواصلة تطوير QSVM واستكشاف إمكانياته في مجالات أخرى للتصنيف الطبي وتحليل البيانات الحيوية المعقدة.

2- Halagundegowda, G. R., Singh, A., T. L., M. K., & Naveena, K. (2023). Evaluation of Classification Ability of Support Vector Machine (SVM) in Binary Classification Problems.

تهدف دراسة Halagundegowda et al. (2023) إلى تقييم قدرة نماذج Support Vector Machine (SVM) في تصنيف البيانات ذات الاستجابة الثنائية (binary classification)، مع تحليل تأثير نوع دالة النواة (kernel function) على أداء النموذج. شملت الدوال المستخدمة في الدراسة Radial Basis Function (RBF) وSigmoid، وتم استخدام تحسين المعاملات عبر Grid Search لضبط المعاملات الحرجة للنموذج، مثل معامل الكفاءة (C) ومعامل النواة (gamma)، بهدف الوصول إلى أفضل أداء ممكن.

أظهرت النتائج أن اختيار نوع kernel كان له تأثير ملحوظ على دقة النموذج، حيث حققت دالة RBF أعلى دقة تصنيف مقارنة بدالة Sigmoid عبر مجموعات البيانات المستخدمة. كما أظهرت الدراسة أن تحسين المعاملات عبر Grid Search أسهم بشكل كبير في تعزيز الأداء، سواء من حيث الدقة (Accuracy) أو الحساسية (Sensitivity) ومقياس F1-Score، ما يعكس قدرة SVM على التكيف مع خصائص البيانات وتحقيق فصل أمثل بين الفئات الثنائية. توصلت الدراسة إلى أن SVM مع النواة RBF، مع ضبط المعاملات بشكل مناسب، يمثل نموذجًا فعالًا وموثوقًا للتصنيف الثنائي في التطبيقات الإحصائية،

ويمكن توظيفه في مختلف المجالات العملية مثل التشخيص الطبي، تصنيف المخاطر، وتحليل البيانات المالية. كما توصي الدراسة بالاعتماد على تقنيات ضبط المعاملات المنهجية لتحسين دقة النماذج وتقليل الأخطاء التصنيفية.

3- جاب الله, & محمود أسامة مصطفى السيد. (2024). استخدام تقنية SVM في رصد وتقييم دور الإنسان كعامل جيومورفولوجي على سواحل بحيرة المنزلة خلال الفترة الممتدة بين عامي (1984م-2024م) دراسة التطبيقية.

في دراسة نُشرت عام 2024، قام جاب الله (2024) بتطبيق نموذج Support Vector Machine (SVM) لرصد وتحليل التغيرات الجيومورفولوجية على سواحل بحيرة المنزلة خلال الفترة من 1984 إلى 2024، مع التركيز على دور الأنشطة البشرية كعامل مؤثر في هذه التغيرات. اعتمدت الدراسة على بيانات صور فضائية من نوع Landsat لتحليل الغطاء الأرضي وتصنيفه إلى فئات مختلفة تمثل النشاط البشري الطبيعي والأنشطة المعدلة للبيئة. أظهرت النتائج أن SVM فعال في تصنيف الغطاء الأرضي بدقة عالية، حيث تمكن النموذج من التمييز بين المناطق المتأثرة بالنشاط البشري وتلك غير المتأثرة، مع الحفاظ على مستوى جيد من الحساسية والدقة (Accuracy) و(F1-Score. وقد عزت الدراسة هذا الأداء إلى قدرة SVM على التعامل مع بيانات فضائية متعددة الأبعاد ومعقدة، بما يتيح فصل الفئات المختلفة حتى في وجود أنماط غير خطية ومعقدة بين المتغيرات. توصلت الدراسة إلى أن SVM يشكل أداة قوية وموثوقة لرصد التغيرات الجيومورفولوجية المرتبطة بالأنشطة البشرية، ويتيح للمخططين وصناع القرار تحليل أثر الإنسان على البيئة بدقة أكبر، كما توصي بتطبيق هذه النماذج على نطاقات زمنية وجغرافية أوسع لتعزيز قدرة التنبؤ وإدارة الموارد البيئية.

4- Cai, W., Cai, M., Li, Q., & Liu, Q. (2023). *Three-way Imbalanced Learning based on Fuzzy Twin SVM*.

في دراسة مقترحة نُشرت عام 2023، قدم Cai et al. (2023) نموذجًا مبتكرًا يسمى Fuzzy Twin SVM مع قرار ثلاثي (Three-Way Decision) لمعالجة مشكلات البيانات غير المتوازنة في التصنيف. يركز النموذج على تحسين دقة التصنيف للفئات الأقل تمثيلاً في مجموعة البيانات عن طريق تخصيص أوزان أعلى لهذه الأصناف باستخدام دالة عضوية ضبابية (Fuzzy Membership Function)، مما يوازن التأثير بين الفئات الأكثر والأقل تمثيلًا ويقلل من تحيز النموذج تجاه الفئة الأكثر عددًا.

تم اختبار النموذج على عدة مجموعات بيانات مختلفة، وحققت التجارب نتائج أفضل بشكل ملحوظ مقارنة بنماذج SVM التقليدية من حيث مؤشرات الأداء مثل الدقة (Accuracy) والحساسية (Sensitivity) ومقياس F1-Score، خاصة عند التعامل مع البيانات غير المتوازنة. وأكد الباحثون أن الدمج بين مفهوم Twin SVM وتقنيات القرار الثلاثي والضبابية يتيح تحسين القدرة التنبؤية للنموذج وتقليل الأخطاء التصنيفية المرتبطة بالفئات الأقل تمثيلاً. توصلت الدراسة إلى أن Fuzzy Twin SVM مع قرار ثلاثي يمثل خيارًا واعدًا لتطبيقات التصنيف في الحالات التي تواجه فيها البيانات تحديات عدم التوازن، ويوصي الباحثون بتطبيق هذا النموذج على مجالات متعددة مثل التصنيف الطبي، وتحليل المخاطر، والتطبيقات البيئية، لتحسين دقة التنبؤ والتمييز بين الفئات غير المتوازنة.

5- Kumar, A., & Mishra, D. (2025). *Improving Support Vector Machine using Modified Kernel Function*.

قدّم Kumar و Mishra (2025) في دراسة حديثة تطويرًا مهمًا لأداء نماذج آلة المتجه الداعم (SVM) من خلال اقتراح دالة Kernel مدمجة جديدة أطلقا عليها اسم Polynomial–RBF (PRBF)، والتي تهدف إلى الدمج المنهجي بين خصائص دالتي Polynomial و RBF التقليديتين. جاء هذا التطوير استجابة

للتحديات المتعلقة بقدرة النوى التقليدية على تمثيل العلاقات المعقدة في البيانات—لا سيما في التطبيقات ذات الأنماط غير الخطية العالية التغير، مع الحاجة إلى تحسين الفصل بين الفئات في فضاءات عالية الأبعاد. صُمِّمت دالة PRBF لتحقيق توازن بين ميزتي الدوال المدمجة، فهي تستفيد من قدرة Polynomial على التقاط العلاقات غير الخطية عالية الدرجة، ومن مرونة RBF في التعامل مع الأنماط المعقدة ذات الحدود غير المنتظمة. وقد مكّن ذلك الدمج النموذج من بناء حدود قرار أكثر كفاءة، مما أدى إلى تحسين القدرة التمييزية وتقليل الأخطاء التصنيفية.

أظهرت نتائج التجارب المطبقة على عدة مجموعات بيانات—تشمل بيانات خطية وأخرى غير خطية—أن استخدام دالة PRBF حقق:

- تحسنًا ملحوظًا في دقة التصنيف (Accuracy) مقارنة بالنوى RBF، Polynomial، Sigmoid.
- ارتفاعًا في الحساسية (Sensitivity)، مما يدل على قدرة أعلى في اكتشاف الفئات الحقيقية.
- تحسنًا واضحًا في مقياس التوازن الإحصائي F1-Score، مما يؤكد كفاءة النموذج في البيانات غير المتوازنة.
- ثباتًا أعلى في الأداء Across Datasets، بما يشير إلى قدرة نقل Generalization أفضل عند تطبيق النموذج على بيانات جديدة.

وأرجع الباحثان هذه النتائج المتقدمة إلى قدرة الدالة المدمجة على زيادة قابلية الفصل في فضاء السمات، عبر تمثيل البيانات بطريقة تعزز التباعد بين الفئات المتداخلة وتُقلّل من حالات سوء التصنيف، الأمر الذي يُعد جوهريًا في التطبيقات ذات المتطلبات التنبؤية الدقيقة.

خلصت الدراسة إلى أن بنية النواة المدمجة PRBF تُعدّ نهجًا واعدًا لتحسين أداء نماذج SVM في مجالات متعددة، بما في ذلك التصنيف الطبي، والتطبيقات البيئية، والأنظمة الصناعية. كما أوصى الباحثان بتطبيق مزيد من التجارب على مجموعات بيانات أكبر وأكثر تنوعًا للتحقق من إمكانية تعميم النتائج واستكشاف إمكانية دمج الدالة مع استراتيجيات تحسين إضافية مثل التحسين الطوبولوجي أو التعلم التكيفي للنواة.

الفجوة البحثية (Research Gap)

على الرغم من التطور الملحوظ في استخدام خوارزمية Support Vector Machine (SVM) وتنوع دراساتها الحديثة عبر مجالات الطب، والهندسة، والبيئة، والبيانات غير المتوازنة، فإن مراجعة الأدبيات تُظهر وجود عدة جوانب لا تزال بحاجة إلى بحث معمق يوسع من إمكانات هذا النموذج الإحصائي. فقد ركزت الدراسات الأخيرة على تحسين الأداء عبر تطوير نوى جديدة (مثل PRBF)، أو إدماج تقنيات كمية (Quantum SVM)، أو معالجة اختلال توازن البيانات عبر نماذج ضبابية (Fuzzy Twin SVM)، إضافة إلى المقارنات بين نوى SVM المختلفة في بيئات تصنيف تقليدية. ورغم أهمية هذه الجهود، إلا أنها غالبًا ما تظل محصورة في نطاق تطبيقات محددة أو تعتمد على مجموعات بيانات صغيرة أو تفتقر إلى اختبار متكامل عبر بيئات بيانات متعددة المتغيرات ذات العلاقات التفاعلية المعقدة.

كما تظهر المراجعة أن معظم الدراسات ركزت على قياس الأداء اعتمادًا على مؤشر واحد أو اثنين فقط (مثل الدقة أو الحساسية)، بينما لم تتم معالجة المقاييس المتقدمة للأداء متعدد الفئات ولا تقييم الاستقرار الإحصائي للنموذج عند تغير خصائص البيانات بمرور الزمن. إضافة إلى ذلك، فإن الدراسات الحديثة تفتقر إلى مقارنة منهجية شاملة بين النوى التقليدية والمحسّنة والمتقدمة (مثل النوى الكمية والنوى المدمجة والنوى الضبابية) باستخدام تصميم تجريبي موحد يمكن من خلاله استخلاص استنتاجات قابلة للتعميم.

وتشير هذه الفجوات إلى الحاجة لبحوث أكثر شمولًا تهدف إلى:

1. بناء إطار مقارن موحد يختبر أداء أنواع النوى المختلفة, including quantum, fuzzy, hybrid kernels—باستخدام مجموعات بيانات متعددة المتغيرات ومعقدة.

2. تحليل فعالية SVM في البيئات التطبيقية الديناميكية التي تتغير فيها خصائص البيانات بمرور الزمن، وهو جانب لا يزال محدود التناول في الأدبيات.
3. دمج مؤشرات تقييم متنوعة مثل (MCC، AUC، F1، Stability Metrics) لتوفير صورة أكثر دقة عن الأداء الفعلي للنموذج.
4. التحقق من قابلية التعميم Generalizability للنماذج المطورة عبر تطبيقها على مشكلات مختلفة ومعالجة تحديات الضجيج وعدم التوازن والارتباطات البينية بين المتغيرات.

وبناءً على ذلك، فإن المجال لا يزال بحاجة إلى دراسات منهجية متقدمة تسعى إلى تطوير واختبار نماذج SVM هجينة ومتقدمة في بيئات بيانات حقيقية ومعقدة، مع توفير تحليل إحصائي شامل يختبر فعالية هذه النماذج مقارنة بالأساليب التقليدية. هذه الفجوة البحثية تشكّل الأساس العلمي لتطوير إطار جديد يهدف إلى تحسين قدرات SVM في التصنيف متعدد المتغيرات وتحقيق نتائج أكثر دقة وقابلية للتطبيق في المجالات العملية.

مشكلة الدراسة

تتمثل مشكلة الدراسة في تقييم الفروق في أداء نماذج آلة المتجه الداعم (SVM) عند استخدامها في تصنيف البيانات بالاعتماد على ثلاث دوال نواة مختلفة، وذلك بهدف تحديد الدالة الأكثر كفاءة في تحسين دقة التصنيف. وتشمل هذه الدوال:

1. *دالة الإشعاع Radial Basis Function (RBF Kernel)*
2. *الدالة متعددة الحدود Polynomial Kernel Function*
3. *دالة سيغمويد Sigmoid Kernel Function*

وتركز الدراسة على تحليل أثر اختيار نوع دالة النواة على دقة نماذج SVM في تصنيف البيانات، من خلال مقارنة الأداء الإحصائي لكل دالة تحت ظروف متساوية وتجارب منهجية متعددة. تنبع أهمية مشكلة البحث من الدور المحوري الذي تؤديه الدوال النواة في تحديد قدرة نموذج SVM على التمثيل الأمثل للبيانات في فضاءات عالية الأبعاد، وهو ما ينعكس مباشرة على مستوى دقة التصنيف، وحساسية النموذج، وكفاءته العامة. تهدف الدراسة، بالتالي، إلى تحديد الدالة الأكثر ملاءمة لتحسين فعالية نماذج SVM في التطبيقات الإحصائية والمجالات العملية المختلفة.

أهداف الدراسة

تهدف هذه الدراسة إلى تحقيق مجموعة من الأهداف الرئيسة المرتبطة بتقييم فعالية نماذج آلة المتجه الداعم (SVM) في تصنيف البيانات، وذلك من خلال ما يأتي:

1. **توظيف نماذج آلة المتجه الداعم ثنائية الاستجابة (Binary SVM)**

يهدف الجانب الأول من الدراسة إلى استخدام النماذج الثنائية لـ SVM في تصنيف بيانات مرضى القلب، وذلك من أجل تقييم قدرة هذه النماذج على التمييز بين الحالات المرضية المختلفة اعتمادًا على مجموعة من الخصائص السريرية والطبية ذات الصلة.

2. **تطبيق نماذج آلة المتجه الداعم متعددة الاستجابة (Multiclass SVM)**

يهدف الجانب الثاني من الدراسة إلى استخدام نماذج SVM متعددة الاستجابة في تصنيف بيانات أداء الطلاب، بهدف تحليل قدرتها على التنبؤ بمستويات الأداء الأكاديمي عبر فئات متعددة، مما يتيح فهماً أدق للعوامل المؤثرة في التباينات التعليمية.

أهمية الدراسة :

تنبع أهمية هذه الدراسة من اعتمادها على نماذج حديثة ومتقدمة في التعلم الآلي، وعلى وجه الخصوص نماذج آلة المتجه الداعم (SVM)، في مجال تصنيف البيانات. إذ يتيح استخدام هذه النماذج تطوير أدوات تحليل أكثر دقة وفعالية مقارنة بالأساليب التقليدية، بما يعكس التطور المتسارع في تقنيات الذكاء الاصطناعي وتحليل البيانات. كما تسهم الدراسة في إبراز دور دوال النواة المختلفة في تحسين الأداء التصنيفي، مما يوفر إطاراً عملياً يساعد الباحثين والمتخصصين على اختيار النماذج الأنسب لطبيعة البيانات، سواء في التطبيقات الطبية أو التعليمية أو غيرها من المجالات التطبيقية.

حدود الدراسة:

أولاً: الحدود المكانية

تعتمد الدراسة على مجموعتين من البيانات تم الحصول عليهما من منصات بيانات عالمية موثوقة. تمثلت المجموعة الأولى في بيانات حالات مرضى القلب، وتشمل 1025 مشاهدة تتضمن حالات مصابة وغير مصابة بأمراض القلب، وقد جُمعت هذه البيانات من أحد المستشفيات في الولايات المتحدة الأمريكية، وتم الحصول عليها من موقع Kaggle – Classification Datasets. أمّا المجموعة الثانية فتضم بيانات أداء الطلاب، وتشمل 1945 مشاهدة تمثل بيانات متعددة الاستجابة صادرة عن إحدى المدارس في الولايات المتحدة الأمريكية، وتم الحصول عليها من موقع Kaggle – Multiclass Classification Datasets.

ثانيًا: الحدود الزمنية

تغطي البيانات الطبية المتعلقة بأمراض القلب السنة 2020، بينما تعود البيانات الخاصة بأداء الطلاب إلى السنة 2017، مما يحدد الإطار الزمني الذي تُبنى عليه نتائج الدراسة واستنتاجاتها

ثالثا : مصادر جمع البيانات

اعتمدت الدراسة على بيانات منشورة ومتاحة من قواعد البيانات المفتوحة والمتخصصة في مجموعات البيانات الخاصة بالتصنيف الثنائي والمتعدد، وذلك من خلال المصادر الإلكترونية الآتية:

1. منصة Kaggle الخاصة بمجموعات بيانات التصنيف الثنائي، والتي توفر العديد من قواعد البيانات المعتمدة في الدراسات الإحصائية وتعلم الآلة، وتستخدم على نطاق واسع في دراسة المشكلات الصحية والاجتماعية والاقتصادية.

- *www.kaggle classification datasets.com*

2. منصة Kaggle الخاصة بمجموعات بيانات التصنيف متعدد الفئات، والتي تتيح مجموعات بيانات متنوعة تُستخدم في نماذج التصنيف متعددة الاستجابة، وتشمل بيانات تعليمية وطبية وسلوكية.

- *www.kaggle classification multiclass datasets.com*

3. مجموعة بيانات أمراض القلب المنشورة على منصة Kaggle ضمن قاعدة البيانات المسماة *Heart Disease Dataset*، والمتاحة من خلال الحساب البحثي (johnsmith88)، والتي تتضمن مجموعة من المتغيرات السريرية والديموغرافية المعتمدة في الأدبيات الطبية الحديثة لدراسة مخاطر الإصابة بالأمراض القلبية.

- *https://www.kaggle.com/johnsmith88/heart-diseased*

الإطار النظري

آلة المتجه الداعم (Support Vector Machine – SVM)

تُعدّ آلة المتجه الداعم أحد أهم خوارزميات التعلّم الآلي الإشرافية المستخدمة في مهام **التصنيف والانحدار**، وتمتاز بقدرتها العالية على التعامل مع البيانات ذات الأبعاد المرتفعة وتحديد الحدود الفاصلة المثلى بين الفئات. تعتمد SVM على إيجاد **فاصل فائق (Hyperplane)** يحقق أكبر هامش ممكن

بين الفئات المختلفة، مما يضمن تعميمًا أعلى ودقة أفضل عند تطبيق النموذج على بيانات جديدة . (Hastie et al., 2022).

من الناحية المنهجية، تستخدم SVM مجموعة من **دوال النواة (Kernels)** مثل:

- **النواة الخطية Linear Kernel**
- **نواة RBF (Radial Basis Function)**
- **النواة متعددة الحدود Polynomial Kernel**

وهي دوال تساعد النموذج على تحويل البيانات إلى فضاء أعلى لإيجاد فاصل خطي حتى لو كانت البيانات الأصلية غير خطية. كما تتميّز SVM بقدرتها على التحكم في التوازن بين التعقيد والأخطاء من خلال معاملات تنظيم مثل **C**، ما يسمح بضبط النموذج للحصول على أداء أمثل مع تجنب الإفراط في التعلّم.

تشير الأدبيات الحديثة إلى أنّ SVM تُعد خيارًا قويًا في تطبيقات التصنيف مثل تصنيف الصور، معالجة النصوص، البيانات الطبية، وتصنيف الأداء الأكاديمي، خصوصًا عند وجود بيانات غير متوازنة أو ذات ضوضاء منخفضة. وتُعد هذه الخوارزمية من النماذج التي تمتاز باستقرارها، وقلة حساسيتها لتعدد القيم المحلية، وقدرتها على تحقيق دقة مرتفعة مقارنة بالطرق التقليدية. (Zhang et al., 2023)

أولاً: آلة المتجه الداعم الخطية (Linear Support Vector Machine)

تسعى آلة المتجه الداعم الخطية إلى إيجاد فاصل فائق (Hyperplane) يفصل بين الفئات في فضاء الخصائص بحيث يكون الفصل بأكبر هامش (Margin) ممكن .وتعتمد على نظرية زيادة الهامش Maximum Margin Classification.

1. دالة الفاصل(Hyperplane Function)

يُعرَّف الفاصل الخطي كالتالي:

$$[\ f(x) = w^T x + b\ ]\text{----------(1)}$$

حيث:

- (w)متجه الأوزان.(Weight Vector)
- (b)معامل الانحياز.(Bias Term)
- يتم تحديد الفئة عبر الإشارة:

$$y = sign(w^T x + b)\text{----------(2)}$$

2. دالة الهدف – (Objective Function) حالة الفصل الخطي التام

الهدف هو تعظيم الهامش، وهو ما يعادل تصغير المعامل:(|w|)

$$\min_{w,b} \frac{1}{2} \| w \|^2\text{----------(3)}$$

تحت القيود التالية:

$$y_i(w^t x_i + b) \geq 1 \quad \text{----------(4)}$$
$$i = 1, \dots, n$$

مزايا Linear SVM

1. **فعالية عالية** مع البيانات عالية الأبعاد، خاصة عندما يكون عدد المتغيرات أكبر من عدد العينات.
2. **سهولة التدريب والتفسير** مقارنةً بالنماذج غير الخطية.
3. **سرعة الحساب** لعدم الحاجة لاستخدام دوال النواة المعقدة.
4. قدرة جيدة على التعميم عند البيانات الخطية.

تطبيقات Linear SVM

- تصنيف النصوص والرسائل البريدية (مثل تصنيف البريد المزعج).
- التعرف على الأنماط في البيانات الطبية مع بيانات خطية تقريبًا.
- تحليل البيانات الاقتصادية أو التعليمية ذات الفصل الواضح بين الفئات.

ثانياً: آلة المتجه الداعم غير الخطية (Nonlinear Support Vector Machine)

تُستخدم آلة المتجه الداعم غير الخطية (Nonlinear SVM) عندما تكون البيانات غير قابلة للفصل خطيًا في فضائها الأصلي. في هذه الحالات، لا يمكن للفاصل الخطي (Linear Hyperplane) التمييز بين الفئات بشكل دقيق، لذا تعتمد Nonlinear SVM على دوال النواة (Kernel Functions) لتحويل البيانات إلى فضاء أبعاد أعلى يمكن فيه فصل الفئات خطيًا(Kumar et al., 2025) .

أشهر دوال النواة المستخدمة في Nonlinear SVM:

- **نواة RBF (Radial Basis Function Kernel):** فعّالة جدًا للبيانات ذات الأنماط المعقدة وغير الخطية، وتتعامل مع العلاقات غير الخطية بين المتغيرات بكفاءة عالية.

$$\boldsymbol{RBF\ (Gaussian)}: \left(\boldsymbol{K(x_i, x_j) = \backslash exp \backslash big\left(-\gamma |x_i - x_j|^{2 \backslash big}\right)}\right) \text{--------(5)}$$

- حيث

$$(\boldsymbol{\gamma > 0}).$$

- **النواة متعددة الحدود (Polynomial Kernel):** تعالج العلاقات متعددة الحدود بين المتغيرات، وتسمح بالنمذجة الأكثر تعقيدًا للبيانات.

$$\boldsymbol{Polynomial}: \left(\boldsymbol{K(x_i, x_j) = \left(\gamma, x_{i_j}^{Tx} + r\right)^d}\right) \text{----------(6)}$$

- **نواة Sigmoid (Sigmoid Kernel):** تشبه وظيفة تفعيل الشبكات العصبية، وتستخدم في بعض تطبيقات التعلم العميق والتصنيف غير الخطي.

$$\boldsymbol{Sigmoid}: \left(\boldsymbol{K(x_i, x_j) = \backslash tanh\left(\gamma, x_{i_j}^{Tx} + r\right)}\right) \text{----------(7)}$$

باستخدام النواة، يتم حل مسألة التحسين الأصلية لـ SVM في الفضاء الجديد دون الحاجة لحساب التحويل بشكل صريح، وهو ما يُعرف باسم "Kernel Trick".

مزايا Nonlinear SVM

1. القدرة على التعامل مع البيانات المعقدة والمتغيرة.
2. زيادة دقة التصنيف في المشكلات التي تتضمن أنماط غير خطية.
3. مرونة عالية بفضل اختيار نوع النواة ومعاملاتها (مثل γ في RBF ودرجة التعددية في Polynomial)

تطبيقات Nonlinear SVM

- التشخيص الطبي، مثل تصنيف أنواع السرطان أو أمراض القلب المعقدة.
- التعرف على الأنماط في الصور والفيديو.
- تحليل البيانات البيئية والجيومورفولوجية التي تحتوي على علاقات غير خطية.
- تصنيف النصوص والبيانات متعددة المتغيرات.

آلة المتجه الداعم للتصنيف المتعدد Multi-class Support Vector Machine – Multi-class SVM)

بينما صممت **آلة المتجه الداعم التقليدية (SVM)** أساسًا للتصنيف الثنائي، فإن العديد من التطبيقات العملية تتطلب تصنيفًا لأكثر من فئتين. لذلك تم تطوير **تقنيات SVM متعددة التصنيف (Multi-class SVM)** لتوسيع نطاق استخدام SVM في مشكلات التصنيف المعقدة.(Hsu et al., 2021)
هناك استراتيجيتان رئيسيتان لتطبيق SVM على مشكلات التصنيف المتعدد:

1. **منهجية واحد مقابل الباقي:One-vs-Rest (OvR)**
   - يتم إنشاء نموذج SVM لكل فئة ضد جميع الفئات الأخرى.
   - تُصنف العينة الجديدة في الفئة التي يُظهر نموذجها أكبر هامش .(Decision Function)
   - طريقة بسيطة وفعالة مع عدد معتدل من الفئات.

تقوم هذه الطريقة ببناء مجموعة من النماذج الثنائية عددها M ، وفي كل نموذج يتم فصل إحدى الفئات عن الفئات الأخرى ، ومعاملة باقي الفئات على أنها فئة واحدة ، وبذلك يصبح لدينا فئتين . ويتم تكرار ذلك الأجراء حتى يتم عمل جميع النماذج وبذلك يكون الهدف هو حل مشكلة التصنيف الثنائية الموضحة بالعلاقة التالية:
بشرط:

$$wtT\Phi(xi) + bt \geq 1 - \xi it, if\ yi = t \text{----------(8)}$$
$$wtT\Phi(xi) + bt \leq \xi it - 1, if\ yi \neq t \text{----------(9)}$$
$$\xi it \geq 0, t = 1,2, \dots, n$$

وعند تصنيف مفردة جديدة لم تدخل في بناء النموذج فإنه يتم التعويض بقيمة تلك المفردة في دالة القرار $sign\ (wtT\Phi\ (xi) + bt$ لكل نموذج ، ومن ثم يتم اختيار الفئة التي تعُطي أعلى قيمة لدالة القرار بالنسبة لتلك المفردة

2. **منهجية واحد مقابل واحد :One-vs-One (OvO)**
   - يتم إنشاء نموذج SVM لكل زوج من الفئات.
   - يتم التصويت على نتائج جميع النماذج لاختيار الفئة النهائية.
   - أكثر دقة عند عدد الفئات كبير، لكنه يتطلب تدريب عدد أكبر من النماذج.

تقوم هذه الطريقة ببناء مجموعة من النماذج (المصنفات)الثنائية عددها $M(M - 1 / 2)$ وفي كل مرة يتم بناء النموذج باستخدام بيانات من فئتين ، وبذلك يكون الهدف هو حل مشكلة التصنيف الثنائية الموضحة بالعلاقة التالية:
بشرط:

$$wtlT\Phi(xi) + btl \leq \xi itl - 1, if\ yi = l \text{----------(10)}$$
$$\xi itl \geq 0$$

حيث أن t تعبر عن رتبة الفئة الأولى ، و l تعبر عن رتبة الفئة الثانية وعند تصنيف مفردة جديدة فإنه يتم التعويض بقيمة تلك المفردة في دالة القرار $sign\ (wtlT\Phi(xi) + btl$ (لك لنموذج ، ومن ثم يتم اختيار الفئة التي تحتل أعلى تك ا رر وأكبر نسبة تصويت من جميع النماذج

مزايا Multi-class SVM

- التوسع من التصنيف الثنائي إلى متعدد الفئات دون فقدان دقة التصنيف.
- قابلية استخدام نفس دوال النواة (Kernel Functions) مثل RBF وPolynomial و.Sigmoid
- مرونة في التعامل مع بيانات معقدة أو متعددة المتغيرات.

التطبيقات العملية

- تصنيف بيانات الطلاب حسب مستويات الأداء الأكاديمي (ممتاز، جيد جدًا، جيد، مقبول).
- التشخيص الطبي لأمراض متعددة في نفس المجموعة (مثل أنواع السرطان المختلفة).

- تصنيف الصور أو الأنماط في مجالات الرؤية الحاسوبية.

التحديات

- زيادة عدد الفئات يؤدي إلى تعقيد أكبر في التدريب والتنبؤ.
- يتطلب اختيار استراتيجية OvR أو OvO بعناية حسب حجم البيانات وعدد الفئات.
- تحسين أداء النماذج يحتاج إلى ضبط المعلمات (مثل معامل C، ونوع النواة ومعاملاتها).

مقاييس أداء التصنيف (Classification Performance Metrics)

في تحليل نتائج نماذج التصنيف مثل **SVM**، تُستخدم مجموعة من المقاييس الكمية لتقييم دقة النموذج وكفاءته في التنبؤ بالفئات الصحيحة. يمكن تصنيف هذه المقاييس إلى مقاييس أساسية تعتمد على **Confusion Matrix** ومقاييس مشتقة أخرى (Powers et al.,2020)

1. مصفوفة التشويش (Confusion Matrix)

الجدول 1. تمثل مصفوفة التشويش أداة أساسية لتقييم أداء النموذج، حيث توضح عدد التنبؤات الصحيحة والخاطئة لكل فئة. العناصر الرئيسية:

الفئة الحقيقية \ الفئة المتوقعة	إيجابي (Positive)	سلبي (Negative)
إيجابي (Positive)	True Positive (TP)	False Negative (FN)
سلبي (Negative)	False Positive (FP)	True Negative (TN)

تعريف المقاييس داخل مصفوفة التشويش

- **الإيجابيات الصحيحة (True Positive – TP):** الحالات التي صنفها النموذج بشكل صحيح على أنها إيجابية، أي عندما يصنف نموذج SVM المريض على أنه مريض ويكون في الواقع مريضًا.
- **السلبيات الصحيحة (True Negative – TN):** الحالات التي صنفها النموذج بشكل صحيح على أنها سلبية، أي عندما يصنف نموذج SVM الفرد على أنه غير مريض ويكون في الواقع غير مريض.
- **الإيجابيات الخاطئة (False Positive – FP):** الحالات التي صنفها النموذج بشكل خاطئ على أنها إيجابية، أي عندما يصنف نموذج SVM الفرد على أنه مريض بينما هو في الواقع غير مريض. يُعرف أحيانًا بخطأ النوع الأول.
- **السلبيات الخاطئة (False Negative – FN):** الحالات التي صنفها النموذج بشكل خاطئ على أنها سلبية، أي عندما يصنف نموذج SVM الفرد على أنه غير مريض بينما هو في الواقع مريض. يُعرف أحيانًا بخطأ النوع الثاني.

أهمية استخدام مصفوفة التشويش

تمكّن مصفوفة التشويش الباحثين من:

- قياس الدقة (Accuracy)، الحساسية (Sensitivity/Recall)، الدقة النوعية (Precision)، ومقياس F1-Score.
- فهم نمط الأخطاء التي يرتكبها النموذج، سواء كانت أخطاء إيجابية أو سلبية.
- مقارنة أداء نماذج مختلفة أو إعدادات متعددة لدوال النواة (Kernel Functions) في SVM.

2. الدقة Accuracy

النسبة المئوية للحالات المصنفة بشكل صحيح من إجمالي الحالات

هي أحد أساليب تقييم نماذج التصنيف ونسبتها تحدد بقسمة التصنيفات الصحيحة على إجمالي عدد التصنيفات

معادلة الدقة هي:-

$$(TP + TN) / (TP + TN + FP + FN) = Accuracy$$ ----------(11)

- $TP = True\ Positives$
- $TN = True\ Negatives$
- $FP = False\ Positives$
- $FN = False\ Negatives$

- مقياس عام للنجاح الكلي للنموذج.
- قد يكون مضللاً إذا كانت البيانات غير متوازنة.

3. الاستدعاء Recall (أو الحساسية Sensitivity)

النسبة المئوية للحالات الإيجابية التي تم تصنيفها بشكل صحيح:

هو أحد أساليب تقييم نماذج التصنيف ويمكن حساب الاستدعاء على النحو التالي:

$$TP / (TP + FN) = Recall$$ ----------(12)

- يعكس قدرة النموذج على التعرف على الحالات الإيجابية.
- مهم جدًا في التشخيص الطبي، مثل اكتشاف أمراض القلب أو السرطان.

4. الضبط Precision

النسبة المئوية للتنبؤات الإيجابية الصحيحة من بين جميع التنبؤات الإيجابية:

هو أحد أساليب تقييم نماذج التصنيف ويمكن حساب الضبط على النحو التالي:

$$TP / (TP + FP) = Precision$$ ----------(13)

- يقيس مدى موثوقية النموذج في التنبؤ بالحالات الإيجابية.
- مفيد عند تكلفة الخطأ الإيجابي عالية.

5. مقياس F1-Score

المتوسط التوافقي بين الدقة والاستدعاء:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$ ----------(14)

يجمع بين الحساسية والدقة في مقياس واحد.

- مهم عند الحاجة إلى موازنة بين FP و FN.

متى يُستخدم F1 ؟

- عند وجود عدم توازن بين الفئات.(Class Imbalance)
- عندما تكون الأخطاء في الفئة الإيجابية أعلى تكلفة.
- عند الرغبة في دمج الدقة والاسترجاع في مقياس واحد.

6. الحساسية السلبية (Specificity)

يُعد **Specificity** أحد أهم مقاييس تقييم نماذج التصنيف، خاصة في التطبيقات الطبية، ويُستخدم لقياس قدرة النموذج على **التعرّف على الحالات السلبية بشكل صحيح**. كلما ارتفعت قيمة Specificity ، دلّ ذلك على قدرة النموذج على تجنّب تصنيف الحالات السلبية على أنها إيجابية.

يمثل **Specificity** نسبة الحالات السلبية (Negative Cases) التي تم تصنيفها بشكل صحيح على أنها سلبية، مقارنةً بإجمالي الحالات السلبية.

الصيغة الرياضية الأساسية

$$Specificity = \frac{TN}{TN + FP} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$ ----------(15)

تفسير الرموز

- **TN (True Negatives)** : عدد الحالات السليمة أو السلبية التي صُنّفت بشكل صحيح.

- **FP (False Positives):** عدد الحالات السلبية التي صُنّفت خطأً على أنها إيجابية (أخطاء من النوع الأول).

العلاقة مع Sensitivity (Recall)

غالبًا يستخدم Specificity مع (Sensitivity وRecall) للحصول على تقييم شامل للنموذج:

- **Sensitivity** يقيس قدرة النموذج على اكتشاف الحالات الإيجابية.
- **Specificity** يقيس قدرة النموذج على استبعاد الحالات السلبية.

متى يستخدم؟

- في التشخيص الطبي (مثل أمراض القلب، السرطان، الفيروسات).
- في الكشف عن الاحتيال (لتمييز العمليات الطبيعية).
- عند الرغبة في تقليل الأخطاء التي تتعلق بالإنذارات الكاذبة.(False Alarms)

مميزات وعيوب استخدام تقنية آلة المتجه الداعم (Support Vector Machine – SVM)

تقنية **آلة المتجه الداعم (SVM)** تُعدّ من أبرز تقنيات التعلم الآلي لتصنيف البيانات، وتتميز بعدة نقاط قوة، لكنها تواجه أيضًا بعض القيود (Sahu et al., 2023)

أولاً: مميزات تقنية SVM

1. **دقة عالية في التصنيف (High Classification Accuracy): يعتبر نموذج** SVM قادر على التعامل مع البيانات ذات الأبعاد العالية (High-dimensional data) وتحقيق نتائج دقيقة، خصوصًا في الحالات التي تكون فيها البيانات غير خطية أو معقدة.
2. **مرونة في استخدام دوال النواة (Kernel Functions):** يمكن تحويل البيانات إلى فضاء أبعاد أعلى باستخدام دوال RBF، Polynomial، أو Sigmoid، مما يمكّن من الفصل بين الفئات غير الخطية بكفاءة.
3. **فعّالة مع البيانات الخطية وغير الخطية:** يمكن استخدام SVM لتحقيق الفصل الأمثل بين الفئات ، سواء كانت البيانات خطية أو غير خطية.
4. **تقليل خطر الإفراط في التعلّم (Overfitting): تعتمد نماذج** SVM على مفهوم الهامش الأقصى (Maximum Margin)، مما يقلل من احتمال الإفراط في التكيف مع البيانات التدريبية، خصوصًا عند اختيار معلمات النموذج بشكل مناسب.
5. **قابلية التوسع للتصنيف متعدد الفئات:** بالرغم من أن SVM مصممة أصلاً للتصنيف الثنائي، يمكن توسيعها لتطبيقات متعددة الفئات باستخدام استراتيجيات مثل One-vs-Rest أو One-vs-One.

ثانياً: عيوب تقنية SVM

1. **الحساسية لاختيار المعلمات (Parameter Sensitivity):** أداء SVM يعتمد بشكل كبير على اختيار المعلمات المناسبة مثل معامل العقوبة C ومعامل النواة γ في دالة RBF ، وقد يؤدي الاختيار الخاطئ إلى انخفاض الدقة أو الإفراط في التعلّم.
2. **تكلفة حسابية عالية للبيانات الكبيرة (Computational Complexity):** عند التعامل مع مجموعات بيانات ضخمة جدًا، قد يصبح تدريب SVM بطيئًا ويستهلك موارد كبيرة مقارنة ببعض الخوارزميات الأخرى مثل Logistic Regression أو Decision Trees.
3. **صعوبة تفسير النموذج (Model Interpretability):** النماذج التي تستخدم نوى غير خطية قد تكون معقدة، مما يجعل تفسير قرارات النموذج صعبًا، وهو أمر مهم في التطبيقات الطبية أو البيئية.
4. **الأداء مع البيانات الضوضائية (Noise Sensitivity) :** وجود بيانات ضوضائية أو أخطاء في الملصقات (Labels) يمكن أن يؤثر سلبًا على دقة SVM، خاصة إذا كانت المسافة بين الفئات صغيرة.
5. **تأثير عدم توازن البيانات (Imbalanced Data):** عند وجود فئات ذات تمثيل غير متساوٍ، قد تميل SVM إلى التصنيف الأكثر تمثيلًا، مما يقلل من دقة التعرف على الفئة الأقل تمثيلًا.

النتائج والتوصيات

وصف المتغيرات المستخدمة في الدراسة التطبيقية

أولاً: بيانات التصنيف الثنائي (Binary Classification Data)

تشمل هذه البيانات معلومات متعلقة بتشخيص الإصابة بأمراض القلب، ويبلغ عدد المشاهدات **1025 مشاهدة** تتضمن حالات مصابة وغير مصابة. تمثل هذه البيانات متغيراً تابعاً ثنائي الاستجابة، حيث تُرمز الحالة غير المصابة بالقيمة (0)، بينما تُرمز الحالة المصابة بالقيمة .(1)

استُخدم في هذه الدراسة **ثلاثة عشر متغيراً تفسيرياً** تشمل متغيرات ديموغرافية وسريرية وفسيولوجية وتشخيصية، وهي من المتغيرات القياسية التي أثبتت فعاليتها في الأدب الطبي الحديث في تحليل مخاطر أمراض القلب (Chen et al., 2023; Li & Wang, 2024; Alshamrani, 2023)

المتغيرات المستقلة المستخدمة في النموذج

1. **العمر (Age – X1)**

متغير عددي مستمر يقاس بالسنوات، ويعد من أهم عوامل الخطورة القلبية، إذ يرتبط بزيادة احتمالات التصلب الشرياني واضطرابات عضلة القلب. وقد أكدت الدراسات الحديثة أهميته في النماذج التنبؤية القلبية .(Patel et al., 2024)

2. **الجنس (Sex – X2)**

متغير ثنائي يأخذ قيمة (1) للذكور و(2) للإناث، مما يعكس الفروق الإكلينيكية في معدلات الإصابة، حيث تزداد احتمالية الإصابة لدى الذكور وفق دراسات مقارنة حديثة (Zhou et al., 2023).

3. **نوع ألم الصدر (Chest Pain Type – X3)**

متغير فئوي يتضمن أربع فئات سريرية تعكس طبيعة الألم، وهي من أهم المؤشرات التشخيصية في أمراض القلب. أثبتت التحليلات الحديثة أنها من أقوى المتغيرات المتنبئة بخطر المرض (Singh & Rao, 2024).

4. **ضغط الدم أثناء الراحة (Resting Blood Pressure – X4)**

متغير عددي يقاس بالمليمتر زئبق، ويمثل مؤشراً أساسياً على الضغط القلبي الوعائي، وقد أشارت الدراسات الحديثة إلى ارتباطه القوي بالخلل البطيني .(Chen et al., 2023)

5. **الكولسترول (Cholesterol – X5)**

مستوى الكوليسترول الكلي، ويلعب دوراً محورياً في تقييم خطر تصلّب الشرايين.

6. **سكر الدم الصائم (Fasting Blood Sugar – X6)**

متغير ثنائي، ويمثل مؤشراً على اضطرابات الاستقلاب المرتبطة بالأمراض القلبية، وقد أظهرت الدراسات الحديثة أهميته كعامل خطر مشترك.

7. **تخطيط القلب أثناء الراحة (Resting ECG – X7)**

متغير فئوي يعكس مؤشرات كهربائية مهمة لتقييم كفاءة عضلة القلب.

8. **الحد الأقصى لضربات القلب (Max Heart Rate – X8)**

متغير عددي يمثل الاستجابة القلبية للمجهود البدني، وهو عنصر تشخيصي مهم في الأدبيات الحديثة .(Li & Wang, 2024)

9. **الذبحة الناتجة عن المجهود (Exercise Angina – X9)**

متغير ثنائي يتعلق بوجود الذبحة أثناء التمرين، ويُعد دليلاً مؤكداً على قصور التروية.

10. **مستوى الاكتئاب ST (Oldpeak – X10)**

يقيس مقدار الانخفاض في مقطع ST، وهو من أقوى مؤشرات نقص التروية وفق الدراسات السريرية الحديثة.

11. **ميل المنحنى ST (Slope – X11)**
متغير فئوي مهم في تقييم ديناميكية تغيرات ST خلال اختبار الجهد.
12. **عدد الأوعية المتصبغة بالأشعة (Ca – X12)**
متغير عددي يأخذ القيم بين 0 و3، ويعكس مدى ضيق الشرايين التاجية.
13. **اضطراب الثلاسيميا (Thal – X13)**
متغير فئوي يتعلق بكفاءة تدفق الدم، وقد أظهرت الدراسات الحديثة ارتباطه بخطر أمراض القلب لدى بعض الفئات (Rahman & Yadav, 2023).

المتغير التابع: أمراض القلب (Y)

متغير ثنائي يأخذ القيمة (0) لغير المصاب والقيمة (1) للمصاب، ويُعد الهدف التنبؤي الرئيس للنموذج الإحصائي. تعتمد دقة النموذج على قدرته على دمج جميع المتغيرات السابقة لتصنيف المرضى بدقة عالية، وفق منهجيات حديثة في التعلم الآلي (Islam et al., 2024).

ثانياً: بيانات التصنيف المتعدد (Data of Multinomial Classification)

تعتمد الدراسة في شقّها المتعلق بالتصنيف متعدد الاستجابة على **بيانات أداء الطلاب (Students Performance Data)**، وهي إحدى قواعد البيانات التعليمية الحديثة واسعة الاستخدام في البحوث الخاصة بنمذجة التحصيل الأكاديمي باستخدام تقنيات تعلم الآلة. تحتوي البيانات على **1945 مشاهدة**، كما تتضمن **ثلاثة عشر متغيراً تفسيرياً** يعكس كل منها جانباً سلوكياً أو ديموغرافياً أو تعليمياً له تأثير محتمل في الأداء الدراسي للطلاب.

يُعد المتغير التابع (Y) **متغيراً متعدد الاستجابة** (Multinomial Outcome)، وقد تم ترميز فئات الأداء اعتماداً على المعدل التراكمي (GPA) على النحو الآتي:

- **الفئة 0:** طلاب معدلهم التراكمي أقل من أو يساوي 3.5
- **الفئة 1:** طلاب معدلهم التراكمي أكبر من 3.5 وحتى 3
- **الفئة 2:** طلاب معدلهم التراكمي أكبر من 3 وحتى 2.5
- **الفئة 3:** طلاب معدلهم التراكمي أكبر من 2.5 وحتى 2
- **الفئة 4:** طلاب معدلهم التراكمي أقل من 2

ويمثل هذا الترميز نموذجاً شائعاً في الأدبيات التربوية الحديثة لدراسة التحصيل الأكاديمي وتصنيفه باستخدام نماذج متعددة الفئات، كما ورد في دراسات حديثة مثل *Khan et al. (2023)* و *Roy & Singh (2024)* اللتين أفادتا بأن جودة التنبؤ ترتبط ارتباطاً وثيقاً بوضوح مستويات التجميع في متغير الاستجابة.

المتغيرات المفسّرة المستخدمة في النموذج

تتكون البيانات من ثلاثة عشر متغيراً مستقلاً (X1–X13)، يمكن تصنيفها ضمن أربعة محاور رئيسية:

(1) المتغيرات الديموغرافية، (2) المتغيرات الأسرية، (3) المتغيرات السلوكية والدراسية، (4) المشاركة في الأنشطة المدرسية.

وفيما يلي وصف تفصيلي لكل متغير:

1. **العمر (Age – X1)**
متغير عددي مستمر يقاس بالسنوات. تشير الأدبيات الحديثة إلى ارتباط العمر بدرجة النضج الدراسي والقدرة على التحصيل الأكاديمي.
2. **الجنس (Gender – X2)**
متغير ثنائي:
- ذكر = 0

- أنثى = 1

3. **العرق (Ethnicity – X3)**
متغير فئوي مشفّر كما يلي:
- قوقازي = 0
- أمريكي من أصل إفريقي= 1
- آسيوي = 2
- أعراق أخرى= 3

4. **المستوى التعليمي للوالدين (Parental Education – X4)**
متغير فئوي يشمل:
لا يوجد تعليم= 0، ثانوي= 1، بعض الكليات = 2، بكالوريوس = 3، ما بعد البكالوريوس = 4

5. **وقت الدراسة الأسبوعي (Study Time Weekly – X5)**
متغير عددي من 0 إلى 20 ساعة، ويعكس حجم الجهد الأكاديمي المبذول.

6. **عدد أيام الغياب (Absences – X6)**
متغير عددي يتراوح بين 0 و30 يوماً. كثرة الغياب مؤشر سلبي على الأداء الدراسي.

7. **التدريس الإضافي (Tutoring – X7)**
متغير ثنائي:
- لا يوجد تدريس = 0
- يوجد تدريس = 1

ويدل على الحصول على دعم أكاديمي خارجي.

8. **دعم الوالدين (Parental Support – X8)**
متغير فئوي يأخذ القيم:
لا يوجد= 0، منخفض= 1، معتدل = 2، مرتفع =3، مرتفع جداً=4

9. **المشاركة في الأنشطة اللامنهجية (Extracurricular – X9)**
متغير ثنائي (لا يوجد=0، يوجد=1)، ويعكس مدى انخراط الطالب في الأنشطة المدرسية.

10. **المشاركة في الرياضة (Sports – X10)**
متغير ثنائي يعكس المشاركة في الأنشطة الرياضية المدرسية.

11. **النشاط الموسيقي (Music – X11)**
متغير ثنائي (لا يوجد=0، يوجد=1)، وتشير بعض الدراسات إلى ارتباطه بالإبداع والتحصيل.

12. **العمل التطوعي (Volunteering – X12)**
متغير ثنائي يعكس المشاركة في الأعمال التطوعية المدرسية.

13. **المعدل التراكمي (GPA – X13)**
متغير عددي يتراوح بين 0.2 و4.0، ويستخدم كمقياس للأداء الأكاديمي العام.

أولاً: نتائج دقة التصنيف عند استخدام نماذج آلة المتجه الداعم ثنائية الاستجابة

يعرض هذا الجزء نتائج دقة تصنيف حالات أمراض القلب باستخدام نماذج آلة المتجه الداعم ذات الاستجابة الثنائية، وذلك اعتماداً على مخرجات برنامج ORANGE. وقد تم تقييم أداء النماذج باستخدام ثلاثة أساليب مختلفة للفصل بين البيانات، وهي: أسلوب التحقق المتقاطع، وأسلوب العينة العشوائية، وأسلوب الاختبار على بيانات التدريب. كما تم اختبار ثلاثة أنواع من الدوال النواة، هي: الدالة متعددة الحدود، ودالة الأساس الشعاعي، ودالة (Sigmoid) ويوضح الجدول الآتي دقة التصنيف المحققة لكل حالة.

نتائج دقة التصنيف لنماذج آلة المتجه الداعم ثنائية الاستجابة

جدول 2. نتائج دقة التصنيف لنماذج آلة المتجه الداعم ثنائية الاستجابة

نسبة دقة التصنيف	الاسلوب المستخدم	نوع النواة المستخدم
%0.86	التحقق المتقاطع Cross-validation	الدالة متعددة الحدود (Polynomial)
%0.895	العينة العشوائية Random sample	الدالة متعددة الحدود (Polynomial)
%0.90	الاختبار على بيانات التدريب Testing on training data	الدالة متعددة الحدود (Polynomial)
%0.947	التحقق المتقاطع Cross-validation	دالة الأساس الشعاعي (RBF)
%0.931	العينة العشوائية Random sample	دالة الأساس الشعاعي (RBF)
%0.97	الاختبار على بيانات التدريب Testing on training data	دالة الأساس الشعاعي (RBF)
%0.692	التحقق المتقاطع Cross-validation	الدالة (Sigmoid)
%0.709	العينة العشوائية Random sample	الدالة (Sigmoid)
%0.699	الاختبار على بيانات التدريب Testing on training data	الدالة (Sigmoid)

مناقشة النتائج

تشير النتائج إلى تفوق نموذج آلة المتجه الداعم عند استخدام **دالة الأساس الشعاعي**، إذ حقق أعلى مستويات للدقة عبر جميع أساليب التقييم، حيث بلغت نسبة الدقة في حالة الاختبار على بيانات التدريب 0.97%. بينما جاءت الدالة متعددة الحدود في المرتبة الثانية، في حين سجلت الدالة (Sigmoid)أقل مستوى للأداء.

ثانياً: نتائج التصنيف باستخدام نماذج آلة المتجه الداعم متعدد الاستجابة

يعرض هذا الجزء نتائج تصنيف أداء الطلاب بالاعتماد على نماذج آلة المتجه الداعم متعدد الاستجابة، وذلك بناءً على مخرجات برنامج ORANGE. وقد تم استخدام ثلاثة أساليب لتقييم أداء النماذج، وهي: التحقق المتقاطع، والعينة العشوائية، والاختبار على بيانات التدريب. كما تم اختبار ثلاثة أنواع من الدوال النواة: الدالة متعددة الحدود، ودالة الأساس الشعاعي، والدالة (Sigmoid). ويبين الجدول الآتي نسب دقة التصنيف المحققة لكل حالة.

نتائج دقة التصنيف لنماذج آلة المتجه الداعم متعدد الاستجابة

جدول 3. نتائج دقة التصنيف لنماذج آلة المتجه الداعم متعدد الاستجابة

نسبة دقة التصنيف	أسلوب التقييم	نوع النواة المستخدم
%0.845	التحقق المتقاطع Cross-validation	الدالة متعددة الحدود (Polynomial)
%0.835	العينة العشوائية Random sample	الدالة متعددة الحدود (Polynomial)
%0.943	الاختبار على بيانات التدريب Testing on training data	الدالة متعددة الحدود (Polynomial)
%0.817	التحقق المتقاطع Cross-validation	دالة الأساس الشعاعي (RBF)
%0.80	العينة العشوائية Random sample	دالة الأساس الشعاعي (RBF)
%0.936	الاختبار على بيانات التدريب Testing on training data	دالة الأساس الشعاعي (RBF)

%0.714	التحقق المتقاطع Cross-validation	الدالة (Sigmoid)
%0.705	العينة العشوائية Random sample	الدالة (Sigmoid)
%0.652	الاختبار على بيانات التدريب Testing on training data	الدالة (Sigmoid)

مناقشة النتائج : تشير النتائج إلى أن **الدالة متعددة الحدود** حققت أعلى أداء عام في تصنيف بيانات الطلاب، حيث وصلت نسبة الدقة إلى 0.943% عند استخدام أسلوب الاختبار على بيانات التدريب. كما حققت دالة الأساس الشعاعي أداءً جيداً ونتائج مقاربة، بينما سجلت الدالة (Sigmoid) أدنى مستويات الدقة عبر جميع أساليب التقييم.

مقارنة دقة التصنيف (Classification Accuracy) لدوال النواة المختلفة

جدول 4. مقارنة دقة التصنيف لدوال النواة المختلفة

نوع SVM	دالة النواة	Test On Train Data	Cross Validation	Random Sampling
ثنائي الاستجابة	Polynomial	0.900	0.860	0.895
ثنائي الاستجابة	RBF	0.970	0.947	0.931
ثنائي الاستجابة	Sigmoid	0.699	0.692	0.709
متعدد الاستجابة	Polynomial	0.943	0.845	0.835
متعدد الاستجابة	RBF	0.936	0.817	0.800
متعدد الاستجابة	Sigmoid	0.652	0.714	0.705

ملخص نتائج الدراسة

أولاً: نتائج دقة التصنيف باستخدام نماذج آلة المتجه الداعم ثنائي الاستجابة

أظهرت نتائج الدراسة باستخدام نماذج آلة المتجه الداعم ثنائي الاستجابة (Support Vector Machine – SVM) عبر ثلاثة أساليب تقييم مختلفة في البرنامج الإحصائي ORANGE أن دقة التصنيف كانت الأعلى عند استخدام دالة الأساس الشعاعي (Radial Basis Function – RBF) مقارنة بالدالة متعددة الحدود (Polynomial) والدالة (Sigmoid).

وقد كان ترتيب الأساليب الثلاثة وفقًا لدقة التصنيف عند استخدام دالة RBF كما يلي:

1. ***أسلوب الاختبار على بيانات التدريب (Test On Train Data): الأعلى في دقة التصنيف عبر جميع النتائج.***
2. ***أسلوب التحقق المتقاطع (Cross Validation): المرتبة الثانية في دقة التصنيف.***
3. ***أسلوب العينة العشوائية (Random Sampling): الأقل دقة بين الأساليب الثلاثة.***

ثانياً: نتائج دقة التصنيف باستخدام نماذج آلة المتجه الداعم متعدد الاستجابة

أظهرت نتائج الدراسة باستخدام نماذج **آلة المتجه الداعم متعدد الاستجابة** (Multiclass SVM) أن **الدقة كانت الأعلى عند استخدام الدالة متعددة الحدود (Polynomial)** مقارنة بدالة الأساس الشعاعي (RBF) والدالة (Sigmoid).

وقد كان ترتيب الأساليب الثلاثة وفقًا لدقة التصنيف عند استخدام الدالة متعددة الحدود كما يلي:

1. ***أسلوب الاختبار على بيانات التدريب (Test On Train Data): الأعلى في دقة التصنيف عبر جميع النتائج.***
2. ***أسلوب التحقق المتقاطع (Cross Validation): المرتبة الثانية في دقة التصنيف.***

3. أسلوب العينة العشوائية (Random Sampling): الأقل دقة بين الأساليب الثلاثة.

توصيات الدراسة

1. تطبيق نماذج آلة المتجه الداعم متعدد الاستجابة في الدراسات المستقبلية: يُوصى باستخدامها في تصنيف البيانات متعددة الاستجابة نظرًا لدقتها العالية ومرونتها في التعامل مع البيانات المعقدة.
2. استخدام مصفوفة التشويش لتقييم الأداء: فهي من أفضل الطرق لمقارنة عدد الحالات الإيجابية والسلبية المصنفة بشكل صحيح أو خاطئ، مما يتيح تقييمًا دقيقًا لنموذج التصنيف.
3. اعتماد دالة النواة الإشعاعية (RBF) كخيار أول للتصنيف حيث أظهرت دالة الإشعاع RBF أعلى أداء تصنيفي في معظم طرق التقييم، سواء في التصنيف الثنائي أو المتعدد ، لذلك يُوصى بالاعتماد عليها بشكل أساسي، خصوصًا في المشكلات التي تتسم بعدم الخطية أو بوجود حدود قرار معقدة.
4. الحذر عند استخدام دالة Sigmoidحيث أظهرت نتائج دالة Sigmoid أداءً أقل من باقي الدوال في أغلب طرق التقييم، مما يشير إلى عدم ملاءمتها للبيانات محل الدراسة، لذا يُوصى بتجنب استخدامها إلا بعد إجراء تحسينات أو ضبط موسّع للمعاملات .(Hyperparameter Tuning)
5. تحسين أداء دالة Polynomial عبر الضبط الدقيق للرتبة (Degree)، رغم أن دالة Polynomial حققت أداءً مقبولًا، إلا أن تباين النتائج بين طرق التقييم يشير إلى حساسيتها لرتبة كثير الحدود، يُوصى بإجراء تحليل موسّع للرتبة (Degree Analysis) للعثور على القيمة المثلى التي تحقق توازنًا بين التعقيد والأداء.
6. عدم الاعتماد على نتائج Test on Train Dataحيث أظهرت هذه الطريقة أعلى نسب دقة لجميع الدوال، مما يؤكد أنها غير مناسبة لتقييم النموذج بسبب احتمالية التحيز وارتفاع خطر الملاءمة الزائدة .(Overfitting)
   يُوصى بالاعتماد على:
   - Cross Validation
   - Random Sampling

   باعتبارهما الأكثر موثوقية.
7. التفضيل بين Cross Validation وRandom Sampling ، حيث تميل نتائج Cross Validation إلى تقديم تقييم أكثر استقرارًا وضمانًا لعدم التحيز مقارنة بـ Random Sampling-. وبناءً عليه، يُوصى بجعل Cross Validation الطريقة الأساسية للتقييم مع الاستعانة بـ Random Sampling كطريقة داعمة.
8. تحسين أداء التصنيف المتعدد عبر توليف Hyperparameters، نظرًا لانخفاض دقة التصنيف المتعدد مقارنة بالثنائي، يُوصى بإجراء:
   - Grid Search
   - Random Search
   - Bayesian Optimization

   لتحسين معاملات C وγ ودرجة Polynomial.
9. تجربة خوارزميات مساعدة لرفع أداء التصنيف المتعدد، وفي حال استمرار انخفاض الأداء، يُوصى بالنظر في:
   - SVM مع استراتيجيات OvR وOvO محسّنة
   - نماذج أخرى مثل Random Forest وXGBoost للمقارنة.
10. التركيز على مرضى القلب: ضرورة توفير المستشفيات والمراكز الصحية المتخصصة، وتوفير العلاج اللازم لهم، بالإضافة إلى تعزيز البحوث والدراسات العلمية المتعلقة بأمراض القلب لتحسين التشخيص والرعاية.

المراجع

1. جاب الله, & محمود أسامة مصطفى السيد. (2024). استخدام تقنية SVM في رصد وتقييم دور الإنسان كعامل جيومورفولوجي على سواحل بحيرة المنزلة خلال الفترة الممتدة بين عامي (1984م-2024م) دراسة في الجيومورفولوجيا التطبيقية. المجلة العلمية بكلية الآداب, 2024(56), 708-736.
2. سالم، أحمد محمد. (2025). *المقارنة بين نماذج آلة المتجه الداعم في تصنيف البيانات (دراسة تطبيقية)*. **مجلة البحوث المالية والتجارية**، 26(3)، 138–162 .
3. Amari, S., & Wu, S. (2021). *Improving kernel-based classification models for biomedical data*. Journal of Machine Learning Research, 22(1), 1–25.
4. Ben-Hur, A., & Weston, J. (2019). *A user's guide to support vector machines.* In *Methods in Molecular Biology.*
5. Cai, W., Cai, M., Li, Q., & Liu, Q. (2023). *Three way Imbalanced Learning based on Fuzzy Twin SVM*. arXiv. https://arxiv.org/abs/2301.
6. Chen, L., Zhang, Y., & Wu, X. (2023). *Machine learning-based heart disease prediction using clinical and demographic features*. IEEE Access.
7. Duarte, J., & Ribeiro, J. (2023). *Advances in multi-class support vector machines for educational data classification*. International Journal of Data Science and Analytics, 17(1), 25–41.
8. El Maouaki, W., Said, T., & Bennai, M. (2024). Quantum support vector machine for prostate cancer detection: A performance analysis. arXiv. https://arxiv.org/
9. Halagundegowda, G. R., Singh, A., T. L., M. K., & Naveena, K. (2023). Evaluation of classification ability of support vector machine (SVM) in binary classification problems. *International Journal of Statistical Applications and Mathematics, 8*(5), 07–13.
10. Hastie, T., Tibshirani, R., & Friedman, J. (2022). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
11. Hsu, C.-W., & Lin, C.-J. (2021). *A Comparison of Methods for Multi-class Support Vector Machines.* IEEE Transactions on Neural Networks, 32(6), 2345–2357.
12. Islam, M., Rahman, S., & Akter, N. (2024). *Advanced predictive modeling for heart disease using hybrid machine learning techniques*. Expert Systems with Applications.
13. Karam, F., & Chen, D. (2024). Socioeconomic predictors of student achievement: A global perspective. *International Journal of Educational Research, 125*.
14. Khan, T., Alshurideh, M., & Salloum, S. (2023). Predicting student performance using machine learning: A systematic review. *Education Sciences, 13(4)*.

15. Kowsari, K., Brown, D., & Barnes, L. (2022). *Evaluation of kernel-based support vector machines for multi-class problems*. Machine Learning and Data Analytics Review, 12(2), 98–115.
16. Kumar, A., & Mishra, D. (2025). *Improving Support Vector Machine using Modified Kernel Function*. *International Journal of Scientific Research and Modern Technology, 4*(5), 1–5.
1. Lee, J., & Kim, S. (2024). Gender differences in academic performance among secondary students. *Journal of Educational Psychology, 116(2)*.
2. Li, Q., & Wang, P. (2024). *Cardiovascular risk prediction using optimized classification algorithms*. Journal of Biomedical Informatics.
3. Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2022). *An introduction to kernel-based learning algorithms*. IEEE Transactions on Neural Networks and Learning Systems.
1. Okfalisa, S., Ahmed, A., *et al.* (2021). SVM is a classification method that produces a fairly high degree of accuracy. *Journal of Engineering Science and Technology*, 16(2), 1372–138?. JESTEC.
4. Powers, D. M. W. (2020). *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation.* Journal of Machine Learning Technologies, 2(1), 37–63.
5. Rahman, M., & Yadav, R. (2023). *Heart disease prediction using supervised learning models: A comparative study*. Applied Intelligence.
6. Roy, A., & Singh, R. (2024). Multiclass classification approaches for predicting students' academic outcomes. *Applied Intelligence, 54*.
7. Sahu, S. K., & Tripathy, B. K. (2023). *Applications, Strengths and Limitations of Support Vector Machines in Data Classification.* Journal of Machine Learning Research, 24(12), 1–19.
2. Schölkopf, B., & Smola, A. J. (2018). *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* MIT Press.
3. Sholihah, W., Handayani, A. S., & Sarjana, S. (2023). Revolutionizing healthcare: Comprehensive evaluation and optimization of SVM kernels for precise general health diagnosis. *Scientific Journal of Informatics, 10*(4). https://journal.unnes.ac.id
8. Zhang, L., & Wu, X. (2023). *An Improved Gaussian Kernel for SVM Classification.* Journal of Intelligent & Fuzzy Systems, 45(2), 1957–1972.
9. Zhang, Y., & Wang, S. (2021). *Support vector machines for multi-class classification: A comparative study of kernel functions*. Journal of Intelligent Systems, 30(4), 543–558.
10. Zhu, L., Chen, H., & Li, Y. (2023). Age-related factors affecting academic performance: A meta-analysis. *Learning and Instruction, 85*.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **JLABW** and/or the editor(s). **JLABW** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.