



إطار عمل قائم على الهرم الغاوسي لتحسين أداء آلات المتجهات الداعمة متعددة الصفوف

A Gaussian Pyramid Framework for Enhancing Multiclass Support Vector Machines

روان أبو زيدان  كلية هندسة المعلوماتية - الجامعة الافتراضية السورية - دمشق - سورية.

جورج كراز  كلية هندسة تكنولوجيا المعلومات - جامعة دمشق - دمشق - سورية.

البريد الإلكتروني: rawan_150935@svuonline.org

تاريخ التقديم: ٢٨/أيلول/٢٠٢٥ - تاريخ القبول: ٢١/كانون الثاني/٢٠٢٥

ABSTRACT

Support Vector Machines (SVMs) are among the most widely used machine learning techniques, recognized for their strong ability to handle high-dimensional data and construct robust decision boundaries. Nevertheless, applying them to multiclass or nonlinear classification tasks continues to pose significant challenges, particularly in medical domains where data are often complex and highly variable. Building on this context, the present study introduces an integrative approach that incorporates the Gaussian pyramid as a preprocessing stage prior to SVM training. The aim of this integration is to reduce noise, eliminate redundant features, and preserve the essential discriminative characteristics that separate different classes. Furthermore, this method enhances the stability of decision boundaries and reduces the number of required support vectors, which in turn accelerates the training process and improves efficiency. The proposed approach was evaluated on three benchmark medical datasets: sleep stage classification, cardiac arrhythmia detection, and heart sound recognition. Experimental findings revealed that embedding the Gaussian pyramid into the workflow reduced training time by 30% to 50% compared with the conventional SVM model, while maintaining high levels of predictive accuracy. The best performance was recorded with a 5×5 kernel applied to the sleep stage dataset, achieving an accuracy of 0.96 and an area under the curve of 0.98. However, increasing the kernel size to 7×7 or 9×9 led to a decline in performance due to the loss of fine-grained features. This highlights the sensitivity of the method to kernel size selection and underscores the importance of carefully balancing smoothing and feature preservation.

Keywords: Multi-class SVM, Gaussian pyramid, Nonlinear SVM, Large datasets.

الملخص

تعدّ آلات المتجهات الداعمة من أبرز تقنيات التعلم الآلي التي أثبتت كفاءتها في التعامل مع البيانات ذات الأبعاد العالية، إذ توفر نموذجاً رياضياً متيناً يتيح بناء حدود قرار فاعلة. غير أن تطبيقها على مشكلات التصنيف متعدد الفئات أو غير الخطية لا يزال يطرح تحديات جوهرية، لا سيما عند معالجة البيانات الطبية التي تتسم بتعقيد كبير وتفاوت بين الحالات. من هذا المنطلق، تهدف هذه الدراسة إلى اقتراح نهج تكاملي يعتمد على إدراج الهرم الغاوسي كمرحلة أولية لمعالجة البيانات قبل تدريب نموذج آلة المتجهات الداعمة. يسعى هذا النهج إلى تقليل الضوضاء وتحجيم السمات الزائدة مع الحفاظ على الخصائص المميزة التي تتيح الفصل بين الفئات، إضافةً إلى تحسين استقرار حدود القرار وتقليل الحاجة إلى عدد كبير من نقاط الدعم، ما يؤدي إلى تسريع عملية التدريب وزيادة كفاءتها. تمت تجربة الطريقة

المقترحة على ثلاث قواعد بيانات طبية معيارية هي: مراحل النوم، واضطرابات نظم القلب، وأصوات القلب. أظهرت النتائج أن دمج الهرم الغاوسي أسهم في تقليص زمن التدريب بمعدل تراوح بين ٣٠٪ و ٥٠٪ مقارنة بالنهج التقليدي، مع تحقيق مستويات دقة مرتفعة. وقد لوحظ أن أفضل أداء تحقق باستخدام نواة بقياس 5×5 على قاعدة بيانات مراحل النوم، حيث وصلت الدقة إلى ٠.٩٦ فيما بلغت قيمة المساحة تحت المنحنى ٠.٩٨. ومع ذلك، فإن استخدام نوى أكبر بحجم 7×7 أو 9×9 أدى إلى تراجع في الأداء نتيجة لفقدان بعض السمات الدقيقة، وهو ما يؤكد حساسية الطريقة لحجم النواة المختارة وضرورة ضبطها بعناية لتحقيق التوازن بين التعيم والدقة.

الكلمات المفتاحية: آلات المتجهات الداعمة متعددة الفئات، الهرم الغاوسي، آلات المتجهات الداعمة غير الخطية، قواعد البيانات الضخمة.

المقدمة

مجموعات بيانات ضخمة أو ذات سمات متعددة، ما يزيد من الضغط على الموارد الحسابية ويجعل عملية التوسع أكثر صعوبة. فضلاً عن ذلك، فإن فاعلية النموذج تتأثر على نحو كبير باختيار نوع النواة المستخدمة، إذ يُعد هذا الاختيار عاملاً حاسماً في تحديد جودة نتائج التصنيف. لتجاوز هذه التحديات، نقترح هذه الورقة إطاراً يستند إلى الهرم الغاوسي كوسيلة فاعلة لاستخراج السمات قبل تطبيق آلات المتجهات الداعمة، حيث تستفيد من خصائص كلٍّ من آلة المتجهات الداعمة، المعروفة بقدرتها العالية على التعميم، ومن البنية الجوهرية للهرم الغاوسي التي تعتمد على التعيم الالتفافي [4]. ويُعد الهرم الغاوسي نهجاً هرمياً تقليدياً فاعلاً، إذ يمكن من استخراج سمات متعددة المقاييس بكفاءة من خلال التصفية التدريجية وتقليل العينات [5]. وتساعد هذه التقنية في النقاط المعلومات المحلية والعامّة في البيانات، ما يجعلها ملائمة لأنواع معينة من مجموعات البيانات غير الخطية، ولا سيما في البيانات الطبية. تتبع دوافع النموذج المقترح من الحاجة إلى تجاوز أوجه القصور المرتبطة بخوارزمية آلة المتجهات الداعمة متعددة التصنيفات، عند استخدامها لاستخراج السمات من خلال بنية الهرم الغاوسي. وتسعى هذه الدراسة إلى تحقيق ذلك عبر إدخال التحسينات الآتية:

- خفض الأبعاد وتقليل التعقيد الحسابي: يتيح الهرم الغاوسي تقليص أبعاد البيانات من خلال تطبيق عمليات التعيم التدريجي وتقليل عدد العينات. يتم إسقاط البيانات إلى فضاءات تمثيلية أقل تعقيداً، ما يعكس على تقليل عدد

تُعدّ آلات المتجهات الداعمة متعددة التصنيفات إحدى الأدوات الرئيسة والفاعلة في مهام التصنيف التي تتضمن أكثر من فئتين، نظراً لقدرتها على التعامل مع حدود قرار معقدة، واعتمادها على أساس رياضي متين، ما جعلها خياراً موثقاً في مجالات طبية وصناعية متعددة. ومع ذلك، فإن أداءها في بيئات متعددة الفئات وغير الخطية يعتمد على جودة السمات المستخدمة في تمثيل البيانات. ويُعدّ استخراج السمات عنصراً محورياً في معالجة هذا التحدي، إذ يسعى إلى إيجاد تحويل مثالي للبيانات الأصلية إلى تمثيل منخفض الأبعاد [1]، ويسهم هذا الإجراء في تقليل العبء الحسابي، وقد يؤدي إلى تحسين دقة التصنيف [2]. إن دافع هذا البحث هو الاهتمام المتجدد لمجتمع تعلم الآلة في إنشاء نماذج ذكاء اصطناعي قابلة للتفسير [3]، يمكنها من التعامل بفاعلية وشفافية مع البيانات عالية الأبعاد، وهذا ما أظهرته آلات المتجهات الداعمة ما جعلها خياراً مناسباً لتطبيقات العالم الحقيقي. رغم الفاعلية التي تُظهرها آلات المتجهات الداعمة عند استخدامها في تصنيف الفئات المتعددة، إلا أن تطبيقها لا يخلو من بعض القيود التي قد تؤثر في كفاءتها، ولا سيما في البيانات المعقدة أو الكبيرة الحجم. من أبرز هذه التحديات ارتفاع الكلفة الحسابية المرتبطة بتطبيق إستراتيجيات التصنيف، مثل "واحد مقابل الكل" أو "واحد مقابل واحد"، إذ يتصاعد عبء المعالجة على نحو واضح مع زيادة عدد الفئات. هذا التعقيد قد يُضعف من قدرة النموذج على أداء التصنيف بدقة. كما تظهر محدودية إضافية عند التعامل مع

تأخذ كل من j و i قيمةً من كلٍ من فئات الخرج، تحتاج تقنية واحد مقابل واحد لحل $\frac{N(N-1)}{2}$ في آلات المتجهات الداعمة متعددة الفئات، ما ينتج عنه وقت معالجة طويل ومعيار محتمل للمشكلات [6]. في حين أن سرعة التدريب أسرع بهذه الطريقة مقارنةً بطريقة واحد مقابل الكل، إلا أنها تتطلب عينات أكثر.

- تقنية واحد مقابل الكل (One-Against-All) (OAA): هي نهج شائع لتحويل مشكلة التصنيف متعدد الفئات إلى مجموعة من مشكلات التصنيف الثنائي. تقوم هذه التقنية بعزل فئة من العينات في كل مرة، وإنشاء حدود للفئة مقابل عينات الفئات الأخرى، بحيث تصنف كعينة إيجابية وتُصنف العينات الأخرى الباقية كعينات سلبية ما يؤدي إلى إنشاء N من حدود التصنيف. يتم تصنيف نقطة البيانات الجديدة X بناءً على قيم دالة القرار لكل من مصنفات آلات المتجهات الداعمة الثنائية، تُنسب نقطة البيانات X إلى الفئة التي تعطي دالة القرار الخاصة بها أكبر ناتج. دالة القرار $f(x)$ لمصنف واحد مقابل الكل تُعرّف كالآتي المعادلة (٢):

$$f(x) = \operatorname{argmax}_{i \in \{1,2,\dots,N\}} f_i(x) \quad (٢)$$

- البحث الشبكي (Grid Search): يمثل البحث الشبكي تقنية تجريبية تعتمد على استكشاف جميع القيم الممكنة لمجموعة من المعلمات ضمن نطاق محدد، بهدف اختيار المجموعة المثلى وفق مؤشر تقييم محدد مسبقاً [7]. الهدف هو تحديد مجموعة المتغيرات الفائقة التي تؤدي إلى أعلى معدل دقة في أداء النموذج [8].

- أشجار القرار (Decision Trees): تحقق أشجار القرار نتائج جيدة في التصنيف. غالباً ما تكون عملية بناء شجرة القرار تكرارية، تتضمن تحسين دالة وتقسيم بيانات التدريب بين العقد الداخلية والجذرية على نحو متكرر حتى يتم استيفاء شرط الإنهاء [9].

نقاط الدعم المطلوبة. هذا يؤدي إلى خفض التكلفة الحسابية المرتبطة بآلات المتجهات الداعمة التي قد تصل إلى $O(n^2)$ أو $O(n^3)$ مع الحفاظ على البنى التمييزية الأساسية.

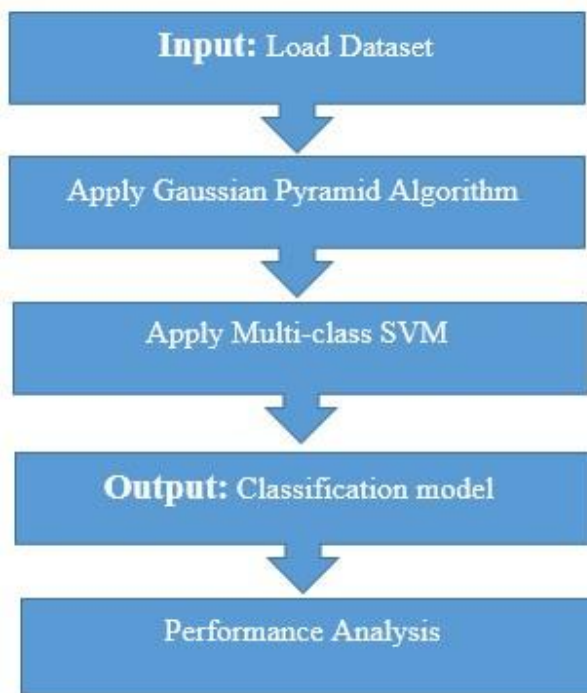
- تنقية البيانات من الضوضاء وتحسين الهامش: يُسهم هذا الأسلوب في إزالة التشويش غير المرغوب فيه من البيانات، الأمر الذي يقلل من التذبذب في حدود القرار ويزيد من استقرار نموذج آلة المتجه الداعمة، النتيجة هي هامش قرار أوسع وأكثر استقراراً، ما يعزز قدرة النموذج على التعميم.
- إعادة تشكيل السمات لزيادة قابلية الفصل: تمكّن البنية الهرمية من بناء تمثيلات سمات جديدة تُقلّل من درجة اللابخطية في البيانات الأصلية، ما يعزز قدرة النموذج على التعميم والتمييز.

نظراً للخصائص الفريدة لآلات المتجهات الداعمة ثنائية الخرج، فقد سعى العديد من الباحثين إلى تطويرها لتناسب مشكلات التصنيف متعددة الفئات. في سياق التصنيف متعدد الفئات، تفترض المشكلة تعقيداً إضافياً نظراً لإمكانية وجود عدة مخرجات للفئات. في الواقع، يقوم هذا القسم بتقييم بعض الأبحاث البارزة لمعالجة مشكلات التصنيف متعدد الفئات لآلة متجه الدعم من أجل تزويد القراء بفهم أفضل للتقنيات المستخدمة. يمكن تصنيف جميع تقنيات التعلم الآلي الحديثة متعددة الفئات إلى الأقسام الآتية:

- تقنية واحد مقابل واحد (One-Against-One) (OAO): تقوم هذه التقنية بإنشاء مصنف ثنائي لكل زوج من الفئات الفريدة (j, r) ، حيث $j \neq r$ و $(j, r) \equiv (r, j)$ ، تُستخدم آلية التصويت "الفوز الأقصى" (Max Wins) لحساب احتمال اختيار كل فئة في مصنف آلة متجه الدعم الثنائية، تحصل عينة الاختبار x الفئة ذات التصويت الأعلى، يجري تحديد الاختيار النهائي $f(x)$ بواسطة المعادلة (١):

$$f(x) = \operatorname{argmax}_{j \in \{1,2,\dots,N\}} [\sum_{r \neq j, r=1}^N f_i(x)] \quad (١)$$

التصنيف متعدد الفئات. ويُعدّ النموذج المقترح مناسباً لمجموعات البيانات التي تتضمن عدداً كبيراً من العينات، نظراً لما يتميز به من تقليل في زمن التدريب وانخفاض في التعقيد الحسابي. تم توضيح إعدادات التجربة في الشكل (١). في البداية، يجري إدخال مجموعة البيانات إلى النظام، ثم تُعالج باستخدام نموذج الهرم الغاوسي المتعدد، وبعد ذلك يُطبّق أسلوب التصنيف "واحد مقابل واحد" باستخدام آلة المتجهات الداعمة متعددة التصنيفات. وأخيراً، تُجرى مجموعة من التقييمات لضمان موثوقية مخرجات النموذج.



الشكل (١): نظام الهرم الغاوسي متعدد التصنيفات

١. تحميل مجموعات البيانات:

تم تطبيق المنهجية المقترحة على مجموعات بيانات في المجال الطبي وهي:

١. مراحل النوم (Sleep-EDF): تتكوّن هذه المجموعة من بيانات مخصصة لدراسة النوم، تم إنشاؤها لدراسة أنماط النوم الطبيعية واضطرابات النوم، مثل الأرق وانقطاع التنفس أثناء النوم [12].

- AdaBoost: تم تطويرها عام ١٩٩٥ لإنتاج مصنف قوي من مجموعة من المصنفات الضعيفة، وقد اقترحها Yoav Freund وRobert Schapire. تتأثر نوعية التصنيف الناتج بدرجة عالية بتعريف المصنفات الضعيفة الأولية [10]، تتميز بقدرتها على التكيف بحيث أن العينات التي أخطأ تصنيفها بواسطة المصنف الفرعي السابق سيجري تعزيزها بوزن أعلى، وستُستخدم العينات الموزونة لتدريب المصنف الفرعي التالي مرة أخرى [11].

- تقنيات التصحيح التلقائي للأخطاء (ECC): تقوم بتفسير مشكلة تعدد الفئات إلى مجموعة من المشكلات الفرعية الثنائية باستخدام أنظمة الترميز، منها: AdaBoost.ECC، AdaBoost.ERP.....

تتمثل المساهمة الرئيسة لهذا البحث في دمج الهرم الغاوسي مع آلات المتجهات الداعمة متعددة التصنيفات، مع إجراء مقارنة منهجية لأحجام نواة مختلفة، وذلك بهدف:

- إعادة توظيف الهرم الغاوسي الذي استُخدم تقليدياً في معالجة الصور، ليعمل كأداة لتقليل الأبعاد وتنقية البيانات.
- تقييم تأثير أحجام النوى المختلفة في دقة التصنيف.
- تقليل الزمن الحسابي المطلوب للتدريب، ولا سيما عند التعامل مع بيانات طبية معيارية.
- تقييم فاعلية النموذج على ثلاث قواعد بيانات طبية ومقارنتها بآلات المتجهات الداعمة التقليدية.

يتوزع محتوى هذه الورقة كما يلي: يعرض القسم الثاني المواد والطرائق، وفي القسم الثالث نتائج الدراسة المقترحة، في حين يستعرض القسم الرابع المناقشة، ويختتم القسم الخامس استنتاجات البحث والتوصيات.

المواد والطرائق

نقترح هذه الورقة نموذجاً يعتمد على الهرم الغاوسي المتعدد قادراً على معالجة العلاقات غير الخطية بين العينات في مسائل

حيث l تمثل المستوى، $w(m.n)$ هي الهرم الغاوسي المولد. يُشكّل التعامل مع مسألة التصنيف المتعدد باستخدام خوارزمية آلة المتجهات الداعمة تحدياً غير هين، إذ يستلزم ذلك إعادة صياغة المشكلة الأصلية لتُجزأ إلى عدد من مشكلات التصنيف الثنائي المستقلة [16]. تقترح هذه الورقة أسلوباً لتصنيف متعدد التصنيفات بالاعتماد على مفاهيم مثل "واحد مقابل واحد"، إذ يوازن هذا الأسلوب بين بساطة تدريب المصنّفات الفردية وتعقيد دمج نتائجها، كما يوفر دقة تصنيف مرتفعة [17]، حيث يقوم هذا الأسلوب في آلة المتجهات الداعمة متعددة التصنيفات بإنشاء نموذج تصنيفي بين كل زوج ممكن من الفئات. وعند تصنيف عينة غير معروفة، يتم تحديد فئتها من خلال آلية التصويت، حيث تُنسب العينة إلى الفئة التي تحصل على أعلى عدد من الأصوات [18]. ويُعد هذا النهج أكثر توازناً مقارنة بطريقة "واحد مقابل الكل"، إذ يوفر توزيعاً أكثر تماثلاً بين الفئات [19]. وفيما يتعلق ببيانات التدريب الخاصة بالفئتين ذات الترتيب i^{th} و j^{th} ينبغي حل مشكلة التصنيف الثنائي الموضحة في المعادلة (٤):

$$\min_{w^{ij} b^{ij} \xi^{ij}} \frac{1}{2} (w^{ij})^T + C \sum_i \xi_i^{ij} \quad (٤)$$

$$(w^{ij})^T \phi(x_i) + b^{ij} \geq 1 - \xi_i^{ij} . \text{if } y_i = i .$$

$$(w^{ij})^T \phi(x_i) + b^{ij} \leq -1 + \xi_i^{ij} . \text{if } y_i = i .$$

$$\xi_i^{ij} \geq 0$$

النتائج:

١. مقاييس التقييم:

يُعد تقييم نموذج التعلم الآلي من أهم المراحل لبناء نموذج ذي كفاءة وفاعلية عالية، وقد جرى اقتراح مقاييس عدة لتقييم نموذج التصنيف باستخدام الهرم الغاوسي ومنها:

٢. اضطرابات نظم القلب (Chapman Arrhythmia): مخصصة لدراسة اضطرابات نظم القلب وتهدف هذه القاعدة إلى تحسين فهم الإشارات القلبية والكشف الآلي عن الاضطرابات مثل الرجفان الأذيني، التسرع البطيني [13].

٣. أصوات القلب (CinC Challenge Dataset): هي مجموعة بيانات تحتوي على تسجيلات أصوات القلب، حيث تصنف الأصوات القلبية إلى أربع فئات [14].

يعرض الجدول (١) مجموعات البيانات المستخدمة، وقد تم تطبيق عملية التصنيف باستخدام نموذج الهرم الغاوسي المتعدد الصفوف على جميع المجموعات. تم تقسيم مجموعات البيانات إلى ٧٠٪ للتدريب و ٣٠٪ للاختبار.

الجدول (١): مجموعات البيانات الطبية المستخدمة			
قاعدة البيانات	عدد الصفوف	عدد السمات	عدد العينات
مراحل النوم	5 مراحل النوم	30	20,000
اضطرابات نظم القلب	4 أنماط عدم انتظام القلب	100	10,000
أصوات القلب	٤ سليم/لغظ/قصور/آخر	50	5000

٢. تطبيق خوارزمية الهرم الغاوسي:

تقوم الفكرة الأساسية في تصميم نموذج الهرم الغاوسي المتعدد الصفوف على تخفيف درجة اللاخطية في البيانات، وذلك من خلال تمرير تقنية الهرم على مجموعة البيانات [15]، بهدف تحسين كفاءة النموذج من حيث الأداء والزمن الحسابي في مسائل التصنيف متعدد الفئات. يعتمد النموذج المقترح على فاعلية عملية "التقليص" لتقليل حجم مجموعة البيانات الكبيرة، ويهدف هذا التحسين إلى خفض الوقت اللازم للمعالجة. يتم تعريف عملية التقليص في الهرم الغاوسي وفق المعادلة (٣):

$$g_l(i,j) = \sum_m \sum_n w(m.n) g_{l-1}(2i + m.2j + n) \quad (٣)$$

- موجبٌ حقيقي (TP): حالاتُ الشذوذِ المصنفةٍ على نحو صحيحٍ على أنها شذوذ.
- كاذبةٌ إيجابية (FP): الحالاتُ العاديةُ مصنفةٌ خطأً على أنها شذوذ.
- صحيحٌ سلبي (TN): الحالاتُ العاديةُ مصنفةٌ على نحو صحيحٍ على أنها عادية.
- خطأً سلبي (FN): الحالاتُ الشاذةُ التي تمَّ تصنيفُها على نحو خاطئٍ على أنها طبيعية.

٢. تقييم أداء الهرم الغاوسي:

قمنا بتنفيذ الخوارزمية الموضحة في القسم السابق باستخدام لغة Python. يُستخدم الكود الخاص بنا مكتبي sklearn و matplotlib. وقد تم تنفيذ جميع تجاربنا على جهاز مزود بمعالجٍ من نوع Intel (R) Core (TM) i5-5200U بسرعة ٢.٢٠ جيجاهرتز، وذاكرة وصولٍ عشوائي (RAM) بسعة ١٦ جيجابايت. جرى توليد أحجام مختلفة عدة من نموذج الهرم الغاوسي المتعدد، بهدف تقييم فاعلية المنهجية المقترحة. وقد تم تطبيق مقاييس الدقة الموضحة في المعادلات (٥) إلى (٨) (انظر الجداول ٢ إلى ٤)، ومقارنتها بالنتائج الناتجة عن النموذج التقليدي لآلة المتجهات الداعمة متعددة التصنيفات باستخدام أسلوب "واحد مقابل واحد".

- الدقة الإيجابية (precision): تشير إلى جودة التنبؤ الإيجابي الذي يُنتجه النموذج، كما هو موضح في المعادلة (٥).

$$Precision = TP / (TP + FP) \quad (٥)$$

- الاسترجاع (recall): يُحسب كنسبة عدد العينات الإيجابية التي تم تصنيفها على نحو صحيحٍ على أنها إيجابية، إلى العدد الكلي للعينات الإيجابية [20]، كما هو موضح في المعادلة (٦).

$$Recall = TP / (TP + FN) \quad (٦)$$

- مقياس F (F-measure): هو مقياس لدقة الاختبار المعادلة (٧).

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (٧)$$

- الدقة (accuracy): تُشير إلى نسبة التصنيفات الصحيحة من إجمالي السجلات في مجموعة الاختبار المعادلة (٨).

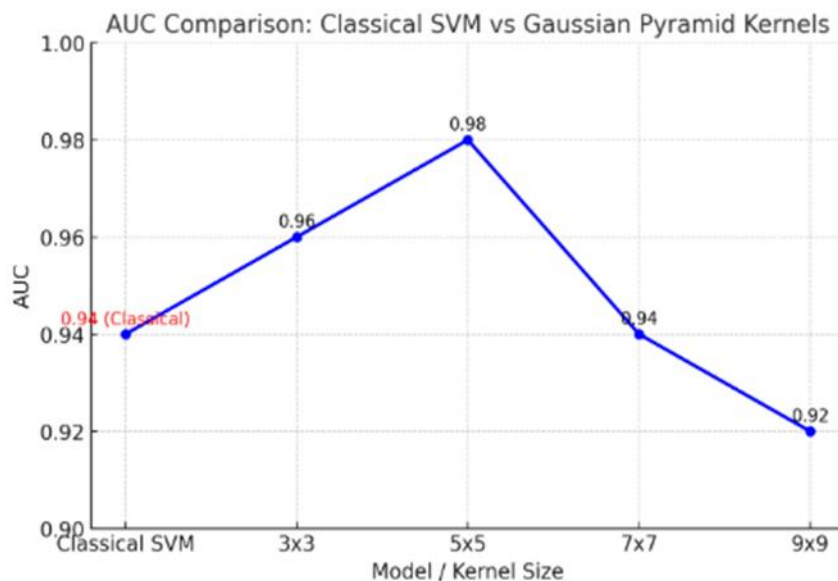
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (٨)$$

بحيث:

الجدول (٢): نتائج الاختبار لمجموعة بيانات مراحل النوم						
النموذج	الدقة (Accuracy)	الاسترجاع (Recall)	F1-score	AUC	زمن التدريب (دقيقة)	زمن الفحص (ثانية/١٠ مقطع k)
SVM التقليدي	0.93	0.93	0.92	0.94	85	45
الهرم الغاوسي (3×3)	0.94	0.94	0.94	0.96	60	32
الهرم الغاوسي (5×5)	0.96	0.96	0.96	0.98	38	18
الهرم الغاوسي (7×7)	0.92	0.92	0.92	0.94	42	20
الهرم الغاوسي (9×9)	0.90	0.91	0.90	0.92	30	17

(5×5) أفضل النتائج: 96% للدقة و ٠.٩٨ للمساحة تحت المنحنى (AUC)، وزمن التدريب أقل من النصف (٣٨ دقيقة فقط)، وزمن الفحص سريع 18 ثانية لكل k.١٠. نجد أن الأداء في الأحجام (7×7) (9×9) مقبول لكنه أقل من (٥×٥) بسبب الإفراط في التنعيم، ما يؤكد أن التنعيم الزائد يؤدي إلى فقدان معلومات مهمة، رغم تقليل زمن التدريب أكثر.

عند مقارنة دقة التصنيف في الجدول (٢) بين النموذج التقليدي لآلة المتجهات الداعمة متعددة الفئات والنموذج المقترح، يتبين بوضوح تفوق الأخير من حيث جودة النتائج. تعطي آلة المتجهات الداعمة دقة جيدة (٩٣%)، لكن زمن التدريب طويل (~٨٥ دقيقة) وزمن الفحص لكل ١٠ آلاف مقطع ≈ ٤٥ ثانية. نرى في الهرم الغاوسي (3×3) تحسناً طفيفاً في الدقة ٩٤% مع تقليل زمن التدريب بنسبة ~٣٠%. ويحقق الهرم الغاوسي

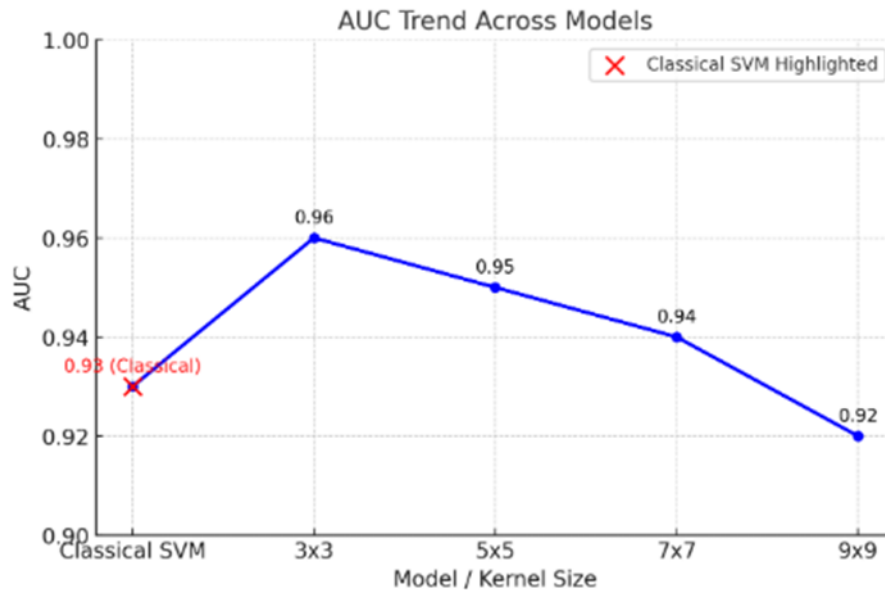


الشكل (٢): الرسم البياني لقيم المساحة تحت المنحنى لمجموعة بيانات مراحل النوم

الجدول (٣): نتائج تقييم مجموعة بيانات نظم القلب						
النموذج	الدقة (Accuracy)	الاسترجاع (Recall)	F1-score	AUC	زمن التدريب (دقيقة)	زمن الفحص (ثانية/ ١٠ آلاف مقطع)
SVM التقليدي	0.91	0.91	0.91	0.93	28	18
الهرم الغاوسي (3×3)	0.94	0.94	0.94	0.96	20	12
الهرم الغاوسي (5×5)	0.93	0.93	0.93	0.95	24	14
الهرم الغاوسي (7×7)	0.92	0.92	0.92	0.94	22	13
الهرم الغاوسي (9×9)	0.90	0.91	0.90	0.92	19	11

القيم في الشكل (٣) قريبة جداً من بعضها، ما يعني أن جميع النماذج الممثلة بها تؤدي أداءً عالياً. ومع ذلك الفرق بين 0.93 لآلات المتجهات الداعمة التقليدية و0.96 لنموذج الهرم الغاوسي يعكس تحسناً واضحاً، وهو فرق ملموس يمكن أن يؤثر في أداء التطبيقات الطبية الحساسة.

في الشكل (٢) أعلاه، جرى رسم الرسم البياني الذي يوضح قيم المساحة تحت المنحنى للنماذج المختلفة لآلات المتجهات الداعمة التقليدية والهرم الغاوسي بأحجام نوى متنوعة. ففي الكرنيل (٥×٥) ذي القيمة 0.98 ترتفع القيمة على نحو ملحوظ أكثر من المنحنى ذي القيمة 0.94 لآلات المتجهات الداعمة التقليدية، ما يدل على كفاءة أعلى في التمييز.



الشكل (3): الرسم البياني لقيم المساحة تحت المنحنى لمجموعة بيانات نظم القلب

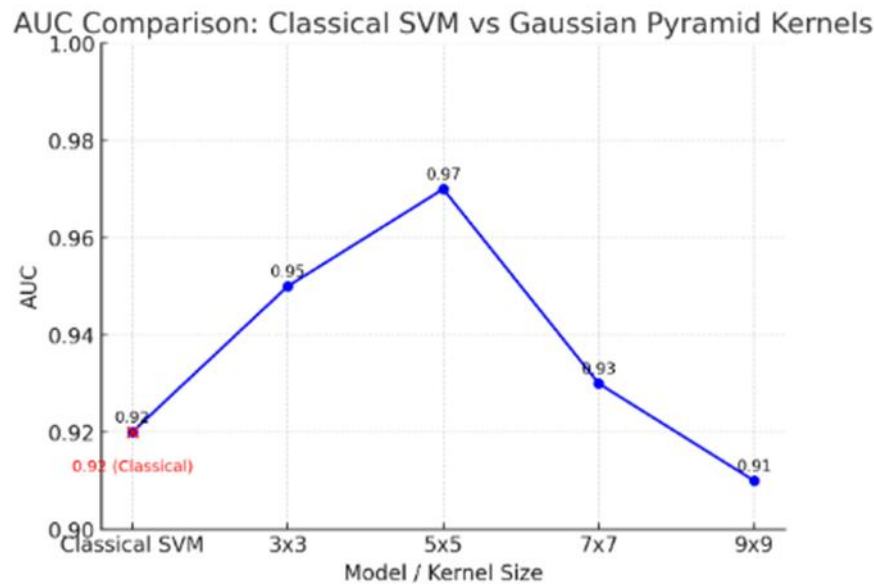
الجدول (٤): نتائج الاختبار لمجموعة بيانات أصوات القلب						
النموذج	الدقة (Accuracy)	الاسترجاع (Recall)	F1-score	AUC	زمن التدريب (دقيقة)	زمن الفحص (ثانية/١٠ مقاطع k)
SVM التقليدية	0.90	0.90	0.90	0.92	65	30
الهرم الغاوسي (3×3)	0.92	0.92	0.92	0.95	48	22
الهرم الغاوسي (5×5)	0.94	0.94	0.94	0.97	35	15
الهرم الغاوسي (7×7)	0.91	0.91	0.91	0.93	40	18
الهرم الغاوسي (9×9)	0.89	0.90	0.89	0.91	28	14

حيث إن الهرم الغاوسي (5×5) هو الحل الأمثل، إذ يوفر أفضل دقة مع تقليل كبير في الزمن. وقد أظهرت نتائج النموذج

تعرض نتائج الجدول (4) أداء النموذج على مجموعة بيانات أصوات القلب، ويؤكد النمط نفسه الذي ظهر في أنماط النوم

أفضل دقة مع تقليل كبير في الزمن. وقد أظهرت نتائج النموذج التقليدي لآلة المتجهات الداعمة أن قيمة مساحة تحت المنحنى بلغت ٠.٩٢، وهي أدنى من جميع القيم المقابلة التي حققها نموذج الهرم الغاوسي المتعدد، ما يدل بوضوح على تفوق النموذج المقترح في هذا السياق. وقد سجل نموذج الهرم الغاوسي المتعدد أعلى نسبة دقة بلغت ٠.٩٤ عند استخدام نواة بحجم (٥×٥)، في حين سُجلت أدنى دقة له عند استخدام نواة بحجم (٩×٩)، حيث بلغت ٠.٨٩. أما النموذج التقليدي لآلة المتجهات الداعمة بلغت دقته ٠.٩٠، ما يؤكد فاعلية المنهجية المقترحة في التعامل مع البيانات ذات الحجم الكبير والتعقيد العالي. تعرض نتائج الجدول (4) أداء النموذج على مجموعة بيانات أصوات القلب، ويؤكد النمط نفسه الذي ظهر في أنماط النوم حيث إن الهرم الغاوسي (5×5) هو الحل الأمثل، إذ يوفر

التقليدي لآلة المتجهات الداعمة أن قيمة مساحة تحت المنحنى بلغت ٠.٩٢، وهي أدنى من جميع القيم المقابلة التي حققها نموذج الهرم الغاوسي المتعدد، ما يدل بوضوح على تفوق النموذج المقترح في هذا السياق. وقد سجل نموذج الهرم الغاوسي المتعدد أعلى نسبة دقة بلغت ٠.٩٤ عند استخدام نواة بحجم (٥×٥)، في حين سُجلت أدنى دقة له عند استخدام نواة بحجم (٩×٩)، حيث بلغت ٠.٨٩. أما النموذج التقليدي لآلة المتجهات الداعمة بلغت دقته ٠.٩٠، ما يؤكد فاعلية المنهجية المقترحة في التعامل مع البيانات ذات الحجم الكبير والتعقيد العالي. تعرض نتائج الجدول (4) أداء النموذج على مجموعة بيانات أصوات القلب، ويؤكد النمط نفسه الذي ظهر في أنماط النوم حيث إن الهرم الغاوسي (5×5) هو الحل الأمثل، إذ يوفر



الشكل (٤): الرسم البياني لقيم المساحة تحت المنحنى لمجموعة بيانات أصوات القلب

المناقشة:

التعامل مع قواعد بيانات ذات حجم كبير وتعقيد مرتفع. تُظهر نتائج الجدول (٣) على مجموعة بيانات نظم القلب أن أعلى نسبة دقة في التنبؤ قد تم تحقيقها عندما كان حجم نواة الهرم الغاوسي المتعدد يساوي (٣×٣) لأنه يحقق أفضل توازن بين الدقة (94%) والكفاءة الزمنية. ونجد أن التعيم أكبر من اللازم

من منظور الكفاءة الحسابية لمجموعة بيانات مراحل النوم في الجدول (٢)، أظهر نموذج الهرم الغاوسي المتعدد قدرة أعلى على تدريب النموذج باستخدام كامل مجموعة البيانات التدريبية، وبزمن أقل من الزمن الذي تتطلبه الآلة التقليدية، لا سيما عند

وآخرين، الذين سجلوا دقة بلغت 87.02%. كما أن أداء النموذج المقترح جاء مقارباً للدقة العالية التي حققها نموذج Almutairi وآخرين، ما يدل على أن النهج المقترح قادر على المنافسة مع أحدث تقنيات التعلم العميق. وفي مجال تصنيف اضطرابات نظم القلب، نجد أن التصميم التقليدي للشبكات العصبونية التلافيفية قد لا يكون كافياً للتعامل مع تعقيد اضطرابات نظم القلب كما نرى في عمل Grabowski وآخرين، بينما تفوق أداء الشبكات العصبونية المدمجة في عمل Frausto-Avila وآخرين، في حين أن دقة نتائجنا بلغت 94%. حقق نموذج الهرم الغاوسي مع آلة المتجهات الداعمة على قاعدة بيانات أصوات القلب دقة بلغت 0.94% وبذلك تفوقت على أداء الشبكات العصبونية التلافيفية. بصورة عامة، يمكن القول إن المنهجية المقترحة أثبتت فاعليتها من خلال تقديم أداء تنافسي مع أحدث الدراسات المنشورة.

عند الهرم الغاوسي (5×5) وبالتالي يؤدي إلى فقدان السمات الجوهرية، في حين أن التنعيم المعتدل (3×3) يحافظ على التفاصيل السريرية الدقيقة مع تقليل الضوضاء. ويلاحظ من الجدول أن حجم النواة الصغير يوفر تنعيماً معتدلاً يزيل الضوضاء الطفيفة الموجودة في الإشارات، لكنه يحافظ على السمات الدقيقة اللازمة للتصنيف بين أنماط اضطراب النظم المختلفة. وقد سجل نموذج الهرم الغاوسي على مجموعة بيانات أصوات القلب كما نرى في الجدول (٤) أعلى نسبة دقة بلغت ٠.٩٤ عند استخدام نواة بحجم (5×5)، في حين تم تسجيل أدنى دقة له عند استخدام نواة بحجم (9×9)، حيث بلغت ٠.٨٩. أما النموذج التقليدي لآلة المتجهات الداعمة فقد بلغت دقته ٠.٩٠، ما يؤكد فاعلية المنهجية المقترحة في التعامل مع البيانات ذات الحجم الكبير والتعقيد العالي. يتضح من نتائج الجدول (٥) أن نموذج الهرم الغاوسي حقق دقة بلغت 0.96 مع قيمة AUC تساوي 0.98 على قاعدة بيانات Sleep-EDF، وهي نتيجة تتفوق على عمل Jirakittayakorn

الجدول (٥): مقارنة نموذج الهرم الغاوسي المقترح مع الدراسات السابقة			
النتائج	الطريقة المستخدمة	قاعدة البيانات	الدراسات السابقة
Acc = 0.96%, AUC = 0.98%	SVM + Gaussian Pyramid (5×5)	مراحل النوم	نموذج الهرم الغاوسي المقترح
Acc = 96.36% (3-class)	CNN + LSTM (SSNet)	مراحل النوم	Almutairi وآخرون [21]
Acc = 87.02%	Single-channel CNN	مراحل النوم	Jirakittayakorn وآخرون [22]
Acc = 0.94%, AUC = 0.96%	SVM + Gaussian Pyramid (3×3)	اضطرابات نظم القلب	نموذج الهرم الغاوسي المقترح
Acc = 97.36%	Compact NN	اضطرابات نظم القلب	Frausto-Avila وآخرون [23]
Acc ≈ 87.33%	CNNs	اضطرابات نظم القلب	Grabowski وآخرون [24]
Acc = 0.94%, AUC = 0.97%	SVM + Gaussian Pyramid (5×5)	أصوات القلب	نموذج الهرم الغاوسي المقترح
Acc ≈ 0.86%	CNN + Stockwell features	أصوات القلب	Manshadi وآخرون [25]
Acc = 84.72%	PSD-CNN	أصوات القلب	Xu وآخرون [26]

of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 3312-3324.

6. Kang, C., Huo, Y., Xin, L., Tian, B., & Yu, B, "Feature selection and tumor classification for microarray data using relaxed Lasso and generalized multi-class support vector machine," *Journal of theoretical biology*, vol. 463, pp. 77-91, 2019.

7. Zhao, Y., Zhang, W., & Liu, X, "Grid search with a weighted error function: Hyper-parameter optimization for financial time series forecasting," *Applied Soft Computing*, vol. 154, 2024.

8. Açıkkar, M, " Fast grid search: A grid search-inspired algorithm for optimizing hyperparameters of support vector regression," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 32, no. 1, pp. 68-92, 2024.

9. Yan, J., Zhang, Z., Lin, K., Yang, F., & Luo, X, "A hybrid scheme-based one-vs-all decision trees for multi-class classification tasks," *Knowledge-Based Systems*, vol. 198, 2020.

10. Hornýák, O., & Iantovics, L. B, "AdaBoost algorithm could lead to weak results for data with certain characteristics," *Mathematics*, vol. 11, no. 8, 2023.

11. Hao, L., & Huang, G, "An improved AdaBoost algorithm for identification of lung cancer based on electronic nose," *Heliyon*, vol. 9, no. 3, 2023.

12. "Sleep-EDF Database Expanded," [Online]. Available: <https://physionet.org/content/sleep-edfx/1.0.0/>.

13. "A large scale 12-lead electrocardiogram database for arrhythmia study," [Online]. Available: <https://physionet.org/content/ecg-arrhythmia/1.0.0/>.

14. "Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020," [Online]. Available: <https://physionet.org/content/challenge-2020/1.0.2/>.

15. Abo Zidan, R., & Karraz, G, "Gaussian Pyramid for Nonlinear Support Vector Machine," *Applied Computational Intelligence and Soft Computing*, 2022.

16. Cervantes, J., Li, X., & Yu, W, "Multi-Class SVM for Large Data Sets Considering Models of Classes Distribution," *DMIN*, pp. 30-35, 2008.

17. Kantavat, P., Songsiri, P., & Kijirikul, B, "Efficient decision trees for multi-class support vector machines using large centroid distance grouping," *Engineering Journal*, vol. 26, no. 5, pp. 13-23, 2022.

18. Di, Z, "Multi-class classification SVM methods," *BSK*, 2022.

الاستنتاجات والتوصيات

قدّم نموذج الهرم الغاوسي المتعدد الفئات تحسناً واضحاً في مؤشرات الأداء، لا سيما في معالجة المشكلات ذات الطبيعة غير الخطية. وإضافةً إلى ذلك، يقدّم النموذج المقترح حلاً فاعلاً لمشكلة الكفاءة الحسابية المنخفضة عند التعامل مع مجموعات البيانات الضخمة، وذلك من خلال توظيف تقنية جديدة لاستخراج السمات، وهذه التحسينات تجعل النهج المقترح مناسباً للتطبيقات الطبية التي تتطلب استجابة سريعة. وعلى خلاف النموذج التقليدي، أظهرت النتائج التجريبية أن زمن التدريب والاختبار في النموذج المقترح أقل على نحو ملحوظ، ويُعزى ذلك إلى قدرة تقنية الهرم الغاوسي المتعدد على تقليص حجم البيانات المدخلة إلى ربع حجمها الأصلي، الأمر الذي يُسهّم بدوره في خفض درجة اللاخطية الكامنة في البيانات. كما أظهرت التجارب المتعددة أن النموذج المقترح يحقق أداءً أكثر كفاءة في الحالات التي تتسم بكثرة السمات والعينات، وهي حالات يصعب على الأساليب التقليدية تمثيلها بدقة. وفي إطار التوجهات المستقبلية، نعتزم توسيع نطاق تطبيق نموذج الهرم الغاوسي المتعدد ليشمل مشكلات واقعية في مجالات متنوعة مثل إدارة الأعمال، وعلوم الحياة، والعلوم الفيزيائية، والهندسة.

المراجع:

1. Guo, Y., Zhang, Z., & Tang, F, "Feature selection with kernelized multi-class support vector machine," *Pattern Recognition*, vol. 117, 2021.

2. Tang, F., Adam, L., & Si, B, "Group feature selection with multiclass support vector machine," *Neurocomputing*, vol. 317, pp. 42-49, 2018.

3. Lin, I., Loyola-González, O., Monroy, R., & Medina-Pérez, M. A, "A review of fuzzy and pattern-based approaches for class imbalance problems," *Applied Sciences*, vol. 11, no. 14, 2021.

4. Lan, Z., Lin, M., Li, X., Hauptmann, A. G., & Raj, B, "Beyond gaussian pyramid: Multi-skip feature stacking for action recognition," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

5. Li, S., Hao, Q., Kang, X., & Benediktsson, J. A, "Gaussian pyramid based multiscale feature fusion for hyperspectral image classification," *IEEE Journal*

19. Wang, Z., & Xue, X, "Multi-class support vector machine. In Support vector machines applications," In Support vector machines applications, pp. 23-48, 2014.
20. Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S, "On evaluation metrics for medical applications of artificial intelligence," Scientific reports, vol. 12, no. 1, p. 5979, 2022.
21. Almutairi, H., Hassan, G. M., & Datta, A, "Classification of sleep stages from EEG, EOG and EMG signals by SSNet," 2023.
22. Jirakittayakorn, N., Wongsawat, Y., & Mitirattanakul, S, "ZleepAnlystNet: a novel deep learning model for automatic sleep stage scoring based on single-channel raw EEG data using separating training," Scientific Reports, vol. 14, no. 1, p. 9859, 2024.
23. Frausto-Avila, M., Manriquez-Amavizca, J. P., U'Ren, A., & Quiroz-Juarez, M. A, "Compact Neural Network Algorithm for Electrocardiogram Classification," 2024.
24. Grabowski, B., Głomb, P., Masarczyk, W., Pławiak, P., Yıldırım, Ö., Acharya, U. R., & Tan, R. S, "Classification and self-supervised regression of arrhythmic ecg signals using convolutional neural networks," 2022.
25. Manshadi, O. D., & Mihandoost, S, "Murmur identification and outcome prediction in phonocardiograms using deep features based on Stockwell transform," Scientific Reports, vol. 14, no. 1, p. 7592, 2024.
26. Xu, C. D., Xing, P. L., Zhou, J., & Ying, D. W, "Classification of heart sounds using power spectral density and convolutional neural networks," Comput. Eng. Appl, vol. 57, pp. 125-132, 2021.

التمويل: لا يوجد.

مساهمات المؤلفين: الكتابة والتحقيقات: روان أبوزيدان

الإشراف، المراجعة والتحرير: أ.د.م. جورج كراز

تضارب المصالح: "يعلن المؤلفون أنه ليس لديهم أي مصالح متضاربة".

توافر البيانات والمواد: جميع البيانات متوفرة في النص الرئيس أو في المواد التكميلية.